

Asymptotically Efficient Estimation of Weighted Average Derivatives with an Interval Censored Variable

Hiroaki Kaido
Boston University*

December, 2013

Abstract

This paper studies the identification and estimation of weighted average derivatives of conditional location functionals including conditional mean and conditional quantiles in settings where either the outcome variable or a regressor is interval-valued. Building on Manski and Tamer (2002) who study nonparametric bounds for mean regression with interval data, we characterize the identified set of weighted average derivatives of regression functions. Since the weighted average derivatives do not rely on parametric specifications for the regression functions, the identified set is well-defined without any parametric assumptions. Under general conditions, the identified set is compact and convex and hence admits characterization by its support function. Using this characterization, we derive the semiparametric efficiency bound of the support function when the outcome variable is interval-valued. We illustrate efficient estimation by constructing an efficient estimator of the support function for the case of mean regression with an interval censored outcome.

Keywords: Partial Identification, Weighted Average Derivative, Semiparametric Efficiency, Support Function, Interval Data

*The author acknowledges excellent research assistance from Michael Gechter. The author also thanks Iván Fernández-Val, Francesca Molinari, Pierre Perron, Zhongjun Qu, Andres Santos, Jörg Stoye, and seminar participants at BU, Cornell, Yale and participants of BU-BC Joint Econometrics Conference, the CEMMAP Workshop (New Developments in the Use of Random Sets in Economics), and the 2012 International Symposium on Econometric Theory and Applications (SETA) for helpful comments. Financial support from the NSF Grant SES-1230071 is gratefully acknowledged.

1 Introduction

Interval censoring commonly occurs in various economic data used in empirical studies. The Health and Retirement Study (HRS), for example, offers wealth brackets to respondents if they are not willing to provide point values for different components of wealth. In real estate data, locations of houses are often recorded by zip codes, which makes the distance between any two locations interval-valued. Analyzing regression models with such interval-valued data poses a challenge as the regression function is not generally point identified. This paper studies the identification and estimation of weighted average derivatives of general regression functions when data include an interval-valued variable.

Let $Y \in \mathcal{Y} \subseteq \mathbb{R}$ denote an outcome variable and let $Z \in \mathcal{Z} \subseteq \mathbb{R}^\ell$ be a vector of covariates. The researcher's interest is often in the regression function defined by

$$m(z) \equiv \operatorname{argmin}_{\tilde{m}} E[\varrho(Y - \tilde{m}) | Z = z], \quad (1)$$

for some loss function $\varrho : \mathbb{R} \rightarrow \mathbb{R}$. For example, m is the conditional mean function of Y given Z when ϱ is the square loss, i.e., $\varrho(\epsilon) = \epsilon^2/2$, while m is the conditional quantile function when $\varrho(\epsilon) = \epsilon(\alpha - 1\{\epsilon \leq 0\})$. Our focus is on estimating the identified features of m when either the outcome variable or one of the covariates is interval-valued. A variable of interest is *interval-valued* when the researcher does not observe the variable of interest W but observes a pair (W_L, W_U) of random variables such that

$$W_L \leq W \leq W_U, \quad \text{with probability 1.} \quad (2)$$

In the presence of an interval-valued variable, data in general do not provide information sufficient for identifying m . Yet, they may provide informative bounds on m . In their pioneering work, Manski and Tamer (2002) derive sharp nonparametric bounds on the conditional mean function when either an outcome or a regressor is interval-valued. Suppose for example that the outcome variable Y is interval-valued. Letting m_L and m_U denote the solutions to (1) with Y_L and Y_U in place of Y respectively and letting $\varrho(\epsilon) = \epsilon^2/2$, the bounds of Manski and Tamer (2002) are given by

$$m_L(Z) \leq m(Z) \leq m_U(Z), \quad \text{with probability 1.} \quad (3)$$

When Y is observed and a component V of the vector of covariates (Z', V) is interval-valued, similar nonparametric bounds can be obtained when the researcher can assume that the regression function is weakly monotonic in V .

Recent developments in the partial identification literature allow us to conduct inference for the identified features of the regression function when inequality restrictions such as (3) are available. For example, when the functional form of m is known up to a finite dimensional parameter, one may construct a confidence set that covers either the identified set of parameters or points inside it with a prescribed probability: (Chernozhukov, Hong, and Tamer, 2007; Andrews and Shi, 2013). One

may also conduct inference for the coefficients of the best linear approximation to the regression function (Beresteanu and Molinari, 2008; Chandrasekhar, Chernozhukov, Molinari, and Schrimpf, 2011). This paper contributes to the literature by studying the estimation of another useful feature of the regression function: the weighted average derivative.

A motivation for studying the weighted average derivative is as follows. A common way to make inference for m is to specify its functional form. For example, one may assume $m(z) = g(z; \gamma_0)$ for some γ_0 , where g is a function known up to a finite dimensional parameter γ . The identified set for γ_0 is then defined as the set of γ 's that satisfy the inequality restrictions: $m_L(Z) \leq g(Z; \gamma) \leq m_U(Z)$ with probability 1. Existing estimation and inference methods for partially identified models can be employed to construct confidence sets for γ_0 or its identified region. However, such inference may be invalid if g is misspecified, a point raised by Ponomareva and Tamer (2011). In contrast, the weighted average derivative is well-defined without functional form assumptions.¹ Suppose m is differentiable with respect to z *a.e.* Letting $w : \mathcal{Z} \rightarrow \mathbb{R}_+$ be a weight function, the *weighted average derivative* of m is defined by

$$\theta \equiv E[w(Z)\nabla_z m(Z)]. \quad (4)$$

Stoker (1986) first analyzed estimation of this parameter. It has also been studied in a variety of empirical studies, including Härdle, Hildenbrand, and Jerison (1991); Deaton and Ng (1998), Coppejans and Sieg (2005), Carneiro, Heckman, and Vytlacil (2010), and Crossley and Pendakur (2010). This parameter allows a simple interpretation: the weighted average of marginal impacts of Z on specific features (e.g., conditional quantiles) of the distribution of Y . Further, under suitable assumptions on the data generating process, it can also serve as a structural parameter associated with the function of interest. For example, if Y is generated as $Y = G(Z) + \epsilon$ with G being a structural function and ϵ being mean independent of Z , the average derivative of the conditional mean summarizes the slope of the structural function G .

In the presence of interval-valued data, the weighted average derivative is generally set identified. This paper's first contribution is to characterize the identified set, the set of weighted average derivatives compatible with the distribution of the observed variables. Specifically, we show that the identified set is compact and convex under mild assumptions. This allows us to represent the identified set by its *support function*, a unique function on the unit sphere that characterizes the location of hyperplanes tangent to the identified set. Support functions have recently been used for making inference for various economic models that involve convex identified sets or have convex predictions (See Beresteanu and Molinari, 2008; Kaido, 2012; Beresteanu, Molchanov, and Molinari, 2011; Chandrasekhar, Chernozhukov, Molinari, and Schrimpf, 2011; Bontemps, Magnac, and Maurin, 2012). Building on the aforementioned studies, we derive a closed form formula for the support function, which in turn gives an explicit characterization of extreme points of the identified set. This characterization also gives closed-form bounds on the weighted average derivative with

¹Another parameter that is also robust to misspecification is the coefficients in the best linear approximation to m . Inference methods for this parameter are studied in Beresteanu and Molinari (2008), Ponomareva and Tamer (2011), and Chandrasekhar, Chernozhukov, Molinari, and Schrimpf (2011).

respect to each covariate.

This paper's second contribution is to characterize the semiparametric efficiency bound for estimating the identified set when the outcome variable is interval-valued. A key insight here is that the support function allows us to interpret the identified set as a parameter taking values in a normed vector space. In recent work, using the theory of semiparametric efficiency for infinite dimensional parameters, Kaido and Santos (forthcoming) characterize the semiparametric efficiency bound for estimating parameter sets defined by convex moment inequalities. Applying their framework, we characterize the semiparametric efficiency bound for the support function of the identified set of the weighted average derivatives. Using mean regression as an example, we further illustrate efficient estimation by showing that an estimator of the identified set for the density weighted average derivative, which builds on Powell, Stock, and Stoker (1989), is semiparametrically efficient. When the interval censoring occurs on a covariate, the nonparametric bounds on the regression function take the form of intersection bounds. We show that the support function of the identified set also depends on these bounds. As pointed out by Hirano and Porter (2009), intersection bounds are not generally pathwise differentiable, which implies that the identified set does not generally admit regular estimation when a covariate is interval-valued. We then discuss a possibility of regular estimation of the support function of another parameter set, which conservatively approximates the true identified set.

This paper is related to the broad literature on semiparametric estimation of weighted average derivatives. For the mean regression function, Stoker (1986) and Härdle and Stoker (1989) study estimation of unweighted average derivatives, while Powell, Stock, and Stoker (1989) study estimation of the density weighted average derivative. Chaudhuri, Doksum, and Samarov (1997) study the weighted average derivative of the quantile regression function. Semiparametric efficiency bounds are shown to exist in these settings. See for example, Samarov (1991), Chaudhuri, Doksum, and Samarov (1997), and Severini and Tripathi (2001). This paper's efficiency results build on Newey and Stoker (1993), who characterize the efficiency bound for the average derivative of general regression functions that are defined through minimizations of various loss functions. The optimal bandwidth for point identified average derivatives is studied in Härdle and Tsybakov (1993) and Powell and Stoker (1996). Other work on semiparametric inference on average derivatives include Nishiyama and Robinson (2000, 2005) who study the higher-order properties of Powell, Stock, and Stoker's (1989) estimator through Edgeworth expansions and Cattaneo, Crump, and Jansson (2010, 2013) who study small bandwidth asymptotics for the estimation of density weighted average derivatives and a robust data driven selector of bandwidth.

The rest of the paper is organized as follows. Section 2 presents the model and characterizes the identified sets. Section 3 gives our main results on the efficiency bounds. Section 4 constructs an efficient estimator of $\Theta_0(P)$ for the mean regression example. We examine the finite sample performance of the estimator in Section 5 and conclude in Section 6.

2 General Setting

Throughout, we let $X \in \mathcal{X} \subset \mathbb{R}^{d_x}$ denote the vector of observables that follows a distribution P . We assume that the observable covariates $Z \in \mathcal{Z} \subseteq \mathbb{R}^\ell$ are continuously distributed and let f denote the probability density function of Z with respect to Lebesgue measure. Suppose that $w(z)f(z)$ vanishes on the boundary of \mathcal{Z} . By integration by parts, Eq. (4) can be equivalently written as

$$\theta = \int m(z)l(z)dP(x), \quad l(z) \equiv -\nabla_z w(z) - w(z)\nabla_z f(z)/f(z). \quad (5)$$

This suggests that the weighted average derivative is a bounded (continuous) linear function of m under mild moment conditions on l . Hence, bounds on m can provide informative bounds on θ . This observation is especially useful when no *a priori* bounds on $\nabla_z m$ are available.²

2.1 Motivating examples

To fix ideas, we briefly discuss examples of regression problems with interval censoring. The first example is based on nonparametric demand analysis (See e.g. Deaton and Ng, 1998).

Example 2.1. Let Y be expenditure on the good of interest. Let Z be a vector of prices of ℓ goods. In survey data, expenditures may be reported as brackets, making Y interval-valued. A key element in the analysis of demand is the effect of a marginal change in the price vector Z on expenditure Y . For example, consider the conditional mean $m(z) \equiv E[Y|Z = z]$ of the demand. The (weighted) average marginal impact of price changes is then measured by $\theta \equiv E[w(Z)\nabla_z m(Z)]$. Similarly, one may also study the average marginal impact of price changes on the conditional median or other conditional quantiles of the demand.

The second example is estimation of a hedonic price model using quantile regression.

Example 2.2. Let Y be the price of a house and Z be a ℓ -dimensional vector of house characteristics. Let V be the distance between a house and another location relevant for the home value (e.g. a school or a factory causing air pollution). If data only record locations by zip codes, one may only obtain an interval-valued measurement $[V_L, V_U]$ of the distance, where V_L and V_U are the minimum and maximum distances between two locations. The researcher's interest may be in the upper tail of the house price, in particular in the weighted average effect of the j -th house characteristic (e.g. square footage) on a specific quantile. Here, the weight function can be chosen so that it puts higher weights on the houses that have specific characteristics the researcher considers relevant. The weighted average effect can be measured by $\theta^{(j)}$, the j -th coordinate of $\theta \equiv E[w(Z)\nabla_z g(Z, v)]$.

²On the other hand, the bounds on m do not generally provide useful bounds on its derivative $\nabla_z m(z)$ evaluated at a point z .

2.2 Identification when Y is interval-valued

Suppose Y is interval-valued. Throughout this section, we let $x \equiv (y_L, y_U, z)'$. Given (5), the sharp identified set for θ is given by

$$\Theta_0(P) \equiv \left\{ \theta \in \Theta : \theta = \int m(z)l(z)dP(x), m \text{ satisfies (1) with } Y_L \leq Y \leq Y_U, P - a.s. \right\}. \quad (6)$$

In order to characterize the identified set, we make the following assumptions on the covariates Z , loss function ϱ , and weight function w .

Assumption 2.1. (i) *The distribution of Z and $l(Z)$ are absolutely continuous with respect to the Lebesgue measure on \mathbb{R}^ℓ . \mathcal{Z} is compact and convex with nonempty interior; (ii) ϱ is convex and satisfies $\varrho(0) = 0$, $\varrho(\epsilon) \geq 0, \forall \epsilon$, and $\varrho(\epsilon) \rightarrow \infty$ as $|\epsilon| \rightarrow \infty$. A measurable function $q(\epsilon) \equiv d\varrho(\epsilon)/d\epsilon$ exists a.e., and q is bounded and continuous a.e.*

Assumption 2.2. (i) *$w(z)$ and $\nabla_z w(z)$ are bounded and continuous on \mathcal{Z} .*

As is standard in the literature, Assumption 2.1 (i) requires Z to be a continuous random vector, where no component of Z can be functionally determined by other components of Z . We also assume that this holds for $l(Z)$. Assumption 2.1 (ii) imposes regularity condition on the derivative q of the loss function, which may require compactness of the supports of $\epsilon_j = Y_j - m_j(Z), j = L, U$ for some specification of ϱ (e.g. square loss) while it is trivially satisfied for the conditional quantile, where $q(\epsilon) = \alpha - 1\{\epsilon \leq 0\}$. We further add the following assumption on P .

Assumption 2.3. (i) *There is a compact set $D \subset \mathbb{R}$ containing the support of Y_j in its interior for $j = L, U$; (ii) $w(z)f(z) = 0$ on the boundary $\partial\mathcal{Z}$ of \mathcal{Z} , $\nabla_z f(z)/f(z)$ is continuous, and $E[\|l(Z)\|^2] < \infty$; (iii) For each $z \in \mathcal{Z}$, $z \mapsto E[q(Y_j - \tilde{m})|Z = z] = 0$ has a unique solution at $\tilde{m} = m_j(z) \in D$ for $j = L, U$; (iv) m_L, m_U are continuously differentiable a.e. with bounded derivatives.*

Assumption 2.3 (ii) is a key condition, which together with other conditions, allows us to write the average derivative as in (5). This, in turn, enables us to obtain bounds on θ from the non-parametric bounds on m . Assumption 2.3 (ii)-(iii) then impose regularity conditions on the density function f and require that the regression functions $m_j, j = L, U$ are well defined. When Assumption 2.3 (iv) holds, we can use it to ensure the sharpness of the identified set.

Under Assumptions 2.1-2.3, $\Theta_0(P)$ is a compact convex set. Hence, it can be uniquely characterized by its support function. Let $\mathbb{S}^\ell = \{p \in \mathbb{R}^\ell : \|p\| = 1\}$ denote the unit sphere in \mathbb{R}^ℓ . For a bounded convex set F , the *support function* of F is defined by

$$v(p, F) \equiv \sup_{x \in F} \langle p, x \rangle. \quad (7)$$

Theorem 2.1 is our first main result, which characterizes the identified set through its support function. Let $\Gamma : \mathbb{R}^3 \rightarrow \mathbb{R}$ be defined pointwise by $\Gamma(w_1, w_2, w_3) = 1\{w_3 \leq 0\}w_1 + 1\{w_3 > 0\}w_2$.

Theorem 2.1. *Suppose Assumptions 2.1, 2.2, and 2.3 (i)-(iii) hold. Suppose further that for each $z \in \mathcal{Z}$, $E[q(Y - \tilde{m})|Z = z] = 0$ has a unique solution at $m(z) \in D$, and m is differentiable a.e. with a bounded derivative. Then, (a) $\Theta_0(P)$ is compact and strictly convex; (b) the support function of $\Theta_0(P)$ is given pointwise by:*

$$v(p, \Theta_0(P)) = \int m_p(z) p'l(z) dP(x), \quad (8)$$

where $m_p(z) = \Gamma(m_L(z), m_U(z), p'l(z)) = 1\{p'l(z) \leq 0\}m_L(z) + 1\{p'l(z) > 0\}m_U(z)$; (c) If Assumption 2.3 (iv) holds, additionally, $\Theta_0(P)$ is sharp.

Theorem 2.1 suggests that the support function is given by the inner product between p and an extreme point $\theta^*(p)$, a unique point such that $\langle p, \theta^*(p) \rangle = v(p, \Theta_0(P))$, which can be expressed as:

$$\theta^*(p) = \int m_p(z) l(z) dP(x), \quad (9)$$

where m_p switches between m_L and m_U depending on the sign of $p'l(z)$. Heuristically, this comes from the fact that the support function of $\Theta_0(P)$ evaluated at p is the maximized value of the map $m \mapsto E[m(Z)'p'l(Z)]$ subject to the constraint $m_L(Z) \leq m(Z) \leq m_U(Z), P - a.s.$ The maximum is then achieved by setting m to m_U when $p'l(z) > 0$ and to m_L otherwise. The form of the support function given in (8) belongs to the general class of functions of the form $E[\Gamma(\delta_L(Z), \delta_U(Z), h(p, Z))h(p, Z)]$ for some functions δ_L, δ_U, h . This functional form is common in the literature on the best linear predictor of m . (See e.g. Stoye, 2007; Beresteanu and Molinari, 2008; Bontemps, Magnac, and Maurin, 2012; Chandrasekhar, Chernozhukov, Molinari, and Schrimpf, 2011).

Theorem 2.1 also gives closed-form bounds on the weighted average derivative with respect to the j -th variable. Let $\theta^{(j)} \equiv E[w(Z)\partial m(Z)/\partial z^{(j)}]$. The upper bound on $\theta^{(j)}$ can be obtained by setting p to ι_j , a vector whose j -th component is 1 and other components are 0. The lower bound can be obtained similarly with $p = -\iota_j$. Therefore, the bounds on $\theta^{(j)}$ are given as $[\theta_L^{(j)}, \theta_U^{(j)}]$ with

$$\theta_L^{(j)} = \int [1\{l^{(j)}(z) > 0\}m_L^{(j)}(z) + 1\{l^{(j)}(z) \leq 0\}m_U^{(j)}(z)]l^{(j)}(z)dP(x) \quad (10)$$

$$\theta_U^{(j)} = \int [1\{l^{(j)}(z) \leq 0\}m_L^{(j)}(z) + 1\{l^{(j)}(z) > 0\}m_U^{(j)}(z)]l^{(j)}(z)dP(x). \quad (11)$$

2.3 Identification when a regressor is interval-valued

We now consider the setting where one of the regressors is interval-valued. Let the vector of covariates be (Z, V) , where Z is fully observed but $V \in \mathcal{V} \subseteq \mathbb{R}$ is unobserved. Suppose that there exists a pair (V_L, V_U) of observables such that $V_L \leq V \leq V_U$ with probability 1. Our interest lies in the average derivative of the regression function defined by:

$$g(z, v) \equiv \operatorname{argmin}_u E[\varrho(Y - u)|Z = z, V = v]. \quad (12)$$

Assuming g is differentiable with respect to z *a.e.*, we define the weighted average derivative pointwise by

$$\theta_v \equiv E[w(Z)\nabla_z g(Z, v)], \quad (13)$$

where the expectation in (13) is with respect to the distribution of Z . θ_v is the average derivative with respect to the observable covariates, fixing V at a given value v . In order to characterize the identified set for θ_v , we make use of the regression function of Y given all observable variables $\tilde{Z} \equiv (Z', V_L, V_U)'$. Specifically, for each (z', v_L, v_U) , define

$$\gamma(z, v_L, v_U) \equiv \operatorname{argmin}_u E[\varrho(Y - u)|Z = z, V_L = v_L, V_U = v_U]. \quad (14)$$

We make the following assumptions to characterize the identified set.

Assumption 2.4. (i) For each $z \in \mathcal{Z}$, $g(z, v)$ is weakly increasing in v . For each $v \in \mathcal{V}$, $g(z, v)$ is differentiable in z with a bounded derivative; (ii) For each $v \in \mathcal{V}$, it holds that

$$E[q(Y - g(Z, V))|\tilde{Z} = \tilde{z}, V = v] = E[q(Y - g(Z, V))|Z = z, V = v]. \quad (15)$$

Following Manski and Tamer (2002), Assumption 2.4 (i) imposes a weak monotonicity assumption on the map $v \mapsto g(z, v)$. Without loss of generality, we here assume that $g(z, \cdot)$ is weakly increasing. Assumption 2.4 (ii) is a conditional mean independence assumption of the “regression residual” $q(Y - g(Z, v))$ from (V_L, V_U) , which means that (V_L, V_U) do not provide any additional information if V is observed. In the case of mean regression, this condition reduces to the mean independence (MI) assumption in Manski and Tamer (2002).

For each v , let $\Xi_L(v) \equiv \{(v_L, v_U) : v_L \leq v_U \leq v\}$ and $\Xi_U(v) \equiv \{(v_L, v_U) : v \leq v_L \leq v_U\}$. Under Assumptions 2.4, one may show that the following functional inequalities hold:

$$g_L(Z, v) \leq g(Z, v) \leq g_U(Z, v), \quad P - a.s., \text{ for all } v, \quad (16)$$

where

$$g_L(z, v) \equiv \sup_{(v_L, v_U) \in \Xi_L(v)} \gamma(z, v_L, v_U), \quad \text{and} \quad g_U(z, v) \equiv \inf_{(v_L, v_U) \in \Xi_U(v)} \gamma(z, v_L, v_U). \quad (17)$$

We then assume the following regularity conditions.

Assumption 2.5. (i) There is a compact set $D \subset \mathbb{R}$ containing the support of Y in its interior. (ii) $E[q(Y - u)|\tilde{Z} = \tilde{z}] = 0$ has a unique solution at $u = \gamma(\tilde{z}) \in D$; (iii) For each $v \in \mathcal{V}$, $g_j(z, v)$ is differentiable in z with a bounded derivative for $j = L, U$.

This assumption is an analog of Assumption 2.3. When Assumption 2.5 (iii) holds, we may use it to ensure the sharpness of the identified set. It requires that the functional bounds $g_j, j = L, U$ are differentiable in z . Since $g_j, j = L, U$ defined in (17) are optimal value functions of parametric optimization problems (indexed by (z, v)), this means that the value functions are assumed to obey an envelope theorem. Various sufficient conditions for such results are known (see for example

Bonnans and Shapiro, 2000; Milgrom and Segal, 2002), but this condition may not hold for some settings, in which case the obtained identified set gives possibly non-sharp bounds on the average derivatives.

Using an argument similar to the one used to establish Theorem 2.1, we now characterize the identified set $\Theta_{0,v}(P)$ for θ_v through its support function.

Theorem 2.2. *Suppose Assumptions 2.1-2.2, 2.3 (ii), 2.4, and 2.5 (i)-(ii) hold. Suppose further that for each $z \in \mathcal{Z}$ and $v \in \mathcal{V}$, $E[q(Y - u)|Z = z, V = v] = 0$ has a unique solution at $u = g(z, v) \in D$. Then, (a) $\Theta_{0,v}(P)$ is compact and strictly convex; (b) its support function is given pointwise by*

$$v(p, \Theta_{0,v}(P)) = \int g_p(z, v) p'l(z) dP(x), \quad (18)$$

where $g_p(z, v) = \Gamma(g_L(z, v), g_U(z, v), p'l(z)) = 1\{p'l(z) \leq 0\}g_L(z, v) + 1\{p'l(z) > 0\}g_U(z, v)$; (c) If, additionally, Assumption 2.5 (iii) holds, $\Theta_{0,v}(P)$ is sharp.

3 Efficiency Bound

In this section, we show that a semiparametric efficiency bound exists for estimation of the support function when Y is interval-valued. Throughout, we assume that observed data $\{X_i\}_{i=1}^n$ are independently and identically distributed (i.i.d.).

3.1 Efficiency bound for an infinite dimensional parameter

We first discuss elements of efficiency analysis.³ Suppose that P is absolutely continuous with respect to a σ -finite measure μ on \mathcal{X} , which we denote by $P \ll \mu$. Let \mathbf{M} denote the set of Borel probability measures on \mathcal{X} and let $\mathbf{M}_\mu \equiv \{P \in \mathbf{M} : P \ll \mu\}$. The set \mathbf{M}_μ then may be identified with a subset of the space L_μ^2 , where:

$$L_\mu^2 \equiv \{s : \mathcal{X} \rightarrow \mathbf{R} : \|s\|_{L_\mu^2} < \infty\}, \quad \|s\|_{L_\mu^2}^2 = \int s^2(x) d\mu(x). \quad (19)$$

via the mapping $s \equiv \sqrt{dP/d\mu}$. A model \mathbf{P} is then a subset of \mathbf{M}_μ . We define curves and the tangent space in the usual manner.

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

The tangent space of a model is then defined as follows.

³The theory of semiparametric efficiency for infinite dimensional parameters is discussed in detail in Chapter 5 in Bickel, Klassen, Ritov, and Wellner (1993). See also Kaido and Santos (forthcoming).

Definition 3.2. For $\mathbf{S} \subseteq L_\mu^2$ and a function $s_0 \in \mathbf{S}$, the tangent set of \mathbf{S} at s_0 is defined as:

$$\dot{\mathbf{S}}^0 \equiv \{\dot{h}_\eta : \eta \mapsto h_\eta \text{ is a curve in } L_\mu^2 \text{ with } h_0 = s_0 \text{ and } h_\eta \in \mathbf{S} \text{ for all } \eta\} . \quad (20)$$

The tangent space of \mathbf{S} at s_0 , denoted by $\dot{\mathbf{S}}$, is the closure of the linear span of $\dot{\mathbf{S}}^0$ (in L_μ^2).

The support function $v(\cdot, \Theta_0(P))$ is a continuous function on the unit sphere. Following Kaido and Santos (2011), we view it as a function-valued parameter taking values in $\mathcal{C}(\mathbb{S}^\ell)$, the set of continuous functions on \mathbb{S}^ℓ . A parameter defined on $\mathcal{C}(\mathbb{S}^\ell)$ is then a mapping $\rho : \mathbf{P} \rightarrow \mathcal{C}(\mathbb{S}^\ell)$ that assigns to each probability measure $Q \in \mathbf{P}$ a corresponding function in $\mathcal{C}(\mathbb{S}^\ell)$. In order to derive a semiparametric efficiency bound for estimating $\rho(P)$, we require that the mapping $\rho : \mathbf{P} \rightarrow \mathcal{C}(\mathbb{S}^\ell)$ 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 $\rho : \mathbf{P} \rightarrow \mathcal{C}(\mathbb{S}^\ell)$ we say ρ is pathwise weak-differentiable at P if there is a continuous linear operator $\dot{\rho} : \dot{\mathbf{S}} \rightarrow \mathcal{C}(\mathbb{S}^\ell)$ such that

$$\lim_{\eta \rightarrow 0} \left| \int_{\mathbb{S}^{d_\theta}} \left\{ \frac{\rho(h_\eta)(p) - \rho(h_0)(p)}{\eta} - \dot{\rho}(h_0)(p) \right\} dB(p) \right| = 0 , \quad (21)$$

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

Given these definitions, we now present the semiparametric efficiency notion for estimating $\rho(P)$. When ρ is pathwise weak-differentiable at P , the Hájek-LeCam convolution theorem (See Bickel, Klassen, Ritov, and Wellner (1993)) implies that any regular estimator of $\rho(P)$ converges in law to \mathbb{G} :

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

where $\stackrel{L}{=}$ denotes equality in law and Δ_0 is a tight Borel measurable random element in $\mathcal{C}(\mathbb{S}^\ell)$ independent of \mathbb{G}_0 .⁴ Thus, a regular estimator may be considered efficient if its asymptotic distribution equals that of \mathbb{G}_0 . We characterize the semiparametric efficiency bound by computing the covariance kernel for the Gaussian process \mathbb{G}_0 , denoted:

$$I^{-1}(p_1, p_2) \equiv \text{Cov}(\mathbb{G}_0(p_1), \mathbb{G}_0(p_2)) \quad (22)$$

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

3.2 Semiparametric efficiency bounds for support functions

In this section, we derive the semiparametric efficiency bound for estimating the support function obtained in Theorem 2.1. Toward this end, we introduce some additional notation and add regularity conditions. For each $z \in \mathcal{Z}$ and $j = L, U$, let

$$r_j(z) \equiv -\frac{d}{d\tilde{m}} E[q(Y_j - \tilde{m}) | Z = z] \Big|_{\tilde{m}=m_j(z)} . \quad (23)$$

⁴ $\{T_n\}$ is regular if there is a tight Borel measurable \mathbb{G} on $\mathcal{C}(\mathbb{S}^\ell)$ such that for every curve $\eta \mapsto h_\eta$ in \mathbf{P} passing through $p_0 \equiv \sqrt{dP_0/d\mu}$ and every $\{\eta_n\}$ with $\eta_n = O(n^{-\frac{1}{2}})$, $\sqrt{n}(T_n - \rho(h_{\eta_n})) \xrightarrow{L_\mathbb{G}} \mathbb{G}$ where L_n is the law under $P_{\eta_n}^n$.

For the mean regression, $r_j(z)$ equals 1, and for the quantile regression, $r_j(z) = f_{Y_j|Z}(m_j(z)|z)$, where $f_{Y_j|Z}$ is the conditional density functions of Y_j given Z for $j = L, U$.

Assumption 3.1. (i) $\mu \in \mathbf{M}$ satisfies $\mu(\{(y_L, y_U, z) : y_L \leq y_U\}) = 1$; (ii) $\mu(\{(y_L, y_U, z) : F(z) = 0\}) = 0$ for any measurable function $F : \mathbb{R}^\ell \rightarrow \mathbb{R}$.

Assumption 3.2. (i) There exists $\bar{\epsilon} > 0$ such that $|r_L(z)| > \bar{\epsilon}$ and $|r_U(z)| > \bar{\epsilon}$ for all $z \in \mathcal{Z}$. (ii) For any $\varphi : \mathcal{X} \rightarrow \mathbb{R}$ that is bounded and continuously differentiable in z with bounded derivatives, $E[\varphi(X)|Z = z]$ is continuously differentiable in z on \mathcal{Z} with bounded derivatives; (iii) $E[q(Y_j - \tilde{m})\varphi(X)|Z = z]$ is continuously differentiable in (z, \tilde{m}) on $\mathcal{Z} \times D$ with bounded derivatives for $j = L, U$.

Since $P \ll \mu$, Assumption 3.1 (i) implies $Y_L \leq Y_U$, $P - a.s.$ Similarly, $P \ll \mu$ and Assumption 3.1 (ii) implies $P(F(Z) = 0) = 0$, which is used to establish pathwise differentiability of the support function. Assumption 3.2 gives additional regularity conditions on P . Assumption 3.2 (i) is trivially satisfied for the conditional mean because $r_L(z) = r_U(z) = 1$. For the conditional α -quantile, Assumption 3.2 (i) requires the conditional densities of Y_L and Y_U to be positive on neighborhoods of $m_L(z)$ and $m_U(z)$ respectively. Assumption 3.2 (ii)-(iii) are regularity conditions invoked in Newey and Stoker (1993), which we also impose here. It should be noted that these conditions exclude the setting where either Y_L or Y_U is discrete and q admits a point of discontinuity on the support of ϵ_L or ϵ_U .

Given these assumptions, we now define our model as the set of distributions that satisfy Assumptions 2.3 and 3.2:

$$\mathbf{P} \equiv \{P \in \mathbf{M} : P \ll \mu, P \text{ satisfies Assumptions 2.3 and 3.2}\}. \quad (24)$$

For each $Q \in \mathbf{P}$, define $\rho(Q) \equiv v(\cdot, \Theta_0(Q))$. The following theorem characterizes the efficiency bound for the support function.

Theorem 3.1. *Suppose Assumptions 2.1-2.2, and 3.1 hold, and suppose $P \in \mathbf{P}$. Then, the inverse information covariance functional is given by*

$$I^{-1}(p_1, p_2) = E[\psi_{p_1}(X)\psi_{p_2}(X)] , \quad (25)$$

where for each $p \in \mathbb{S}^\ell$, the efficient influence function ψ_p is

$$\psi_p(x) \equiv w(z)p'\nabla_z m_p(z) - v(p, \Theta_0(P_0)) + p'l(z)\zeta_p(x), \quad (26)$$

and $\nabla_z m_p$ and ζ_p are given by $\nabla_z m_p(z) = \Gamma(\nabla_z m_L(z), \nabla_z m_U(z), p'l(z))$ and $\zeta_p(x) = \Gamma(r_L^{-1}(z)q(y_L - m_L(z)), r_U^{-1}(z)q(y_U - m_U(z)), p'l(z))$.

Theorem 3.1 naturally extends Theorem 3.1 in Newey and Stoker (1993) to the current setting. When there is no interval censoring, i.e. $m_L(Z) = m(Z) = m_U(Z)$, the obtained semiparametric efficiency bound reduces to that of Newey and Stoker (1993), i.e. $\psi_p = p'\psi$, where ψ is the efficient

influence function for point identified θ . Theorem 3.1 also shows that the variance bound for estimating the support function $v(p, \Theta_0(P))$ at p is given by

$$E[|\psi_p(X)|^2] = \text{Var}(w(Z)p'\nabla_z m_p(Z)) + E[|p'l(Z)\zeta_p(x)|^2]. \quad (27)$$

The first term in (27) can be interpreted as the variance bound when m_p is known but f is unknown as this is the asymptotic variance of $\frac{1}{n} \sum_{i=1}^n w(Z_i)p'\nabla_z m_p(Z_i)$, while the second term can be interpreted as the variance bound when f is known but m_p is unknown. (See Newey and Stoker, 1993, page 1205 for a more detailed discussion.)

Theorem 3.1 establishes that the support function of the identified set has a finite efficiency bound. In the next section, we show that it is possible to construct an estimator that achieves this bound in a leading example. Efficient estimation of the support function also has an important consequence on estimation of the identified set. Namely, an estimator of the identified set which is constructed from the efficient estimator of the support function is also asymptotically optimal for a wide class of loss functions based on the Hausdorff distance. (See Remark 4.1.)

Remark 3.1. Powell, Stock, and Stoker (1989) study the setting where m is the conditional mean, and the weight function is the *density weight*: $w(z) = f(z)$. The efficiency bound in Theorem 3.1 can be extended to this setting. For this choice of the weight function, the efficient influence function differs slightly from Eq. (26) due to f being unknown. Taking into account the pathwise derivative of unknown f , one can show that the inverse covariance functional for this case is given as in (25) with

$$\psi_p(x) \equiv 2\{f(z)p'\nabla_z m_p(z) - v(p, \Theta_0(P))\} - 2p'\nabla_z f(z)(y_p - m_p(z)), \quad (28)$$

where $y_p = \Gamma(y_L, y_U, p'l(z))$.

Remark 3.2. For the setting where an explanatory variable is interval-valued, Theorem 2.2 shows that the support function of $\Theta_{0,v}(P)$ involves functions that are defined as the supremum (or the infimum) of functions indexed by (v_L, v_U) , e.g. $g_L(z, v) = \sup_{(v_L, v_U) \in \Xi_L(v)} \gamma(z, v_L, v_U)$. These types of bounds are known as the *intersection bounds* (Chernozhukov, Lee, and Rosen, 2013). In particular, for parametric submodels $\eta \mapsto P_\eta$ passing through P , one may show that the support function depends on the intersection bounds in the following way:

$$v(p, \Theta_{0,v}(P_\eta)) = \int [1\{p'l_\eta(z) \leq 0\}g_{L,\eta}(z, v) + 1\{p'l_\eta(z) > 0\}g_{U,\eta}(z, v)]p'l_\eta(z)dP_\eta(x), \quad (29)$$

where l_η is defined as in (5) and $g_{L,\eta}, g_{U,\eta}$ are defined as in (17) under P_η . Hirano and Porter (2012) give general conditions under which intersection bounds are not pathwise differentiable therefore do not admit regular estimation. When the set of z 's on which $g_{L,\eta}$ or $g_{U,\eta}$ is pathwise non-differentiable has a positive probability mass, the support function is pathwise non-differentiable as well. Hence, $v(p, \Theta_{0,v}(P))$ does not generally admit regular estimation. Therefore, for optimal inference on $v(p, \Theta_{0,v}(P))$, an alternative optimal criterion would be needed. (See for example Song, 2010; Chernozhukov, Lee, and Rosen, 2013).

There is, however, a possibility for regular estimation of a function that approximates $v(p, \Theta_{0,v}(P))$. For simplicity, suppose that V_L and V_U have discrete supports so that $\Xi_L(v)$ and $\Xi_U(v)$ are finite sets. Then for a given $\kappa > 0$, define

$$\mathbf{g}_L(z, v; \kappa) \equiv \sum_{(v_L, v_U) \in \Xi_L(v)} \gamma(z, v_L, v_U) \frac{\exp(\kappa\gamma(z, v_L, v_U))}{\sum_{(v_L, v_U) \in \Xi_L(v)} \exp(\kappa\gamma(z, v_L, v_U))} \quad (30)$$

$$\mathbf{g}_U(z, v; \kappa) \equiv \sum_{(v_L, v_U) \in \Xi_U(v)} \gamma(z, v_L, v_U) \frac{\exp(-\kappa\gamma(z, v_L, v_U))}{\sum_{(v_L, v_U) \in \Xi_U(v)} \exp(-\kappa\gamma(z, v_L, v_U))}, \quad (31)$$

where the smooth weighted averages on the right hand side of the equations above conservatively approximate the maximum and minimum respectively, where the approximation errors are inversely proportional to κ (Chernozhukov, Kocatulum, and Menzel, 2012). Suppose that the researcher chooses a fixed $\kappa > 0$. Define $\mathbf{u}(p; \kappa) \equiv \int \mathbf{g}_p(z, v; \kappa) p'l(z) dP(x)$, where $\mathbf{g}_p(z, v; \kappa) = \Gamma(\mathbf{g}_L(z, v; \kappa), \mathbf{g}_U(z, v; \kappa), p'l(z))$. $\mathbf{u}(p; \kappa)$ is then a conservative approximation of the support function $v(p, \Theta_{0,v}(P))$ whose approximation bias can be bounded as follows:

$$|\mathbf{u}(p; \kappa) - v(p, \Theta_{0,v}(P))| \leq CE[|l(Z)|^2]\kappa^{-2} \quad \text{uniformly in } p \in \mathbb{S}^\ell, \quad (32)$$

where C is a positive constant that depends on the cardinality of the support of (V_L, V_U) . Note that $\mathbf{u}(p; \kappa)$ depends smoothly on the underlying distribution. This is because, as oppose to the maximum and minimum, the smooth weighted averages in (30)-(31) relate $\mathbf{g}_L, \mathbf{g}_U$ and γ in a differentiable manner. This suggests that, although regular estimation of $v(p, \Theta_{0,v}(P))$ is not generally possible, it may be possible to estimate $\mathbf{u}(p; \kappa)$ in a regular manner, which we leave as future work.

4 Estimation of Density Weighted Average Derivative in Mean Regression

In this section, we illustrate efficient estimation by studying a well-known example. We focus on the case where Y is interval-valued, and the parameter of interest is the density weighted average derivative of the mean regression function as in Powell, Stock, and Stoker (1989). That is, $\theta = E[f(Z)\nabla_z m(Z)]$, where m is the conditional mean of Y . The density weight is attractive for practical purposes as it does not require the choice of a trimming parameter to control for the stochastic denominator $\hat{f}_{i,h}$ defined below, which would otherwise arise upon estimating the ‘‘score function’’ l .⁵

Theorem 2.1 and the law of iterated expectations imply that the support function of the identified

⁵Powell, Stock, and Stoker (1989) are interested in estimating index coefficients up to scale in a single index model. The scale normalization implicit in the density weighted average derivative may not yield most easily interpretable estimates. For this reason, Powell, Stock, and Stoker (1989) also suggest estimating the rescaled parameter $\tilde{\theta} \equiv \theta/E[f(Z)]$. We consider an estimator of the identified set of θ in Section 5.

set in this setting is given by

$$v(p, \Theta_0(P)) = E[Y_p p' l(Z)], \quad (33)$$

where $Y_p = \Gamma(Y_L, Y_U, p'l(Z))$ and $l(z) = -2\nabla_z f(z)$. Our estimator of the support function is a natural extension of Powell, Stock, and Stoker (1989) and replaces unknown objects in (33) with nonparametric kernel estimators and expectations with sample averages. Let $K : \mathbb{R}^\ell \rightarrow \mathbb{R}$ be a kernel function. For each $z \in \mathcal{Z}$, i and bandwidth h , define the “leave-one-out” kernel density estimator by

$$\hat{f}_{i,h}(z) \equiv \frac{1}{(n-1)h^\ell} \sum_{j=1, j \neq i}^n K\left(\frac{z - Z_j}{h}\right). \quad (34)$$

Our estimator of l is then defined by $\hat{l}_{i,h}(z) \equiv -2\nabla_z \hat{f}_{i,h}(z)$. The support function of $\Theta_0(P)$ is then estimated by

$$\hat{v}_n(p) \equiv \frac{1}{n} \sum_{i=1}^n p' \hat{l}_{i,h}(Z_i) \hat{Y}_{p,i}, \quad (35)$$

where $\hat{Y}_{p,i}$ is an estimator of $Y_{p,i}$, which is not observed. For this, we let $\hat{Y}_{p,i} = \Gamma(Y_{L,i}, Y_{U,i}, p' \hat{l}_{i,\tilde{h}})$, where \tilde{h} is another bandwidth parameter. Computing the estimator in (35) only involves kernel density estimation and taking averages. Hence, it can be implemented quite easily. When the researcher is only interested in the average derivative with respect to a particular variable, the required computation simplifies further. For example, suppose the parameter of interest is the average derivative $\theta^{(j)}$ with respect to the j -th variable. An estimate of the upper bound on $\theta^{(j)}$ can be obtained by computing the support function in (35) only for one direction $p = \iota_j$, i.e., $\hat{\theta}_{U,n}^{(j)} = \hat{v}_n(\iota_j)$. The lower bound can be computed similarly with $p = -\iota_j$.

We now add regularity conditions required for efficient estimation of the support function. Let $J \equiv (\ell + 4)/2$ if ℓ is even and $J \equiv (\ell + 3)/2$ if ℓ is odd.

Assumption 4.1. (i) *There exists $M : \mathcal{Z} \rightarrow \mathbb{R}_+$ such that*

$$\|\nabla_z f(z+e) - \nabla_z f(z)\| < M(z)\|e\| \quad (36)$$

$$\|\nabla_z(f(z+e) \times m_j(z+e)) - \nabla_z(f(z) \times m_j(z))\| < M(z)\|e\|, \quad j = L, U, \quad (37)$$

and $E[|M(Z)|^2] < \infty$. (ii) *All partial derivatives of f of order $J+1$ exist. $E[Y_L(\partial^k f / \partial z^{(j_1)} \dots \partial z^{(j_k)})]$ and $E[Y_U(\partial^k f / \partial z^{(j_1)} \dots \partial z^{(j_k)})]$ exist for all $k \leq J+1$.*

Assumption 4.2. (i) *The support \mathcal{S}_K of K is a convex subset of \mathbb{R}^ℓ with nonempty interior with the origin as an interior point. Let $\partial\mathcal{S}_K$ be the boundary of \mathcal{S}_K ; (ii) K is a bounded, continuously differentiable function with bounded derivatives, and $\int K(u)du = 1$, $\int uK(u)du = 0$; (iii) $K(u) = 0$*

for all $u \in \partial\mathcal{S}_K$; (iv) $K(u) = K(-u)$ for all $u \in \mathcal{S}_K$. (v) $K(u)$ is of order J :

$$\int u^{j_1} u^{j_2} \cdots u^{j_k} K(u) du = 0, \quad j_1 + \cdots + j_k < J \quad (38)$$

$$\int u^{j_1} u^{j_2} \cdots u^{j_k} K(u) du \neq 0, \quad j_1 + \cdots + j_k = J. \quad (39)$$

(vi) The J moments of K exist.

Assumptions 4.1 and 4.2 are based on the assumptions in Powell, Stock, and Stoker (1989). Assumption 4.1 imposes suitable smoothness conditions on f , m_L and m_U . Assumption 4.2 then gives standard regularity conditions on the kernel. A higher-order kernel is used to remove an asymptotic bias. With these additional assumptions, the next theorem establishes the asymptotic efficiency of the estimator.

Theorem 4.1. *Suppose the conditions of Theorem 3.1 hold and Assumptions 4.1 and 4.2 hold. Suppose further that $h \rightarrow 0$, $nh^{\ell+2+\delta} \rightarrow \infty$ for some $\delta > 0$, $nh^{2J} \rightarrow 0$, $\tilde{h} \rightarrow 0$, and $n\tilde{h}^{4(\ell+1)} \rightarrow \infty$ as $n \rightarrow \infty$. Then, (a) $\{\hat{v}_n(\cdot)\}$ is a regular estimator for $v(\cdot, \Theta_0(P))$; (b) Uniformly in $p \in \mathbb{S}^\ell$:*

$$\begin{aligned} \sqrt{n}\{\hat{v}_n(p) - v(p, \Theta_0(P))\} &= \frac{2}{\sqrt{n}} \sum_{i=1}^n [f(Z_i) p' \nabla_z m_p(Z_i) - v(p, \Theta_0(P))] \\ &\quad - \frac{2}{\sqrt{n}} \sum_{i=1}^n p' \nabla_z f(Z_i) (Y_{p,i} - m_p(Z_i)) + o_p(1); \end{aligned} \quad (40)$$

(c) As a process in $\mathcal{C}(\mathbb{S}^\ell)$,

$$\sqrt{n}\{\hat{v}_n(\cdot) - v(\cdot, \Theta_0(P))\} \xrightarrow{L} \mathbb{G}_0, \quad (41)$$

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

Remark 4.1. Each extreme point $\theta^*(p)$ can be estimated by its sample analog estimator, $\hat{\theta}_n(p) = \frac{1}{n} \sum_{i=1}^n \hat{l}_{i,n}(Z_i) \hat{Y}_{p,i}$. Using this estimator, we may also define an estimator of the identified set as follows:

$$\hat{\Theta}_n \equiv \text{co}(\{\theta \in \Theta : \theta = \hat{\theta}_n(p), p \in \mathbb{S}^\ell\}), \quad (42)$$

where $\text{co}(A)$ denotes the convex hull of A . Theorem 4.1 has direct consequences on the consistency and asymptotic optimality of this set estimator. For any two compact convex sets A, B , let the Hausdorff distance be $d_H(A, B) \equiv \max\{\sup_{a \in A} \inf_{b \in B} \|a - b\|, \sup_{b \in B} \inf_{a \in A} \|a - b\|\}$. Due to the equality of the Hausdorff distance between sets and the supremum distance between the corresponding support functions $d_H(A, B) = \sup_{p \in \mathbb{S}^\ell} |v(p, A) - v(p, B)|$ for any compact convex sets A, B , we have $d_H(\hat{\Theta}_n, \Theta_0(P)) = O_p(n^{-1/2})$ by Hörmander's embedding theorem (Li, Ogura, and Kreinovich, 2002). Further, Kaido and Santos (forthcoming) show that if $\hat{\Theta}_n$ is associated with the efficient estimator of the support function and $L : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is a subconvex continuous function, it holds under regularity conditions that

$$\liminf_{n \rightarrow \infty} E[L(\sqrt{n}d_H(C_n, \Theta_0(P)))] \geq \limsup_{n \rightarrow \infty} E[L(\sqrt{n}d_H(\hat{\Theta}_n, \Theta_0(P)))] = E[L(\|\mathbb{G}_0\|_\infty)], \quad (43)$$

for any regular convex compact valued set estimator $\{C_n\}$ for $\Theta_0(P)$. That is, among regular set estimators, $\hat{\Theta}_n$ asymptotically minimizes a wide class of estimation risks based on the Hausdorff distance.

Remark 4.2. Efficient estimators of the support function can also be used to conduct inference for $\Theta_0(P)$ and points inside it using the score-based weighted bootstrap of Kline and Santos (2012). Specifically, let W_i be a mean zero random scalar with variance 1 and let $\{W_i\}$ be a sample independent of $\{X_i\}_{i=1}^n$. For each $p \in \mathbb{S}^\ell$, define the process:

$$G_n^*(p) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^n W_i \left\{ \sum_{j=1, j \neq i}^n \frac{-2}{(n-1)h^{\ell+1}} \nabla_z K \left(\frac{Z_i - Z_j}{h} \right) (\hat{Y}_{p,i} - \hat{Y}_{p,j}) - \hat{v}_n(p) \right\}, \quad (44)$$

where the process G_n^* is a U -process which is first-order asymptotically equivalent to the process $\frac{1}{\sqrt{n}} \sum_{i=1}^n W_i \psi_p(X_i)$.⁶ In practice, the distribution of G_n^* can be simulated by generating random samples of $\{W_i\}$, which weakly converges to \mathbb{G}_0 conditional on $\{X_i, i = 1, \dots, n\}$. This distribution, in turn, can be used to make inference. For example, a level $1 - \alpha$ one-sided confidence set as in Beresteanu and Molinari (2008) can be constructed as $\mathcal{C}_{1n} \equiv \hat{\Theta}_n^{c_{1n}^*/\sqrt{n}}$, where $\hat{\Theta}_n^\epsilon \equiv \{\theta \in \Theta : \inf_{\theta' \in \hat{\Theta}_n} \|\theta - \theta'\| \leq \epsilon\}$ and c_{1n}^* is the $1 - \alpha$ quantile of $\sup_{p \in \mathbb{S}^\ell} \{-G_n^*(p)\}_+$ (see also Kaido, 2012; Kaido and Santos, forthcoming).

Remark 4.3. One may also construct a confidence set for a particular coordinate $\theta^{(j)}$ of θ or its identified set $\Theta_0^{(j)}(P)$. For example, a symmetric confidence set for $\Theta_0^{(j)}(P)$ can be constructed as

$$\mathcal{C}_n^{(j)} \equiv [\hat{\theta}_{L,n}^{(j)} - c_n^{(j)}/\sqrt{n}, \hat{\theta}_{U,n}^{(j)} + c_n^{(j)}/\sqrt{n}],$$

where $c_n^{(j)}$ is the $1 - \alpha$ quantile of $\max_{p \in \{\iota_j, -\iota_j\}} \{-G_n^*(p)\}_+$.

5 Simulation evidence

In this section, we examine the finite sample performance of an estimator of the support function through Monte Carlo experiments. Throughout, we let $Z_i \equiv (Z_{1,i}, Z_{2,i}, Z_{3,i})'$, where $Z_{1,i} = 1$ is a constant, and $Z_{2,i}$ and $Z_{3,i}$ are continuously distributed. For $\beta = (1, 1)'$, we generate:

$$Y_i = Z_i' \beta + \epsilon_i \quad i = 1, \dots, n, \quad (45)$$

where ϵ_i is a standard normal random variable independent of Z_i . We then generate $(Y_{L,i}, Y_{U,i})$ as:

$$\begin{aligned} Y_{L,i} &= Y_i - c - e_{2i} Z_{2i}^2 - e_{3i} Z_{3i}^2 \\ Y_{U,i} &= Y_i + c + e_{2i} Z_{2i}^2 + e_{3i} Z_{3i}^2, \end{aligned} \quad (46)$$

where $c > 0$ and e_{2i} and e_{3i} are independently uniformly distributed on $[0, 0.2]$ independently of (Y_i, Z_i) . Here, c is a design parameter that controls the diameter of the identified set. The identified sets under three different values of c are plotted in Figure 1.

⁶This can be shown following an argument similar to the one in Section 3.4 in Powell, Stock, and Stoker (1989). The proof is omitted for brevity.

We report estimation results for two different estimators of the support function. Since scale normalization implicit in θ may not allow a simple interpretation of estimation results, we follow Powell, Stock, and Stoker (1989) and renormalize the weighted average derivative as follows:

$$\tilde{\theta} \equiv E[f(Z)]^{-1} E[f(Z) \nabla_z m(Z)]. \quad (47)$$

Integrating by parts, it holds that $I_\ell E[f(Z)] = E[\nabla_z Z f(Z)] = E[Z l(Z)]$, where I_ℓ is the identity matrix of dimension ℓ . Thus, $\tilde{\theta}$ can be rewritten as $\tilde{\theta} = E[l(Z) Z]^{-1} E[l(Z) m(Z)]$. Our first estimator of the support function applies this renormalization to the sample counterpart and is defined by

$$\hat{v}_n^{IV} \equiv p' \left(\frac{1}{n} \sum_{i=1}^n \hat{l}_{i,h}(Z_i) Z_i \right)^{-1} \frac{1}{n} \sum_{i=1}^n \hat{l}_{i,h}(Z_i) Y_{p,i}, \quad (48)$$

where $\hat{l}_{i,h}$ uses a Gaussian kernel. This estimator may be interpreted as the inner product between p and a boundary point estimated by an instrumental variable (IV) estimator, which regresses $Y_{p,i}$ on Z_i using $\hat{l}_{i,h}$ as an instrument. Our second estimator replaces the Gaussian kernel with a higher order kernel.⁷

Tables 1-2 report the average losses of these estimators, measured in the Hausdorff distance measures: $R_H \equiv E[d_H(\hat{\Theta}_n, \Theta_0(P))]$, $R_{IH} \equiv E[\vec{d}_H(\hat{\Theta}_n, \Theta_0(P))]$ and $R_{OH} \equiv E[\vec{d}_H(\Theta_0(P), \hat{\Theta}_n)]$. We call them the *Hausdorff risk*, *inner Hausdorff risk* and *outer Hausdorff risk* respectively. The directed Hausdorff distance \vec{d}_H is defined by $\vec{d}_H(A, B) \equiv \sup_{a \in A} \inf_{b \in B} \|a - b\|$, which has the property that $\vec{d}_H(\Theta_0(P), \hat{\Theta}_n) = 0$ if $\Theta_0(P) \subseteq \hat{\Theta}_n$ but takes a positive value otherwise. Hence, R_{OH} penalizes $\hat{\Theta}_n$ when it is a “small set” that does not cover $\Theta_0(P)$. On the other hand, R_{IH} penalizes $\hat{\Theta}_n$ when it is a “large set” that does not fit inside $\Theta_0(P)$. The Hausdorff risk R_H then penalizes $\hat{\Theta}_n$ for both types of deviations from $\Theta_0(P)$.

Table 1 reports R_H , R_{IH} , and R_{OH} for the first estimator under different values of c , h , and n . Throughout simulations, we have set $h = \tilde{h}$ for simplicity. One observation is that, for any value of n when $c = 0.5$ or 1 , R_{IH} is increasing in h , which suggests that a larger bandwidth (oversmoothing) may introduce an outward bias to the set estimator. This is consistent with the outer hausdorff risk R_{OH} being decreasing in h when identified sets are relatively large ($c = 0.5, 1$). However, R_{IH} is not increasing in h when the identified set is small ($c = 0.1$) suggesting that there may be different sources of risk that could affect R_{IH} in this setting. For example, even if one uses a small bandwidth and the estimated set itself $\hat{\Theta}_n$ is small, its location may still be biased so that it does not stay inside $\Theta_0(P)$. The Hausdorff risk R_H takes both errors into account and seems to have a well-defined minimum as a function of the bandwidth. For example, when $c = 1$ and $n = 1,000$, the Hausdorff risk is minimized when the bandwidth is about 0.6.

Table 2 reports results for the bias-corrected (second) estimator. Again, for $c = 0.5$ and 1 , R_{IH} is increasing in h , and R_{OH} is decreasing in h , which suggests an outward bias with oversmoothing, but this tendency is not clear when the identified region is relatively small ($c = 0.1$). We also note

⁷Detailed description of the construction of the higher-order kernels is in Powell, Stock, and Stoker (1989) Appendix 2.

that the bias correction through the higher-order kernel improves the lowest Hausdorff risk but not in a significant manner. In sum, the simulation results show a tradeoff between the inner and outer Hausdorff risks. The optimal bandwidth in terms of the Hausdorff risk seems to exist, which makes these two risks roughly of the same order.

6 Concluding remarks

This paper studies the identification and estimation of weighted average derivatives in the presence of interval censoring on either an outcome or on a covariate. We show that the identified set of average derivatives is compact and convex under general assumptions and further show that it can be represented by its support function. This representation is used to characterize the semiparametric efficiency bound for estimating the identified set when the outcome variable is interval-valued. For mean regression with an interval censored outcome, we construct a semiparametrically efficient set estimator.

For practical purposes, an important avenue for future research is to develop a theory of optimal bandwidth choice. The simulation results suggest the Hausdorff risks vary with the choice of bandwidth. It is an open question how to trade off different types of biases (inward, outward, and shift) and variance. Another interesting direction for future research would be to study the higher order properties of first-order efficient estimators, which would require an asymptotic expansion as in Nishiyama and Robinson (2000) extended to the context of interval censoring.

References

- ALIPRANTIS, C. D., AND K. C. BORDER (2006): *Infinite Dimensional Analysis – A Hitchhiker’s Guide*. Springer-Verlag, Berlin.
- ANDREWS, D. W. K., AND X. SHI (2013): “Inference based on conditional moment inequalities,” *Econometrica*, 81(2), 609–666.
- BERESTEANU, A., I. MOLCHANOV, AND F. MOLINARI (2011): “Sharp Identification Regions in Models with Convex Predictions,” *Econometrica*, 79(6).
- BERESTEANU, A., AND F. MOLINARI (2008): “Asymptotic Properties for a Class of Partially Identified Models,” *Econometrica*, 76(4), 763–814.
- BICKEL, P. J., C. A. KLASSEN, Y. RITOV, AND J. A. WELLNER (1993): *Efficient and Adaptive Estimation for Semiparametric Models*. Springer, New York.
- BONNANS, J. F., AND A. SHAPIRO (2000): *Perturbation analysis of optimization problems*. Springer.

- BONTEMPS, C., T. MAGNAC, AND E. MAURIN (2012): “Set Identified Linear Models,” *Econometrica*, 80(3), 1129–1155.
- CARNEIRO, P., J. J. HECKMAN, AND E. VYTLACIL (2010): “Evaluating marginal policy changes and the average effect of treatment for individuals at the margin,” *Econometrica*, 78(1), 377–394.
- CATTANEO, M. D., R. K. CRUMP, AND M. JANSSON (2010): “Robust Data-Driven Inference for Density-Weighted Average Derivatives,” *J Am Stat Assoc*, 105(491), 1070–1083.
- CATTANEO, M. D., R. K. CRUMP, AND M. JANSSON (2013): “Small Bandwidth Asymptotics for Density-Weighted Average Derivatives,” *Econometric Theory*, pp. 1–25.
- CHANDRASEKHAR, A., V. CHERNOZHUKOV, F. MOLINARI, AND P. SCHRIMPF (2011): “Inference for Best Linear Approximations to Set Identified Functions,” Discussion Paper, University of British Columbia.
- CHAUDHURI, P., K. DOKSUM, AND A. SAMAROV (1997): “On average derivative quantile regression,” *Ann. Statist.*, pp. 715–744.
- CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2007): “Estimation and Confidence Regions for Parameter Sets in Econometric Models,” *Econometrica*, 75(5), 1243–1284.
- CHERNOZHUKOV, V., E. KOCATULUM, AND K. MENZEL (2012): “Inference on sets in finance,” *arXiv preprint arXiv:1211.4282*.
- CHERNOZHUKOV, V., S. LEE, AND A. ROSEN (2013): “Intersection bounds: Estimation and Inference,” *Econometrica*, 81(2), 667–737.
- COPPEJANS, M., AND H. SIEG (2005): “Kernel estimation of average derivatives and differences,” *Journal of Business and Economic Statistics*, 23(2), 211–225.
- CROSSLEY, T. F., AND K. M. PENDAKUR (2010): “The Common-Scaling Social Cost of Living Index,” *Journal of Business and Economic Statistics*, 28(4), 523–538.
- DEATON, A., AND S. NG (1998): “Parametric and Nonparametric Approaches to Price and Tax Reform,” *Journal of the American Statistical Association*, 93(443), 900–909.
- HÄRDLE, W., W. HILDENBRAND, AND M. JERISON (1991): “Empirical Evidence on the Law of Demand,” *Econometrica*, 59(6), pp. 1525–1549.
- HÄRDLE, W., AND T. STOKER (1989): “Investigating smooth multiple regression by the method of average derivatives,” *Journal of the American Statistical Association*, pp. 986–995.
- HÄRDLE, W., AND A. TSYBAKOV (1993): “How sensitive are average derivatives?,” *Journal of Econometrics*, 58(1-2), 31–48.

- HIRANO, K., AND J. PORTER (2009): “Asymptotics for statistical treatment rules,” *Econometrica*, 77(5), 1683–1701.
- HIRANO, K., AND J. R. PORTER (2012): “Impossibility results for nondifferentiable functionals,” *Econometrica*, 80(4), 1769–1790.
- KAIDO, H. (2012): “A Dual Approach to Inference for Partially Identified Models,” Discussion Paper, Boston University.
- KAIDO, H., AND A. SANTOS (forthcoming): “Asymptotically Efficient Estimation of Models Defined by Convex Moment Inequalities,” *Econometrica*.
- KLINE, P., AND A. SANTOS (2012): “A Score Based Approach to Wild Bootstrap Inference,” *Journal of Econometric Methods*, 1(1), 23–41, NBER Working Paper 16127.
- KOSOROK, M. R. (2008): *Introduction to Empirical Processes and Semiparametric Inference*. Springer, New York.
- LI, S., Y. OGURA, AND V. KREINOVICH (2002): *Limit Theorems and Applications of Set-Valued and Fuzzy Set-Valued Random Variables (Theory and Decision Library B)*. Berlin: Springer.
- MANSKI, C. F., AND E. TAMER (2002): “Inference on Regressions with Interval Data on a Regressor or Outcome,” *Econometrica*, 70(2), 519–546.
- MILGROM, P., AND I. SEGAL (2002): “Envelope Theorems for Arbitrary Choice Sets,” *Econometrica*, 70(2), 583–601.
- NEWHEY, W. (1991): “Efficient estimation of Tobit models under conditional symmetry,” *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, pp. 291–336.
- NEWHEY, W., AND T. STOKER (1993): “Efficiency of weighted average derivative estimators and index models,” *Econometrica: Journal of the Econometric Society*.
- NISHIYAMA, Y., AND P. ROBINSON (2000): “Edgeworth expansions for semiparametric averaged derivatives,” *Econometrica*, pp. 931–979.
- (2005): “The Bootstrap and the Edgeworth Correction for Semiparametric Averaged Derivatives*,” *Econometrica*, 73(3), 903–948.
- PAKES, A., AND D. POLLARD (1989): “Simulation and the asymptotics of optimization estimators,” *Econometrica: Journal of the Econometric Society*, pp. 1027–1057.
- PONOMAREVA, M., AND E. TAMER (2011): “Misspecification in moment inequality models: Back to moment equalities?,” *The Econometrics Journal*, 14(2), 186–203.
- POWELL, J., J. STOCK, AND T. STOKER (1989): “Semiparametric estimation of index coefficients,” *Econometrica: Journal of the Econometric Society*, pp. 1403–1430.

- POWELL, J., AND T. STOKER (1996): “Optimal bandwidth choice for density-weighted averages,” *Journal of Econometrics*, 75(2), 291–316.
- ROCKAFELLAR, R. T., AND R. J.-B. WETS (2005): *Variational Analysis*. Berlin: Springer, 2nd edn.
- SAMAROV, A. (1991): “On asymptotic efficiency of average derivative estimates,” *Nonparametric functional estimation and related topics*, 335, 167.
- SCHNEIDER, R. (1993): *Convex bodies: the Brunn-Minkowski theory*, vol. 44. Cambridge Univ Pr.
- SEVERINI, T., AND G. TRIPATHI (2001): “A simplified approach to computing efficiency bounds in semiparametric models,” *Journal of Econometrics*, 102(1), 23–66.
- SHERMAN, R. P. (1994a): “Maximal inequalities for degenerate U-processes with applications to optimization estimators,” *The Annals of Statistics*, pp. 439–459.
- (1994b): “U-processes in the analysis of a generalized semiparametric regression estimator,” *Econometric Theory*, 10, 372–372.
- SONG, K. (2010): “Robust estimation of some nonregular parameters,” Discussion paper, Penn Institute for Economic Research, Department of Economics, University of Pennsylvania.
- STOKER, T. (1986): “Consistent estimation of scaled coefficients,” *Econometrica: Journal of the Econometric Society*.
- STOYE, J. (2007): “Bounds on Generalized Linear Predictors wiith Incomplete Outcome Data,” *Reliable Computing*, 13(3), 293–302.
- VAN DER VAART, A. W., AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes: with Applications to Statistics*. Springer, New York.

SUPPLEMENTAL APPENDIX

In this supplemental appendix, we include the proofs of results stated in the main text. The contents of the supplemental appendix are organized as follows. Appendix A contains notations and definitions used throughout the appendix. Appendix B contains the proof of Theorems 2.1 and 2.2. Appendix C contains the proof of Theorem 3.1 and auxiliary lemmas. Appendix D contains the proof of Theorem 4.1. Appendix E then reports the Monte Carlo results.

APPENDIX A: Notation and Definitions

Let $\Pi : \mathcal{X} \rightarrow \mathcal{Z}$ be the projection map pointwise defined by $x = (y_L, y_U, z) \mapsto z$. Let $\nu = \Pi_{\#}\mu$ be the pushforward measure of μ on \mathcal{Z} . We then denote the marginal density of P with respect to ν by $\phi_0^2(z)$. By the disintegration theorem, there exists a family $\{\mu_z : z \in \mathcal{Z}\}$ of probability measures on \mathcal{X} . Throughout, we assume that μ_z is absolutely continuous with respect to some σ -finite measure λ for all $z \in \mathcal{Z}$. We then denote the conditional density function of P with respect to λ by $v_0^2(y_L, y_U|z)$.

For any $1 \leq p \leq \infty$, we let $\|\cdot\|_{L^p_\pi}$ be the usual L^p -norm with respect to a measure π , where $\|\cdot\|_{L^\infty_\pi}$ denotes the essential supremum.

APPENDIX B: Proof of Theorems 2.1 and 2.2

Proof of Theorem 2.1. We first show that the identified set can be written as

$$\Theta_0(P) = \{\theta : \theta = E[m(Z)l(Z)], P(m_L(Z) \leq m(Z) \leq m_U(Z)) = 1\}. \quad (\text{B.1})$$

For this, we note that, by Assumptions 2.1-2.3 and arguing as in the proof of Theorem 1 in Stoker (1986), we have

$$E[w(Z)\nabla_z m(Z)] = E[m(Z)l(Z)]. \quad (\text{B.2})$$

Further, the distribution of Y first-order stochastically dominates that of Y_L . Similarly, the distribution of Y_U first-order stochastically dominates that of Y . Since q is nondecreasing by the convexity of ϱ , it then follows that, for each $u \in \mathbb{R}$,

$$E[q(Y_L - u)|Z] \leq E[q(Y - u)|Z] \leq E[q(Y_U - u)|Z], P - a.s. \quad (\text{B.3})$$

Eq. (3) then follows by (B.3), Assumption 2.3 (iii), and the hypothesis that $E[q(Y - u)|Z = z] = 0$ has a unique solution at $u = m(z)$ on D .

For the convexity of $\Theta_0(P)$, observe that for any $\theta_1, \theta_2 \in \Theta_0(P)$, there exist $m_1, m_2 : \mathcal{Z} \rightarrow \mathbb{R}$ such that $\theta_j = E[m_j(Z)l(Z)]$ and $m_L(Z) \leq m_j(Z) \leq m_U(Z), P - a.s.$ for $j = 1, 2$. Let $\alpha \in [0, 1]$ and let $\theta_\alpha \equiv \alpha\theta_1 + (1 - \alpha)\theta_2$. Then,

$$\theta_\alpha = E[m_\alpha(Z)l(Z)],$$

where $m_\alpha \equiv \alpha m_1 + (1 - \alpha)m_2$. Since $m_L(Z) \leq m_\alpha(Z) \leq m_U(Z), P - a.s.$, it follows that $\theta_\alpha \in \Theta_0(P)$. Therefore, $\Theta_0(P)$ is convex.

We show compactness of $\Theta_0(P)$ by showing $\Theta_0(P)$ is bounded and closed. By Assumption 2.3 (i)-(ii), for any $\theta \in \Theta_0(P)$,

$$|\theta^{(j)}| \leq \sup_{z \in \mathcal{Z}} |m(z)| E[|l^{(j)}(Z)|] \leq \sup_{x \in D} |x| E[|l^{(j)}(Z)|] < \infty, \quad \text{for } j = 1, \dots, \ell. \quad (\text{B.4})$$

Hence, $\Theta_0(P)$ is bounded. To see that $\Theta_0(P)$ is closed, consider the following maximization problem:

$$\text{maximize } E[m(Z)p'l(Z)], \quad (\text{B.5})$$

$$\text{s.t. } m_L(Z) \leq m(Z) \leq m_U(Z), P - a.s. \quad (\text{B.6})$$

Arguing as in the proof of Proposition 2 in Bontemps, Magnac, and Maurin (2012), the objective function is maximized by setting $m(z) = m_L(z)$ when $p'l(z) \leq 0$ and setting $m(z) = m_U(z)$ otherwise. This and (B.2) give the support function of $\Theta_0(P)$ in (8) and also shows that, for each $p \in \mathbb{S}$, there exists $m_p(z) \equiv 1\{p'l(z) \leq 0\}m_L(z) + 1\{p'l(z) > 0\}m_U(z)$ such that $v(p, \Theta_0(P)) = \langle p, \theta^*(p) \rangle$, where $\theta^*(p) = E[m_p(Z)l(Z)]$. Since m_p satisfies $m_L(Z) \leq m_p(Z) \leq m_U(Z), P - a.s.$, we have $\theta^*(p) \in \Theta_0(P)$. By Proposition 8.29 (a) in Rockafellar and Wets (2005), the boundary of $\Theta_0(P)$ is $\{\tilde{\theta} : \langle p, \tilde{\theta} \rangle = v(p, \Theta_0(P)), p \in \mathbb{S}^\ell\}$. Therefore, $\Theta_0(P)$ contains its boundary, and hence it is closed.

For the strict convexity of $\Theta_0(P)$, we show it through the differentiability of the support function. The proof is similar to that of Lemma A.8 in Beresteanu and Molinari (2008) and Lemma 23 in Bontemps, Magnac, and Maurin (2012). To this end, we extend the support function and define $s(p, \Theta_0(P))$ as in (8) for each $p \in \mathbb{R}^\ell \setminus \{0\}$.

For each $z \in \mathcal{Z}$, let $\xi(z) \equiv (m_L(z) - m_U(z))l(z)$. For each $p \in \mathbb{R}^\ell \setminus \{0\}$, let $\zeta(p) \equiv E[1\{p'\xi(Z) \geq 0\}p'\xi(Z)]$. Then, since $m_L(Z) - m_U(Z) \leq 0$ almost surely, it holds that $v(p, \Theta_0(P)) = \zeta(p) + E[m_U(Z)p'l(Z)]$ for all $p \in \mathbb{R}^\ell \setminus \{0\}$. For any $p, q \in \mathbb{R}^\ell \setminus \{0\}$, it then follows by the Cauchy-Schwarz inequality that

$$\begin{aligned} & |\zeta(q) - \zeta(p) - (q - p)'E[\xi(Z)1\{p'\xi(Z) \geq 0\}]| \\ &= |E[(1\{q'\xi(Z) \geq 0\} - 1\{p'\xi(Z) \geq 0\})q'\xi(Z)]| \leq \|1\{q'\xi \geq 0\} - 1\{p'\xi \geq 0\}\|_{L^2_P} \|q'\xi\|_{L^2_P}. \end{aligned} \quad (\text{B.7})$$

By Assumptions 2.1 (i), the distribution of $\xi(Z)$ does not assign a positive measure to any proper subspace of \mathbb{R}^ℓ with dimension $\ell - 1$, which ensures $P(p'\xi(Z) = 0) = 0$. Thus, for any sequence $\{q_n\}$ such that $q_n \rightarrow p$, it follows that $1\{q_n'\xi(Z) \geq 0\} \xrightarrow{a.s.} 1\{p'\xi(Z) \geq 0\}$ as $n \rightarrow \infty$. Note that $1\{p'\xi(Z)\}$ is bounded for all p . Thus, the function class $\{1^2\{p'\xi(\cdot)\} : p \in \mathbb{R}^\ell \setminus \{0\}\}$ is uniformly integrable. These results ensure that $\|1\{q'\xi \geq 0\} - 1\{p'\xi \geq 0\}\|_{L^2_P} \rightarrow 0$ as $q \rightarrow p$. This and $\|q'\xi\|_{L^2_P} < \infty$ imply that the right hand side of (B.7) is $o(1)$. Hence, ζ is differentiable at every point on $\mathbb{R}^\ell \setminus \{0\}$ with the derivative $E[\xi(Z)1\{p'\xi(Z) \geq 0\}]$. Note that, for each $z \in \mathcal{Z}$, $p \mapsto m_U(z)p'l(z)$ is differentiable with respect to p and $m_U l$ is integrable with respect to P by Assumption 2.3 (i)-(ii). This ensures that $p \mapsto E[m_U(Z)p'l(Z)]$ is differentiable with respect to p at every $p \in \mathbb{R} \setminus \{0\}$. Therefore, the map $p \mapsto v(p, \Theta_0(P))$ is differentiable for all $p \in \mathbb{R}^\ell \setminus \{0\}$. By Corollary 1.7.3 in Schneider (1993), the support set $H(p, \Theta_0(P)) \equiv \{\theta : \langle p, \theta \rangle = v(p, \Theta_0(P))\} \cap \Theta_0(P)$ for each p then contains only one point, which ensures the strict convexity of $\Theta_0(P)$.

To see that $\Theta_0(P)$ is sharp, take any $\theta \in \Theta_0(P)$. Then, by convexity, there exist $p, q \in \mathbb{S}^\ell$ and $\alpha \in [0, 1]$ such that $\theta = \alpha\theta^*(p) + (1 - \alpha)\theta^*(q)$, which further implies

$$\theta = E[(\alpha m_p(Z) + (1 - \alpha)m_q(Z))l(Z)] = E[w(Z)\nabla_z m_{\alpha,p,q}(Z)], \quad (\text{B.8})$$

where $m_{\alpha,p,q} \equiv \alpha m_p + (1 - \alpha)m_q$, and the last equality follows from integration by parts and Assumptions 2.1 (i) and 2.3 (iv) ensuring the almost everywhere differentiability of $m_{\alpha,p,q}$. Since $m_{\alpha,p,q}$ satisfies (3) in place of m with $m_{\alpha,p,q}$ and $m_{\alpha,p,q}$ is almost everywhere differentiable, θ is the weighted average derivative of a regression function consistent with some data generating process. Hence, $\Theta_0(P)$ is sharp. ■

Proof of Theorem 2.2. We first show (16). By the first order condition for (12) and Assumption 2.4 (ii), $E[q(Y - g(Z, V))|Z, V] = E[q(Y - g(Z, V))|\tilde{Z}, V] = 0, P - a.s.$ By Assumption 2.4 (i) and the monotonicity of q , for $v_L \leq v \leq v_U$, we have

$$\int q(y - g(z, v_U))dP(y|\tilde{z}, v) \leq 0 \leq \int q(y - g(z, v_L))dP(y|\tilde{z}, v), P - a.s. \quad (\text{B.9})$$

Taking expectations with respect to V , we obtain

$$\int q(y - g(z, v_U))dP(y|\tilde{z}) \leq 0 \leq \int q(y - g(z, v_L))dP(y|\tilde{z}), \quad P - a.s. \quad (\text{B.10})$$

Further, by Assumption 2.5 (ii),

$$\int q(y - \gamma(z, v_L, v_U))dP(y|\tilde{z}) = 0. \quad (\text{B.11})$$

By (B.10)-(B.11) and the monotonicity of q , we then have

$$g(z, v_L) \leq \gamma(z, v_L, v_U) \leq g(z, v_U). \quad (\text{B.12})$$

Let $\Xi_L(v) \equiv \{(v_L, v_U) : v_L \leq v_U \leq v\}$ and $\Xi_U(v) \equiv \{(v_L, v_U) : v \leq v_L \leq v_U\}$. To prove the lower bound on $g(z, v)$, take any $v_U \leq v$. Then by Assumption 2.4 (i) and (B.12), we have $\gamma(z, v_L, v_U) \leq g(z, v)$ for any $(v_L, v_U) \in \Xi(v)$. Hence, it follows that $g_L(z, v) \equiv \sup_{(v_L, v_U) \in \Xi_L(v)} \gamma(z, v_L, v_U) \leq g(z, v)$. Note that $g_L(z, v)$ is weakly increasing in v by construction and differentiable in z with a bounded derivative by Assumption 2.5 (iii). Hence, it is consistent with Assumption 2.4 (i). Thus, the bound is sharp. A similar argument gives the upper bound. Hence, (16) holds. This and integration by parts imply that the sharp identified set can be written as

$$\Theta_{0,v}(P) = \{\theta : \theta = E[g(Z, v)l(Z)], P(g_L(Z, v) \leq g(Z, v) \leq g_U(Z, v)) = 1\}. \quad (\text{B.13})$$

The rest of the proof is then similar to that of Theorem 2.1. It is therefore omitted. ■

APPENDIX C: Proof of Theorem 3.1

This Appendix contains the proof of Theorem 3.1 and auxiliary lemmas needed to establish the main result.

To characterize the efficiency bound, it proves useful to study a parametric submodel of \mathbf{P} . We define a parametric submodel through a curve in L_μ^2 . Let $h_0 \equiv \sqrt{dP/d\mu}$. Let $\tilde{v} : \mathcal{X} \rightarrow \mathbb{R}$ and $\tilde{\phi} : \mathcal{Z} \rightarrow \mathbb{R}$ be bounded functions that are continuously differentiable in z with bounded derivatives. We then define

$$\bar{v}(x) \equiv \tilde{v}(x) - E[\tilde{v}(X)|Z = z], \quad \text{and} \quad \bar{\phi}(z) \equiv \tilde{\phi}(z) - E[\tilde{\phi}(Z)], \quad (\text{C.1})$$

where expectations are with respect to P . For each $\eta \in \mathbb{R}$, define $v_\eta : \mathcal{X} \rightarrow \mathbb{R}$ and $\phi_\eta : \mathcal{Z} \rightarrow \mathbb{R}$ by

$$v_\eta^2(y_L, y_U|z) = v_0^2(y_L, y_U|z)(1 + 2\eta\bar{v}(x)), \quad \text{and} \quad \phi_\eta^2(z) = \phi_0^2(z)(1 + 2\eta\bar{\phi}(z)). \quad (\text{C.2})$$

We then let h_η^2 be defined pointwise by

$$h_\eta^2(x) \equiv v_\eta^2(y_L, y_U|z)\phi_\eta^2(z). \quad (\text{C.3})$$

It is straightforward to show that $\eta \mapsto h_\eta^2$ is a curve in L_μ^2 with $\dot{h}_0 = \dot{v}_0\phi_0 + v_0\dot{\phi}_0$, where $\dot{v}_0(y_L, y_U|z) \equiv \bar{v}(x)v_0(y_L, y_U|z)$ and $\dot{\phi}_0(z) = \bar{\phi}(z)\phi_0(z)$. We also note that for any η and η_0 in a neighborhood of 0, it holds that

$$v_\eta^2(y_L, y_U|z) = v_{\eta_0}^2(y_L, y_U|z)(1 + 2(\eta - \eta_0)\bar{v}_\eta(x)), \quad \text{and} \quad \phi_\eta^2(z) = \phi_{\eta_0}^2(z)(1 + 2\eta\bar{\phi}_\eta(z)). \quad (\text{C.4})$$

where $\bar{v}_\eta = \bar{v}v_0^2/v_{\eta_0}^2$ and $\bar{\phi}_\eta = \bar{\phi}\phi_0^2/\phi_{\eta_0}^2$. We then define $\dot{v}_\eta(y_L, y_U|z) = \bar{v}_\eta(x)v_\eta(y_L, y_U|z)$ and $\dot{\phi}_\eta(z) = \bar{\phi}_\eta(z)\phi_\eta(z)$.

We further introduce notation for population objects along this curve. Let $f_\eta(z) \equiv \phi_\eta^2(z)$ and $l_\eta \equiv -\nabla_z w(z) - w(z)\nabla_z f_\eta(z)/f_\eta(z)$. Lemma C.1 will show that there exists a neighborhood N of 0 such that the equations $\int q(y_L - \tilde{m})v_\eta^2(y_L, y_U|z)d\lambda(y_L, y_U) = 0$ and $\int q(y_U - \tilde{m})v_\eta^2(y_L, y_U|z)d\lambda(y_L, y_U) = 0$ have unique solutions on D for all $\eta \in N$. We denote these solutions by $m_{L,\eta}$ and $m_{U,\eta}$ respectively. We then let $m_{p,\eta}(z) \equiv \Gamma(m_{L,\eta}(z), m_{U,\eta}(z), p'l_\eta(z))$.

Further, we define

$$r_{j,\eta}(z) \equiv -\frac{d}{d\tilde{m}} E_\eta [q(y_j - \tilde{m}) | Z = z] \Big|_{\tilde{m}=m_{j,\eta}(z)}, \quad j = L, U, \quad (\text{C.5})$$

where expectations are taken with respect to P_η . Finally, define

$$\zeta_{p,\eta} \equiv \Gamma(r_{L,\eta}^{-1}(z)q(y_L - m_{L,\eta}(z)), r_{U,\eta}^{-1}(z)q(y_U - m_{U,\eta}(z)), p'l_\eta(z)). \quad (\text{C.6})$$

Given these definitions, we give an outline of the general structure of the proof. The proof of Theorem 3.1 proceeds by verifying the conditions of Theorem 5.2.1 in Bickel, Klassen, Ritov, and Wellner (1993), which requires (i) the characterization of the tangent space at P , which we accomplish in Theorem C.1 and (ii) the pathwise weak differentiability of the map $Q \mapsto v(\cdot, \Theta_0(Q))$, which is established by Theorem C.2.

TANGENT SPACE (Theorem C.1)

Step 1: Lemmas C.1-C.5 show that for some neighborhood N of 0, Assumptions 2.3 and 3.2 hold with P_η in place of P for all $\eta \in N$, where $\sqrt{dP_\eta/d\mu} = h_\eta$ defined in (C.2)-(C.3). This means that the restrictions on P in Assumptions 2.3 and 3.2 do not restrict the neighborhood in such a way that affects the tangent space derived in the next step. In this step, we exploit the fact that \mathcal{Z} is determined by the dominating measure μ instead of each distribution in the model.

Step 2: Theorem C.1 then establishes that the tangent space $\dot{\mathbf{S}}$ equals $\mathbf{T} \equiv \{h \in L_\mu^2 : \int h(x)s(x)d\mu(x) = 0\}$ by showing that (i) $\dot{\mathbf{S}} \subseteq \mathbf{T}$ generally and (ii) due to Step 1, $\{P_\eta, \eta \in N\}$ is a regular parametric submodel of \mathbf{P} whose tangent set $\dot{\mathbf{U}} \subset \dot{\mathbf{S}}$ is dense in \mathbf{T} implying $\mathbf{T} \subseteq \dot{\mathbf{S}}$.

DIFFERENTIABILITY (Theorem C.2)

Step 1: Lemmas C.1 and C.7 explicitly characterize the pathwise derivatives of $m_{j,\eta}, j = L, U$ along the curve $\eta \mapsto h_\eta$ defined in (C.2)-(C.3).

Step 2: Based on Step 1 and Lemma C.8, Lemma C.9 then characterizes the pathwise derivative of the support function $v(p, \Theta_0(P_\eta))$ at a point p along the curve $\eta \mapsto h_\eta$ defined in (C.2)-(C.3). Lemmas C.10 and C.11 further show that this pathwise derivative is uniformly bounded and continuous in $(p, \eta) \in \mathbb{S}^\ell \times N$.

Step 3: Based on Step 2, Theorem C.2 first characterizes the pathwise weak derivative of $\rho(P_\eta) = v(\cdot, \Theta_0(P_\eta))$ on the tangent space of the curve $\eta \mapsto h_\eta$ and further extends it to $\dot{\mathbf{S}}$.

Lemma C.1. *Let $\eta \mapsto h_\eta$ be a curve in L_μ^2 defined in (C.2)-(C.3). Suppose Assumption 2.1 holds. Suppose $P \in \mathbf{P}$. Then, there exists a neighborhood N of 0 such that (i) $\int q(y_j - \tilde{m})v_\eta^2(y_L, y_U | z)d\lambda(y_L, y_U) = 0$ has a unique solution at $\tilde{m} = m_{j,\eta}(z)$ on D for $j = L, U$ and for all $\eta \in N$; (ii) For each $(z, \eta) \in \mathcal{Z} \times N$, $m_{\eta,L}$ and $m_{\eta,U}$ are continuously differentiable a.e. on the interior of $\mathcal{Z} \times N$ with bounded derivative. In particular, it holds that*

$$\frac{\partial}{\partial \eta} m_{L,\eta}(z) \Big|_{\eta=\eta_0} = 2r_{L,\eta_0}^{-1}(z) \int q(y_L - m_{L,\eta_0}(z))\dot{v}_{\eta_0}(y_L, y_U | z)v_{\eta_0}(y_L, y_U | z)d\lambda(y_L, y_U) \quad (\text{C.7})$$

$$\frac{\partial}{\partial \eta} m_{U,\eta}(z) \Big|_{\eta=\eta_0} = 2r_{U,\eta_0}^{-1}(z) \int q(y_U - m_{U,\eta_0}(z))\dot{v}_{\eta_0}(y_L, y_U | z)v_{\eta_0}(y_L, y_U | z)d\lambda(y_L, y_U), \quad (\text{C.8})$$

for all $\eta_0 \in N$.

Proof of Lemma C.1. The proof builds on the proof of Theorem 3.1 in Newey and Stoker (1993). By Eq. (C.2), it follows that

$$\begin{aligned} \int q(y_L - \tilde{m})v_\eta^2(y_L, y_U | z)d\lambda(y_L, y_U) &= \int q(y_L - \tilde{m})v_0^2(y_L, y_U | z)d\lambda(y_L, y_U) \\ &\quad + 2\eta \int q(y_L - \tilde{m})\bar{v}(x)v_0^2(y_L, y_U | z)d\lambda(y_L, y_U). \end{aligned} \quad (\text{C.9})$$

Since $P \in \mathbf{P}$, Assumption 3.2 and Lemma C.2 in Newey (1991) imply that $\int \tilde{v}(x)v_0^2(y_L, y_U|z)d\lambda(y_L, y_U)$ is continuously differentiable with bounded derivatives, and hence so is \bar{v} . Therefore, $\int q(y_j - \tilde{m})\bar{v}(x)v_0^2(y_L, y_U|z)d\lambda(y_L, y_U)$ is continuously differentiable in (z, \tilde{m}) on $\mathcal{Z} \times D$ for $j = L, U$. Hence, by (C.9), there is a neighborhood N' of 0 such that the map $(z, \tilde{m}, \eta) \mapsto \int q(y_L - \tilde{m})v_\eta^2(y_L, y_U|z)d\lambda(y_L, y_U)$ is continuously differentiable on $\mathcal{Z} \times D \times N'$ with bounded derivatives. By continuity, we may take a neighborhood N of 0 small enough so that $\int q(y_L - \tilde{m})v_\eta^2(y_L, y_U|z)d\lambda(y_L, y_U) = 0$ admits a unique solution $m_{L,\eta}(z)$ for all $\eta \in N$. A similar argument can be made for $m_{U,\eta}$.

By the implicit function theorem, there is a neighborhood of $(z, 0)$ on which $\nabla_z m_{j,\eta_0}$ and $\frac{\partial}{\partial \eta} m_{j,\eta}(z)|_{\eta=\eta_0}$ exist and are continuous in their arguments on that neighborhood for $j = L, U$. By the compactness of \mathcal{Z} , N can be chosen small enough so that $\nabla_z m_{\eta,L}$ and $\frac{\partial}{\partial \eta} m_{j,\eta}(z)|_{\eta=\eta_0}$ are continuous and bounded on $\mathcal{Z} \times N$ and

$$\frac{\partial}{\partial \eta} m_{j,\eta}(z)|_{\eta=\eta_0} = 2r_{j,\eta_0}^{-1}(z) \int q(y_j - m_{j,\eta_0}(z))\dot{v}_{\eta_0}(y_L, y_U|z)v_{\eta_0}(y_L, y_U|z)d\lambda(y_L, y_U), \quad j = L, U. \quad (\text{C.10})$$

This completes the proof of the lemma. ■

Lemma C.2. *Let $\eta \mapsto h_\eta$ be a curve in L_μ^2 defined in (C.2)-(C.3). Suppose Assumption 2.1 (i) holds. Suppose further that $P \in \mathbf{P}$. Then, there exists a neighborhood N of 0 such that the conditional support of (Y_L, Y_U) given Z is in $D^\circ \times D^\circ$, $w(z)f_\eta(z) = 0$ on $\partial\mathcal{Z}$, P_η - a.s., $\nabla_z f_\eta/f_\eta(z)$ is continuous a.e., and $\int \|l_\eta(z)\|^2 \phi_\eta^2(z) d\nu(z) < \infty$ for all $\eta \in N$.*

Proof of Lemma C.2. By (C.2), $\{(y_L, y_U) : v_0^2(y_L, y_U|z) = 0\} \subseteq \{(y_L, y_U) : v_\eta^2(y_L, y_U|z) = 0\}$ for all $z \in \mathcal{Z}$ and $\eta \in \mathbb{R}$. This implies $\{(y_L, y_U) : v_\eta^2(y_L, y_U|z) > 0\} \subseteq \{(y_L, y_U) : v_0^2(y_L, y_U|z) > 0\} \subset D^\circ \times D^\circ$ for all $z \in \mathcal{Z}$ and $\eta \in \mathbb{R}$, where the last inclusion holds by Assumption 2.3. This establishes the first claim. Similarly, the second claim follows immediately from Eq. (C.2) and Assumption 2.3 (ii).

For the third claim, using Eq. (C.2), we write

$$\frac{\nabla_z f_\eta(z)}{f_\eta(z)} = \frac{\nabla_z f(z)}{f(z)} + \frac{2\eta \nabla_z \bar{\phi}(z)}{1 + 2\eta \bar{\phi}(z)}. \quad (\text{C.11})$$

By Assumption 2.3 (ii), (C.11), and $\bar{\phi}$ being bounded and continuously differentiable in z , $(\eta, z) \mapsto \nabla_z f_\eta(z)/f_\eta(z)$ is continuous. This and Assumption 2.2 in turn imply that the map $(\eta, z) \mapsto \|l_\eta(z)\|^2$ is continuous. Hence, by Assumption 2.1 (i), it achieves a finite maximum on $N \times \mathcal{Z}$ for some N small enough. Therefore, $\int \|l_\eta(z)\|^2 \phi_\eta^2(z) d\nu(z) < \infty$ for all $\eta \in N$. ■

Lemma C.3. *Let $\eta \mapsto h_\eta$ be a curve in L_μ^2 defined in (C.2)-(C.3). Suppose further that $P \in \mathbf{P}$. Then, there exists a neighborhood N of 0 such that $|r_{L,\eta}(z)| > \bar{\epsilon}$ and $|r_{U,\eta}(z)| > \bar{\epsilon}$, for all $z \in \mathcal{Z}$ and $\eta \in N$.*

Proof of Lemma C.3. By (C.2) and (C.5), $r_{L,\eta}$ can be written as

$$\begin{aligned} r_{L,\eta}(z) &\equiv -\frac{d}{d\tilde{m}} E_\eta [q(y_L - \tilde{m})|Z = z] \Big|_{\tilde{m}=m_{L,\eta}(z)} \\ &= -\frac{d}{d\tilde{m}} \left(E [q(y_L - \tilde{m})|Z = z] + 2\eta \int q(y_L - \tilde{m})\bar{v}(x)v_0^2(y_L, y_U|z)d\lambda(y_L, y_U) \right) \Big|_{\tilde{m}=m_{L,\eta}(z)} \\ &= r_L(z) - 2\eta \frac{d}{d\tilde{m}} \int q(y_L - \tilde{m})\bar{v}(x)v_0^2(y_L, y_U|z)d\lambda(y_L, y_U) \Big|_{\tilde{m}=m_{L,\eta}(z)}. \end{aligned} \quad (\text{C.12})$$

Since \bar{v} is bounded and continuously differentiable, the second term on the right hand side of (C.12) is well-defined and is bounded because Assumption 3.2 (iii) holds for $P \in \mathbf{P}$. By Assumption 3.2 (i) and Eq. (C.12), we may take a neighborhood N of 0 small enough so that $|r_{L,\eta}(z)| > \bar{\epsilon}$ for all $\eta \in N$ and $z \in \mathcal{Z}$. A similar argument can be made for $r_{U,\eta}(z)$. Thus, the claim of the lemma follows. ■

Lemma C.4. *Let $\eta \mapsto h_\eta$ be a curve in L_μ^2 defined in (C.2)-(C.3). Suppose further that $P \in \mathbf{P}$. Then, there exists a neighborhood N of 0 such that (i) for any $\varphi : \mathcal{X} \rightarrow \mathbb{R}$ that is bounded and continuously differentiable in z with*

bounded derivatives, $\int \varphi(x)v_\eta^2(y_L, y_U|z)d\lambda(y_L, y_U)$ is continuously differentiable in z on \mathcal{Z} with bounded derivatives; (ii) $\int q(y_L - \tilde{m})\varphi(x)v_\eta^2(y_L, y_U|z)d\lambda(y_L, y_U)$ and $\int q(y_U - \tilde{m})\varphi(x)v_\eta^2(y_L, y_U|z)d\lambda(y_L, y_U)$ are continuously differentiable in (z, \tilde{m}) on $\mathcal{Z} \times D$ with bounded derivatives for all $\eta \in N$.

Proof of Lemma C.4. Let $\varphi : \mathcal{X} \rightarrow \mathbb{R}$ be bounded and continuously differentiable in z with bounded derivatives. By (C.2), we may write

$$\int \varphi(x)v_\eta^2(y_L, y_U)d\lambda(y_L, y_U) = \int \varphi(x)v_0^2(y_L, y_U|z)d\lambda(y_L, y_U) + 2\eta \int \varphi(x)\bar{v}(x)v_0^2(y_L, y_U|z)d\lambda(y_L, y_U) \quad (\text{C.13})$$

Note that \bar{v} is bounded and continuously differentiable in z with bounded derivatives. Thus, by Assumption 3.2 (ii), $\int \varphi(x)v_\eta^2(y_L, y_U)d\lambda(y_L, y_U)$ is bounded and continuously differentiable in z with bounded derivatives. Similarly, we may write

$$\begin{aligned} & \int q(y_L - \tilde{m})\varphi(x)v_\eta^2(y_L, y_U|z)d\lambda(y_L, y_U) \\ &= \int q(y_L - \tilde{m})\varphi(x)v_0^2(y_L, y_U|z)d\lambda(y_L, y_U) + 2\eta \int q(y_L - \tilde{m})\varphi(x)\bar{v}(x)v_0^2(y_L, y_U|z)d\lambda(y_L, y_U). \end{aligned} \quad (\text{C.14})$$

By Assumption 3.2 (iii), $\int q(y_L - \tilde{m})\varphi(x)v_0^2(y_L, y_U|z)d\lambda(y_L, y_U)$ is bounded and continuously differentiable in z with bounded derivatives. Further, since \bar{v} is bounded and continuously differentiable in z with bounded derivatives, again by Assumption 3.2 (iii), the same is true for the second term in the right hand side of (C.14). The argument for $\int q(y_U - \tilde{m})\varphi(x)v_\eta^2(y_L, y_U|z)d\lambda(y_L, y_U)$ is similar. Thus the claim of the lemma follows. ■

Lemma C.5. *Let $\eta \mapsto h_\eta$ be a curve in L_μ^2 defined in (C.2)-(C.3). Suppose that $P \in \mathbf{P}$. Then, there exists a neighborhood N of 0 such that $m_{L,\eta}$ and $m_{U,\eta}$ are continuously differentiable a.e. on \mathcal{Z} with bounded derivatives. Further, the maps $(z, \eta) \mapsto \nabla_z m_{L,\eta}(z)$ and $(z, \eta) \mapsto \nabla_z m_{U,\eta}(z)$ are continuous on $\mathcal{Z} \times N$.*

Proof of Lemma C.5. We show the claims of the lemma for $m_{L,\eta}$. By $P \in \mathbf{P}$, Assumption 3.2 (iii) holds, which implies that the maps $(z, \tilde{m}) \mapsto \int q(y_L - \tilde{m})v_0^2(y_L, y_U|z)d\lambda(y_L, y_U)$ and $(z, \tilde{m}) \mapsto \int q(y_L - \tilde{m})\bar{v}(x)v_0^2(y_L, y_U|z)d\lambda(y_L, y_U)$ are continuously differentiable on $\mathcal{Z} \times D$. By (C.14) (with $\varphi(x) = 1$), it then follows that $(z, \tilde{m}, \eta) \mapsto \int q(y_L - \tilde{m})\varphi(x)v_\eta^2(y_L, y_U|z)d\lambda(y_L, y_U)$ is continuously differentiable on $\mathcal{Z} \times D \times N$ for some N that contains 0 in its interior. Following the argument in the proof of Theorem 3.1 in Newey and Stoker (1993), it then follows that $\nabla_z m_{L,\eta}$ exists and is continuous on $\mathcal{Z} \times N$. By the compactness of \mathcal{Z} , N can be chosen small enough so that $\nabla_z m_{L,\eta}$ is bounded on $\mathcal{Z} \times N$. The argument for $m_{U,\eta}$ is similar. Hence it is omitted. ■

Theorem C.1. *Let Assumptions 2.1-2.2 and 3.1 hold and $P \in \mathbf{P}$. Then, the tangent space of \mathbf{S} at $s \equiv \sqrt{dP/d\mu}$ is given by $\dot{\mathbf{S}} = \{h \in L_\mu^2 : \int h(x)s(x)d\mu(x) = 0\}$*

Proof of Theorem C.1. Let $\mathbf{T} \equiv \{h \in L_\mu^2 : \int h(x)s(x)d\mu(x) = 0\}$. $\dot{\mathbf{S}} \subseteq \mathbf{T}$ holds by Proposition 3.2.3 in Bickel, Klassen, Ritov, and Wellner (1993).

For the converse: $\mathbf{T} \subseteq \dot{\mathbf{S}}$, it suffices to show that a dense subset of \mathbf{T} is contained in $\dot{\mathbf{S}}$. For this, let $\dot{\mathbf{U}} \equiv \{\dot{h}_0 \in L_\mu^2 : \int \dot{h}_0(x)s(x)d\mu(x) = 0\}$ denote the tangent space of the curve defined in (C.2)-(C.3). By Lemmas C.1-C.5, there is a neighborhood N of 0 for which Assumptions 2.3 and 3.2 hold for all $\eta \in N$. Therefore, $\eta \mapsto h_\eta^2, \eta \in N$ is a regular parametric submodel of \mathbf{P} whose Fréchet derivative at $\eta = 0$ is given by \dot{h}_0 . Hence, $\dot{\mathbf{U}} \subseteq \dot{\mathbf{S}}$. Further, by Lemma C.7 in Newey (1991) and the argument used in the proof of Theorem 3.1 in Newey and Stoker (1993), $\dot{\mathbf{U}}$ is dense in \mathbf{T} . Thus, $\mathbf{T} \subseteq \dot{\mathbf{S}}$. ■

Lemma C.6. *Let $\eta \mapsto h_\eta$ be a curve in L_μ^2 defined in (C.2)-(C.3). Suppose Assumption 2.1 holds. Suppose further that $P \in \mathbf{P}$. Then, there is a compact set D' and a neighborhood N of 0 such that D' contains the support of $Y_L - m_{L,\eta_0}(Z)$ and $Y_U - m_{U,\eta_0}(Z)$ in its interior for all $\eta_0 \in N$.*

Proof of Lemma C.6. By Lemma C.2, there exists a neighborhood N' of 0 such that the supports of Y_L and Y_U are contained in the interior of D under P_η for all η in N' . Similarly, by Lemma C.1, there is a neighborhood N'' of 0 such that $m_{L,\eta}(Z), m_{U,\eta}(Z)$ are well defined for all $\eta \in N''$ and their supports are contained in the interior of D respectively. Without loss of generality, let $D = [a, b]$ for some $-\infty < a < b < \infty$ and let $N = N' \cap N''$. Then, the support of $Y_L - m_{L,\eta}(Z)$ is contained in $D' \equiv [a - b, b - a]$ for all $\eta \in N$. A similar argument ensures that the support of $Y_U - m_{U,\eta}(Z)$ is contained in D' . This completes the proof. ■

Lemma C.7. *Let $\eta \mapsto h_\eta$ be a curve in L_μ^2 defined in (C.2)-(C.3). Suppose Assumption 2.1 holds. Suppose further that $P \in \mathbf{P}$. Then there is a neighborhood N of 0 such that (i) the functions $(z, \eta_0) \mapsto \frac{\partial}{\partial \eta} m_{L,\eta}(z)|_{\eta=\eta_0}$ and $(z, \eta_0) \mapsto \frac{\partial}{\partial \eta} m_{U,\eta}(z)|_{\eta=\eta_0}$ are bounded on $\mathcal{Z} \times N$; (ii) For each $z \in \mathcal{Z}$, the maps $\eta_0 \mapsto \frac{\partial}{\partial \eta} m_{L,\eta}(z)|_{\eta=\eta_0}$ and $\eta_0 \mapsto \frac{\partial}{\partial \eta} m_{U,\eta}(z)|_{\eta=\eta_0}$ are continuous on N .*

Proof of Lemma C.7. By Lemmas C.1, C.3 and C.6, Assumption 2.1 and \bar{v} being bounded, it follows that

$$\left| \frac{\partial}{\partial \eta} m_{j,\eta}(z) \Big|_{\eta=\eta_0} \right| \leq 2\bar{\epsilon}^{-1} \sup_{u \in D'} |q(u)| \times \sup_{x \in \mathcal{X}} \bar{v}(x) < \infty, \quad j = L, U. \quad (\text{C.15})$$

Hence, the first claim follows. Now let $\eta_n \in N$ be a sequence such that $\eta_n \rightarrow \eta_0$. Then, by the triangle and Cauchy-Schwarz inequalities,

$$\left| \frac{\partial}{\partial \eta} m_{L,\eta}(z) \Big|_{\eta=\eta_n} - \frac{\partial}{\partial \eta} m_{L,\eta}(z) \Big|_{\eta=\eta_0} \right| \leq 2|r_{L,\eta_n}^{-1}(z) - r_{L,\eta_0}^{-1}(z)| \sup_{u \in D'} |q(u)| \times \|\dot{v}_{\eta_0}\|_{L_\lambda^2} \|v_{\eta_0}\|_{L_\lambda^2} \quad (\text{C.16})$$

$$+ 2\bar{\epsilon}^{-1} \sup_{u \in D'} |q(u)| (\|v_{\eta_n} - v_{\eta_0}\|_{L_\lambda^2} \|\dot{v}_{\eta_n}\|_{L_\lambda^2} + \|\dot{v}_{\eta_n} - \dot{v}_{\eta_0}\|_{L_\lambda^2} \|v_{\eta_0}\|_{L_\lambda^2}) \quad (\text{C.17})$$

$$+ 2\bar{\epsilon}^{-1} \int |q(y_L - m_{\eta_n}(z)) - q(y_L - m_{\eta_0}(z))| \dot{v}_{\eta_0}(y_L, y_U | z) v_{\eta_0}(y_L, y_U | z) d\lambda(y_L, y_U). \quad (\text{C.18})$$

Note that $|r_{L,\eta_n}^{-1} - r_{L,\eta_0}^{-1}| \rightarrow 0$, a.e. by (C.12). By the continuous Fréchet differentiability of $\eta \mapsto v_\eta$, it follows that $\|v_{\eta_n} - v_{\eta_0}\|_{L_\lambda^2} = o(1)$ and $\|\dot{v}_{\eta_n} - \dot{v}_{\eta_0}\|_{L_\lambda^2} = o(1)$. Further, (C.18) tends to 0 as $\eta_n \rightarrow \eta_0$ by the dominated convergence theorem, almost everywhere continuity of q ensured by Assumption 2.1 (ii), and $m_{\eta_n} \rightarrow m_{\eta_0}$, a.e. by Lemma C.5. Therefore, $|\frac{\partial}{\partial \eta} m_{L,\eta}(z)|_{\eta=\eta_n} - \frac{\partial}{\partial \eta} m_{L,\eta}(z)|_{\eta=\eta_0}| = o(1)$. The continuity of $\frac{\partial}{\partial \eta} m_{U,\eta}(z)|_{\eta=\eta_0}$ can be shown in the same way. This completes the proof. ■

Lemma C.8. *Let $\eta \mapsto h_\eta$ be a curve in L_μ^2 defined in (C.2)-(C.3). Suppose Assumption 3.1 holds. Suppose further that $P \in \mathbf{P}$. Then, there is a neighborhood N of 0 such that*

$$\frac{\partial}{\partial \eta} \int p'l_{\eta_0}(z) 1\{p'l_\eta(z) > 0\} (m_{U,\eta_0}(z) - m_{L,\eta_0}(z)) \phi_{\eta_0}^2(z) d\nu(z) \Big|_{\eta=\eta_0} = 0, \quad (\text{C.19})$$

for all $\eta_0 \in N$.

Proof of Lemma C.8. By (C.11), there is a neighborhood N of 0 such that for all $\eta_0 \in N$,

$$\frac{\partial p'l_\eta(z)}{\partial \eta} \Big|_{\eta=\eta_0} = \frac{4p'\nabla_z \bar{\phi}(z)(1 + \eta_0 \bar{\phi}(z))}{(1 + 2\eta_0 \bar{\phi}(z))^2}. \quad (\text{C.20})$$

Hence, by compactness of $\mathbb{S}^\ell \times \mathcal{Z}$, the map $(p, z, \eta_0) \mapsto \frac{\partial p'l_\eta(z)}{\partial \eta} \Big|_{\eta=\eta_0}$ is uniformly bounded on $\mathbb{S}^\ell \times \mathcal{Z} \times N$ by some constant $M > 0$. Let η_n be a sequence such that $\eta_n \rightarrow \eta_0 \in N$. By the mean value theorem, there is $\bar{\eta}_n(z)$ such that

$$\begin{aligned} & \lim_{n \rightarrow \infty} \int p'l_{\eta_0}(z) (1\{p'l_{\eta_n}(z) > 0\} - 1\{p'l_{\eta_0}(z) > 0\}) (m_{U,\eta_0}(z) - m_{L,\eta_0}(z)) \phi_{\eta_0}^2(z) d\nu(z) \\ &= \lim_{n \rightarrow \infty} \int p'l_{\eta_0}(z) (1\{p'l_{\eta_0}(z) > (\eta_0 - \eta_n) \frac{\partial p'l_\eta(z)}{\partial \eta} \Big|_{\eta=\bar{\eta}_n(z)}\} - 1\{p'l_{\eta_0}(z) > 0\}) (m_{U,\eta_0}(z) - m_{L,\eta_0}(z)) \phi_{\eta_0}^2(z) d\nu(z) \\ &\leq \lim_{n \rightarrow \infty} \int |p'l_{\eta_0}(z)| 1\{|p'l_{\eta_0}(z)| \leq M|\eta_0 - \eta_n|\} |m_{U,\eta_0}(z) - m_{L,\eta_0}(z)| \phi_{\eta_0}^2(z) d\nu(z), \end{aligned} \quad (\text{C.21})$$

where the last inequality follows from $\frac{\partial p' l_\eta(z)}{\partial \eta} \Big|_{\eta=\bar{\eta}_n(z)}$ being bounded by M . Therefore, from (C.21), we conclude that

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{|\eta_n - \eta_0|} \left| \int p' l_{\eta_0}(z) (1\{p' l_{\eta_n}(z) > 0\} - 1\{p' l_{\eta_0}(z) > 0\}) (m_{U, \eta_0}(z) - m_{L, \eta_0}(z)) \phi_{\eta_0}^2(z) d\nu(z) \right| \\ \leq 2 \sup_{y \in D} |y| \times M \times \lim_{n \rightarrow \infty} \int 1\{|p' l_{\eta_0}(z)| \leq M|\eta_0 - \eta_n|\} \phi_{\eta_0}^2(z) d\nu(z) = 0, \end{aligned} \quad (\text{C.22})$$

where the last equality follows from the monotone convergence theorem and P being in \mathbf{P} ensuring Assumption 3.1 (ii). ■

Lemma C.9. *Let $\eta \mapsto h_\eta$ be a curve in L_μ^2 defined in (C.2)-(C.3). Suppose Assumptions 2.1-2.2, and 3.1 hold. Suppose further that $P \in \mathbf{P}$. Then, there is a neighborhood N of 0 such that for all $\eta_0 \in N$,*

$$\frac{\partial v(p, \Theta_0(P_\eta))}{\partial \eta} \Big|_{\eta=\eta_0} = 2 \int \{w(z) p' \nabla_z m_{p, \eta_0}(z) - v(p, \Theta_0(P_{\eta_0})) + p' l_{\eta_0}(z) \zeta_{p, \eta_0}(y_L, y_U, z)\} \dot{h}_{\eta_0}(x) h_{\eta_0} d\mu(x). \quad (\text{C.23})$$

Proof of Lemma C.9. We first show $\Gamma(\nabla_z m_{L, \eta}(z), \nabla_z m_{U, \eta}(z), p' l_\eta(z))$ is the gradient of $m_{p, \eta}(z)$, $\mu - a.e.$ By Assumption 3.1 (ii), it suffices to show the equality for z such that $p' l_\eta(z) \neq 0$. Write

$$\begin{aligned} m_{p, \eta}(z+h) - m_{p, \eta}(z) - \Gamma(\nabla_z m_{L, \eta}(z), \nabla_z m_{U, \eta}(z), p' l_\eta(z))' h \\ = 1\{p' l_\eta(z+h) > 0\} [(m_{U, \eta}(z+h) - m_{L, \eta}(z+h)) - (m_{U, \eta}(z) - m_{L, \eta}(z)) - (\nabla_z m_{U, \eta}(z) - \nabla_z m_{L, \eta}(z))' h] \\ + (1\{p' l_\eta(z+h) > 0\} - 1\{p' l_\eta(z) > 0\}) \times [(m_{U, \eta}(z) - m_{L, \eta}(z)) - (\nabla_z m_{U, \eta}(z) - \nabla_z m_{L, \eta}(z))' h] \\ + (m_{L, \eta}(z+h) - m_{L, \eta}(z) - \nabla_z m_{L, \eta}(z)' h). \end{aligned} \quad (\text{C.24})$$

$\nabla_z m_{U, \eta}$ and $\nabla_z m_{L, \eta}$ being the gradients of $m_{U, \eta}$ and $m_{L, \eta}$, respectively implies that

$$\begin{aligned} (m_{U, \eta}(z+h) - m_{L, \eta}(z+h)) - (m_{U, \eta}(z) - m_{L, \eta}(z)) - (\nabla_z m_{U, \eta}(z) - \nabla_z m_{L, \eta}(z))' h = o(\|h\|), \\ m_{L, \eta}(z+h) - m_{L, \eta}(z) - \nabla_z m_{L, \eta}(z)' h = o(\|h\|). \end{aligned} \quad (\text{C.25})$$

By the continuity of $z \mapsto l_\eta(z)$, ensured by Assumption 2.2, (C.11), and $\bar{\phi}$ being continuously differentiable, there exists $\epsilon > 0$ such that $1\{p' l_\eta(z+h) > 0\} = 1\{p' l_\eta(z) > 0\}$ for all h such that $\|h\| < \epsilon$. These results and (C.24) ensure that

$$m_{p, \eta}(z+h) - m_{p, \eta}(z) - \Gamma(\nabla_z m_{L, \eta}(z), \nabla_z m_{U, \eta}(z), p' l_\eta(z))' h = o(\|h\|). \quad (\text{C.26})$$

In what follows, we therefore simply write $\nabla_z m_{p, \eta}(z) = \Gamma(\nabla_z m_{L, \eta}(z), \nabla_z m_{U, \eta}(z), p' l_\eta(z))$.

Next, we show that $\eta \mapsto \nabla_z m_{p, \eta}(z)$ is continuous for almost all $z \in \mathcal{Z}$. By Lemma C.7, $\eta \mapsto \nabla_z m_{L, \eta}$ and $\eta \mapsto \nabla_z m_{U, \eta}$ are continuous. Further, if $\eta_n \rightarrow \eta_0$, then

$$\mu(\{x : \lim_{n \rightarrow \infty} 1\{p' l_{\eta_n}(z) > 0\} = 1\{p' l_{\eta_0}(z) > 0\}\}) = 1$$

by the continuity of $\eta \mapsto p' l_\eta(z)$ ensured by (C.11) and $\mu(\{x : p' l_{\eta_0}(z) = 0\}) = 0$ by Assumption 3.1 (ii). Hence, $\eta \mapsto \nabla_z m_{p, \eta}(z)$ is continuous *a.e.*

By Theorem 2.1, integration by parts, and (C.4), we may write

$$v(p, \Theta_0(P_\eta)) = \int w(z) p' \nabla_z m_{p, \eta}(z) \phi_{\eta_0}^2(z) d\nu(z) + 2(\eta - \eta_0) \int w(z) p' \nabla_z m_{p, \eta}(z) \dot{\phi}_{\eta_0}(z) \phi_{\eta_0}(z) d\nu(z) \quad (\text{C.27})$$

$$= \int p' l_{\eta_0}(z) m_{p, \eta}(z) \phi_{\eta_0}^2(z) d\nu(z) + 2(\eta - \eta_0) \int w(z) p' \nabla_z m_{p, \eta}(z) \dot{\phi}_{\eta_0}(z) \phi_{\eta_0}(z) d\nu(z). \quad (\text{C.28})$$

This and $v(p, \Theta_0(P_{\eta_0})) = \int p'l_{\eta_0}(z)m_{p,\eta_0}(z)\phi_{\eta_0}^2(z)d\nu(z)$ by Theorem 2.1 imply

$$\begin{aligned} & \left. \frac{\partial v(p, \Theta_0(P_\eta))}{\partial \eta} \right|_{\eta=\eta_0} \\ &= \lim_{\eta \rightarrow \eta_0} \frac{1}{\eta - \eta_0} \int p'l_{\eta_0}(z)(m_{p,\eta}(z) - m_{p,\eta_0}(z))\phi_{\eta_0}^2(z)d\nu(z) + 2 \lim_{\eta \rightarrow \eta_0} \int w(z)p'\nabla_z m_{p,\eta}(z)\dot{\phi}_{\eta_0}(z)\phi_{\eta_0}(z)d\nu(z), \\ &= \lim_{\eta \rightarrow \eta_0} \frac{1}{\eta - \eta_0} \int p'l_{\eta_0}(z)(m_{p,\eta}(z) - m_{p,\eta_0}(z))\phi_{\eta_0}^2(z)d\nu(z) + 2 \int w(z)p'\nabla_z m_{p,\eta_0}(z)\dot{\phi}_{\eta_0}(z)\phi_{\eta_0}(z)d\nu(z), \end{aligned} \quad (\text{C.29})$$

where the second equality follows from w , $\nabla_z m_{p,\eta}$ and $\bar{\phi}_{\eta_0}$ being bounded by Assumption 2.2 and Lemma C.1, which allows us to apply the dominated convergence theorem, and the almost everywhere continuity of $\eta \mapsto \nabla_z m_{p,\eta}(z)$. The first term on the right hand side of (C.29) may be further rewritten as

$$\lim_{\eta \rightarrow \eta_0} \frac{1}{\eta - \eta_0} \int p'l_{\eta_0}(z)(m_{p,\eta}(z) - m_{p,\eta_0}(z))\phi_{\eta_0}^2(z)d\nu(z) \quad (\text{C.30})$$

$$= \lim_{\eta \rightarrow \eta_0} \frac{1}{\eta - \eta_0} \int p'l_{\eta_0}(z)1\{p'l_{\eta_0}(z) > 0\}(m_{U,\eta}(z) - m_{U,\eta_0}(z))\phi_{\eta_0}^2(z)d\nu(z) \quad (\text{C.31})$$

$$+ \lim_{\eta \rightarrow \eta_0} \frac{1}{\eta - \eta_0} \int p'l_{\eta_0}(z)1\{p'l_{\eta_0}(z) \leq 0\}(m_{L,\eta}(z) - m_{L,\eta_0}(z))\phi_{\eta_0}^2(z)d\nu(z) \quad (\text{C.32})$$

$$+ \lim_{\eta \rightarrow \eta_0} \frac{1}{\eta - \eta_0} \int p'l_{\eta_0}(z)(1\{p'l_{\eta_0}(z) > 0\} - 1\{p'l_{\eta_0}(z) \leq 0\})(m_{U,\eta_0}(z) - m_{L,\eta_0}(z))\phi_{\eta_0}^2(z)d\nu(z). \quad (\text{C.33})$$

For (C.31), by the mean value theorem, we have

$$\begin{aligned} & \lim_{\eta \rightarrow \eta_0} \frac{1}{\eta - \eta_0} \int p'l_{\eta_0}(z)1\{p'l_{\eta_0}(z) > 0\}(m_{U,\eta}(z) - m_{U,\eta_0}(z))\phi_{\eta_0}^2(z)d\nu(z) \\ &= \lim_{\eta \rightarrow \eta_0} \int p'l_{\eta_0}(z)1\{p'l_{\eta_0}(z) > 0\} \frac{\partial}{\partial \eta} m_{U,\eta}(z) \Big|_{\eta=\bar{\eta}(z,\eta)} \phi_{\eta_0}^2(z)d\nu(z) \\ &= \int p'l_{\eta_0}(z)1\{p'l_{\eta_0}(z) > 0\} \frac{\partial}{\partial \eta} m_{U,\eta}(z) \Big|_{\eta=\eta_0} \phi_{\eta_0}^2(z)d\nu(z), \end{aligned} \quad (\text{C.34})$$

where the first equality holds for each p for some $\bar{\eta}(p, \eta)$ a convex combination of η and η_0 . The second equality follows from Lemmas C.2 and C.7, $\|p\| = 1$, and Assumption 2.3 (ii) justifying the use of the dominated convergence theorem. Similarly, for (C.32), we have

$$\begin{aligned} & \lim_{\eta \rightarrow \eta_0} \frac{1}{\eta - \eta_0} \int p'l_{\eta_0}(z)1\{p'l_{\eta_0}(z) \leq 0\}(m_{L,\eta}(z) - m_{L,\eta_0}(z))\phi_{\eta_0}^2(z)d\nu(z) \\ &= \int p'l_{\eta_0}(z)1\{p'l_{\eta_0}(z) \leq 0\} \frac{\partial}{\partial \eta} m_{L,\eta}(z) \Big|_{\eta=\eta_0} \phi_{\eta_0}^2(z)d\nu(z). \end{aligned} \quad (\text{C.35})$$

Hence, by (C.29)-(C.34), integration by parts, and (C.33) being 0 by Lemma C.8, we obtain

$$\left. \frac{\partial v(p, \Theta_0(P_\eta))}{\partial \eta} \right|_{\eta=\eta_0} = 2 \int p'l_{\eta_0}(z)\zeta_{p,\eta_0}(y_L, y_U, z)\dot{v}_{\eta_0}(y_L, y_U|z)v_{\eta_0}(y_L, y_U, z)d\lambda(y_L, y_U)\phi_{\eta_0}^2(z)d\nu(z) \quad (\text{C.36})$$

$$+ 2 \int w(z)p'\nabla_z m_{p,\eta_0}(z)\dot{\phi}_{\eta_0}(z)\phi_{\eta_0}(z)d\nu(z). \quad (\text{C.37})$$

Using $\dot{h}_{\eta_0} = \dot{v}_{\eta_0}\phi_{\eta_0} + v_{\eta_0}\dot{\phi}_{\eta_0}$, $\int v_{\eta_0}^2 d\lambda = 1$, and $\int \dot{h}_{\eta_0} h_{\eta_0} d\mu = 0$, we may rewrite this as

$$\left. \frac{\partial v(p, \Theta_0(P_\eta))}{\partial \eta} \right|_{\eta=\eta_0} = 2 \int \{w(z)p'\nabla_z m_{p,\eta_0}(z) - v(p, \Theta_0(P_{\eta_0})) + p'l_{\eta_0}(z)\zeta_{p,\eta_0}(y_L, y_U, z)\} \dot{h}_{\eta_0}(x)h_{\eta_0} d\mu(x). \quad (\text{C.38})$$

Therefore, the conclusion of the lemma follows. ■

Lemma C.10. *Let $\eta \mapsto h_\eta$ be a curve in L_μ^2 defined in (C.2)-(C.3). Suppose Assumptions 2.1-2.2, and 3.1 hold. Suppose further that $P \in \mathbf{P}$. Then, there is a neighborhood N of $\eta = 0$ such that the map $(p, \eta_0) \mapsto \left. \frac{\partial v(p, \Theta_0(P_\eta))}{\partial \eta} \right|_{\eta=\eta_0}$ is uniformly bounded on $\mathbb{S}^\ell \times N$.*

Proof of Lemma C.10. By Lemma C.9 and the triangle inequality,

$$\begin{aligned} \left| \frac{\partial v(p, \Theta_0(P_\eta))}{\partial \eta} \Big|_{\eta=\eta_0} \right| &= 2 \left| \int \{w(z)p' \nabla_z m_{p, \eta_0}(z) - v(p, \Theta_0(P_{\eta_0}))\} \dot{h}_{\eta_0}(x) h_{\eta_0}(x) d\mu(x) \right| \\ &\quad + 2 \left| \int p' l_{\eta_0}(z) \zeta_{p, \eta_0}(y_L, y_U, z) \dot{h}_{\eta_0}(x) h_{\eta_0}(x) d\mu(x) \right|. \end{aligned} \quad (\text{C.39})$$

By Assumption 2.2 and $\|p\| = 1$, uniformly on N ,

$$\|w(z)p' \nabla_z m_{p, \eta}(z)\|_{L_\mu^\infty} \leq \sup_{z \in \mathcal{Z}} |w(z)| \times \|\nabla_z m_{p, \eta}(z)\|_{L_\mu^\infty} \leq \sup_{z \in \mathcal{Z}} |w(z)| \times (\sup_{z \in \mathcal{Z}} |\nabla_z m_{L, \eta}(z)| + \sup_{z \in \mathcal{Z}} |\nabla_z m_{U, \eta}(z)|) < \infty. \quad (\text{C.40})$$

where the last inequality follows from Lemma C.1. This ensures that $(p, \eta) \mapsto v(p, \Theta_0(P_\eta))$ is uniformly bounded on $\mathbb{S}^\ell \times N$. We therefore have

$$\begin{aligned} \left| \int \{w(z)p' \nabla_z m_{p, \eta_0}(z) - v(p, \Theta_0(P_{\eta_0}))\} \dot{h}_{\eta_0}(x) h_{\eta_0}(x) d\mu(x) \right| \\ \leq \sup_{(p, \eta_0, z) \in \mathbb{S}^\ell \times N \times \mathcal{Z}} |w(z)p' \nabla_z m_{p, \eta_0}(z) - v(p, \Theta_0(P_{\eta_0}))| \|\dot{h}_{\eta_0}\|_{L_\mu^2} \|h_{\eta_0}\|_{L_\mu^2} < \infty. \end{aligned} \quad (\text{C.41})$$

Further, by Assumption 2.1 (ii) and Lemmas C.3 and C.6, it follows that $\sup_{x \in \mathcal{X}} |\zeta_{p, \eta_0}(x)| \leq 2\bar{\epsilon}^{-1} \sup_{u \in D'} |q(u)| < \infty$ for all $(p, \eta_0) \in \mathbb{S}^\ell \times N$. Therefore, by the triangle and Cauchy-Schwarz inequalities and Lemma C.2, we have

$$\begin{aligned} \left| \int p' l_{\eta_0}(z) \zeta_{p, \eta_0}(y_L, y_U, z) \dot{h}_{\eta_0}(x) h_{\eta_0}(x) d\mu(x) \right| \\ \leq 2\bar{\epsilon}^{-1} \sup_{u \in D'} q(u) \int \|l_{\eta_0}(z)\| \|\dot{h}_{\eta_0}(x) h_{\eta_0}(x)\| d\mu(x) \leq 2\bar{\epsilon}^{-1} \sup_{u \in D'} q(u) \|\dot{h}_{\eta_0}\|_{L_\mu^2} E_{\eta_0} [\|l_{\eta_0}(z)\|^2]^{1/2} < \infty. \end{aligned} \quad (\text{C.42})$$

By (C.39), (C.41), and (C.42), the conclusion of the lemma follows. ■

Lemma C.11. *Let $\eta \mapsto h_\eta$ be a curve in L_μ^2 defined in (C.2)-(C.3). Suppose Assumptions 2.1-2.2, and 3.1 hold. Suppose further that $P \in \mathbf{P}$. Then, there is a neighborhood N of 0 such that the map $(p, \eta_0) \mapsto \frac{\partial v(p, \Theta_0(P_\eta))}{\partial \eta} \Big|_{\eta=\eta_0}$ is continuous at all $(p, \eta_0) \in \mathbb{S}^\ell \times N$.*

Proof of Lemma C.11. Let (p_n, η_n) be a sequence such that $(p_n, \eta_n) \rightarrow (p, \eta_0)$. For each $(p, \eta, z) \in \mathbb{S}^\ell \times N \times \mathcal{Z}$, let $\gamma_{p, \eta}(z) \equiv w(z)p' \nabla_z m_{p, \eta}(z) - v(p, \Theta_0(P_\eta))$. We first show that $(p, \eta) \mapsto \gamma_{p, \eta}(z)$ and $(p, \eta) \mapsto \zeta_{p, \eta}(x)$ are continuous *a.e.* By Lemma C.7, $\eta \mapsto \nabla_z m_{L, \eta}(z)$ and $\eta \mapsto \nabla_z m_{U, \eta}(z)$ are continuous for every $z \in \mathcal{Z}$. Further, if $(p_n, \eta_n) \rightarrow (p, \eta_0)$, then

$$\mu(\{x : \lim_{n \rightarrow \infty} 1\{p'_n l_{\eta_n}(z) > 0\} = 1\{p' l_{\eta_0}(z) > 0\}\}) = 1$$

by the continuity of $(p, \eta) \mapsto p' l_\eta(z)$ implied by (C.11) and $\mu(\{x : p' l_{\eta_0}(z) = 0\}) = 0$ by Assumption 3.1 (ii). Hence, $(p, \eta) \mapsto w(z)p' \nabla_z m_{p, \eta}(z)$ is continuous *a.e.* Note that by (C.4), $v(p, \Theta_0(P_\eta)) = \int w(z)p' \nabla_z m_{p, \eta}(z) (1 + 2(\eta - \eta_0) \bar{\phi}_{\eta_0}(z)) \phi_{\eta_0}^2(z) d\nu(z)$. Hence, as $(p_n, \eta_n) \rightarrow (p, \eta_0)$, it follows that

$$\begin{aligned} \lim_{n \rightarrow \infty} v(p_n, \Theta_0(P_{\eta_n})) &= \lim_{n \rightarrow \infty} \int w(z)p'_n \nabla_z m_{p_n, \eta_n}(z) (1 + 2(\eta_n - \eta_0) \bar{\phi}_{\eta_0}(z)) \phi_{\eta_0}^2(z) d\nu(z) \\ &= \int \lim_{n \rightarrow \infty} w(z)p'_n \nabla_z m_{p_n, \eta_n}(z) (1 + 2(\eta_n - \eta_0) \bar{\phi}_{\eta_0}(z)) \phi_{\eta_0}^2(z) d\nu(z) = v(p, \Theta_0(P_{\eta_0})), \end{aligned} \quad (\text{C.43})$$

where the second equality follows from $w(z)p' \nabla_z m_{p, \eta}(z)$ and $\bar{\phi}(z)$ being bounded on $\mathbb{S}^\ell \times N \times \mathcal{Z}$ and an application of the dominated convergence theorem, while the last equality follows from the continuity of $(p, \eta) \mapsto w(z)p' \nabla_z m_{p, \eta}(z)$ for almost all z . Hence, $(p, \eta) \mapsto \gamma_{p, \eta}(z)$ is continuous *a.e.*

The maps $(p, \eta) \mapsto r_{j, \eta}^{-1}(z) q(y_j - m_{j, \eta}(z))$, $j = L, U$ are continuous for almost all x by Assumption 2.1 (ii), (C.12), and $\eta \mapsto m_{j, \eta}(z)$ being continuous for almost all z for $j = L, U$ by Lemma C.7. Since $(p, \eta) \mapsto 1\{p' l_\eta(z) > 0\}$ is continuous for almost all z as shown above, it then follows that $(p, \eta) \mapsto \zeta_{p, \eta}(x)$ is continuous for almost all x .

Given these results, we show $\frac{\partial v(p_n, \Theta_0(P_{\eta}))}{\partial \eta} \Big|_{\eta=\eta_n} - \frac{\partial v(p, \Theta_0(P_{\eta}))}{\partial \eta} \Big|_{\eta=\eta_0} \rightarrow 0$ as $\eta_n \rightarrow \eta_0$. Toward this end, we first note that

$$\frac{\partial v(p_n, \Theta_0(P_{\eta}))}{\partial \eta} \Big|_{\eta=\eta_n} - \frac{\partial v(p, \Theta_0(P_{\eta}))}{\partial \eta} \Big|_{\eta=\eta_0} \quad (\text{C.44})$$

$$= 2 \int \gamma_{p_n, \eta_n}(z) \dot{h}_{\eta_n}(x) h_{\eta_n} d\mu(x) - 2 \int \gamma_{p, \eta_0}(z) \dot{h}_{\eta_0}(x) h_{\eta_0} d\mu(x) \quad (\text{C.45})$$

$$+ 2 \int p'_n l_{\eta_n}(z) \zeta_{p_n, \eta_n}(x) \dot{h}_{\eta_n}(x) h_{\eta_n} d\mu(x) - 2 \int p' l_{\eta_0}(z) \zeta_{p, \eta_0}(x) \dot{h}_{\eta_0}(x) h_{\eta_0} d\mu(x). \quad (\text{C.46})$$

By Lemma C.10, the Cauchy-Schwarz and triangle inequalities,

$$\left| \int \gamma_{p_n, \eta_n}(z) \dot{h}_{\eta_n}(x) h_{\eta_n} d\mu(x) - \int \gamma_{p, \eta_0}(z) \dot{h}_{\eta_0}(x) h_{\eta_0} d\mu(x) \right| \quad (\text{C.47})$$

$$\leq \sup_{(p, \eta, z) \in \mathbb{S}^\ell \times N \times \mathcal{Z}} |\gamma_{p, \eta}(z)| (\|\dot{h}_{\eta_n}\|_{L_\mu^2} \|h_{\eta_n} - h_{\eta_0}\|_{L_\mu^2} + \|\dot{h}_{\eta_n} - \dot{h}_{\eta_0}\|_{L_\mu^2} \|h_{\eta_0}\|_{L_\mu^2}) \quad (\text{C.48})$$

$$+ \left| \int \{\gamma_{p_n, \eta_n}(z) - \gamma_{p, \eta_0}(z)\} \dot{h}_{\eta_0}(z) h_{\eta_0}(z) d\mu(x) \right| = o(1), \quad (\text{C.49})$$

where the last equality follows from $\eta \mapsto h_\eta$ being continuously Fréchet differentiable, $(p, \eta, z) \mapsto \gamma_{p, \eta}(z)$ being bounded on $\mathbb{S}^\ell \times N \times \mathcal{Z}$ as shown in Lemma C.10, the dominated convergence theorem, and $(p, \eta) \mapsto \gamma_{p, \eta}(z)$ being continuous *a.e.*

Further, we may write (C.46) as

$$\int p'_n l_{\eta_n}(z) \zeta_{p_n, \eta_n}(x) \dot{h}_{\eta_n}(x) h_{\eta_n} d\mu(x) - \int p' l_{\eta_0}(z) \zeta_{p, \eta_0}(x) \dot{h}_{\eta_0}(x) h_{\eta_0} d\mu(x) \quad (\text{C.50})$$

$$= \int p'_n l_{\eta_n}(z) \zeta_{p_n, \eta_n}(x) (\dot{h}_{\eta_n}(x) h_{\eta_n}(x) - \dot{h}_{\eta_0}(x) h_{\eta_0}(x)) d\mu(x) \quad (\text{C.51})$$

$$+ \int p'_n l_{\eta_n}(z) (\zeta_{p_n, \eta_n}(x) - \zeta_{p, \eta_0}(x)) \dot{h}_{\eta_0}(x) h_{\eta_0}(x) d\mu(x) \quad (\text{C.52})$$

$$+ \int (p'_n l_{\eta_n}(z) - p' l_{\eta_0}(z)) \zeta_{p, \eta_0}(x) \dot{h}_{\eta_0}(x) h_{\eta_0}(x) d\mu(x). \quad (\text{C.53})$$

By Assumptions 2.2 and 2.3 (ii), (C.11), and $\bar{\phi}$ being continuously differentiable, $(\eta, z) \mapsto \|l_\eta(z)\|$ is continuous on $N \times \mathcal{Z}$. Hence, it achieves a finite maximum on $N \times \mathcal{Z}$. Further, by Lemmas C.3, C.6 and Assumption 2.1 (ii), $\sup_{x \in \mathcal{X}} |\zeta_{p, \eta_0}(x)| \leq 2\bar{\epsilon}^{-1} \sup_{u \in D'} |q(u)| < \infty$ for all $(p, \eta_0) \in \mathbb{S}^\ell \times N$. By Cauchy-Schwarz inequality, and $\|p_n\| \leq 1$ for all n , it then follows that

$$\begin{aligned} & \int p'_n l_{\eta_n}(z) (\dot{h}_{\eta_n}(x) h_{\eta_n}(x) - \dot{h}_{\eta_0}(x) h_{\eta_0}(x)) d\mu(x) \\ & \leq \sup_{(\eta, z) \in N \times \mathcal{Z}} \|l_\eta(z)\| \sup_{(p, \eta) \in \mathbb{S}^\ell \times N} \sup_{x \in \mathcal{X}} |\zeta_{p, \eta}(x)| (\|\dot{h}_{\eta_n}\|_{L_\mu^2} \|h_{\eta_n} - h_{\eta_0}\|_{L_\mu^2} + \|\dot{h}_{\eta_n} - \dot{h}_{\eta_0}\|_{L_\mu^2} \|h_{\eta_0}\|_{L_\mu^2}) = o(1), \end{aligned} \quad (\text{C.54})$$

where the last equality follows from $\eta \mapsto h_\eta$ being continuously Fréchet differentiable. Further, by the almost everywhere continuity of $(p, \eta) \mapsto p' l_\eta(\zeta_{p, \eta} - \zeta_{p, \eta_0})$,

$$\begin{aligned} & \lim_{n \rightarrow \infty} \int p'_n l_{\eta_n}(z) (\zeta_{p_n, \eta_n}(x) - \zeta_{p, \eta_0}(x)) \dot{h}_{\eta_0}(x) h_{\eta_0}(x) d\mu(x) \\ & = \int \lim_{n \rightarrow \infty} p'_n l_{\eta_n}(z) (\zeta_{p_n, \eta_n}(x) - \zeta_{p, \eta_0}(x)) \dot{h}_{\eta_0}(x) h_{\eta_0}(x) d\mu(x) = 0. \end{aligned} \quad (\text{C.55})$$

where the first equality follows from the dominated convergence theorem. Finally, again by the dominated convergence theorem,

$$\begin{aligned} & \lim_{n \rightarrow \infty} \int (p'_n l_{\eta_n}(z) - p' l_{\eta_0}(z)) \zeta_{p, \eta_0}(x) \dot{h}_{\eta_0}(x) h_{\eta_0}(x) d\mu(x) \\ & = \int \lim_{n \rightarrow \infty} (p'_n l_{\eta_n}(z) - p' l_{\eta_0}(z)) \zeta_{p, \eta_0}(x) \dot{h}_{\eta_0}(x) h_{\eta_0}(x) d\mu(x) = 0. \end{aligned} \quad (\text{C.56})$$

By (C.44)-(C.56), we conclude that $\frac{\partial v(p_n, \Theta_0(P_{\eta_n}))}{\partial \eta} \Big|_{\eta=\eta_n} - \frac{\partial v(p, \Theta_0(P_{\eta}))}{\partial \eta} \Big|_{\eta=\eta_0} \rightarrow 0$ as $\eta_n \rightarrow \eta_0$. This establishes the claim of the lemma. ■

Theorem C.2. *Suppose Assumptions 2.1-2.3, and 3.2 hold. Then, the mapping $\rho : \mathbf{P} \rightarrow \mathcal{C}(\mathbb{S}^\ell)$ pointwise defined by $\rho(h_\eta)(p) \equiv v(p, \Theta_0(P_\eta))$ for $h_\eta = \sqrt{dP_\eta/d\mu}$ is then pathwise weak differentiable at $h_0 \equiv \sqrt{dP_0/d\mu}$. Moreover, the derivative $\dot{\rho} : \dot{\mathbf{P}} \rightarrow \mathcal{C}(\mathbb{S}^\ell)$ satisfies:*

$$\dot{\rho}(\dot{h}_0)(p) = 2 \int \{w(z)p' \nabla_z m_p(z) - v(p, \Theta_0(P_0)) + p'l(z)\zeta_p(x)\} \dot{h}_0(x) h_0(x) d\mu(x). \quad (\text{C.57})$$

Proof of Theorem C.2. We first show that $\partial \rho(P_\eta)/\partial \eta \Big|_{\eta=0}$ is the pathwise weak derivative of ρ . For this, $\dot{\rho}(\dot{h}_0) \in \mathcal{C}(\mathbb{S}^\ell)$ for all $\dot{h}_0 \in \dot{\mathbf{S}}$ as implied by Lemmas C.9 and C.11. Linearity of $\dot{\rho}$ is immediate, while continuity follows by noting that by the Cauchy-Schwarz inequality and $\|p\| = 1$,

$$\begin{aligned} & \sup_{\|\dot{h}_0\|_{L_\mu^2} = 1} \|\dot{\rho}(\dot{h}_0)\|_\infty \\ & \leq 2 \left\{ \sup_{p \in \mathbb{S}^\ell} \|w(z)p' \nabla_z m_p(z) - v(p, \Theta_0(P))\|_{L_\mu^\infty} + \sup_{z \in \mathcal{Z}} \|l(z)\| \times \sup_{u \in \mathcal{D}'} |q(u)| \right\} \|\dot{h}_0\|_{L_\mu^2} \|h_0\|_{L_\mu^2} < \infty, \end{aligned} \quad (\text{C.58})$$

where we exploited (C.40), Assumption 2.1 (ii), and the fact that $z \mapsto \|l(z)\|$ being continuous hence achieves a finite maximum on \mathcal{Z} by Assumptions 2.1 (i), 2.2, and $P \in \mathbf{P}$. Let $\eta \mapsto h_\eta$ be a curve in L_μ^2 defined in (C.2)-(C.3). For each $p \in \mathbb{S}^\ell$, by the mean value theorem,

$$\lim_{\eta_0 \rightarrow 0} \int_{\mathbb{S}^\ell} \frac{v(p, P_{\eta_0}) - v(p, P_0)}{\eta_0} dB(p) = \lim_{\eta_0 \rightarrow 0} \int_{\mathbb{S}^\ell} \frac{\partial v(p, P_\eta)(p)}{\partial \eta} \Big|_{\eta=\bar{\eta}(p, \eta_0)} dB(p) \quad (\text{C.59})$$

$$= \int_{\mathbb{S}^\ell} \frac{\partial v(p, P_\eta)(p)}{\partial \eta} \Big|_{\eta=0} dB(p) = \int_{\mathbb{S}^\ell} \dot{\rho}(\dot{h}_0)(\tau) dB(p), \quad (\text{C.60})$$

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 C.10 justifying the use of the dominated convergence theorem, while the final equality follows by Lemma C.11 and the definition of $\dot{\rho} : \dot{\mathbf{P}} \rightarrow \mathcal{C}(\mathbb{S}^\ell)$. Eqs. (C.59)-(C.60) hold for any \dot{h}_0 in the tangent space $\dot{\mathbf{U}}$ of the curve defined in (C.2)-(C.3). As discussed in the proof of Theorem C.1, $\dot{\mathbf{U}}$ is dense in $\dot{\mathbf{S}}$. Since $\dot{\rho}$ is continuous, Eqs. (C.59)-(C.60) then hold for any $\dot{h}_0 \in \dot{\mathbf{P}}$. This completes the proof. ■

Proof of Theorem 3.1. Let $\mathbf{B} \equiv \mathcal{C}(\mathbb{S}^\ell)$ and let \mathbf{B}^* be the set of finite Borel measures on \mathbb{S}^ℓ , which is the norm dual of \mathbf{B} by Corollary 14.15 in Aliprantis and Border (2006). By Theorem C.2, ρ has pathwise weak derivative $\dot{\rho}$. For each $B \in \mathbf{B}^*$, define

$$\dot{\rho}^T(B)(x) \equiv \int_{\mathbb{S}^\ell} 2\{w(z)p' \nabla_z m_p(z) - v(p, \Theta_0(P)) + p'l(z)\zeta_p(x)\} h_0(x) dB(p). \quad (\text{C.61})$$

We show that (i) $\dot{\rho}^T$ 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 operator of $\dot{\rho}$.

We first note that $(p, z) \mapsto p'l(z)$ is continuous in z for each p by Assumption 2.2 and measurable in p for each z . Thus, $(p, x) \mapsto p'l(z)$ is jointly measurable by Lemma 4.51 in Aliprantis and Border (2006). This implies the joint measurability of $(p, x) \mapsto 1\{p'l(z) > 0\}$. A similar argument also ensures the joint measurability of $p' \nabla_z m_L(z)$ and $p' \nabla_z m_U(z)$. By the joint measurability of $(p, x) \mapsto (w(z), 1\{p'l(z) > 0\}, p' \nabla_z m_L(z), p' \nabla_z m_U(z))$ and Assumption 2.2, $(p, x) \mapsto w(z)p' \nabla_z m_p(z)$ is jointly measurable. By the proof of Theorem 2.1, $v(p, \Theta_0(p))$ is differentiable in p and is therefore continuous, implying $(p, x) \mapsto v(p, \Theta_0(p))$ is jointly measurable. Further, r_L and r_U are measurable by $P \in \mathbf{P}$ satisfying Assumption 3.2 (iii). $q(y_L - m_L(z)), q(y_U - m_U(z))$ are measurable by Assumption 2.1 and $P \in \mathbf{P}$ satisfying Assumption 2.3 (iv). Hence, $(p, x) \mapsto \zeta_p(x)$ is jointly measurable. Therefore, the map $(p, x) \mapsto (w(z)p' \nabla_z m_p(z), v(p, \Theta_0(p)), p'l(z), \zeta_p(x), h_0(x))'$ is jointly measurable by Lemma 4.49 in Aliprantis and Border (2006). Hence, the map

$$(p, x) \mapsto 2\{w(z)p' \nabla_z m_p(z) - v(p, \Theta_0(P_0)) + p'l(z)\zeta_p(x)\} \quad (\text{C.62})$$

is jointly measurable by the measurability of the composite map.

Moreover, for $|B|$ the total variation of the measure B , by (C.40), we have

$$\int \left(\int_{\mathbb{S}^\ell} 2\{w(z)p'\nabla_z m_p(z) - v(p, \Theta_0(P_0))\}h_0(z)dB(p) \right)^2 d\mu(x) \leq 16 \times \sup_{p \in \mathbb{S}^\ell} \|w(z)p'\nabla_z m_{p,\eta}(z)\|_{L_\mu^\infty}^2 \times |B|^2 < \infty. \quad (\text{C.63})$$

Further,

$$\int \left(\int_{\mathbb{S}^\ell} 2p'l(z)\zeta_p(x)h_0(x)dB(p) \right)^2 d\mu(x) \leq 16 \int |p'l(z)|^2 h_0(x)^2 d\mu(x) \times \bar{\epsilon}^{-2} \times \sup_{u \in D'} |q(u)|^2 \times |B|^2 < \infty, \quad (\text{C.64})$$

by Assumption 2.1 and $p \in \mathbf{P}$ satisfying Assumptions 2.3 and 3.2. Therefore, $\dot{\rho}^T(B) \in L_\mu^2$ for each $B \in \mathbf{B}^*$.

By Fubini's theorem and Assumption 2.1 (iv), we have

$$\begin{aligned} \int \int_{\mathbb{S}^\ell} 2\{w(z)p'\nabla_z m_p(z) - v(p, \Theta_0(P_0)) + p'l(z)\zeta_p(x)\}h_0(x)dB(p)h_0(x)d\mu(x) \\ = \int_{\mathbb{S}^\ell} \int 2\{w(z)p'\nabla_z m_p(z) - v(p, \Theta_0(P_0)) + p'l(z)\zeta_p(x)\}h_0^2(x)d\mu(x)dB(p) = 0, \end{aligned} \quad (\text{C.65})$$

where we exploited $v(p, \Theta_0(P_0)) = E[w(Z)p'\nabla_z m_p(Z)]$ and $E[\zeta_p(x)|Z = z] = 1\{p'l(z) \leq 0\}E[q(Y_L - m_L(Z))|Z = z] + 1\{p'l(z) > 0\}E[q(Y_U - m_U(Z))|Z = z] = 0$, $P - a.s.$ Thus, by Theorem C.1 and (C.61), $\dot{\rho}^T(B) \in \dot{\mathbf{S}}$ for all $B \in \mathbf{B}^*$. Further, for any $\dot{h}_0 \in \dot{\mathbf{S}}$, again by interchanging the order of integration

$$\int_{\mathbb{S}^\ell} \dot{\rho}(\dot{h}_0)(p)dB(p) = \int_{\mathcal{X}} \dot{h}_0(x)\dot{\rho}^T(B)(x)d\mu(x), \quad (\text{C.66})$$

which ensures that $\dot{\rho}^T : \mathbf{B}^* \rightarrow \dot{\mathbf{P}}$ is the adjoint of $\dot{\rho} : \dot{\mathbf{P}} \rightarrow \mathbf{B}$.

Since $\dot{\mathbf{S}}$ is linear by Theorem C.1, Theorem C.2 and Theorem 5.2.1 in Bickel, Klassen, Ritov, and Wellner (1993) establishes that

$$\begin{aligned} \text{Cov} \left(\int_{\mathbb{S}^\ell} \mathbb{G}(p)dB_1(p), \int_{\mathbb{S}^\ell} \mathbb{G}(p)dB_2(p) \right) &= \frac{1}{4} \int_{\mathcal{X}} \dot{\rho}^T(B_1)(x)\dot{\rho}^T(B_2)(x)d\mu(x) \\ &= \int_{\mathbb{S}^\ell} \int_{\mathbb{S}^\ell} E[\{w(z)p'\nabla_z m_p(z) - v(p, \Theta_0(P_0)) + p'l(z)\zeta_p(x)\} \{w(z)q'\nabla_z m_q(z) - v(q, \Theta_0(P_0)) + q'l(z)\zeta_q(x)\}] dB_1(p)dB_2(q), \end{aligned} \quad (\text{C.67})$$

for any $B_1, B_2 \in \mathbf{B}^*$ by Fubini's theorem. Letting B_1 and B_2 be the degenerate measures at p and q in (C.67) gives the desired result. ■

APPENDIX D: Proof of Theorem 4.1.

In this appendix, we establish Theorem 4.1. Throughout, let $Y_{p,i} \equiv 1\{p'l(Z_i) \leq 0\}Y_{L,i} + 1\{p'l(Z_i) > 0\}Y_{U,i}$ and let $\bar{v}_n(p) \equiv \frac{1}{n} \sum_{i=1}^n p'\hat{l}_{i,h}(Z_i)Y_{p,i}$. The proof of Theorem 4.1 proceeds by decomposing $\sqrt{n}(\hat{v}_n(p) - v(p, \Theta_0(P_0)))$ as follows:

$$\begin{aligned} \sqrt{n}(\hat{v}_n(p) - v(p, \Theta_0(P_0))) &= \sqrt{n}(\hat{v}_n(p) - \bar{v}_n(p)) + \sqrt{n}(\bar{v}_n(p) - E[\bar{v}_n(p)]) + \sqrt{n}(E[\bar{v}_n(p)] - v(p, \Theta_0(P_0))) \\ &\equiv G_{1n}(p) + G_{2n}(p) + G_{3n}(p). \end{aligned} \quad (\text{D.1})$$

G_{1n} is the difference between \hat{v}_n and the infeasible estimator \bar{v}_n which requires the knowledge of $Y_{p,i}$. G_{2n} represents a properly centered version of \bar{v}_n , and G_{3n} is the asymptotic bias of \bar{v}_n . The auxiliary lemmas are then used to show the following results:

Step 1: Lemma D.2 shows $G_{1n} = o_p(1)$ uniformly in $p \in \mathbb{S}^\ell$ using the result of Lemma D.1, while Lemma D.5 shows $G_{3n} = o(1)$ uniformly in $p \in \mathbb{S}^\ell$.

Step 2: Lemmas D.3 and D.4 then establish that $G_{2n} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_p(X_i) + o_p(1)$ uniformly in $p \in \mathbb{S}^\ell$, and Lemma D.6 establishes that $\{\psi_p : p \in \mathbb{S}^\ell\}$ is a P -Donsker class.

Step 3: Combining Steps 1-2 and (D.1) gives the main claim of Theorem 4.1.

Before proceeding further, we introduce one more piece of notation. For each $p \in \mathbb{S}^\ell$, define

$$p_n(x_i, x_j; p) \equiv -\left(\frac{1}{h}\right)^{\ell+1} p' \nabla_z K\left(\frac{z_i - z_j}{h}\right)(y_{p,i} - y_{p,j}), \quad \text{and} \quad r_n(x_i; p) \equiv E[p_n(X_i, X_j; p) | X_i = x_i]. \quad (\text{D.2})$$

Lemma D.1. *Let $\mathcal{H}_n \equiv \{\tilde{p}_n : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R} : \tilde{p}_n(x, x'; p) = -p' \nabla_z K\left(\frac{z - z'}{h}\right)(y_p - y'_p), p \in \mathbb{B}^\ell\}$ and $\mathcal{G}_n \equiv \{\tilde{q}_n : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R} : \tilde{q}_n(x, x'; p) = \tilde{p}_n(x, x'; p) - \tilde{r}_n(x, p) - \tilde{r}_n(x', p) - E[\tilde{r}_n(x, p)], p \in \mathbb{B}^\ell\}$ where $\tilde{r}_n(x; p) \equiv E[\tilde{p}_n(X_i, X_j; p) | X_i = x]$ and $\mathbb{B}^\ell \equiv \{p \in \mathbb{R}^\ell : \|p\| \leq 1\}$ is the unit ball in \mathbb{R}^ℓ . Suppose Assumption 4.2 holds. Then, \mathcal{H}_n and \mathcal{G}_n are Euclidean in the sense of Sherman (1994a) with envelope functions $H : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ and $G : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ such that $E[H(X_i, X_j)^2] < \infty$ and $E[G(X_i, X_j)^2] < \infty$.*

Proof of Lemma D.1. Let $\mathcal{F}_g \equiv \{f : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R} : f(x, x', p) = p' g(x, x'), p \in \mathbb{B}^\ell\}$, where $g : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}^\ell$ is a known function. By the Cauchy-Schwarz inequality, for any $p, q \in \mathbb{B}^\ell$, we therefore have $|f(x, x', p) - f(x, x', q)| \leq \|g(x, x')\| \|p - q\|$. Hence, by Lemma 2.13 in Pakes and Pollard (1989), \mathcal{F}_g is Euclidean with the envelope function $F_g(x, x') = g(x, x')' p_0 + M \|g(x, x')\|$ for some $p_0 \in \mathbb{B}^\ell$, where $M = 2\sqrt{\ell} \sup_{p \in \mathbb{B}^\ell} \|p - p_0\|$, which can be further bounded from above by $4\sqrt{\ell}$. Hence, we may take the envelope function as $F_g(x, x') = (1 + 4\sqrt{\ell}) \|g(x, x')\|$.

By Lemma 2.4 in Pakes and Pollard (1989), the class of sets $\{x \in \mathcal{X} : p'l(z) > 0\}$ is a VC-class, which in turn implies that the function classes $\mathcal{F}_{\phi_1} \equiv \{f : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R} : f(x, x', p) = 1\{p'l(z) > 0\}\}$ and $\mathcal{F}_{\phi_2} \equiv \{f : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R} : f(x, x', p) = 1\{p'l(z') > 0\}\}$ are Euclidean, where $x = (y_L, y_U, z)$ and $x' = (y'_L, y'_U, z')$ with envelope function $F_{\phi_j}(x, x') = 1, j = 1, 2$.

Note that we may write

$$\tilde{p}_n(x, x'; p) = -p' \nabla_z K\left(\frac{z - z'}{h}\right) \{(y_U - y_L) 1\{p'l(z) > 0\} - y_L - (y'_U - y'_L) 1\{p'l(z') > 0\} + y'_L\}.$$

Hence, \mathcal{H}_n can be written as the combination of classes of functions as $\mathcal{H}_n = \mathcal{F}_{g_1} \cdot \mathcal{F}_{\phi_1} + \mathcal{F}_{g_2} + \mathcal{F}_{g_3} \cdot \mathcal{F}_{\phi_2} + \mathcal{F}_{g_4}$, where

$$\begin{aligned} g_1(x, x') &= -\nabla_z K\left(\frac{z - z'}{h}\right)(y_U - y_L), & g_2(x, x') &= \nabla_z K\left(\frac{z - z'}{h}\right)y_L \\ g_3(x, x') &= \nabla_z K\left(\frac{z - z'}{h}\right)(y'_U - y'_L), & g_4(x, x') &= -\nabla_z K\left(\frac{z - z'}{h}\right)y'_L. \end{aligned}$$

By Lemma 2.14 in Pakes and Pollard (1989) and \mathcal{F}_{ϕ_1} and \mathcal{F}_{ϕ_2} having constant envelope functions, \mathcal{H}_n is Euclidean with the envelope function $F_{g_1} + F_{g_2} + F_{g_3} + F_{g_4}$. Hence, \mathcal{H}_n is Euclidean with the envelope function $H(x, x') \equiv 8(1 + 4\sqrt{\ell}) \sup_{y \in D} |y| \sup_{h > 0} \|\nabla_z K\left(\frac{z - z'}{h}\right)\|$. By Assumption 4.2, $E[\sup_{h > 0} \|\nabla_z K\left(\frac{z_i - z_j}{h}\right)\|^2] < \infty$, which in turn implies $E[H(X_i, X_j)^2] < \infty$. This shows the claim of the lemma for \mathcal{H}_n . Showing \mathcal{G}_n is Euclidean is similar. Hence, the rest of the proof is omitted. ■

Lemma D.2. *Suppose Assumptions 2.3, 3.1, and 4.2 hold. Suppose further that $\tilde{h} \rightarrow 0$ and $n\tilde{h}^{4(\ell+1)} \rightarrow \infty$. Then, uniformly in $p \in \mathbb{S}^\ell$, $\hat{v}_n(p) - \bar{v}_n(p) = o_p(n^{-1/2})$.*

Proof of Lemma D.2. By (35), we may write

$$\begin{aligned} \hat{v}_n(p) - \bar{v}_n(p) &= \frac{1}{n} \sum_{i=1}^n p' \hat{l}_{i,h}(Z_i) (\hat{Y}_{p,i} - Y_{p,i}) \\ &= \frac{1}{n} \sum_{i=1}^n p' \hat{l}_{i,h}(Z_i) [(1\{p' \hat{l}_{i,\tilde{h}}(Z_i) > 0\} - 1\{p'l(Z_i) > 0\}) (Y_{U,i} - Y_{L,i})] = \frac{1}{n} \sum_{i=1}^n u_{i,n}(p), \quad (\text{D.3}) \end{aligned}$$

where $u_{i,n}(p) = p' \hat{l}_{i,h}(Z_i) [(1\{p' \hat{l}_{i,\tilde{h}}(Z_i) > 0\} - 1\{p'l(Z_i) > 0\}) (Y_{U,i} - Y_{L,i})]$. For the conclusion of the lemma, it therefore suffices to show $E[\sup_{p \in \mathbb{S}^\ell} |u_{i,n}(p)|^2] = o(n^{-1})$.

By $\|p\| = 1$, the Cauchy-Schwarz inequality, and Assumption 2.3,

$$\begin{aligned} E[\sup_{p \in \mathbb{S}^\ell} |u_{i,n}(p)|^2] &\leq 2 \sup_{y \in D} |y| E[\|\hat{l}_{i,h}(Z_i)\|^2] \sup_{p \in \mathbb{S}^\ell} |1\{p' \hat{l}_{i,\tilde{h}}(Z_i) > 0\} - 1\{p'l(Z_i) > 0\}|^2 \\ &\leq 2 \sup_{y \in D} |y| E[\|\hat{l}_{i,h}(Z_i)\|^4]^{1/2} P(\text{sgn}(p' \hat{l}_{i,\tilde{h}}(Z_i)) \neq \text{sgn}(p'l(Z_i)), \exists p \in \mathbb{S}^\ell), \end{aligned} \quad (\text{D.4})$$

where $E[\|\hat{l}_{i,h}(Z_i)\|^4] < \infty$ under our choice of h . Hence, for the desired result, it suffices to show that $P(\text{sgn}(p' \hat{l}_{i,\tilde{h}}(Z_i)) \neq \text{sgn}(p'l(Z_i)), \exists p \in \mathbb{S}^\ell) = o(n^{-1})$. By Assumption 3.1, it follows that

$$\begin{aligned} P(\text{sgn}(p' \hat{l}_{i,\tilde{h}}(Z_i)) \neq \text{sgn}(p'l(Z_i)), \exists p \in \mathbb{S}^\ell) \\ \leq P(p' \hat{l}_{i,\tilde{h}}(Z_i) > 0 \text{ and } p'l(Z_i) < 0, \exists p \in \mathbb{S}^\ell) + P(p' \hat{l}_{i,\tilde{h}}(Z_i) < 0 \text{ and } p'l(Z_i) > 0, \exists p \in \mathbb{S}^\ell). \end{aligned} \quad (\text{D.5})$$

Without loss of generality, suppose that $p' \hat{l}_{i,\tilde{h}}(Z_i) > 0$ and $p'l(Z_i) < 0$ for some $p \in \mathbb{S}^\ell$. Then, there must exist $\epsilon > 0$ such that $\sup_{p \in \mathbb{S}^\ell} |p' \hat{l}_{i,\tilde{h}}(Z_i) - E[p' \hat{l}_{i,\tilde{h}}(Z_i)|Z_i] + E[p' \hat{l}_{i,\tilde{h}}(Z_i)] + p'l(Z_i)| > \epsilon$. This is also true if $p' \hat{l}_{i,\tilde{h}}(Z_i) < 0$ and $p'l(Z_i) > 0$. Therefore, by the triangle inequality and the law of iterated expectations, we may write

$$\begin{aligned} P(\text{sgn}(p' \hat{l}_{i,\tilde{h}}(Z_i)) \neq \text{sgn}(p'l(Z_i)), \exists p \in \mathbb{S}^\ell) \\ \leq 2 \left\{ E \left[P \left(\sup_{p \in \mathbb{S}^\ell} |p' \hat{l}_{i,\tilde{h}}(Z_i) - E[p' \hat{l}_{i,\tilde{h}}(Z_i)|Z_i]| > \epsilon/2 |Z_i \right) + P \left(\sup_{p \in \mathbb{S}^\ell} |E[p' \hat{l}_{i,\tilde{h}}(Z_i)|Z_i] - p'l(Z_i)| > \epsilon/2 |Z_i \right) \right] \right\}, \end{aligned} \quad (\text{D.6})$$

where the second term in (D.6) vanishes for all n sufficiently large because the bias satisfies $\|E[\hat{l}_{i,\tilde{h}}(Z_i)|Z_i] - l(Z_i)\| \rightarrow 0$ with probability 1 as $\tilde{h} \rightarrow 0$. Hence, we focus on controlling the first term in (D.6) below.

Let $\bar{M} \equiv \sup_{z \in \mathcal{Z}} \|\nabla_z K(z)\|$ and define

$$W_n(p) \equiv \frac{\tilde{h}^{(\ell+1)}}{2\bar{M}} (p' \hat{l}_{i,\tilde{h}}(z) - E[p' \hat{l}_{i,\tilde{h}}(Z_i)|Z_i = z]) = \frac{1}{2\bar{M}(n-1)} \sum_{j=1, j \neq i}^n p' \left\{ \nabla_z K \left(\frac{z - Z_j}{\tilde{h}} \right) - E[\nabla_z K \left(\frac{z - Z_j}{\tilde{h}} \right)] \right\} \quad (\text{D.7})$$

$$\bar{\sigma}^2 \equiv E \left[\sup_{p \in \mathbb{S}^\ell} \left(\frac{1}{2\bar{M}(n-1)} \sum_{j=1, j \neq i}^n p' \left\{ \nabla_z K \left(\frac{z - Z_j}{\tilde{h}} \right) - E[\nabla_z K \left(\frac{z - Z_j}{\tilde{h}} \right)] \right\} \right)^2 \right]. \quad (\text{D.8})$$

Define $\mathcal{W} \equiv \{f : \mathcal{X} \rightarrow \mathbb{R} : f(z_j) = \frac{1}{2\bar{M}} p' \{ \nabla_z K \left(\frac{z - z_j}{\tilde{h}} \right) - E[\nabla_z K \left(\frac{z - Z_j}{\tilde{h}} \right)] \}, p \in \mathbb{S}^\ell\}$. Then by \mathbb{S}^ℓ being finite dimensional and Lemma 2.6.15 in van der Vaart and Wellner (1996), \mathcal{W} is a VC-subgraph class, which in turn implies that $\sup_Q N(\epsilon, \mathcal{W}, L_2(Q)) \leq (\frac{K}{\epsilon})^V$ for all $0 < \epsilon < K$ for some positive constants V and K by Lemma 2.6.7 in van der Vaart and Wellner (1996). Then, by W_n being independent of Z_i and Theorem 2.14.16 in van der Vaart and Wellner (1996), we have

$$\begin{aligned} P(\sup_{p \in \mathbb{S}^\ell} |p' \hat{l}_{i,\tilde{h}}(Z_i) - E[p' \hat{l}_{i,\tilde{h}}(Z_i)|Z_i]| > \epsilon/2 |Z_i = z) &= P(\|W_n\|_{\mathcal{W}} > \frac{\epsilon \tilde{h}^{(\ell+1)}}{4\bar{M}}) \\ &\leq C \left(\frac{1}{\bar{\sigma}} \right)^{2V} \left(1 \vee \frac{\epsilon \tilde{h}^{(\ell+1)}}{4\bar{M}\bar{\sigma}} \right)^{3V+1} \exp \left(- \frac{1}{2} \frac{(\frac{\epsilon \tilde{h}^{(\ell+1)}}{4\bar{M}})^2}{\bar{\sigma}^2 + (3 + \frac{\epsilon \tilde{h}^{(\ell+1)}}{4\bar{M}})/\sqrt{n}} \right), \end{aligned} \quad (\text{D.9})$$

where C is a constant that depends on V and K . Note that under the imposed conditions on \tilde{h} , we have

$$\frac{(\frac{\epsilon \tilde{h}^{(\ell+1)}}{4\bar{M}})^2}{\bar{\sigma}^2 + (3 + \frac{\epsilon \tilde{h}^{(\ell+1)}}{4\bar{M}})/\sqrt{n}} = \frac{1}{S_{1,n} + S_{2,n}}, \quad (\text{D.10})$$

where $S_{1,n} \equiv \bar{\sigma}^2 / (\frac{\epsilon \tilde{h}^{(\ell+1)}}{4\bar{M}})^2$ and $S_{2,n} \equiv (3 + \frac{\epsilon \tilde{h}^{(\ell+1)}}{4\bar{M}}) / [(\frac{\epsilon \tilde{h}^{(\ell+1)}}{4\bar{M}})^2 \sqrt{n}]$. By (D.8), $\bar{\sigma}^2 = o(n^{-2})$, which implies that $S_{1,n} = o(1/(n\tilde{h}^{(\ell+1)}))^2 = o(1)$ under our choice of \tilde{h} . Further, under our assumption, $\sqrt{n}\tilde{h}^{2(\ell+1)} \rightarrow \infty$, which in turn implies $S_{2,n} = o(1)$. This ensures that, by (D.9), $P(\sup_{p \in \mathbb{S}^\ell} |p' \hat{l}_{i,\tilde{h}}(Z_i) - E[p' \hat{l}_{i,\tilde{h}}(Z_i)|Z_i]| > \epsilon/2 |Z_i = z)$ decays

exponentially as $n \rightarrow \infty$. Hence, by (D.4), (D.5), and (D.6), we have $E[\sup_{p \in \mathbb{S}^\ell} |u_{i,n}(p)|^2] = o(n^{-1})$ as desired. This establishes the claim of the Lemma. ■

Lemma D.3. *Let $U_n(p) \equiv \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n p_n(X_i, X_j; p)$ and $\hat{U}_n(p) = \frac{2}{n} \sum_{i=1}^n r_n(X_i; p)$. Suppose Assumptions 2.1 and 4.2 hold. Suppose further that $nh^{\ell+2+\delta} \rightarrow \infty$ for some $\delta > 0$ as $h \rightarrow 0$. Then, $\sqrt{n}(\hat{U}_n(p) - U_n(p)) = o_p(1)$ uniformly in $p \in \mathbb{S}^\ell$.*

Proof of Lemma D.3. Following the same argument as in the proof of Lemma 3.1 in Powell, Stock, and Stoker (1989), we may write

$$\hat{U}_n(p) - U_n(p) = \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n q_n(X_i, X_j; p), \quad (\text{D.11})$$

where $q_n(x_i, x_j; p) = p_n(x_i, x_j; p) - r_n(x_i, p) - r_n(x_j, p) - E[r_n(X_i; p)]$. Recall that $\tilde{q}_n = h^{(\ell+1)}q_n$. By (D.2), we may then obtain the following bound:

$$\begin{aligned} E[\sup_{p \in \mathbb{S}^\ell} |\tilde{q}_n(X_i, X_j; p)|^2] &\leq 16E[\sup_{p \in \mathbb{S}^\ell} |\tilde{p}_n(X_i, X_j; p)|^2] \leq 64(\sup_{y \in D} |y|)^2 E[\|\nabla_z K((Z_i - Z_j)/h)\|^2] \\ &\leq 64(\sup_{y \in D} |y|)^2 h^\ell \int \|\nabla_z K(u)\|^2 f(z_i) f(z_i + hu) dz_i du = O(h^\ell), \end{aligned} \quad (\text{D.12})$$

where the second inequality follows from Assumption 2.1, $\|p\| = 1$ for all p , and the Cauchy-Schwarz inequality, while the third inequality uses the change of variables from (z_i, z_j) to $(z_i, u = (z_i - z_j)/h)$ with Jacobian $h^{-\ell}$. By Lemma D.1, \mathcal{G}_n is Euclidean. By Theorem 3 in Sherman (1994b) applied with $\delta_n = 1$ and $\gamma_n^2 = h^\ell$, it then follows that for some $0 < \alpha < 1$, which can be made arbitrarily close to 1, we have

$$\begin{aligned} \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n q_n(X_i, X_j; p) &= h^{-(\ell+1)} \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \tilde{q}_n(X_i, X_j; p) \\ &\leq O(h^{-(\ell+1)}) O_p(h^{\alpha\ell/2}/n) = O_p(h^{\ell(\frac{\alpha}{2}-1)-1}/n) \end{aligned} \quad (\text{D.13})$$

uniformly over \mathbb{B}^ℓ . Since α can be made arbitrarily small and $nh^{\ell+2+\delta} \rightarrow \infty$, we have $h^{\ell(\frac{\alpha}{2}-1)-1} \leq O(h^{-\frac{\ell}{2}-1-\frac{\delta}{2}}) = o(\sqrt{n})$. Therefore, $\binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n q_n(X_i, X_j; p) = o(n^{-1/2})$. This, together with $\mathbb{S}^\ell \subset \mathbb{B}^\ell$ and (D.11), establishes the claim of the lemma. ■

Lemma D.4. *Suppose Assumptions 2.1-2.3, 3.1, and 4.1-4.2 hold. Suppose further that $nh^{\ell+2+\delta} \rightarrow \infty$ for some $\delta > 0$ as $h \rightarrow 0$. Then, uniformly in $p \in \mathbb{S}^\ell$, $\sqrt{n}(\bar{v}_n(p) - E[\bar{v}_n(p)]) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_p(Z_i) + o_p(1)$.*

Proof of Lemma D.4. Note that, for each $p \in \mathbb{S}^\ell$, one may write

$$\begin{aligned} \bar{v}_n(p) &= \frac{1}{n} \sum_{i=1}^n p' \hat{l}_{i,h}(Z_i) Y_{p,i} = \frac{-2}{n} \sum_{i=1}^n p' \nabla_z \hat{f}_{i,h}(Z_i) Y_{p,i} = \frac{-2}{n(n-1)h^{\ell+1}} \sum_{i=1}^n \sum_{j=1, j \neq i}^n p' \nabla_z K\left(\frac{Z_i - Z_j}{h}\right) Y_{p,i} \\ &= -\binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \left(\frac{1}{h}\right)^{\ell+1} p' \nabla_z K\left(\frac{Z_i - Z_j}{h}\right) (Y_{p,i} - Y_{p,j}) = \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n p_n(X_i, X_j; p). \end{aligned} \quad (\text{D.14})$$

By Lemma D.3 and $nh^{\ell+2+\delta} \rightarrow \infty$, it then follows that

$$\begin{aligned} \sqrt{n}(\bar{v}_n(p) - E[\bar{v}_n(p)]) &= \sqrt{n} \left(\binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n p_n(X_i, X_j; p) - E[p_n(X_i, X_j; p)] \right) \\ &= \frac{2}{\sqrt{n}} \sum_{i=1}^n r_n(X_i; p) - E[r_n(X_i; p)] + o_p(1), \end{aligned} \quad (\text{D.15})$$

uniformly over \mathbb{S}^ℓ . By Eq. (3.15) in Powell, Stock, and Stoker (1989), we may write $r_n(x; p) = \psi_p(x) + t_n(x; p)$ with

$$t_n(x; p) \equiv \int \{\nabla_z(m_p f)(z + hu) - \nabla_z(m_p f)(z)\}K(u)du - y_p \int \{\nabla_z f(z + hu) - \nabla_z f(z)\}K(u)du, \quad (\text{D.16})$$

Hence, for the conclusion of the lemma, it suffices to show $t_n(x, \cdot)$ converges in probability to 0 uniformly in $p \in \mathbb{S}^\ell$. By Assumptions 3.1, 4.1 and $m_p(z) = 1\{p'l(z) \leq 0\}m_L(z) + 1\{p'l(z) > 0\}m_U(z)$, we have uniformly in $p \in \mathbb{S}^\ell$,

$$|\nabla_z(m_p f)(z + hu) - \nabla_z(m_p f)(z)| \leq 2M(z)\|hu\|. \quad (\text{D.17})$$

Further, by Assumptions 2.1 and 4.1, uniformly in $p \in \mathbb{S}^\ell$,

$$|y_p|\|\nabla_z f(z + hu) - \nabla_z f(z)\| \leq \sup_{y \in \mathcal{D}} |y| \times M(z)\|hu\|. \quad (\text{D.18})$$

Assumption 4.1 and (D.17)-(D.18) then imply $E[\sup_{p \in \mathbb{S}^\ell} |t_n(x; p)|^2] \leq |h|^2(2 + \sup_{y \in \mathcal{D}} |y|)^2 E[|M(Z)|^2] \int \|u\| |K(u)| du)^2 = O(h^2)$, which in turn implies $\frac{1}{\sqrt{n}} \sum_{i=1}^n t_n(x, \cdot) - E[t_n(x, \cdot)]$ converges in probability to 0 uniformly in $p \in \mathbb{S}^\ell$. This establishes the claim of the lemma. ■

Lemma D.5. *Suppose Assumptions 4.1-4.2 hold. Suppose that $nh^{2J} \rightarrow 0$. Then, uniformly in $p \in \mathbb{S}^\ell$, $E[\bar{v}_n(p)] - v(p, \Theta_0(P_0)) = o(n^{-1/2})$.*

Proof of Lemma D.5. The proof is based on the proof of Theorem 3.2 in Powell, Stock, and Stoker (1989). Hence, we briefly sketch the argument. By (D.14), the law of iterated expectations, and arguing as in (3.19) in Powell, Stock, and Stoker (1989), we obtain

$$\begin{aligned} E[\bar{v}_n(p)] &= -2E\left[\left(\frac{1}{h}\right)^{\ell+1} p' \nabla_z K\left(\frac{Z_i - Z_j}{h}\right) Y_{p,i}\right] \\ &= \frac{-2}{h} \int \int p' \nabla_z K(u) m_p(z) f(z) f(z + hu) dz du = \frac{-2}{h} \int \int K(u) m_p(z) f(z) p' \nabla_z f(z + hu) dz du, \end{aligned} \quad (\text{D.19})$$

where the second equality follows from change of variables. By Assumptions 4.1, 4.2, and Young's version of Taylor's theorem, for each $p \in \mathbb{S}^\ell$, we then obtain the expansion:

$$\sqrt{n}(E[\bar{v}_n(p)] - v(p, \Theta_0(P_0))) = b_1(p)\sqrt{nh} + b_2(p)\sqrt{nh^2} + \dots + b_{J-1}(p)\sqrt{nh^{J-1}} + O(\sqrt{nh^J}), \quad (\text{D.20})$$

where b_k is given by

$$b_k(p) = \frac{-2}{k!} \sum_{j_1, \dots, j_k} \int u^{j_1} \dots u^{j_k} K(u) du \times \int m_p(z) \sum_{i=1}^{\ell} p^{(i)} \frac{\partial^{k+1} f(z)}{\partial z_{j_1} \dots \partial z_{j_k} \partial z_i} f(z) dz, \quad p \in \mathbb{S}^\ell, \quad k = 1, \dots, J, \quad (\text{D.21})$$

which shows that the map $p \mapsto b_k(p)$ is continuous on \mathbb{S}^ℓ for $k = 1, 2, \dots, J$. This implies that the expansion in (D.20) is valid uniformly over the compact set \mathbb{S}^ℓ . By Assumption 4.2 (v) and (D.21), $b_k(p) = 0$ for all $k \leq J$ but $b_k \neq 0$ for $k = J$. By the hypothesis that $nh^{2J} \rightarrow 0$, we obtain $\sqrt{n}(E[\bar{v}_n(p)] - v(p, \Theta_0(P_0))) = O(\sqrt{nh^J}) = o(1)$. This establishes the claim of the lemma. ■

Lemma D.6. *Suppose Assumptions 2.1-2.3, and 3.2 hold. Then, $\mathcal{F} \equiv \{\psi_p : \mathcal{X} \rightarrow \mathbb{R} : \psi_p(x) = w(z)p' \nabla_z m_p(z) - v(p, \Theta_0(P)) + p'l(z)\zeta_p(x)\}$ is Donsker in $\mathcal{C}(\mathbb{S}^\ell)$.*

Proof of Lemma D.6. Let $\mathcal{F}_g \equiv \{f : \mathcal{X} \rightarrow \mathbb{R} : f(x) = p'g(x), p \in \mathbb{S}^\ell\}$, where $g : \mathcal{X} \rightarrow \mathbb{R}^\ell$ is a known function. Then by \mathbb{S}^ℓ being finite dimensional and Lemma 2.6.15 in van der Vaart and Wellner (1996), \mathcal{F}_g is a VC-subgraph class of index $\ell + 2$ with an envelope function $F(x) \equiv \|g(x)\|$. Define

$$g_1(x) \equiv w(z)(\nabla_z m_U(z) - \nabla_z m_L(z)), \quad g_2(x) \equiv w(z)\nabla_z m_L(z), \quad (\text{D.22})$$

$$g_3(x) \equiv l(z)\{r_U^{-1}(z)q(y_U - m_U(z)) - r_L^{-1}(z)q(y_L - m_L(z))\}, \quad g_4(x) \equiv l(z)r_L^{-1}(z)q(y_L - m_L(z)), \quad g_5(x) \equiv l(z). \quad (\text{D.23})$$

Then $\mathcal{F}_{g_j}, j = 1, \dots, 5$ are VC-subgraph classes. Further, let $\mathcal{F}_v \equiv \{f : \mathcal{X} \rightarrow \mathbb{R} : f(x) = v(p, \Theta_0(P)), p \in \mathbb{S}^\ell\}$. This is also finite dimensional. Hence, \mathcal{F}_v is a VC-subgraph class. Finally, let $\mathcal{F}_\phi \equiv \{f : \mathcal{X} \rightarrow \mathbb{R} : 1\{p'l(z) > 0\}, p \in \mathbb{S}^\ell\}$. Then, $F_\phi = \phi \circ \mathcal{F}_{g_5}$, where $\phi : \mathbb{R} \rightarrow \mathbb{R}$ is the monotone map $\phi(w) = 1\{w > 0\}$. By Lemma 2.6.18 in van der Vaart and Wellner (1996), \mathcal{F}_ϕ is also a VC-subgraph class.

Note that ψ_p can be written as

$$\begin{aligned} \psi_p(x) &= w(z)p'\{1\{p'l(z) > 0\}(\nabla_z m_U(z) - \nabla_z m_L(z)) + \nabla_z m_L(z)\} - v(p, \Theta_0(P)) \\ &\quad + p'l(z)\{1\{p'l(z) > 0\}\{r_U^{-1}(z)q(y_U - m_U(z)) - r_L^{-1}(z)q(y_L - m_L(z))\} + r_L^{-1}(z)q(y_L - m_L(z))\}. \end{aligned} \quad (\text{D.24})$$

Therefore, $\mathcal{F} = \mathcal{F}_{g_1} \cdot \mathcal{F}_\phi + \mathcal{F}_{g_2} + (-\mathcal{F}_v) + \mathcal{F}_{g_3} \cdot \mathcal{F}_\phi + \mathcal{F}_{g_4}$, which is again a VC-subgraph class with some index $V(\mathcal{F})$ by Lemma 2.6.18 in van der Vaart and Wellner (1996). By Assumptions 2.1-2.3 and 3.2, we may take $F(x) \equiv \sup_{p \in \mathbb{S}^\ell} \|w(z)p'\nabla_z m_{p,\eta}(z)\|_{L_\mu^\infty} + \|l(z)\| \times \bar{\epsilon}^{-1} \times \sup_{u \in D'} |q(u)|$ as an envelope function such that $E[F(x)^2] < \infty$. Then, by Theorems 2.6.7 and 2.5.1 in van der Vaart and Wellner (1996), \mathcal{F} is a Donsker class. This establishes the claim of the lemma. ■

Proof of Theorem 4.1. For each $p \in \mathbb{S}^\ell$, we have the following decomposition:

$$\sqrt{n}(\hat{v}_n(p) - v(p, \Theta_0(P_0))) = \sqrt{n}(\hat{v}_n(p) - \bar{v}_n(p)) + \sqrt{n}(\bar{v}_n(p) - E[\bar{v}_n(p)]) + \sqrt{n}(E[\bar{v}_n(p)] - v(p, \Theta_0(P_0))) \quad (\text{D.25})$$

$$\equiv G_{1n}(p) + G_{2n}(p) + G_{3n}(p). \quad (\text{D.26})$$

By Lemmas D.2-D.5, uniformly in $p \in \mathbb{S}^\ell$, $G_{1n}(p) = G_{3n}(p) = o_p(1)$, and $G_{2n}(p) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_p(Z_i) + o_p(1)$. This establishes the second claim of the Theorem. By Theorem 3.1, ψ_p is the efficient influence function, and hence regularity of $\{\hat{v}_n(\cdot)\}$ follows from Lemma D.6 and Theorem 18.1 in Kosorok (2008), which establishes the first claim. The stated convergence in distribution is then immediate from (D.26) and Lemma D.6, while the limiting process having the efficient covariance kernel is a direct result of the characterization of $I(p_1, p_2)$ obtained in Theorem 3.1, which establishes the third claim. ■

APPENDIX E: Figures and Tables

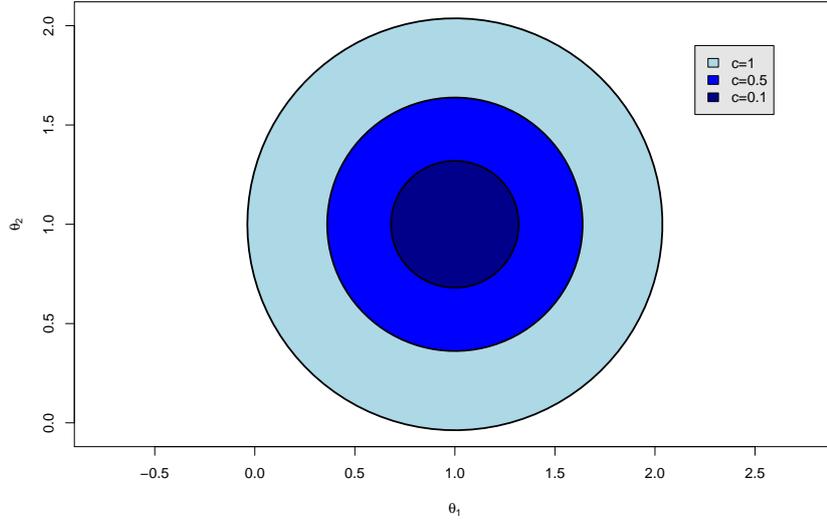


Figure 1: Identified sets for the density weighted average derivatives

Table 1: Risk of \hat{v}_n^{IV} (Gaussian kernel)

Sample Size	h	R_H	$c=0.1$		$c=0.5$			$c=1$		
			R_{IH}	R_{OH}	R_H	R_{IH}	R_{OH}	R_H	R_{IH}	R_{OH}
$n=1000$										
	0.4	0.0608	0.0477	0.0600	0.0834	0.0709	0.0673	0.1229	0.1037	0.0801
	0.5	0.0588	0.0468	0.0578	0.0785	0.0749	0.0485	0.1212	0.1185	0.0437
	0.6	0.0572	0.0452	0.0564	0.0809	0.0804	0.0351	0.1305	0.1304	0.0229
	0.7	0.0567	0.0416	0.0563	0.0844	0.0844	0.0263	0.1416	0.1416	0.0086
	0.8	0.0555	0.0386	0.0553	0.0882	0.0882	0.0195	0.1556	0.1556	0.0026
$n=500$										
	0.4	0.0929	0.0703	0.0919	0.1185	0.0877	0.1072	0.1731	0.1203	0.1437
	0.5	0.0836	0.0684	0.0817	0.1091	0.0979	0.0839	0.1555	0.1414	0.0873
	0.6	0.0799	0.0646	0.0786	0.1038	0.0999	0.0640	0.1555	0.1520	0.0530
	0.7	0.0774	0.0607	0.0762	0.1060	0.1051	0.0512	0.1679	0.1677	0.0297
	0.8	0.0775	0.0592	0.0769	0.1098	0.1096	0.0410	0.1785	0.1785	0.0173
$n=250$										
	0.4	0.1357	0.0960	0.1349	0.1820	0.1061	0.1770	0.2480	0.1256	0.2339
	0.5	0.1189	0.0941	0.1169	0.1517	0.1231	0.1289	0.2013	0.1638	0.1446
	0.6	0.1133	0.0914	0.1112	0.1413	0.1299	0.1053	0.1954	0.1818	0.1084
	0.7	0.1121	0.0910	0.1098	0.1365	0.1317	0.0890	0.1974	0.1949	0.0725
	0.8	0.1086	0.0864	0.1068	0.1374	0.1360	0.0737	0.2069	0.2061	0.0500

Table 2: Risk of \hat{v}_n^{IV} (Higher-order kernel)

Sample Size	h	$c=0.1$			$c=0.5$			$c=1$		
		R_H	R_{IH}	R_{OH}	R_H	R_{IH}	R_{OH}	R_H	R_{IH}	R_{OH}
$n=1000$										
	0.5	0.0722	0.0549	0.0714	0.1267	0.0461	0.1256	0.2038	0.0494	0.2017
	0.6	0.0654	0.0551	0.0637	0.0912	0.0532	0.0872	0.1384	0.0636	0.1312
	0.7	0.0600	0.0511	0.0583	0.0760	0.0631	0.0645	0.1020	0.0835	0.0745
	0.8	0.0564	0.0470	0.0553	0.0759	0.0741	0.0444	0.1093	0.1085	0.0370
	0.9	0.0565	0.0446	0.0559	0.0802	0.0801	0.0313	0.1302	0.1302	0.0134
$n=500$										
	0.5	0.1104	0.0744	0.1101	0.1867	0.0587	0.1861	0.2887	0.0604	0.2870
	0.6	0.0947	0.0753	0.0930	0.1308	0.0745	0.1267	0.1993	0.0857	0.1914
	0.7	0.0869	0.0737	0.0846	0.1080	0.0843	0.0970	0.1453	0.1085	0.1184
	0.8	0.0802	0.0668	0.0783	0.1019	0.0958	0.0747	0.1373	0.1308	0.0683
	0.9	0.0772	0.0635	0.0758	0.1042	0.1034	0.0564	0.1513	0.1508	0.0372
$n=250$										
	0.5	0.1788	0.1038	0.1787	0.2832	0.0630	0.2831	0.4316	0.0511	0.4313
	0.6	0.1374	0.1034	0.1359	0.1959	0.0979	0.1925	0.2802	0.1061	0.2716
	0.7	0.1212	0.1001	0.1187	0.1571	0.1147	0.1460	0.2063	0.1401	0.1773
	0.8	0.1143	0.0948	0.1118	0.1385	0.1231	0.1133	0.1811	0.1620	0.1201
	0.9	0.1107	0.0910	0.1085	0.1342	0.1292	0.0903	0.1865	0.1820	0.0819