WINPEC Working Paper Series No.E1723
March 2018

Optimal Bid Selection for Differentiable Functionals of a Semi-parametric Binary Response Model

Hisatoshi Tanaka

Waseda INstitute of Political EConomy
Waseda University
Tokyo,Japan

Optimal Bid Selection for Differentiable Functionals of a Semi-parametric Binary Response Model*

Hisatoshi Tanaka†

February 26, 2018

Abstract

This study develops a method of optimal survey design for the dichotomous choice contingent valuation method to estimate differentiable functionals of a semi-parametric binary response model. The method uses an iterative procedure to select the survey design that minimizes semi-parametric efficiency bounds of the estimation. An efficient method of implementation is discussed.

1 Introduction

The *dichotomous-choice contingent valuation method* (DC-CVM), which is a survey method using hypothetical discrete response (yes/no) valuation questions, is one of the most widely used techniques to measure the welfare value of non-marketed goods such as environmental resources. The idea was introduced by Bishop and Heberlein (1979), who asked the respondent to accept or reject a suggested cost (*bid price*) for a certain given environmental change. In the method, the researcher chooses m different bid prices, x^1, …, x^m, and gives them to n_1, …, n_m individuals within a hypothetical valuation question. By counting the number of people accepting the given prices within each subgroup, the distribution F of willingness to pay (WTP) ω for the good is estimated. Bishop and Heberlein (1979) estimate F by logistic regression. Kristrom (1990) introduces a non-parametric maximum likelihood method to estimate F directly. Once F is estimated, its functional values, such as the mean $\int \omega\, dF(\omega)$ or the median $F^{-1}(0.5)$ of the WTP, are obtained. For more details of the contingent valuation method (CVM), see *e.g.,* Carson and Hanemann (2006).

It is noteworthy that the accuracy of the inference depends not only on the method of estimation, but also on the bid design. Cooper (1993) proposes a method to choose the number of bid prices m, the locations of bid prices, $x^1, \ldots, x^m$, and the sizes of the subgroups, n_1, …, n_m. The design is determined to minimize the mean squared error (MSE) of the mean estimation. By thorough numerical experiments, Cooper (1993) shows that the optimal design significantly reduces the MSE.

Sharp identification of general functional values of F, however, often requires the bid prices to have infinite variations. Therefore, the finite support assumption of the prices should be relaxed for more general applications. The purpose of this study is to introduce a generalized method to design DC-CVM surveys, and propose a non-parametric efficient estimation under the optimal design.

For this purpose, it is necessary to consider the smoothness of the target of estimation because, according to van der Vaart (1991), a parameter has regular estimators only if it is expressed as a *differentiable functional* of the underlying statistical model. In general, the efficiency bound of estimation is defined as the lower bound of asymptotic variances of regular estimators for the parameter. Therefore, if the target of estimation is not a differentiable functional, the optimal design might become an ill-posed problem, since it does not have well-defined objectives to be minimized.

The rest of this paper is organized as follows. In Section 2, a necessary and sufficient condition for differentiability of functionals of a semi-parametric binary response model is presented. In Section 3, an optimal design for differentiable functionals is proposed. In Section 4, an efficient estimation to implement the method is given.

2 The Model

This section summarizes the theory of differentiable functionals. For more details of the theory, see, for example, van der Vaart (1991), Bickel et al. (1993), and Severini and Tripathi (2001). For differentiability of the binary response model in particular, see Chapter 10 of Groeneboom and Jongbloed (2014).

In the following, $L_2(P)$ is a space of P-square integrable functions with inner product $\langle \alpha_1, \alpha_2 \rangle_P = \int \alpha_1 \alpha_2 \, dP$ and norm $\|\alpha\|_P = \langle \alpha, \alpha \rangle_P^{1/2}$, where P is

*JEL Classification: C14, Q51

†School of Political Science and Economics, Waseda University. Nishiwaseda 1-6-1, Shinjuku-ku, Tokyo 169-8050, Japan. E-mail: hstnk@waseda.jp

a generic distribution; $L_2^0(P)$ is a subset of $L_2(P)$ such that $\int \alpha\, dP = 0$ if $\alpha \in L_2^0(P)$.

Suppose that $\omega \sim F$, and $x \sim G$, where ω and x are independent. Here, ω is the variable of main interest (WTP), and x is a censoring variable (bid price). Assume also that the only observable variables are $z = (x, y)$, where $y = \{\omega \le x\}$. Let $dP_{F,G}(z)$ denote the density of z when (F, G) is given, that is,

$$dP_{F,G}(z) = F(x)^y(1 - F(x))^{1-y}dG(x)d\delta(y), \tag{2.1}$$

where δ is the counting measure on $\{0, 1\}$; $P_{F,G}$ is the probability measure defined by $P_{F,G}D = \int_D dP_{F,G}$ for every measurable $D \subset \mathbf{R} \times \{0, 1\}$; $\mathcal{F} = \mathcal{F}[0, 1]$ is a set of distribution functions defined on $[0, 1]$; and $\mathcal{P}_G = \{P_{F,G} : F \in \mathcal{F}\}$ is the model set, where G is given and fixed; The goal of estimation is a value of parameter $\theta = \varphi(F)$, where $\varphi : \mathcal{F} \mapsto \mathbf{R}$ is a known functional of unknown F.

Let $F_0 \in \mathcal{F}$ be the *true* distribution of ω to be estimated; $\mathcal{F}(F_0)$ is a set of 1-dimensional submodels $t \in (0, \epsilon) \mapsto F_t \in \mathcal{F}$ such that

$$\int \left(\frac{\sqrt{dF_t} - \sqrt{dF_0}}{t} - \frac{1}{2}\alpha\sqrt{dF_0} \right)^2 \to 0 \tag{2.2}$$

as $t \downarrow 0$ for elements $\alpha \in L_2(F_0)$; $T(F_0)$ is the *tangent space* of $\mathcal{F}$ at F_0, which is the set of all α, as in (2.2). It is known that $T(F_0) = L_2^0(F_0)$. See, for example, Groeneboom and Wellner (1992) for the proof.

Let $P_0 = P_{F_0,G}$ be the true model from which data $z_1 = (x_1, y_1)$, ..., $z_n = (x_n, y_n)$ are independently sampled; $\mathcal{P}_G(P_0)$ is a set of 1-dimensional submodels $t \in (0, \epsilon) \mapsto P_t = P_{F_t,G} \in \mathcal{P}_G$ induced by each $F_t \in \mathcal{F}(F_0)$; $\beta \in L_2^0(P_0)$ such that

$$\int \left(\frac{\sqrt{dP_t} - \sqrt{dP_0}}{t} - \frac{\beta}{2}\sqrt{dP_0} \right)^2 \to 0 \tag{2.3}$$

as $t \downarrow 0$ is the *score function* of P_t; $T(P_0)$ is the tangent space of $\mathcal{P}_G$ at P_0, which is the set of all β, as in (2.3).

The *score operator* A maps each $\alpha \in T(F_0)$ into corresponding $\beta \in T(P_0)$ by

$$(A\alpha)(x, y) = \frac{y - F_0(x)}{F_0(x)(1 - F_0(x))} \int_0^x \alpha dF_0. \tag{2.4}$$

The tangent space of $\mathcal{P}_G$ is equivalent to the range of A: that is, $T(P_0) = AT(F_0) = R(A)$. Define $\mathcal{G} \subset \mathcal{F}$ by

$$\mathcal{G} = \left\{ G \in \mathcal{F} : \int \frac{dG}{F_0(1 - F_0)} < \infty \right\}, \tag{2.5}$$

and assume as follows.

(A1) $G \in \mathcal{G}$,

so that A becomes a linear continuous operator because $(\|A\alpha\|_{P_0}/\|\alpha\|_{F_0})^2 \le \int \frac{dG}{F_0(1-F_0)} < \infty$. Under **(A1)**, the adjoint operator of A is given by

$$(A^*\beta)(\omega) = \int_\omega^1 \beta(x, 1)\, dG(x) + \int_0^\omega \beta(x, 0)\, dG(x). \tag{2.6}$$

Let $\theta = \varphi(F) \in \mathbf{R}$ be a parameter of interest. In particular, $\theta_0 = \varphi(F_0)$ is to be estimated. Assume that φ is smooth in the following sense.

(A2) A *gradient* $\tilde{\varphi}_0 \in T(F_0)$ exists such that

$$\frac{\varphi(F_t) - \varphi(F_0)}{t} \to \langle \tilde{\varphi}_0, \alpha \rangle_{F_0} \quad (t \downarrow 0) \tag{2.7}$$

for every $F_t \in \mathcal{F}(F_0)$.

In addition, it is assumed that the parameter is identified, as follows.

(A3) A functional $\kappa : \mathcal{P}_G \mapsto \mathbf{R}$ exists such that $\kappa(P_{F,G}) = \varphi(F)$ for every $F \in \mathcal{F}$ and $G \in \mathcal{G}$.

Assumption **(A3)** together with **(A2)** implies that $\lim_{t\downarrow 0} t^{-1}(\kappa(P_t) - \kappa(P_0))$ exists for every path P_t in $\mathcal{P}_G(P_0)$. The functional κ is said to be *differentiable at* P_0 *relative to* $\mathcal{P}_G(P_0)$ if there is a linear continuous functional $\kappa_0' : T(P_0) \mapsto \mathbf{R}$ such that

$$\frac{\kappa(P_t) - \kappa(P_0)}{t} \to \kappa_0'\beta \quad (t \downarrow 0) \tag{2.8}$$

for every $P_t \in \mathcal{P}_G(P_0)$.

If (2.8) holds, the Riesz representation theorem shows the existence of the *influence function* $\tilde{\kappa}_0 \in \overline{T(P_0)}$ such that $\kappa_0'\beta = \langle \tilde{\kappa}_0, \beta \rangle_{P_0}$ for every $\beta \in T(P_0)$. Since $\langle \tilde{\kappa}_0, \beta \rangle_{P_0} = \langle \tilde{\kappa}_0, A\alpha \rangle_{P_0} = \langle A^*\tilde{\kappa}_0, \alpha \rangle_{F_0} = \langle \tilde{\varphi}_0, \alpha \rangle_{F_0}$, $\tilde{\varphi}_0 \in R(A^*)$ is necessary for differentiability of κ. The differentiability theorem by van der Vaart (1991) proves that $\tilde{\varphi}_0 \in R(A^*)$ is sufficient for the differentiability of κ. Consequently, the following theorem is obtained.

Theorem 2.1 Assume **(A1)-(A3)**. Assume that G has a density function $g = G'$. Then, κ is differentiable at P_0 relative to $\mathcal{P}_G(P_0)$ if and only if the gradient $\tilde{\varphi}_0$ has the first derivative $\tilde{\varphi}_0'$ such that

$$\int (\tilde{\varphi}_0')^2 \cdot \frac{F_0(1 - F_0)}{g} < \infty. \tag{2.9}$$

(Proof) Assume that $\tilde{\varphi}_0$ has the first derivative $\tilde{\varphi}_0'$ satisfying (2.9). Define $\tilde{\kappa}_0$ by

$$\tilde{\kappa}_0(x, y) = -\tilde{\varphi}_0'(x) \cdot \frac{y - F_0(x)}{g(x)}. \tag{2.10}$$

Then, (2.9) implies $\tilde{\kappa}_0 \in L_2^0(P_0)$ and

$$\begin{aligned}(A^*\tilde{\kappa}_0)(\omega) &= -\int_\omega^1 \tilde{\varphi}_0'(x) \cdot \frac{1 - F_0(x)}{g(x)} dG(x) \\ &\quad + \int_0^\omega \tilde{\varphi}_0'(x) \cdot \frac{F_0(x)}{g(x)} dG(x) = \tilde{\varphi}_0(\omega).\end{aligned}$$

Hence, $\tilde{\varphi}_0 \in R(A^*)$ is shown.

On the other hand, if $\tilde{\varphi}_0 \in R(A^*)$ holds, $\tilde{\kappa}_0 \in L_2^0(P_0)$ exists such that

$$\begin{aligned}\tilde{\varphi}_0(\omega) &= (A^*\tilde{\kappa}_0)(\omega) \\ &= \int_\omega^1 \tilde{\kappa}_0(x,1)\,dG(x) + \int_0^\omega \tilde{\kappa}_0(x,0)\,dG(x),\end{aligned}$$

whose first derivative, $\tilde{\varphi}_0'(\omega) = -[\tilde{\kappa}_0(\omega,1) - \tilde{\kappa}_0(\omega,0)] \cdot g(\omega)$, satisfies (2.9), because

$$\int (\tilde{\varphi}_0')^2 \cdot \frac{F_0(1-F_0)}{g} = \int \tilde{\kappa}_0^2\,dP_0 < \infty.$$

■

3 Optimal Bid Design for Differentiable Functionals

Suppose that $G \in \mathcal{G}$ is given and fixed. Let $\{\hat{\theta}_n\}$ be a regular estimator for $\theta_0 = \kappa(P_0)$; L is a tight Borel law on $\mathbf{R}$ such that

$$\sqrt{n}(\hat{\theta}_n - \kappa(P_{h_n/\sqrt{n}})) \overset{w}{\Rightarrow}_{P_{h_n/\sqrt{n}}} L$$

for every path $P_t \in \mathcal{P}_G(P_0)$ and sequence $h_n \to h \in \mathbf{R}$, where "$\overset{w}{\Rightarrow}_{P_{h_n/\sqrt{n}}}$" means weak convergence under $P_{h_n/\sqrt{n}}$. In particular, $\sqrt{n}(\hat{\theta}_n - \theta_0) \overset{w}{\Rightarrow}_{P_0} L$ as $n \to \infty$ when $\{\hat{\theta}_n\}$ is regular, and hence, the asymptotic variance of $\{\hat{\theta}_n\}$ equals

$$\mathrm{avar}\{\hat{\theta}_n\} = \int \left(u - \int u\,dL(u)\right)^2 dL(u). \qquad (3.1)$$

Let $\hat{\Theta}$ be a set of regular estimators for θ_0. Given G, the efficiency bound $v_0(G)$ for estimating θ_0 is defined by

$$v_0(G) = \inf\ \mathrm{avar}\{\hat{\theta}_n\} \quad \text{subject to} \quad \{\hat{\theta}_n\} \in \hat{\Theta}. \qquad (3.2)$$

The *optimal bid distribution* G_0 minimizes $v_0(G)$ over $\mathcal{G}$. Now, the main result of the study is ready to be stated.

Theorem 3.1 Assume **(A1)-(A3)**. Assume that $\tilde{\varphi}_0$ is differentiable and

$$\int \frac{|\tilde{\varphi}_0'|}{\sqrt{F_0(1-F_0)}} < \infty. \qquad (3.3)$$

Then, the optimal bid distribution is $G_0(x) = \int_0^x g_0$, where

$$g_0 = \frac{|\tilde{\varphi}_0'|\sqrt{F_0(1-F_0)}}{\int |\tilde{\varphi}_0'|\sqrt{F_0(1-F_0)}}. \qquad (3.4)$$

The optimized efficiency bound equals

$$v_0 = \left(\int |\tilde{\varphi}_0'|\,\sqrt{F_0(1-F_0)}\right)^2. \qquad (3.5)$$

(Proof) When κ is differentiable at P_0 relative to $\mathcal{P}_G(P_0)$, the efficiency bound to estimate $\theta_0 = \kappa(P_0)$ equals $\|\tilde{\kappa}_0\|_{P_0}^2$ (van der Vaart (1991); Bickel et al. (1993); Severini and Tripathi (2001)). Equation (2.10) implies

$$\|\tilde{\kappa}_0\|_{P_0}^2 = \int (\tilde{\varphi}_0')^2 \cdot \frac{F_0(1-F_0)}{g}. \qquad (3.6)$$

The Cauchy–Schwartz inequality implies that, for an arbitrary square integrable function ψ,

$$\begin{aligned}\int |\tilde{\varphi}_0'|\,\sqrt{F_0(1-F_0)} &\le \sqrt{\int (\tilde{\varphi}_0')^2 \frac{F_0(1-F_0)}{\psi^2} \int \psi^2} \\ &= \sqrt{\int (\tilde{\varphi}_0')^2 \frac{F_0(1-F_0)}{g}},\end{aligned}$$

where $g = \psi^2 / \int \psi^2$ is an arbitrary density. Therefore, the lower bound for $v_0(G)$ is given by $v_0 = \left(\int |\tilde{\varphi}_0'|\,\sqrt{F_0(1-F_0)}\right)^2$. The bound is attained if and only if $\psi^2 = |\tilde{\varphi}_0'|\sqrt{F_0(1-F_0)}$, and hence, (3.4) minimizes $v_0(G)$. Note that condition (3.3) implies that

$$\int |\tilde{\varphi}_0'|\sqrt{F_0(1-F_0)} \le \frac{1}{4}\int \frac{|\tilde{\varphi}_0'|}{\sqrt{F_0(1-F_0)}} < \infty$$

and that

$$\int \frac{g_0}{F_0(1-F_0)} = \frac{\int \frac{|\tilde{\varphi}_0'|}{\sqrt{F_0(1-F_0)}}}{\int |\tilde{\varphi}_0'|\sqrt{F_0(1-F_0)}} < \infty,$$

and hence, $G_0 \in \mathcal{G}$. ■

Example 3.1 Suppose that the m-th moment of ω is a parameter of interest; that is, $\theta = \kappa(P_{F,G}) = \int \omega^m\,dF(\omega)$. To observe whether the functional is differentiable, let $dF_t = (1 + t \cdot \alpha)dF_0$ for arbitrary $\alpha \in L_2^0(F_0)$. Then,

$$\frac{\varphi(F_t) - \varphi(F_0)}{t} = \int \omega^m \cdot \alpha(\omega)\,dF_0(\omega) = \langle \tilde{\varphi}_0, \alpha \rangle_{F_0},$$

where $\tilde{\varphi}_0(\omega) = \omega^m - \int \omega^m\,dF_0(\omega)$. Since the gradient has the first derivative $\tilde{\varphi}_0'(\omega) = m \cdot \omega^{m-1}$, the functional κ is differentiable if $\int \frac{F_0(1-F_0)}{g} < \infty$.

Assume that $\int F_0^{-1/2}(1-F_0)^{-1/2} < \infty$. Then, $\tilde{\varphi}_0(\omega) = \omega^m - \int \omega^m\,dF_0(\omega)$ satisfies (3.3). The optimal bid density is

$$g_0(x) = \frac{x^{m-1}\sqrt{F_0(x)(1-F_0(x))}}{\int u^{m-1}\sqrt{F_0(u)(1-F_0(u))}\,du}$$

with the optimized efficiency bound,

$$v_0 = \left(\int u^{m-1}\sqrt{F_0(u)(1-F_0(u))}\,du\right)^2.$$

Figure 3.1 depicts graphs of the optimal bid densities for $m = 1$, 2, and 5, where $F_0 = U[0, 1]$. Note that $\int F_0^{-1/2}(1 - F_0)^{-1/2} < \infty$ holds in this case. The graphs show that the optimal density tends to skew to the left as m grows.

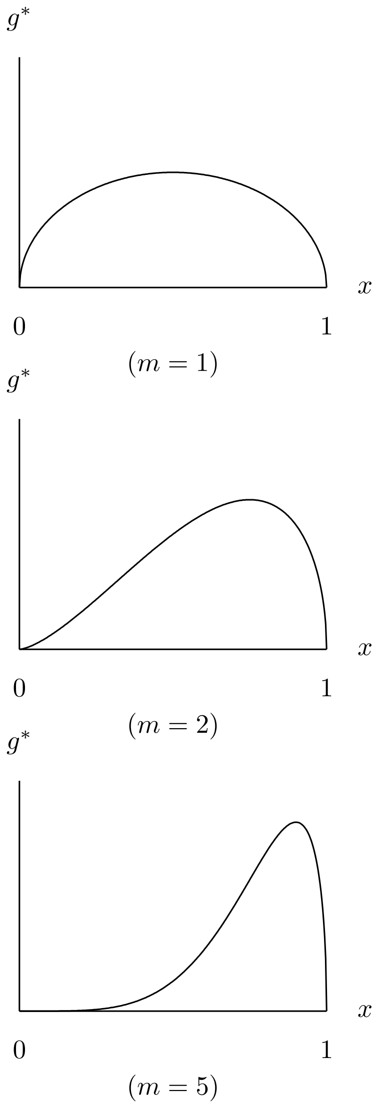

Figure 3.1: The optimal bid densities to estimate $E\omega^m$ for the cases of $m = 1$, 2, and 5, where $F_0 = U[0, 1]$.

Example 3.2 Suppose that the variance of ω is a parameter of interest; that is, $\kappa(P_{F,G}) = \int \omega^2\, dF(\omega) - \left(\int \omega\, dF(\omega)\right)^2$. Let $dF_t = (1 + t \cdot \alpha) dF_0$ for some $\alpha \in L_2^0(F_0)$. Then, $t^{-1} \cdot (\varphi(F_t) - \varphi(F_0)) \to \langle \tilde{\varphi}_0, \alpha \rangle_{F_0}$ as $t \downarrow 0$, where

$$\begin{aligned} \tilde{\varphi}_0(\omega) &= \omega^2 - \int \omega^2\, dF_0(\omega) \\ &\quad -2\left(\int \omega\, dF_0(\omega)\right)\left(\omega - \int \omega\, dF_0(\omega)\right). \end{aligned}$$

The gradient has the first derivative $\tilde{\varphi}_0'(\omega) = 2(\omega - \int \omega\, dF_0(\omega))$, and hence, κ becomes differentiable if $\int \omega^2 \cdot \frac{F_0(\omega)(1-F_0(\omega))}{g(\omega)}\, d\omega < \infty$.

Assume that $\int F_0^{-1/2}(1 - F_0)^{-1/2} < \infty$. Then,

$$\int \frac{|\tilde{\varphi}_0'|}{\sqrt{F_0(1 - F_0)}} \leq 4 \int F_0^{-1/2}(1 - F_0)^{-1/2} < \infty.$$

Therefore, the optimal bid density for estimating $Var(\omega)$ is given by

$$g_0(x) = \frac{|x - \mu_0|\sqrt{F_0(x)(1 - F_0(x))}}{\int |u - \mu_0|\sqrt{F_0(u)(1 - F_0(u))}\, du},$$

where $\mu_0 = \int \omega\, dF_0(\omega)$. The optimized efficiency bound equals

$$v_0 = \left(\int |u - \mu_0|\sqrt{F_0(u)(1 - F_0(u))}\, du\right)^2.$$

Figure 3.2 depicts a graph of the optimal bid density for estimation of the variance, where $F_0 = U[0, 1]$. The graph shows strong bimodality around the mean, $E\omega = 0.5$.

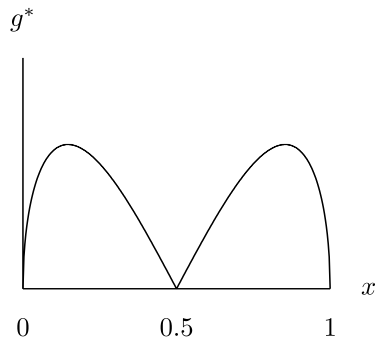

Figure 3.2: The optimal bid densities for estimation of the variance of ω, where $F_0 = U[0, 1]$.

Example 3.3 Suppose that the value of F_0 at $\omega_0 \in (0, 1)$ is a parameter of interest; that is, $\kappa(P_{F,G}) = F(\omega_0)$. For arbitrary $F_t \in \mathcal{F}(F_0)$,

$$\frac{F_t(\omega_0) - F_0(\omega_0)}{t} = \int_{-\infty}^{\omega_0} \alpha\, dF_0 = \langle \tilde{\varphi}_0, \alpha \rangle_{F_0},$$

where $\tilde{\varphi}_0(\omega) = \{\omega \leq \omega_0\} - F_0(\omega_0)$. Since $\tilde{\varphi}_0$ is discontinuous at ω_0, $\kappa(p_{F,G}) = F(\omega_0)$ fails to be differentiable and Theorem 3.1 is not applied to this case.

Example 3.4 Let $\theta = median(F)$. Assume that every $F \in \mathcal{F}$ has a positive density $f = F'$ at θ, so that $\varphi(F) = F^{-1}(1/2)$. For arbitrary $F_t \in \mathcal{F}(F_0)$,

$$\lim_{t \downarrow 0} \frac{\varphi(F_t) - \varphi(F_0)}{t} = -\frac{\int_{-\infty}^{\theta_0} \alpha\, dF_0}{f_0(\theta_0)} = \langle \tilde{\varphi}_0, \alpha \rangle_{F_0},$$

where $\tilde{\varphi}_0(\omega) = f_0(\theta_0)^{-1} \cdot (1/2 - \{\omega \leq \theta_0\})$. Since $\tilde{\varphi}_0$ is discontinuous at θ_0, $\kappa(p_{F,G}) = F^{-1}(1/2)$ fails to be differentiable and Theorem 3.1 is not applied to the median estimation.

4 Estimation

Since G_0 given by (3.4) contains unknown F_0, the optimal bid design is implemented only by an iterative procedure. In the first step, a preliminary survey is undertaken to obtain a prior estimate of F_0. Let F^0 be the first-step estimate, and assume that

(A4) $^\exists \lambda > 0$ such that $\lambda < F^0(0) < F^0(1) < 1 - \lambda$

and that

(A5) $F^0 \circ (F_0^{-1}) : [0,1] \mapsto (\lambda, 1-\lambda)$ is Lipschitz continuous.

Assumption **(A4)** means that, in the preliminary survey, the range of bid amounts is set large enough to contain the entire support of the WTP. Assumption **(A5)** is satisfied if F_0 has a positive density on $(0,1)$, and if, say, the probit estimation is employed to approximate F_0 in the first step.

In the second step, a main part of the survey is conducted by using the feasible bid distribution $G^0 = \int g^0$, where g^0 is the feasible bid density based on F^0, that is,

$$g^0 = \frac{|\tilde{\varphi}_0'| \sqrt{F^0(1-F^0)}}{\int |\tilde{\varphi}_0'| \sqrt{F^0(1-F^0)}}. \tag{4.1}$$

Using data $z_1 = (x_1, y_1)$, $\ldots$, $z_n = (x_n, y_n)$ sampled from P_{F_0, G^0}, F_0 is consistently estimated by the non-parametric maximum likelihood estimator, $\hat{F}_n = \ell_n(F)$ subject to $F \in \mathcal{F}$, where

$$\ell_n(F) = \sum_{i=1}^n \left[y_i \log F(x_i) + (1-y_i) \log(1 - F(x_i)) \right].$$

The method is introduced by Kristrom (1990) in the CVM literature. The pool-adjust-violate (PAV) algorithm is employed to solve the maximization problem. For details of non-parametric maximum likelihood estimation, see, for example, Cosslett (1983), Huang and Wellner (1995), van de Geer (2000), van der Vaart and Wellner (2007), or Groeneboom and Jongbloed (2014) among many others.

The parameter is estimated by $\hat{\theta}_n = \varphi(\hat{F}_n)$. To derive asymptotic distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$, the first-order expansion of φ is assumed as follows:

(A6) $\tilde{\varphi}_0 \in L_2(F_0)$ exists such that, for every $F \in \mathcal{F}$,

$$\varphi(F) = \varphi(F_0) + \int \tilde{\varphi}_0 \, d(F - F_0) + O(h^0(F, F_0)^2),$$

where $h^0(\cdot,\cdot)$ is the metric defined by $h^0(F,F')^2 = \int (\sqrt{F} - \sqrt{F'})^2 dG^0$ for $F, F' \in \mathcal{F}$.

In the case of estimating the non-centered moment $E\omega^m$, for example, $\varphi(F) - \varphi(F_0) = \int \omega^m \, d(F - F_0)$, hence the condition is trivially satisfied.

Assumption **(A6)** implies **(A1)** if $\tilde{\varphi}_0$ is bounded on $[0,1]$. To observe this, note that for every $P_t = P_{F_t, G^0}$, $\beta \in T(P_0)$ exists such that

$$\begin{aligned} h^0(F_t, F_0) &\leq \sqrt{\int (\sqrt{dP_t} - \sqrt{dP_0})^2} \\ &\leq \sqrt{\int \left(\sqrt{dP_t} - \sqrt{dP_0} - t \cdot \frac{\beta}{2} \sqrt{dP_0} \right)^2} \\ &\quad + \frac{t}{2} \sqrt{\int \beta^2 dP_0} \end{aligned}$$

and therefore, $\lim_{t \downarrow 0} \frac{h^0(F_t, F_0)}{t} \leq \frac{1}{2} \|\beta\|_{P_0} < \infty$. Hence, for any $F_t \in \mathcal{F}(F_0)$,

$$\lim_{t \downarrow 0} \frac{\varphi(F_t) - \varphi(F_0)}{t} = \lim_{t \downarrow 0} \frac{1}{t} \int \tilde{\varphi}_0 \, d(F_t - F_0).$$

Set $\xi_t = \frac{2}{t} \left(\sqrt{\frac{dF_t}{dF_0}} - 1 \right) - \alpha$, so that $\int \xi_t^2 dF_0 \to 0$ as $t \downarrow 0$ and that

$$\frac{1}{t} \int \tilde{\varphi}_0 \, d(F_t - F_0) = \left\langle \tilde{\varphi}_0, \alpha + \xi_t + \frac{t}{4} (\alpha + \xi_t)^2 \right\rangle_{F_0}.$$

Hence, $\frac{1}{t} \int \tilde{\varphi}_0 \, d(F_t - F_0) \to \langle \tilde{\varphi}_0, \alpha \rangle_{F_0}$ as $t \downarrow 0$ if $\tilde{\varphi}_0$ is bounded on $[0,1]$.

It is also assumed that

(A7) $\tilde{\varphi}_0 : [0,1] \mapsto \mathbf{R}$ is a monotone function.

The assumption is technical and satisfied in the case of estimation of the non-centered moments $E\omega^m$, but not in the case of the variance estimation.

Theorem 4.1 Assume **(A2)-(A7)**, then $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, v^0)$ as $n \to \infty$, where

$$v^0 = \left(\int |\tilde{\varphi}_0'| \sqrt{F^0(1-F^0)} \right)^2 \int \frac{F_0(1-F_0)}{F^0(1-F^0)} dG^0.$$

(Proof) The following proof employs the strategy proposed in Huang and Wellner (1995). By **(A6)**, $\sqrt{n}(\hat{\theta}_n - \theta_0) = \sqrt{n} \int \tilde{\varphi}_0 \, d(\hat{F}_n - F_0) + \sqrt{n}\, O(h^0(\hat{F}_n, F_0)^2)$. Let $\hat{P}_n = P_{\hat{F}_n, G^0}$. Then,

$$\int (\sqrt{d\hat{P}_n} - \sqrt{dP_0})^2 = O_p(n^{-2/3})$$

as $n \to \infty$ (van de Geer (1993); Patilea (2001)). Since $h^0(\hat{F}_n, F_0)^2 \leq \int (\sqrt{d\hat{P}_n} - \sqrt{dP_0})^2$,

$$\begin{aligned} &\sqrt{n}(\hat{\theta}_n - \theta_0) \\ &\quad = -\sqrt{n} \int \tilde{\varphi}_0' \sqrt{F^0(1-F^0)} \int \frac{\hat{F}_n - F_0}{\sqrt{F^0(1-F^0)}} dG^0 \\ &\qquad + o_p(1), \end{aligned}$$

by **(A7)**.

Set $\Lambda = F^0 \circ (F_0^{-1})$. By **(A4)** and **(A5)**, a constant C exists such that

$$\left| \frac{1}{\sqrt{\Lambda(\hat{F}_n)(1-\Lambda(\hat{F}_n))}} - \frac{1}{\sqrt{\Lambda(F_0)(1-\Lambda(F_0))}} \right| \leq \frac{3}{2} \cdot \frac{C}{(\lambda(1-\lambda))^{3/2}} |\hat{F}_n - F_0|.$$

Since $\int |\hat{F}_n - F_0|^2 dG^0 \leq 4 \int (\sqrt{d\hat{P}_n} - \sqrt{dP_0})^2$,

$$\begin{aligned}\sqrt{n} \int \frac{\hat{F}_n - F_0}{\sqrt{F^0(1-F^0)}} dG^0 &= \sqrt{n} \int \frac{\hat{F}_n(x) - y}{\sqrt{\Lambda(\hat{F}_n(x))(1-\Lambda(\hat{F}_n(x)))}} dP_{F_0,G^0}(x,y) \\ &\quad + o_p(1)\end{aligned}$$

by the law of iterated expectations.

Let $\mathbb{P}_n$ be the empirical measure based on observations $z_1 = (x_1, y_1), \ldots, z_n = (x_n, y_n)$. Since

$$\int \frac{\hat{F}_n(x) - y}{\sqrt{\Lambda(\hat{F}_n(x))(1-\Lambda(\hat{F}_n(x)))}} d\mathbb{P}_n(x,y) \equiv 0$$

(Huang and Wellner (1995)), and since $\|\Lambda(\hat{F}_n) - F^0\|_{G^0} = O_p(n^{-1/3})$, it is concluded that

$$\sqrt{n} \int \frac{\hat{F}_n - F_0}{\sqrt{F^0(1-F^0)}} dG^0 \xrightarrow{d} N(0, \sigma^2)$$

as $n \to \infty$, where

$$\begin{aligned}\sigma^2 &= \int \left(\frac{y - F_0(x)}{\sqrt{F^0(x)(1-F^0(x))}} \right)^2 dP_{F_0,G^0}(x,y) \\ &= \int \frac{F_0(1-F_0)}{F^0(1-F^0)} dG^0.\end{aligned}$$

■

As shown in Theorem 4.1, the efficiency of the estimation depends on the choice of the first-step estimate, F^0. The following corollary shows that the asymptotic variance of $\hat{\theta}_n$ can reach its lower bound only in the limit of $F^0 \to F_0$.

Corollary 4.2 Assume that $\tilde{\varphi}_0'$ is bounded on $[0,1]$. Then, $v^0 \to v_0$ as $\int |F^0 - F_0| \to 0$.

(Proof) The L_1-convergence $\int |F^0 - F_0| \to 0$ of the distribution function implies pointwise convergence $F^0(x) \to F_0(x)$ at almost every $x \in [0,1]$. Since $\frac{F_0(1-F_0)}{F^0(1-F^0)} \leq \frac{1}{4\lambda(1-\lambda)} < \infty$, the dominated convergence theorem shows that $\sigma^2 \to 1$ as $F^0(x) \to F_0(x)$ almost everywhere. Moreover,

$$\begin{aligned}&\left| \int |\tilde{\varphi}_0'| \sqrt{F^0(1-F^0)} - \int |\tilde{\varphi}_0'| \sqrt{F_0(1-F_0)} \right| \\ &\leq \frac{3 \sup_{0 \leq x \leq 1} |\tilde{\varphi}_0'(x)|}{\sqrt{\lambda(1-\lambda)}} \int |F^0 - F_0|,\end{aligned}$$

and thus, $v^0 \to v_0$ as $\int |F^0 - F_0| \to 0$. ■

References

Bickel, P. J., C. A. J. Klaassen, Y. Ritov, and J. A. Wellner (1993): *Efficient and Adaptive Estimation for Semiparametric Models*, Johns Hopkins University Press.

Bishop, R. C. and T. A. Heberlein (1979): "Measuring values of extramarket goods: Are indirect measures biased?" *American journal of agricultural economics*, 61, 926–930.

Carson, R. and M. Hanemann (2006): "Contingent Valuation," in *Handbook of Environmental Economics*, ed. by K. G. Maler and J. R. Vincent, Elsevier, vol. 2, chap. 17, 821–936, 1 ed.

Cooper, J. C. (1993): "Optimal Bid Selection for Dichotomous Choice Contingent Valuation Surveys," *Journal of Environmental Economics and Management*, 24, 25–40.

Cosslett, S. R. (1983): "Distribution-Free Maximum Likelihood Estimator of the Binary Choice Model," *Econometrica*, 51, 765–782.

Groeneboom, P. and G. Jongbloed (2014): *Nonparametric Estimation under Shape Constraints*, Cambridge University Press.

Groeneboom, P. and J. A. Wellner (1992): *Information Bounds and Nonparametric Maximum Likelihood Estimation*, Birkhäuser Verlag.

Huang, J. and J. A. Wellner (1995): "Asymptotic normality of the NPMLE of linear functionals for interval censored data, case 1," *Statistica Neerlandica*, 49, 153–163.

Kristrom, B. (1990): "A Non-Parametric Approach to the Estimation of Welfare Measures in Discrete Response Valuation Studies," *Land Economics*, 66, 135–139.

Patilea, V. (2001): "Convex Models, MLE, and Misspecification," *Annals of Statistics*, 29, 94–123.

Severini, T. A. and G. Tripathi (2001): "A Simplified Approach to Computing Efficiency Bounds in Semiparametric Models," *Journal of Econometrics*, 102, 23–66.

van de Geer, S. (1993): “Hellinger-Consistency of Certain Nonparametric Maximum Likelihood Estimators,” *The Annals of Statistics*, 21, 14–44.

——— (2000): *Empirical Processes in M-Estimation*, Cambridge University Press.

van der Vaart, A. W. (1991): “On Differentiable Functionals,” *Annals of Statistics*, 19, 178–204.

van der Vaart, A. W. and J. A. Wellner (2007): “Empirical Processes Indexed by Estimated Functions,” in *Asymptotics: particles, processes and inverse problems*, Institute of Mathematical Statistics, vol. 55 of *IMS Lecture Notes-Monograph Series*, 234–252.