Econometric Theory, **33**, 2017, 1218–1241.
doi:10.1017/S0266466616000384

ASYMPTOTICALLY EFFICIENT ESTIMATION OF WEIGHTED AVERAGE DERIVATIVES WITH AN INTERVAL CENSORED VARIABLE

Hiroaki Kaido
Boston University

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, *Econometrica* 70(2), 519–546) 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 functional-form 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. Using mean regression as an example, we further demonstrate that the support function can be estimated in a regular manner by a computationally simple estimator and that the efficiency bound can be achieved.

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.

The author thanks the editor and anonymous referees for comments that helped improve this paper. The author thanks Iván Fernández-Val, Arthur Lewbel, Francesca Molinari, Pierre Perron, Zhongjun Qu, Andres Santos, Jörg Stoye, and seminar participants at BU, Cornell, Brown, Yale and participants of BU-BC Joint Econometrics Conference, the CEMMAP Workshop (New Developments in the Use of Random Sets in Economics), the 2012 International Symposium on Econometric Theory and Applications, and CEME 2014 for helpful comments. The author acknowledges excellent research assistance from Michael Gechter. Financial support from NSF Grants SES-1230071 and SES-1357643 is gratefully acknowledged. Address correspondence to Hiroaki Kaido, Dept. of Economics, Boston University 270 Bay State Road, Boston, MA 02215, USA; e-mail: hkaido@bu.edu.

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

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

for some criterion function $\varrho : \mathbb{R} \to \mathbb{R}_{+}$. 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 is interval-valued when the researcher does not observe the variable itself W but observes a pair (W_L, W_U) of random variables such that

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

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

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

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). One may also conduct inference for the coefficients of the best linear approximation to the regression function (Beresteanu and Molinari, 2008). 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

https://doi.org/10.1017/S0266466616000384 Published online by Cambridge University Press

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 $w : \mathcal{Z} \to \mathbb{R}_+$ be a weight function, the *weighted average derivative* of m is defined by

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

Stoker (1986) first analyzed estimation of this parameter. It has also been studied in a variety of empirical studies, including Deaton and Ng (1998), 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 a specific feature (e.g., conditional quantile) 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, e.g., Beresteanu and Molinari, 2008; Beresteanu, Molchanov, and Molinari, 2011). 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 respect to each covariate. We further show that these bounds are useful for obtaining bounds on parameters in commonly used semiparametric models, the semiparametric index models.

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 (2014) (KS henceforth) 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. This result is useful in two respects. First, as shown in KS, an estimator of the identified set associated with the efficient estimator of the support function asymptotically minimizes a broad class of risk functions based on the

Hausdorff distance. Hence, it gives a benchmark against which the asymptotic relative efficiency of any estimator of the identified set can be measured. Efficient estimation of the support function also implies efficient estimation of the upper and lower bounds on each average partial derivative. Second, the derived semiparametric efficiency bound has an implication on the optimality of inference. Specifically, for testing the hypothesis that a given value θ of average derivative belongs to the identified set, i.e., $H_0 : \theta \in \Theta_0(P)$, a test based on the efficient estimator of the support function can be shown to maximize power against local alternatives. This test can also be used to construct a confidence region through a test inversion.

Using mean regression as an example, we further illustrate estimation by showing that the support function can be estimated in a regular manner, and a computationally simple estimator which builds on Stoker (1986), Härdle and Stoker (1989), Powell, Stock, and Stoker (1989) is semiparametrically efficient. A contribution of this paper on the technical side is that the characterization of the efficiency bound and construction of an efficient estimator are done in a model with infinite dimensional parameters that are explicitly estimated. In particular, some of the steps in the analysis of the asymptotic distribution of the estimator of support function requires tools from the semiparametric inference literature.[2] This feature was not present in KS, and hence these results may be of independent interest.

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 (2012), 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 et al. (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. 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 rest of the paper is organized as follows. Section 2 presents the model, discusses examples, 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. The proofs of all theoretical results and details of Monte Carlo experiments are provided online at Cambridge Journals Online (journals.cambridge.org/ect) in supplementary material to this article.

2. GENERAL SETTING

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. Throughout, we let $\varrho : \mathbb{R} \to \mathbb{R}_+$ be a *criterion function*, which we use to define the *regression function* m through the minimization problem in (1). This paper's results apply to the general class of regression functions defined this way. For expositional purposes, however, we mainly focus on mean and quantile regression functions. Throughout, we let $q : \mathbb{R} \to \mathbb{R}$ be the derivative of the criterion function, which is assumed to be well-defined almost everywhere and is used to define the *regression residual* $q(Y - m(Z))$, which coincides with $Y - m(Z)$ in the case of mean regression and $\alpha - 1\{Y - m(Z) \le 0\}$ in the case of quantile regression.

Suppose that $w(z)f(z)$ vanishes on the boundary of $\mathcal{Z}$. By integration by parts, equation (4) can be equivalently written as

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

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.[3]

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.

https://doi.org/10.1017/S0266466616000384 Published online by Cambridge University Press

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 $\theta \equiv E[w(Z)\nabla_z g(Z,v)]$.

The analysis of the weighted average derivative is useful for *semiparametric index models*. Let $\beta \in B \subseteq \mathbb{R}^\ell$. In the case of interval-censoring on the outcome Y, the semiparametric index model is characterized by the regression function:

$$m(z) = M(z'\beta), \tag{6}$$

for $M : \mathbb{R} \to \mathbb{R}$. Similarly, in the case of interval-censoring on a covariate V, the regression function is

$$g(z,v) = G(z'\beta, v), \tag{7}$$

for $G : \mathbb{R}^2 \to \mathbb{R}$. This class nests a variety of semiparametric models, including binary choice, transformation and proportional hazards, and Tobit-type censoring models (Ichimura, 1993; Klein and Spady, 1993; Horowitz, 1996). In many of the settings where these models are applied, it is also common for the researcher to face an interval-valued variable. For example, an individual's age is often an important control variable. However, for privacy reasons, the date of birth may not be used for analysis. Instead, only the quarter or year of birth may be used. This is a common form of interval censoring. Average derivatives are useful for obtaining bounds on parameters in such settings. Below, we take a semiparametric binary choice model with an interval censored regressor as an example of the semiparametric index model.

Example 2.3
Let $Y \in \{0,1\}$ be a binary outcome generated as

$$Y = 1\{Z'\beta + \gamma V + \epsilon > 0\}. \tag{8}$$

Then, the conditional mean of Y can be written as $E[Y|Z=z, V=v] = G(z'\beta, v)$ for a function G determined by the distribution of ϵ. Let $G_u(u,v) \equiv \partial G(u,v)/\partial u$. The index coefficients and average derivatives are related to each other through

$$\theta = E[w(Z)G_u(Z'\beta, v)]\beta. \tag{9}$$

The bounds on the average derivatives can be interpreted as the bounds on the scaled index coefficients. We will show that, with a monotonicity assumption on G, bounds on the index coefficients can be obtained from this relationship.

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

$$\Theta_0(P) \equiv \left\{\theta \in \Theta : \theta = \int m(z)l(z)dP(x),\ m \text{ satisfies (1) with } Y_L \leq Y \leq Y_U,\ P-a.s.\right\}. \quad \textbf{(10)}$$

To characterize the identified set, we make the following assumptions on the primitives of the model.

Assumption 2.1. (i) The distribution of Z and $l(Z)$ are absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^\ell$. $\mathcal{Z}$ is compact and convex with nonempty interior; (ii) $\varrho : \mathbb{R} \to \mathbb{R}_+$ is convex and satisfies $\varrho(0) = 0$ and $\varrho(\epsilon) \to \infty$ as $|\epsilon| \to \infty$. The function $q(\epsilon) \equiv d\varrho(\epsilon)/d\epsilon$ exists and is continuous $a.e.$, and it is bounded on the support of $\epsilon_j = Y_j - m_j(Z)$ for $j = L, U$.

Assumption 2.2. (i) $w : \mathcal{Z} \to \mathbb{R}$ and $\nabla_z w : \mathcal{Z} \to \mathbb{R}^\ell$ 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 $l(Z)$. This excludes discrete regressors. One way to relax this assumption would be to partition Z as (Z_1, Z_2), where $Z_1 \in \mathbb{R}^{\ell_1}$ is continuously distributed and let $\nabla m(z) = \partial m(z)/\partial z_1$ to accommodate discrete or mixed variables as components in Z_2. To keep a tight focus, we leave this extension elsewhere. Assumption 2.1(ii) requires that q is bounded and continuous. This condition is implied by other assumptions in the case of our leading examples (mean and quantile regression) but is maintained here as a high-level condition for the general setting. In the case of mean regression, the boundedness of q requires the compactness of the support of Y_j for $j = L, U$, which is ensured by Assumption 2.3 below. In the case of quantile regression, $q(\epsilon) = \alpha - 1\{\epsilon \leq 0\}$ trivially satisfies the condition. Assumption 2.2 then assumes that the weight function and its gradient are bounded and continuous. These conditions allow a broad class of smooth weight functions.

We further add the following assumption on the data generating process 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 assumes that Y_j, $j = L, U$ are bounded random variables. This excludes cases where $Y_U = \infty$ or $Y_L = -\infty$ with positive probability. However, in many empirical examples, the researcher may set the upper bound of Y_U (or the lower bound of Y_L) to a finite value, which is large enough for practical purposes (e.g., setting a finite upper bound on the total expenditure in Example 2.1).

In Assumption 2.3(ii), we require that $w(z)f(z)$ vanishes on the boundary of $\mathcal{Z}$. As is well known from Stoker (1986), this is the key condition for writing the average derivative as in (5). In our setting, this is also important for obtaining bounds on θ from the nonparametric bounds on m. This condition still allows the possibility of the density f to be positive for some $z \in \partial\mathcal{Z}$, in which case $w(z)$ must be 0. Assumption 2.3(iii) requires that the regression functions $m_j, j = L, U$ are well-defined. This condition is not restrictive. It trivially holds with $m_j(z) = E[Y_j|Z = z]$ in the case of mean regression. For quantile regression, a sufficient condition is that, for each $z \in \mathcal{Z}$, the conditional density of Y_j is bounded away from 0 at the α-th quantile for $j = L, U$.[4] Assumption 2.3(iv) then assumes that the regression functions of Y_L and Y_U are continuously differentiable for almost all $z \in \mathcal{Z}$. With these assumptions, $\Theta_0(P)$ 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 $\mathbb{R}^\ell$. For a bounded convex set F, the *support function* of F is defined by

$$\upsilon(p, F) \equiv \sup_{x \in F} \langle p, x \rangle. \quad \textbf{(11)}$$

Under Assumption 2.3(iv), the sharp identified set defined in (10) coincides with the set characterized by the support function in (13) in Theorem 2.1 below. In this case, for any θ belonging to the latter set, there is a regression function m^* compatible with the model whose average derivative coincides with θ. If this assumption is violated, the set characterized by the support function in (13) still provides valid bounds on the average derivatives but may also contain points that do not belong to $\Theta_0(P)$.

In what follows, we let $\Gamma : \mathbb{R}^3 \to \mathbb{R}$ be defined pointwise by

$$\Gamma(w_1, w_2, w_3) = 1\{w_3 \le 0\}w_1 + 1\{w_3 > 0\}w_2. \quad \textbf{(12)}$$

Theorem 2.1 is our first main result, which characterizes the identified set through its support function.

THEOREM 2.1. *Suppose Assumptions 2.1–2.3 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) the sharp identified set $\Theta_0(P)$ is compact and strictly convex; and (b) the support function of $\Theta_0(P)$ is given pointwise by:*

$$\upsilon(p, \Theta_0(P)) = \int m_p(z) p' l(z) dP(x), \quad \textbf{(13)}$$

where $m_p(z) = \Gamma(m_L(z), m_U(z), p' l(z))$.

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 $\langle p, \theta^*(p) \rangle = \upsilon(p, \Theta_0(P))$, which can be expressed as:

$$\theta^*(p) = \int m_p(z) l(z) dP(x), \quad \textbf{(14)}$$

https://doi.org/10.1017/S0266466616000384 Published online by Cambridge University Press

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 (13) 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 Chandrasekhar, Chernozhukov, Molinari, and Schrimpf, 2011, and references therein).

Remark 2.1. While the closed-form bounds in Theorem 2.1 are derived using the structure of our model, the key insight that the average derivative is a bounded linear map of m may also be useful in other partially identified models. For example, Blundell, Kristensen, and Matzkin (2014) (BKM) study bounds on demand functions $m^{(j)}(p,x,\tau), j = 1,\dots,\ell$, where $p \in \mathbb{R}^{\ell}$ is a vector of prices on ℓ products, $x \in \mathbb{R}$ is individual's income, and $\tau \in [0,1]$ represents a scalar taste heterogeneity. In their setting, the map $(x,\tau) \mapsto m^{(j)}(p_t,x,\tau)$ is identified for $t = 1,\dots,T$ from repeated cross-section data, but the demand $q = m(p_0,x,\tau)$ under a new price p_0 is only identified as a set $Q_I(p_0,x,\tau)$ through revealed preference bounds and expansion paths. Extending this line of analysis, the average slope of demand (in price) under a counterfactual distribution $f(\cdot)$ of prices can be obtained by maximizing/minimizing

$$\beta_j = E_f[m^{(j)}(p,x,\tau)l^{(j)}(p)], \tag{15}$$

subject to the revealed preference constraints $m^{(j)}(p,x,\tau) \in Q_I(p,x,\tau)$ for all p in f's support. This functional may be of interest if the researcher is interested in the average response of demand over a range of prices under a counterfactual distribution f. Note that, in this example, the counterfactual distribution $f(p)$ and therefore $\ell(p) = -\nabla w(p) - w(p)f(p)/\nabla f(p)$ are known. Whether closed-form bounds can be obtained or not is not clear. However, as shown in BKM, the identified set for the counterfactual demand $Q_I(p,x,\tau)$ is characterized through moment inequalities. This may give a way to make inference on β_j using inference methods developed for functionals of parameters satisfying inequality restrictions (see, e.g., Bugni, Canay, and Shi, 2016; Kaido, Molinari, and Stoye, 2016).

Theorem 2.1 implies 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 as $-\upsilon(-\iota_j, \Theta_0(P))$. Therefore, the bounds on $\theta^{(j)}$ are given as the interval $[\theta_L^{(j)}, \theta_U^{(j)}]$ with

$$\theta_L^{(j)} = \int \left[1\{l^{(j)}(z) > 0\}m_L^{(j)}(z) + 1\{l^{(j)}(z) \leq 0\}m_U^{(j)}(z)\right]l^{(j)}(z)dP(x), \tag{16}$$

https://doi.org/10.1017/S0266466616000384 Published online by Cambridge University Press

$$\theta_U^{(j)} = \int \left[1\{l^{(j)}(z) \le 0\} m_L^{(j)}(z) + 1\{l^{(j)}(z) > 0\} m_U^{(j)}(z)\right] l^{(j)}(z) dP(x). \tag{17}$$

Further, these bounds are useful for obtaining bounds on the coefficients in semiparametric index models. Consider the regression function in (6). We normalize the scale of the index by setting the first component $\beta^{(1)}$ of the index coefficient vector β to 1. Note also that Z does not include a constant regressor, and hence the location of the index is normalized as well. We then make the following assumption.

Assumption 2.4. (i) For each $j = 2, \ldots, \ell$, let $-\infty < \underline{\beta}^{(j)} < \overline{\beta}^{(j)} < \infty$. The regression function satisfies $m(z) = M(z'\beta)$ for a function $M : \mathbb{R} \to \mathbb{R}$ and $\beta \in B \subseteq \mathbb{R}^\ell$, where $B = \{1\} \times \prod_{j=2}^{\ell} [\underline{\beta}^{(j)}, \overline{\beta}^{(j)}]$; (ii) M is differentiable with a bounded derivative M', and $M'(u) > 0$ for all u in the support of $Z'\beta$.

Assumption 2.4 makes a monotonicity assumption on the index function. This is satisfied, for example, by a wide class of semiparametric transformation and hazards models (see, e.g., Horowitz, 1996). We assume that the parameter space for β is the product of closed intervals for simplicity, which can be relaxed at the cost of a more complex notation. With this additional assumption, one may obtain bounds on the index coefficients. For this, let $\theta_{L+}^{(1)} = \max\{\theta_L^{(1)}, 0\}$.

COROLLARY 2.1. *Suppose that the conditions of Theorem 2.1 hold. Suppose Assumption 2.4 holds. Then, for each* $j = 2, \ldots, \ell$,

$$\beta_L^{(j)} \le \beta^{(j)} \le \beta_U^{(j)}, \tag{18}$$

where

$$\beta_L^{(j)} = \max\left\{ \frac{\theta_L^{(j)}}{\theta_U^{(1)}}, \underline{\beta}^{(j)} \right\}, \quad \beta_U^{(j)} = \min\left\{ \frac{\theta_U^{(j)}}{\theta_{L+}^{(1)}}, \overline{\beta}^{(j)} \right\}. \tag{19}$$

Provided that the a priori lower bound $\underline{\beta}^{(j)}$ is sufficiently small, the lower bound on each coefficient is obtained as the ratio of the lower bound $\theta_L^{(j)}$ on the j-th component of the average derivative vector θ and the upper bound $\theta_U^{(1)}$ on the first component of θ. This is because (i) each index coefficient is proportional to the average derivative and (ii) the bounds on the first component of θ corresponds to those on the scaling factor $E[w(Z)M'(Z\beta)]$. The upper bound on $\beta^{(j)}$ is obtained similarly. Note that, if $\theta_L^{(1)}$ is negative, this implies that the lower bound on the scaling constant is 0 by the monotonicity of M, which in turn implies $\theta_U^{(j)}/\theta_{L+}^{(1)}$ is unbounded. Hence, in this case, the upper bound on $\beta^{(j)}$ is given by the a priori upper bound $\overline{\beta}^{(j)}$.

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

https://doi.org/10.1017/S0266466616000384 Published online by Cambridge University Press

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:

$$g(z,v) \equiv \text{argmin}_u E[\varrho(Y-u)|Z=z, V=v]. \tag{20}$$

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

$$\theta_v \equiv E[w(Z)\nabla_z g(Z,v)], \tag{21}$$

where the expectation in (21) 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. This is a useful parameter to estimate if one is interested in the marginal effect of covariates (e.g., square footage in Example 2.2) on the outcome (house price) while fixing V (the distance to the city center) to some value. In order to characterize the identified set for θ_v, we make use of the regression function of Y given all observable variables $(Z', V_L, V_U)'$. Specifically, for each (z', v_L, v_U), define

$$\gamma(z, v_L, v_U) \equiv \text{argmin}_u E[\varrho(Y-u)|Z=z, V_L=v_L, V_U=v_U]. \tag{22}$$

We make the following assumptions to characterize the identified set.

Assumption 2.5. (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

$$E[q(Y-g(Z,V))|Z=z, V_L=v_L, V_U=v_U, V=v] = E[q(Y-g(Z,V))|Z=z, V=v]. \tag{23}$$

Following Manski and Tamer (2002), Assumption 2.5(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 2.5(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 2.5, one may show that the following functional inequalities hold:

$$g_L(Z,v) \leq g(Z,v) \leq g_U(Z,v), \;\; P-a.s., \text{ for all } v, \tag{24}$$

where

$$g_L(z,v) \equiv \sup_{(v_L,v_U)\in\Xi_L(v)} \gamma(z,v_L,v_U), \quad \text{and} \quad g_U(z,v) \equiv \inf_{(v_L,v_U)\in\Xi_U(v)} \gamma(z,v_L,v_U). \tag{25}$$

We then assume the following regularity conditions.

Assumption 2.6. (i) There is a compact set $D \subset \mathbb{R}$ containing the support of Y in its interior. (ii) $E[q(Y-u)|Z=z, V_L=v_L, V_U=v_U] = 0$ has a unique

https://doi.org/10.1017/S0266466616000384 Published online by Cambridge University Press

solution at $u = \gamma(z, v_L, v_U) \in D$; (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. We assume that the observed Y is bounded, which is satisfied in various limited dependent variable models including Example 2.3. Assumption 2.6(ii) is again a condition for the regression function to be uniquely defined given all observable covariates. For models with limited dependent variables such as Example 2.3, this assumption does not hold for the quantile regression, and hence for such models, our focus will be on the mean regression function for which this assumption trivially holds.[5] Assumption 2.6(iii) requires that the functional bounds $g_j, j = L, U$ are differentiable in z. Since $g_j, j = L, U$ defined in (25) 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, e.g., Milgrom and Segal, 2002), but this condition may not hold for some settings, in which case the obtained identified set gives possibly nonsharp 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), 2.5, and 2.6 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*

$$\upsilon(p, \Theta_{0,v}(P)) = \int g_p(z,v) p' l(z) dP(x), \tag{26}$$

where $g_p(z,v) = \Gamma(g_L(z,v), g_U(z,v), p'l(z))$.

Again, Theorem 2.2 can be used to give closed-form bounds on the weighted average derivative $\theta_v^{(j)}$ with respect to the j-th variable. The bounds $\left[\theta_L^{(j)}(v), \theta_U^{(j)}(v)\right]$ are given as

$$\theta_L^{(j)}(v) = \int \left[1\{l^{(j)}(z) > 0\} g_L^{(j)}(z,v) + 1\{l^{(j)}(z) \le 0\} g_U^{(j)}(z,v)\right] l^{(j)}(z) dP(x), \tag{27}$$

$$\theta_U^{(j)}(v) = \int \left[1\{l^{(j)}(z) \le 0\} g_L^{(j)}(z,v) + 1\{l^{(j)}(z) > 0\} g_U^{(j)}(z,v)\right] l^{(j)}(z) dP(x). \tag{28}$$

These bounds are useful for obtaining bounds on index coefficients in the semiparametric index model (7). Toward this end, we make the following assumption.

Assumption 2.7. For each $j = 2, \ldots, \ell$, let $-\infty < \underline{\beta}^{(j)} < \overline{\beta}^{(j)} < \infty$. For each $(z,v) \in \mathcal{Z} \times \mathcal{V}$, the regression function satisfies $g(z,v) = G(z'\beta, v)$ for a function $G : \mathbb{R}^2 \to \mathbb{R}$ and $\beta \in B \subseteq \mathbb{R}^\ell$, where $B = \{1\} \times \prod_{j=2}^{\ell} [\underline{\beta}^{(j)}, \overline{\beta}^{(j)}]$. For each v, $G(\cdot, v)$ is differentiable with a bounded derivative $G_u(u,v) = \partial G(u,v)/\partial u$, and $G_u(u,v) > 0$ for all u in the support of $Z'\beta$.

Assumption 2.7 is an analog of Assumption 2.4. We assume that g depends on the covariates through the linear index $Z'\beta$, where the first component of the coefficient β is normalized to 1. For each $v \in \mathcal{V}$, we assume the map $u \mapsto G(u,v)$ is strictly increasing. These assumptions are satisfied, for example, by the semiparametric binary choice model (Example 2.3) with a continuously distributed error.

The bounds on the index coefficients are then given as follows:

COROLLARY 2.2. *Suppose that the conditions of Theorem 2.2 hold. Suppose Assumption 2.7 holds. Then, for each* $j = 2, \dots, \ell$,

$$\beta_L^{(j)} \le \beta^{(j)} \le \beta_U^{(j)}, \tag{29}$$

where

$$\beta_L^{(j)} = \max\left\{\sup_{v\in\mathcal{V}} \frac{\theta_L^{(j)}(v)}{\theta_U^{(1)}(v)}, \underline{\beta}^{(j)}\right\}, \quad \beta_U^{(j)} = \min\left\{\inf_{v\in\mathcal{V}} \frac{\theta_U^{(j)}(v)}{\theta_{L+}^{(1)}(v)}, \overline{\beta}^{(j)}\right\}. \tag{30}$$

A key difference from Corollary 2.1 is that the bounds on the index coefficient can be obtained by combining bounds across different values of v. That is, for each $v \in \mathcal{V}$, the model predicts that $\theta_L^{(j)}(v)/\theta_U^{(1)}(v) \le \beta^{(j)} \le \theta_U^{(j)}(v)/\theta_{L+}^{(1)}(v)$. As $\beta^{(j)}$ does not depend on v, one may intersect these bounds. The result of the corollary is obtained by combining this with the a priori bound on the coefficient.

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 $\{X_i\}_{i=i}^n$ are independently and identically distributed (i.i.d.) according to a distribution P, which is absolutely continuous with respect to a σ-finite measure μ satisfying the following assumption.

Assumption 3.1. (i) μ satisfies $\mu(\{(y_L, y_U, z) : y_L \le y_U\}) = 1$; (ii) $\mu(\{(y_L, y_U, z) : F(z) = 0\}) = 0$ for any measurable function $F : \mathbb{R}^\ell \to \mathbb{R}$.

Assumption 3.1 ensures that $Y_L \le Y_U$, $P - a.s.$ and excludes the setting where components of Z have functional dependence with each other as assumed earlier in Section 2. Here, by considering distributions dominated by μ, we will require that these conditions hold for all distributions in our model specified below.

We introduce an additional piece of notation. For each $z \in \mathcal{Z}$ and $j = L, U$, let

$$r_j(z) \equiv -\frac{d}{d\tilde{m}} E[q(Y_j - \tilde{m})|Z = z]\Big|_{\tilde{m}=m_j(z)}. \tag{31}$$

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 3.2. (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 $\varphi : \mathcal{X} \to \mathbb{R}$ that is bounded and continuously differentiable

https://doi.org/10.1017/S0266466616000384 Published online by Cambridge University Press

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 $j=L,U$.

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. These conditions are generally satisfied when q is continuously differentiable. In such settings, Y_L and Y_U can be either continuously or discretely distributed. A common example is a mean regression problem where $q(\epsilon)=\epsilon$ with an unobserved outcome Y, which falls in a given set of brackets. When q has a point of discontinuity, however, as in the quantile regression where $q(\epsilon)=\alpha-1\{\epsilon\le 0\}$, these conditions require that Y_L and Y_U are continuously distributed.

Given these assumptions, we now define our model as the set of distributions that satisfy Assumptions 2.3 and 3.2. Let $\mathbf{M}_\mu$ be the set of Borel probability measures dominated by μ. Define

$$\mathbf{P}\equiv\{P\in\mathbf{M}_\mu : P \text{ satisfies Assumptions 2.3 and 3.2}\}. \qquad \textbf{(32)}$$

The support function $\upsilon(\cdot,\Theta_0(P))$ is a continuous function on the unit sphere. Following Kaido and Santos (2014), we view it as a function-valued parameter taking values in $\mathcal{C}(\mathbb{S}^\ell)$, the set of continuous functions on $\mathbb{S}^\ell$. For any regular estimator $\tilde{\upsilon}_n(\cdot)$ of $\upsilon(\cdot,\Theta_0(P))$, its asymptotic variance (at p) is bounded below by the semiparametric efficiency bound $E[\psi_p(X_i)^2]$, where $\psi_\cdot:\mathcal{X}\to\mathcal{C}(\mathbb{S}^\ell)$ is called the *efficient influence function*. Moreover, this bound holds uniformly in p, and we refer to Bickel, Klassen, Ritov, and Wellner (1993) for details. The following theorem characterizes this efficiency bound via the efficient influence function.

THEOREM 3.1. *Suppose Assumptions 2.1–2.2 and 3.1 hold, and suppose* $P\in\mathbf{P}$. *Then, the* efficient influence function *for estimating the support function* $\upsilon(\cdot,\Theta_0(P))$ *is*

$$\psi_p(x)\equiv w(z)p'\nabla_z m_p(z)-\upsilon(p,\Theta_0(P))+p'l(z)\zeta_p(x), \qquad \textbf{(33)}$$

where $\nabla_z m_p$ *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. It shows that the variance bound for estimating the support function $\upsilon(p,\Theta_0(P))$ at p is given by

$$E[|\psi_p(X)|^2]=Var(w(Z)p'\nabla_z m_p(Z))+E[|p'l(Z)\zeta_p(x)|^2]. \qquad \textbf{(34)}$$

The first term in (34) 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

https://doi.org/10.1017/S0266466616000384 Published online by Cambridge University Press

but m_p is unknown (see Newey and Stoker, 1993, p. 1205 for a more detailed discussion). 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 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 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 also Remark 4.1). For any two compact convex sets A, B, let the Hausdorff distance be $d_H(A, B) \equiv \max\{\sup_{a\in A}\inf_{b\in B}\|a-b\|, \sup_{b\in B}\inf_{a\in A}\|a-b\|\}$. This distance measure is commonly used to examine consistency of set-valued estimators Chernozhukov et al. (2007). KS show that for any regular convex compact valued set estimator C_n for $\Theta_0(P)$ and a subconvex continuous function $L : \mathbb{R}_+ \to \mathbb{R}_+$, it holds under regularity conditions that

$$\liminf_{n\to\infty} E[L(\sqrt{n}d_H(C_n, \Theta_0(P)))] \geq E[L(\|\mathbb{G}_0\|_\infty)], \tag{35}$$

where $\mathbb{G}_0$ is a Gaussian process on $\mathbb{S}^\ell$ such that $\mathrm{Cov}(\mathbb{G}_0(p), \mathbb{G}_0(q)) = E[\psi_p(X_i)\psi_q(X_i)]$ for all $p, q \in \mathbb{S}^\ell$.[6] Hence, the efficiency bound in Theorem 3.1 also characterizes the asymptotic lower bound for the estimation risk of the identified set.

Remark 3.1. Powell et al. (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 equation (33) due to f being unknown. Taking into account the pathwise derivative of unknown f, one can show that the efficient influence function for this case is

$$\psi_p(x) \equiv 2\{f(z)p'\nabla_z m_p(z) - \upsilon(p, \Theta_0(P))\} - 2p'\nabla_z f(z)(y_p - m_p(z)), \tag{36}$$

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

Remark 3.2. The existence of the semiparametric efficiency bound of the support function in the case of interval censoring on Y is due to the pathwise weak differentiability of the support function Bickel et al. (1993). For the setting where an explanatory variable is interval-valued, however, Theorem 2.2 shows that the support function of $\Theta_{0,\upsilon}(P)$ involves functions that are defined as the supremum (or the infimum) of functions indexed by (υ_L, υ_U), e.g., $g_L(z, \upsilon) = \sup_{(\upsilon_L,\upsilon_U)\in\Xi_L(\upsilon)} \gamma(z, \upsilon_L, \upsilon_U)$. 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:

$$\upsilon(p, \Theta_{0,\upsilon}(P_\eta)) = \int \Big[1\{p'l_\eta(z) \le 0\} g_{L,\eta}(z,\upsilon) + 1\{p'l_\eta(z) > 0\} g_{U,\eta}(z,\upsilon)\Big] p'l_\eta(z) dP_\eta(x), \tag{37}$$

where l_η is defined as in (5) and $g_{L,\eta}, g_{U,\eta}$ are defined as in (25) 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 nondifferentiable has a positive probability mass, the support function is pathwise nondifferentiable either. Hence, $\upsilon(p, \Theta_{0,\upsilon}(P))$ does not generally admit regular estimation. Therefore, for optimal inference on $\upsilon(p, \Theta_{0,\upsilon}(P))$, an alternative optimal criterion would be needed (see, e.g., Song, 2014).

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

$$\mathbf{g}_L(z,\upsilon;\kappa) \equiv \sum_{(\upsilon_L,\upsilon_U)\in\Xi_L(\upsilon)} \gamma(z,\upsilon_L,\upsilon_U) \frac{\exp(\kappa\gamma(z,\upsilon_L,\upsilon_U))}{\sum_{(\upsilon_L,\upsilon_U)\in\Xi_L(\upsilon)} \exp(\kappa\gamma(z,\upsilon_L,\upsilon_U))}, \tag{38}$$

$$\mathbf{g}_U(z,\upsilon;\kappa) \equiv \sum_{(\upsilon_L,\upsilon_U)\in\Xi_U(\upsilon)} \gamma(z,\upsilon_L,\upsilon_U) \frac{\exp(-\kappa\gamma(z,\upsilon_L,\upsilon_U))}{\sum_{(\upsilon_L,\upsilon_U)\in\Xi_U(\upsilon)} \exp(-\kappa\gamma(z,\upsilon_L,\upsilon_U))}, \tag{39}$$

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, 2015). Suppose that the researcher chooses a fixed $\kappa > 0$. Define $\mathbf{u}(p;\kappa) \equiv \int \mathbf{g}_p(z,\upsilon;\kappa) p'l(z) dP(x)$, where $\mathbf{g}_p(z,\upsilon;\kappa) = \Gamma(\mathbf{g}_L(z,\upsilon;\kappa), \mathbf{g}_U(z,\upsilon;\kappa), p'l(z))$. $\mathbf{u}(p;\kappa)$ is then a conservative approximation of the support function $\upsilon(p, \Theta_{0,\upsilon}(P))$ whose approximation bias can be bounded as follows:

$$|\mathbf{u}(p;\kappa) - \upsilon(p, \Theta_{0,\upsilon}(P))| \le CE[|l(Z)|^2]\kappa^{-2} \text{ uniformly in } p \in \mathbb{S}^\ell, \tag{40}$$

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 opposed to the maximum and minimum, the smooth weighted averages in (38)–(39) relate $\mathbf{g}_L, \mathbf{g}_U$, and γ in a differentiable manner. This suggests that, although regular estimation of $\upsilon(p, \Theta_{0,\upsilon}(P))$ is not generally possible, it may be possibile to estimate $\mathbf{u}(p;\kappa)$ in a regular manner, which we leave as future work.

4. ESTIMATION OF WEIGHTED AVERAGE DERIVATIVES IN MEAN REGRESSION

In this section, we illustrate estimation of support functions by studying a leading example. We focus on the case where Y is interval-valued, and the parameter of interest is a weighted average derivative of the mean regression function. That is, $\theta = E[w(Z)\nabla_z m(Z)]$, where m is the conditional mean of Y.

https://doi.org/10.1017/S0266466616000384 Published online by Cambridge University Press

Theorem 2.1 and the law of iterated expectations imply that the support function of the identified set in this setting is given by

$$\upsilon(p,\Theta_0(P)) = E[Y_p p' l(Z)], \tag{41}$$

where $Y_p = \Gamma(Y_L, Y_U, p'l(Z))$ and $l(z) \equiv -\nabla_z w(z) - w(z)\nabla_z f(z)/f(z)$. Our estimator of the support function replaces unknown objects in (41) with nonparametric kernel estimators and expectations with sample averages. Let $K : \mathbb{R}^\ell \to \mathbb{R}$ be a kernel function. For each $z \in \mathcal{Z}$, i, and bandwidth h, define the "leave-one-out" kernel density estimator by

$$\hat{f}_{i,h}(z) \equiv \frac{1}{(n-1)h^\ell} \sum_{j=1, j\neq i}^{n} K\Big(\frac{z - Z_j}{h}\Big). \tag{42}$$

Our estimator of l is then defined by

$$\hat{l}_{i,h}(z) \equiv -\nabla_z w(z) - w(z) \frac{\nabla_z \hat{f}_{i,h}(z)}{\hat{f}_{i,h}(z)} \tau_{n,i}, \tag{43}$$

where $\tau_{n,i}$ is a trimming function that is used to control for the stochastic denominator $\hat{f}_{i,h}(z)$.[7] The support function of $\Theta_0(P)$ is then estimated by

$$\hat{\upsilon}_n(p) \equiv \frac{1}{n} \sum_{i=1}^{n} p' \hat{l}_{i,h}(Z_i) \hat{Y}_{p,i}, \tag{44}$$

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 (44) only involves kernel density estimation and taking averages. Hence, it can be implemented 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 (44) only for one direction $p = \iota_j$, i.e., $\hat{\theta}^{(j)}_{U,n} = \hat{\upsilon}_n(\iota_j)$. 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 $M : \mathcal{Z} \to \mathbb{R}_+$ such that

$$\|\nabla_z f(z+e) - \nabla_z f(z)\| < M(z)\|e\|, \tag{45}$$

$$\|\nabla_z (f(z+e) \times m_j(z+e)) - \nabla_z (f(z) \times m_j(z))\| < M(z)\|e\|, \quad j = L, U, \tag{46}$$

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

Assumption 4.2. (i) The support $\mathcal{S}_K$ of K is a convex subset of $\mathbb{R}^\ell$ with nonempty interior with the origin as an interior point. Let $\partial\mathcal{S}_K$ be the boundary

https://doi.org/10.1017/S0266466616000384 Published online by Cambridge University Press

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:

$$\int u^{j_1}u^{j_2}\cdots u^{j_k}K(u)du = 0,\ j_1+\cdots+j_k < J, \tag{47}$$

$$\int u^{j_1}u^{j_2}\cdots u^{j_k}K(u)du \neq 0,\ j_1+\cdots+j_k = J. \tag{48}$$

(vi) The J moments of K exist.

Assumption 4.3. For each i and n, the trimming term is given by $\tau_{n,i} = \tau_n(Z_i)$ for some function $\tau_n : \mathcal{Z} \to \{0,1\}$ such that (i) $\|\tau_{n,i} - 1\|_{L^2_P} = o(n^{-1/2})$; (ii) $|\frac{\tau_n(z)}{f(z)}| \le b_n$ uniformly in z for some sequence b_n such that $b_n = o(n^\eta)$ for all $\eta > 0$.

Assumptions 4.1 and 4.2 are standard in the literature and based on the assumptions in Powell et al. (1989). Assumption 4.1 imposes suitable smoothness conditions on f, m_L, and m_U. Assumption 4.2 then gives regularity conditions on the kernel. A higher-order kernel is used to remove an asymptotic bias.[8] Assumption 4.3(i) requires that the trimming function tends to 1 fast enough so that the effect of the trimming is negligible. This is satisfied for example when $\tau_n(z) = 0$ when z is in a distance of ϵ_n of the boundary of the support and $\tau_n(z) = 1$ otherwise, and $n^{1/2}\epsilon_n \to 0$ (Lewbel, 2000). Assumption 4.3(ii) requires that the trimming function allows us to bound $\tau_n(z)/f(z)$ by a slowing diverging sequence. For example, if f has a sub-Gaussian tail, one may construct a trimming function that satisfies this condition (see Sherman, 1994 for details).

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 4.1–4.3 hold. Suppose further that* $h \to 0$, $nh^{\ell+2+\delta} \to \infty$ *for some* $\delta > 0$, $nh^{2J} \to 0$, $\tilde{h} \to 0$, *and* $n\tilde{h}^{\ell+2} \to \infty$ *as* $n \to \infty$. *Then,* (*a*) $\{\hat{\upsilon}_n(\cdot)\}$ *is a regular estimator for* $\upsilon(\cdot, \Theta_0(P))$; (*b*) *Uniformly in* $p \in \mathbb{S}^\ell$:

$$\sqrt{n}\{\hat{\upsilon}_n(p) - \upsilon(p,\Theta_0(P))\} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \big[w(z_i)p'\nabla_z m_p(Z_i) - \upsilon(p,\Theta_0(P)) + p'l(Z_i)\zeta_p(x_i)\big] + o_p(1); \tag{49}$$

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

$$\sqrt{n}\{\hat{\upsilon}_n(\cdot) - \upsilon(\cdot,\Theta_0(P))\} \xrightarrow{L} \mathbb{G}_0, \tag{50}$$

where $\mathbb{G}_0$ *is a tight mean zero Gaussian process on* $\mathcal{C}(\mathbb{S}^\ell)$ *with* $\mathrm{Cov}(\mathbb{G}_0(p_1), \mathbb{G}_0(p_2)) = E[\psi_{p_1}(X_i)\psi_{p_2}(X_i)']$.

Theorem 4.1 shows that $\hat{\upsilon}_n$ is semiparametrically efficient in the sense that it is asymptotically linear and has an efficient influence function. The limiting distribution of the estimator is given by a Gaussian process on the unit sphere whose covariance kernel coincides with the semiparametric efficiency bound obtained in Theorem 3.1.

Remark 4.1. Theorem 4.1 has an immediate consequence on set estimation. Note that 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:

$$\hat{\Theta}_n \equiv \mathrm{co}(\{\theta \in \Theta : \theta = \hat{\theta}_n(p),\ p \in \mathbb{S}^\ell\}), \tag{51}$$

where $\mathrm{co}(A)$ denotes the convex hull of A. Due to the equality of the Hausdorff distance between sets and the supremum distance between the corresponding support functions, we have that $\hat{\Theta}_n$ is $\sqrt{n}$-consistent: $d_H(\hat{\Theta}_n, \Theta_0(P)) = O_p(n^{-1/2})$. Further, as shown in KS (Section 4.1), if $\hat{\Theta}_n$ is associated with the efficient estimator of the support function, it holds under regularity conditions that

$$\limsup_{n\to\infty} E[L(\sqrt{n}d_H(\hat{\Theta}_n, \Theta_0(P)))] = E[L(\|\mathbb{G}_0\|_\infty)]. \tag{52}$$

From (35) and (52), it then follows that, among regular set estimators, $\hat{\Theta}_n$ asymptotically minimizes the risk based on the Hausdorff distance.

Remark 4.2. Efficient estimation of the support function also has implications on optimal inference. Specifically, consider testing

$$H_0 : \theta \in \Theta_0(P) \quad vs \quad H_1 : \theta \notin \Theta_0(P). \tag{53}$$

A natural test statistic for testing this hypothesis is the scaled directed Hausdorff distance $J_n(\theta) = \sqrt{n}\vec{d}_H(\theta, \hat{\Theta}_n) = \sqrt{n}\sup_{\tilde{\theta}\in\hat{\Theta}_n}\|\theta - \tilde{\theta}\|$, which may also be written as

$$J_n(\theta) = \sqrt{n}\sup_{p\in\mathbb{S}^\ell}\{\langle p,\theta\rangle - \hat{\upsilon}_n(p)\}_+. \tag{54}$$

Kaido (2016) shows that the appropriate critical value for this statistic is

$$c_{1-\alpha}(\theta) \equiv \inf\{c : P(\sup_{p\in\Psi(\theta)}(-\mathbb{G}_0(p))_+ \le c) \ge 1-\alpha\}, \tag{55}$$

where $\Psi(\theta) = \mathrm{argmax}_p\langle p,\theta\rangle - \upsilon_n(p, \Theta_0(P))$. This critical value can be estimated using a (score-based) weighted bootstrap (see, e.g., Lewbel, 1995). 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:

$$G_n^*(p) \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n W_i\Big\{\sum_{j=1, j\neq i}\frac{-2}{(n-1)h^{\ell+1}}\nabla_z K\Big(\frac{Z_i - Z_j}{h}\Big)(\hat{Y}_{p,i} - \hat{Y}_{p,j}) - \hat{\upsilon}_n(p)\Big\}, \tag{56}$$

where the process G_n^* is a U-process which is first-order asymptotically equivalent to the process $\frac{1}{\sqrt{n}}\sum_{i=1}^n W_i\psi_p(X_i)$.[9] 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 $\{X_i, i = 1, \ldots, n\}$. A consistent bootstrap estimator of $c_{1-\alpha}(\theta)$ is then

$$\hat{c}_{1-\alpha}(\theta) \equiv \inf\{c : P(\sup_{p \in \hat{\Psi}_n(\theta)} \big(-G_n^*(p)\big)_+ \leq c | \{X_i\}_{i=1}^n) \geq 1-\alpha\}, \quad \textbf{(57)}$$

where $\hat{\Psi}_n(\theta)$ is a consistent estimator of $\Psi(\theta)$.[10] Under regularity conditions, Theorems 5.3 and 5.4 in KS then suggest that a test that rejects the null hypothesis when $J_n(\theta) > \hat{c}_{1-\alpha}(\theta)$ is optimal in the sense that for any path $\{P_\eta, \eta \in \mathbb{R}\}$ of distributions such that $\theta \in \Theta_0(P_\eta)$ for $\eta \leq 0$ and $\theta \notin \Theta_0(P_\eta)$ for $\eta > 0$ with $\theta \in \partial\Theta_0(P)$ being supported by a unique hyperplane (see KS for details) and for any sequence of power functions $\pi_n(P_\eta)$ such that

$$\limsup_{n\to\infty} \pi_n(P_{\eta/\sqrt{n}}) \leq \alpha, \ \forall \eta \leq 0, \quad \textbf{(58)}$$

one has

$$\limsup_{n\to\infty} \pi_n(P_{\eta/\sqrt{n}}) \leq \lim_{n\to\infty} \pi_n^*(P_{\eta/\sqrt{n}}), \ \forall \eta > 0, \quad \textbf{(59)}$$

where $\pi_n^*(P_{\eta/\sqrt{n}}) = P_{\eta/\sqrt{n}}(J_n(\theta) > \hat{c}_{1-\alpha}(\theta))$. Thus, the test based on the efficient estimator of the support function achieves the power envelope for asymptotic level-α tests under local perturbations of the distribution from the null hypothesis where θ is on the smooth boundary of the identified set when $P_\eta = P$. This test can also be inverted to construct a confidence region for θ:

$$\mathcal{C}_{1n} \equiv \{\theta \in \Theta : J_n(\theta) \leq \hat{c}_{1-\alpha}(\theta)\}, \quad \textbf{(60)}$$

which can be shown to have the asymptotic coverage probability $1-\alpha$.

The bootstrap procedure above can be used in other types of inference. For example, a level $1-\alpha$ one-sided confidence set as in Beresteanu and Molinari (2008) can be constructed as $\mathcal{C}_{2n} \equiv \hat{\Theta}_n^{c_{2n}^*/\sqrt{n}}$, where $\hat{\Theta}_n^{\epsilon} \equiv \{\theta \in \Theta : \inf_{\theta' \in \hat{\Theta}_n} \|\theta - \theta'\| \leq \epsilon\}$ and c_{2n}^* is the $1-\alpha$ quantile of $\sup_{p \in \mathbb{S}^\ell}\{-G_n^*(p)\}_+$ (see also Kaido, 2016; Kaido and Santos, 2014). Yet another useful type of inference is the construction of 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

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

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:

$$Y_i = Z_i'\beta + \epsilon_i \qquad i = 1, \ldots, n, \quad \textbf{(61)}$$

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

$$Y_{L,i} = Y_i - c - e_{2i} Z_{2i}^2 - e_{3i} Z_{3i}^2,$$
$$Y_{U,i} = Y_i + c + e_{2i} Z_{2i}^2 + e_{3i} Z_{3i}^2, \quad \textbf{(62)}$$

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 in the supplementary material.

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 et al. (1989) and renormalize the weighted average derivative as follows:

$$\tilde{\theta} \equiv E[f(Z)]^{-1} E[f(Z)\nabla_z m(Z)]. \quad \textbf{(63)}$$

Integrating by parts, it holds that $I_\ell E[f(Z)] = E[\nabla_z Z f(Z)] = E[Zl(Z)]$, 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

$$\hat{\upsilon}_n^{IV} \equiv p' \Big(\frac{1}{n} \sum_{i=1}^n \hat{l}_{i,h}(Z_i) Z_i \Big)^{-1} \frac{1}{n} \sum_{i=1}^n \hat{l}_{i,h}(Z_i) Y_{p,i}, \quad \textbf{(64)}$$

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.[11]

Tables 1–2 in the supplementary material 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 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 that 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.

NOTES

1. Another 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).

2. These steps include applications of the U-process theory in Lemmas D.1 and D.3 in the supplementary material available at Cambridge Journals Online, and Lemma D.4, which uses an argument similar to Powell et al. (1989) to derive the asymptotic linear representation of the estimator of the support function with a pre-estimated infinite dimensional parameter.

3. On 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.

4. For the main partial identification result (Theorem 2.1), Assumption 2.3(iii) is a simplifying assumption. If this assumption does not hold, one may set $m_L(z)$ to the largest element of the solutions,

i.e., $m_L(z) = \sup\{\tilde{m} \in \mathbb{R} : E[q(Y_L - \tilde{m})|Z = z] = 0\}$ and similarly set $m_U(z)$ to the smallest element. For the efficiency result (Theorem 3.1), however, it is not clear if the results hold with this extended definition of m_L and m_U, and hence we maintain Assumption 2.3(iii) throughout.

5. In Example 2.3, this is not a problem as the regression function of interest is the conditional mean of Y. See also the discussions following Assumption 2.3 for sufficient conditions for the existence of the unique root.

6. This result uses the isometry $d_H(A, B) = \sup_{p\in\mathbb{S}^\ell} |\upsilon(p, A) - \upsilon(p, B)|$ between the space of compact convex sets and the space of support functions. See KS and references therein for details.

7. The trimming function can be dropped if w is supported in the interior of the support of Z.

8. Although we do not pursue here, alternative bandwidth asymptotics could provide better approximations to the finite sample distribution of the estimator than the procedure that removes the asymptotic bias through a higher-order kernel. For robust inference using an alternative asymptotic framework for point identified weighted average derivatives, we refer to the recent work by Cattaneo, Crump, and Jansson (2010, 2013).

9. This can be shown following an argument similar to the one in Section 3.4 in Powell et al. (1989). The proof is omitted for brevity.

10. Kaido (2016) proposed $\hat{\Psi}_n(\theta) = \{p \in \mathbb{S}^\ell : \langle p, \theta\rangle - \hat{\upsilon}_n(p) \geq \sup_{\tilde{p}}\{\langle \tilde{p}, \theta\rangle - \hat{\upsilon}_n(\tilde{p})\} - \kappa_n/\sqrt{n}\}$, which is consistent in Hausdorff distance provided $\kappa_n \to \infty$ and $\kappa_n = o(n^{1/2})$.

11. Detailed description of the construction of the higher-order kernels is in Powell et al. (1989) Appendix 2.

REFERENCES

Beresteanu, A., I. Molchanov, & F. Molinari (2011) Sharp identification regions in models with convex predictions. *Econometrica* 79(6), 1785–1821.

Beresteanu, A. & 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, & J.A. Wellner (1993) *Efficient and Adaptive Estimation for Semiparametric Models*. Springer.

Blundell, R., D. Kristensen, & R. Matzkin (2014) Bounding quantile demand functions using revealed preference inequalities. *Journal of Econometrics* 179(2), 112–127.

Bugni, F.A., I.A. Canay, & X. Shi (2016) Inference for subvectors and other functions of partially identified parameters in moment inequality models. *Quantitative Economics*, first published online 12 February 2016.

Carneiro, P., J.J. Heckman, & 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, & M. Jansson (2010) Robust data-driven inference for density-weighted average derivatives. *Journal of the American Statistical Association* 105(491), 1070–1083.

Cattaneo, M.D., R.K. Crump, & M. Jansson (2013) Small bandwidth asymptotics for density-weighted average derivatives. *Econometric Theory* 30, 1–25.

Chandrasekhar, A., V. Chernozhukov, F. Molinari, & P. Schrimpf (2011) Inference for Best Linear Approximations to Set Identified Functions. Discussion paper, University of British Columbia.

Chaudhuri, P., K. Doksum, & A. Samarov (1997) On average derivative quantile regression. *The Annals of Statistics* 25(2), 715–744.

Chernozhukov, V., H. Hong, & E. Tamer (2007) Estimation and confidence regions for parameter sets in econometric models. *Econometrica* 75(5), 1243–1284.

Chernozhukov, V., E. Kocatulum, & K. Menzel (2015) Inference on sets in finance. *Quantitative Economics* 6(2), 309–358.

Chernozhukov, V., S. Lee, & A. Rosen (2013) Intersection bounds: Estimation and inference. *Econometrica* 81(2), 667–737.

Crossley, T.F. & K.M. Pendakur (2010) The common-scaling social cost of living index. *Journal of Business and Economic Statistics* 28(4), 523–538.

Deaton, A. & S. Ng (1998) Parametric and nonparametric approaches to price and tax reform. *Journal of the American Statistical Association* 93(443), 900–909.
Härdle, W. & T. Stoker (1989) Investigating smooth multiple regression by the method of average derivatives. *Journal of the American Statistical Association* 84(408), 986–995.
Hirano, K. & J.R. Porter (2012) Impossibility results for nondifferentiable functionals. *Econometrica* 80(4), 1769–1790.
Horowitz, J.L. (1996) Semiparametric estimation of a regression model with an unknown transformation of the dependent variable. *Econometrica* 64(1), 103–137.
Ichimura, H. (1993) Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *Journal of Econometrics* 58(1–2), 71–120.
Kaido, H. (2016) A dual approach to inference for partially identified models. *Journal of Econometrics* 192, 269–290.
Kaido, H., F. Molinari, & J. Stoye (2016) Confidence Intervals for Projections of Partially Identified Parameters. Working paper, Boston University.
Kaido, H. & A. Santos (2014) Asymptotically efficient estimation of models defined by convex moment inequalities. *Econometrica* 82(1), 387–413.
Klein, R. & R. Spady (1993) An efficient semiparametric estimator for binary response models. *Econometrica* 61(2), 387–421.
Lewbel, A. (1995) Consistent nonparametric hypothesis tests with an application to Slutsky symmetry. *Journal of Econometrics* 67, 379–401.
Lewbel, A. (2000) Asymptotic Trimming for Bounded Density Plug-in Estimators. Discussion paper, Boston College.
Manski, C.F. & E. Tamer (2002) Inference on regressions with interval data on a regressor or outcome. *Econometrica* 70(2), 519–546.
Milgrom, P. & I. Segal (2002) Envelope theorems for arbitrary choice sets. *Econometrica* 70(2), 583–601.
Newey, W. & T. Stoker (1993) Efficiency of weighted average derivative estimators and index models. *Econometrica* 61(5), 1199–1223.
Nishiyama, Y. & P. Robinson (2000) Edgeworth expansions for semiparametric averaged derivatives. *Econometrica* 68(4), 931–979.
Ponomareva, M. & E. Tamer (2011) Misspecification in moment inequality models: Back to moment equalities? *The Econometrics Journal* 14(2), 186–203.
Powell, J., J. Stock, & T. Stoker (1989) Semiparametric estimation of index coefficients. *Econometrica* 57(6), 1403–1430.
Sherman, R.P. (1994) U-processes in the analysis of a generalized semiparametric regression estimator. *Econometric Theory* 10, 372–372.
Song, K. (2014) Local asymptotic minimax estimation of nonregular parameters with translation-scale equivariant maps. *Journal of Multivariate Analysis* 125, 136–158.
Stoker, T. (1986) Consistent estimation of scaled coefficients. *Econometrica* 54(6), 1461–1481.