Waseda Economics
Working Paper Series

Consistency of the Iterative Least Squares Estimator for the Binary Choice Model with Varying Coefficients

TANAKA, Hisatoshi,
(Faculty of Political Science and Economics, Waseda University)
Tadao Hoshino
(Department of Social Engineering, Tokyo Institute of Technology)

Graduate School of Economics
Waseda University

1-6-1, Nishiwaseda, Shinjuku-ku, Tokyo

169-8050 JAPAN

Waseda Economics Working Paper Series

Consistency of the Iterative Least Squares Estimator for the Binary Choice Model with Varying Coefficients

TANAKA, Hisatoshi,

(Faculty of Political Science and Economics, Waseda University)

Tadao Hoshino

(Department of Social Engineering, Tokyo Institute of Technology)

Number 11-3

February 2012

Graduate School of Economics

Waseda University

1-6-1, Nishiwaseda, Shinjuku-ku, Tokyo

169-8050 JAPAN

Consistency of the Iterative Least Squares Estimator for the Binary Choice Model with Varying Coefficients

Hisatoshi Tanaka*

Tadao Hoshino†

February 14, 2012

Abstract

Wang and Zhou (1995) has proposed the iterative least squares estimator of the semiparametric binary choice model. Their method shows outstanding computational advantages over the other semiparametric binary choice estimators, however, it has been only applied to the constant-coefficient model, and its large sample properties have not been rigorously proven yet. This paper considers a semiparametric estimator for the varying-coefficient binary choice model in which the regression coefficients are allowed to vary with observable characteristics. Our estimator is based on Wand and Zhou ' s algorithm. A set of sufficient conditions for consistency of the estimator is also given.

JEL classification: C14

KEYWORDS: EM algorithm; Isotonic Regression; Nonparametric Maximum Likelihood; Binary Choice Model; Varying Coefficient;

*Waseda University, 1-6-1 Nishi-Waseda, Shinjuku-Ku, Tokyo 169-8050, Japan. E-mail: t.hisatoshi@gmail.com.

The author would like to deeply appreciate the financial support by the Seimeikai Foundation at Bank of Tokyo-Mitsubishi UFJ, 2-7-1 Marunouchi, Chiyoda-ku, Tokyo 100-8388, Japan.

†W9-89, Department of Social Engineering, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo, 152-8550, Japan. JSPS Research Fellow (DC2). E-mail: hoshino.t.ai@m.titech.ac.jp.

Hoshino's work is supported by a JSPS grant-in-aid for scientific research (2200-2513).

1 Introduction

In this paper, we study a varying-coefficient binary choice model

$$\begin{cases} D = \{x_1\gamma_1(z) + \cdots + x_k\gamma_k(z) \ge e\} \\ e \perp\!\!\!\perp (x, z),\ Ee = 0 \end{cases} \tag{1.1}$$

where $D\,(= 0 \text{ or } 1)$ is a binary response, $(x_1, \cdots, x_k, z)$ is a vector of covariates and e is an unobserved random term. For every event A, $\{A\} = 1$ if A is true; $\{A\} = 0$ otherwise. We are interested in estimating a value of the coefficient function $\gamma_0 = (\gamma_1(z_0), \cdots, \gamma_k(z_0))'$ at a generic point z_0.

The model has many potential applications in labor economics and other fields of economics and social sciences (Maddala 1983). It is very often in such empirical studies to place parametric specification assumptions on the model structure. If the error distribution function $F(\cdot) = \mathbf{P}\{e \le \cdot\}$ is parameterized by a finite dimensional vector $\theta \in \mathbb{R}^d$, the log likelihood of independent observations $(D_1, x_1, z_1), \cdots, (D_n, x_n, z_n)$ is equal to

$$\ell_n(\gamma, \theta) = \sum_{i=1}^{n} [D_i \log F_\theta(x_i'\gamma) + (1 - D_i)\log(1 - F_\theta(x_i'\gamma))] \times W_h(z_i - z_0),$$

where $W_h(z_i - z_0)$ is a kernel-based weighting function defined by

$$W_h(z_i - z_0) = \frac{K\left(\frac{z_i - z_0}{h}\right)/h}{\sum_{j=1}^{n} K\left(\frac{z_j - z_0}{h}\right)/h}$$

and $K : \mathbb{R} \mapsto \mathbb{R}$ is a standard symmetric kernel function with bandwidth $h(> 0)$ such that

$$K \ge 0,\ \int K(u)\,du = 1,\ \int uK(u)\,du = 0,\ \int u^2 K(u)\,du = \nu < \infty. \tag{1.2}$$

The model is consistently estimated by maximizing ℓ_n with respect to (γ, θ). Recent studies including Horowitz (1992), however, have shown that the popular parametric assumptions on F such as logit or probit may be misleading if the error distribution is misspecified. To avoid such restrictive distributional assumptions, extensive studies have been done: For example, Cosslett (1983), Manski (1985), Han (1987), Horowitz (1992), Ichimura (1993), Ichimura and Thompson (1998), Sherman (1993), Klein and Spady (1993), Wnag and Zhou (1995) and Gautier and Kitamura (2009) among many others.

The Wang-Zhou (hereinafter, WZ) estimator is one of the semiparametric methods to estimate binary choice models without placing finite dimensional parametric assumptions on F. A major advantage of the WZ estimator among other semiparametric methods is its fast and easy-to-implement algorithm. The method is similar to the Expectation-Maximization (EM) algorithm, but does not involve numerical optimization, hence it is computationally cheap and fast. In particular, the computing time is not very affected by the number of explanatory variables, while methods based on numerical optimizations often become computationally intractable as the number of explanatory variables increases.

In spite of its remarkable advantages, the WZ estimator has not been applied to empirical studies mainly because of two reasons: unknown asymptotics and restrictive assumption of homogeneity. Based on a heuristic proof and thorough Monte Carlo simulations, Wang and Zhou has presented strong evidences of consistency and asymptotic normality of their estimator, however, Pagan and Ullah (1999) has pointed out technical gaps in their original proof. In this paper we complete the proof of consistency by means of powerful techniques developed by *empirical process theory*, which is a unified method to study asymptotic properties of statistics defined on a very general (often infinite dimensional) normed space. Basic idea of our proof is similar to that of Dominitz and Sherman (2005), who also study asymptotic properties of a class of iterative semiparametric estimations.

In this paper, another disadvantage of the WZ estimator is also overcome. That is, the original WZ estimation is applied only to the case of fixed coefficients, where γ is known to be constant and independent of z. Our method is now applied to a varying-coefficient model. In recent years, varying-coefficient models, where the regression coefficients are nonparametric continuous functions of observable covariates, z in this paper, have been studied widely (e.g., Hastie and Tibshirani 1993, Fan and Zhang 1999, Cai et al. 2000). These studies have highlighted the usefulness of varying-coefficient models for analyzing time series, spatial, longitudinal, and functional data. Therefore, binary choice models with varying coefficients are useful in applied research to account for heterogeneity in the choice behavior of individuals. In order to capture such heterogeneity, random parameter logit/probit models, also called mixed logit/probit models, which are discussed in details by Train (2009), have often been used in many applied studies. Conventional random parameter models have a serious problem: namely, that researchers usually parametrically assume ad hoc that the distributions of random parameters are normal, log-normal, triangular, or uniform (Revelt and Train 1999, Hensher and Green 2003).[1] Another disadvantage is that even though we can check for the presence of heterogeneity, mixed logit/probit models are not suitable for obtaining a direct structural explanation of heterogeneity.[2] In our model, we do not impose any parametric assumptions on the form of coefficient functions, and estimate the functions as the structural parameters of interest.

The rest of the paper is organized as follows. Section 2 defines the model and states details of our local WZ algorithm. Section 3 gives a list of main assumptions and states the consistency result of the estimator. An idea of the proof is sketched. In Section 4, the properties of the local WZ estimator are examined using a series of Monte Carlo experiments. In addition, as a real data example, we examine the application of the local WZ estimator to a dichotomous contingent valuation analysis of the mountain landscape in Hyogo prefecture, Japan. Section 5 gives the proof in details.

---

[1] As demonstrated by McFadden and Train (2000), when the error term has an i.i.d. extreme value/standard normal distribution, any random utility model can be approximated to any degree of accuracy by a mixed logit/probit with a mixed distribution. However, in practice, there is generally no guidance for finding a mixing distribution that attains a good approximation. If one wants to directly estimate the random parameters without relying on any parametric assumptions, an alternative is to use Ichimura and Thompson (1998)'s or Gautier and Kitamura (2009)'s nonparametric random-coefficient method. However, their methods are based on fairly stringent conditions such as independence between covariates and random coefficients, and the convergence rate of their methods is slow instead of allowing a fully nonparametric specification for the joint distribution of random coefficients.

[2] We can allow the mixed logit/probit random terms to be correlated with z. However, such a methodology does not provide us with a clear structural relationship between z and preference heterogeneity.

2 The Method

For any positive number c,

$$\{x'\gamma(z) \geq e\} = \{x'(c \cdot \gamma(z)) \geq c \cdot e\},$$

hence the model is not identified without normalization. Here, we assume that at least one of the regression coefficients, say $\gamma_1(z)$, is constant over the support of z (namely, $\gamma_1(z) = \gamma_1$). Then, we estimate a vector of varying coefficients at z_0

$$\beta(z_0) = \left(\frac{\gamma_2(z_0)}{\gamma_1}, \frac{\gamma_3(z_0)}{\gamma_1}, \cdots, \frac{\gamma_k(z_0)}{\gamma_1}\right) \in B(\subset \mathbb{R}^{k-1}) \tag{2.1}$$

instead of original $\gamma(z_0)$. Let us replace e with $\varepsilon = e/\gamma_1$, so that

$$\begin{cases} D = \{y \geq 0\} \\ y = x_1 + x_{-1}'\beta(z) - \varepsilon \\ \varepsilon \perp\!\!\!\perp (x, z), \ E\varepsilon = 0, \end{cases} \tag{2.2}$$

where $x_{-1} = (x_2, \cdots, x_k)$. In the following, let us write

$$x * b = x_1 + x_2 b_1 + \cdots x_k b_{k-1} \tag{2.3}$$

for every $x \in \mathbb{R}^k$ and $b \in \mathbb{R}^{k-1}$ to save space, e.g. $D = \{x * \beta(z) \geq \varepsilon\}$.

Our idea of estimation is based on a trivial equation

$$\begin{aligned} \beta(z_0) &= E[x_{-1}x_{-1}'|z = z_0]^{-1}E[x_{-1}(y - x_1)|z = z_0] \\ &= \beta(z_0) - E[x_{-1}x_{-1}'|z = z_0]^{-1}E[x_{-1}\mathcal{E}(D, x, z)|z = z_0], \end{aligned} \tag{2.4}$$

where $\mathcal{E}(D, x, z) := E(\varepsilon|D, x, z)$. By the definition, $\mathcal{E}(D, x, z)$ is equal to

$$\frac{D}{F(x * \beta(z))} \int_{-\infty}^{x*\beta(z)} u \, dF(u) + \frac{1 - D}{1 - F(x * \beta(z))} \int_{x*\beta(z)}^{\infty} u \, dF(u), \tag{2.5}$$

where $F(u) = \mathbf{P}\{D = 1 \,|\, x * \beta(z) = u\}$. By integrating-by-part,

$$\mathcal{E}(D, x, z) = \mu - \frac{[D - F(x * \beta(z))] \int_{-\infty}^{x*\beta(z)} (\mu - s) \, dF(s)}{F(x * \beta(z)) \cdot [1 - F(x * \beta(z))]}, \tag{2.6}$$

where $\mu = \int s \, dF(s) = 0$ by the assumption. The distribution F is also identified by

$$F(\,\cdot\,) = \underset{H \in \mathcal{F}}{\operatorname{argmin}} \ E\left\{[D - H(x * \beta(z))]^2\right\},$$

where $\mathcal{F}$ is a set of distribution functions, *i.e.* nondecreasing, right-continuous functions taking values between 0 and 1. These expressions suggest us to use an iteration $b^{t+1}(z_0) = \mathcal{R}(b^t(z_0))$ to find $\beta(z_0)$, where

$$\begin{cases} \mathcal{R}(b(z_0)) = b(z_0) - E[x_{-1}x_{-1}'|z = z_0]^{-1}E[x_{-1}\mathcal{E}_b(D, x, z)|z = z_0] \\ F_b^*(\,\cdot\,) = \operatorname{argmin}_{H \in \mathcal{F}} E\left\{[D - H(x * b(z))]^2\right\} \\ \mathcal{E}_b(D, x, z) = -\frac{[D - F_b^*(x*b(z))] \int_{-\infty}^{x*b(z)} (\mu_b^* - s) \, dF_b^*(s)}{F_b^*(x*b(z)) \cdot [1 - F_b^*(x*b(z))]} \end{cases} \tag{2.7}$$

with $\mu_b^* = \int s\, dF_b^*(s)$. By the definition, in particular, $\mathcal{R}(\beta(z_0)) = \beta(z_0)$. Moreover, it will be proven that there is an open set B_0 such that

$$\lim_{t\to\infty} \mathcal{R}^t(b(z_0)) = \beta(z_0) \tag{2.8}$$

for arbitrary $b(z_0) \in B_0$, where $\mathcal{R}^t$ means that $\mathcal{R}$ is operated t times. For example, $\mathcal{R}^2(b(z)) = \mathcal{R}(\mathcal{R}(b(z)))$ and so on.

Convergence (2.8) in population implies that, if iid n observations (D_1, x_1, z_1), $\cdots$, (D_n, x_n, z_n) from the model are obtained, and if any initial estimator $\hat{b}^0(z_0)$ is available, $\beta(z_0)$ might be consistently estimated by $\hat{\beta}(z_0) = \lim_{t\to\infty} \hat{\mathcal{R}}^t\left(\hat{b}^0(z_0)\right)$, where

$$\begin{cases} \hat{\mathcal{R}}(b(z_0)) = b(z_0) - \left(\sum_{i=1}^n W_h(z_i - z_0)x_{i,-1}x_{i,-1}'\right)^{-1} \sum_{i=1}^n W_h(z_i - z_0)x_{i,-1}\hat{\mathcal{E}}_b(D_i, x_i, z_i) \\ \hat{F}_b^*(\,\cdot\,) \in \text{argmin}_{H\in\mathcal{F}}\ n^{-1}\sum_{i=1}^n \Big[D_i - H(x_i * b(z_i))\Big]^2 \\ \hat{\mathcal{E}}_b(D_i, x_i, z_i) = -\frac{[D_i - \hat{F}_b^*(x_i * b(z_i))]\int_{-\infty}^{x_i * b(z_i)}(\hat{\mu}_b^* - s)\, d\hat{F}_b^*(s)}{\kappa + \hat{F}_b^*(x_i * b(z_i))\cdot[1-\hat{F}_b^*(x_i * b(z_i))]} \end{cases} \tag{2.9}$$

with $\hat{\mu}_b^* = \int s\, d\hat{F}_b^*(s)$. The parameter κ is introduced to keep the denominator away from zero. It will be proven that κ does not cause any asymptotic bias.

The functional optimization problem

$$\min \quad n^{-1}\sum_{i=1}^n [D_i - H(x_i * b(z_i))]^2 \quad s.t. \quad H \in \mathcal{F} \tag{2.10}$$

is known as *isotonic regression problem* and quickly solved by the Pool-And-Violation (PAV) algorithm developed by Ayer et al. (1955).

As stated above, the WZ estimator is based on the EM-like algorithm, and so is our estimator. Let us point out that, in a special case where the error distribution is known to be normal, such iterative algorithms are equivalent to the EM algorithm. In other words, the suggested iterative method attains the maximum likelihood estimates at its fixed point. To see this, suppose that $e \sim N(0, \sigma^2)$. Setting $\sigma^2 = 1$ for normalization as convention, we have

$$\mathbf{P}\{D = 1|x, z\} = \Phi(x'\gamma(z)),$$

where $\Phi(\cdot)$ is the standard normal distribution function. Then, we have

$$\begin{aligned} \mathcal{E}(D, x, z) := E[\varepsilon|D, x, z] &= \frac{D - \Phi(x'\gamma(z))}{\Phi(x'\gamma(z))\cdot[1 - \Phi(x'\gamma(z))]}\int_{-\infty}^{x'\gamma(z)} u d\Phi(u) \\ &= -\frac{D - \Phi(x'\gamma(z))}{\Phi(x'\gamma(z))\cdot[1 - \Phi(x'\gamma(z))]}\cdot\phi(x'\gamma(z)), \end{aligned} \tag{2.11}$$

where $\phi(\cdot)$ is the standard normal density. The iterative algorithm suggests that, for a given point z_0, $\gamma(z_0)$ can be found by a fixed point of the following mapping:

$$\gamma^{t+1}(z_0) = \gamma^t(z_0) - E[xx'|z = z_0]^{-1}E[x\mathcal{E}_{\gamma^t}(D, x, z)|z = z_0], \tag{2.12}$$

where

$$\mathcal{E}_{\gamma^t}(D,x,z) = -\frac{D-\Phi(x'\gamma^t(z))}{\Phi(x'\gamma^t(z))\cdot[1-\Phi(x'\gamma^t(z))]}\cdot\phi(x'\gamma^t(z)).$$

On the other hand, $\gamma(z_0)$ can be characterized as the maximizer of

$$\ell(\gamma|z=z_0) = E[D\log\Phi(x'\gamma) + (1-D)\log(1-\Phi(x'\gamma))|z=z_0], \tag{2.13}$$

with respect to γ. Since the log-likelihood function for probit models is globally concave, the maximizer is exactly represented as the solution to the first-order condition of the log-likelihood function, i.e.,

$$\frac{\partial}{\partial\gamma}\ell(\gamma(z)|z=z_0) = E\left[x\left\{\frac{D\cdot\phi(x'\gamma(z))}{\Phi(x'\gamma(z))} - \frac{(1-D)\cdot\phi(x'\gamma(z))}{1-\Phi(x'\gamma(z))}\right\}\middle| z=z_0\right] = 0. \tag{2.14}$$

(2.14) implies that $E[x\mathcal{E}(D,x,z)|z=z_0] = 0$ at the true parameter values, which is also the convergence condition of the iterative algorithm in (2.12). It is not difficult to show the global convergence of the iteration under normality. Then, it is shown that, when the error distribution is normal, the iterative estimator coincides with the maximum likelihood estimator.

3 Consistency

In the following part of the paper, $\mathcal{B}$ denotes a set of continuous functions $b:\mathcal{Z}\mapsto B\subset\mathbb{R}^{k-1}$ with bounded variations, and $\|\cdot\|_\infty$ is the norm on $\mathcal{B}$ defined by $\|b\|_\infty = \sum_{j=1}^{k-1}\sup_{z\in\mathcal{Z}}|b^j(z)|$ for every $b=(b^1,\cdots,b^{k-1})\in\mathcal{B}$. In order to save spaces, we often write $b_z = b(z)$ so that

$$F\left(x_1 + \sum\nolimits_{j=1}^{k-1}x_{j+1}b^j(z)\right) = F(x * b_z).$$

A set of functions $h:\mathcal{Z}\mapsto\mathbb{R}^{k-1}$ with $\|h\|_\infty<\infty$ is $L_\infty(\mathcal{Z})^{k-1}$. A set of measurable functions $h:\mathcal{X}\times\mathcal{Z}\mapsto\mathbb{R}^d$ such that $E[h(x,z)^2] = \int h(x,z)^2\,d\mathbf{P}(x,z)<\infty$ is $L_2(\mathbf{P})$.

The main result of the paper is given by the next theorem.

Theorem 3.1 *Let $\{t_n\in\mathbb{N}\}$ be an arbitrary sequence such that $t_n\to\infty$ as $n\to\infty$. Define the local WZ estimator $\hat\beta_n(\cdot)$ by*

$$\hat\beta_n(\cdot) = \hat{\mathcal{R}}^{t_n}(b^0(\cdot)), \tag{3.15}$$

where b^0 is an initial function of the iteration. Assume the following **(A.1)-(A.8)**, *then there is a neighborhood $\mathcal{N}_\infty$ of $\beta(\cdot)$ with respect to the norm $\|\cdot\|_\infty$ such that $\hat\beta_n(z_0)\xrightarrow{p}\beta(z_0)$ as $n\to\infty$ for any $z_0\in\mathcal{Z}$ and $b^0\in\mathcal{N}_\infty$.*

(A.1) *There is a compact set $B\subset\mathbb{R}^{k-1}$ such that $\beta(z)\in B^{int}$ for any $z\in\mathcal{Z}$.*

(A.2) *Let $\mathcal{X}\subset\mathbb{R}^k$ be the support of x such that $\mathbf{P}\{x\in\mathcal{X}|z\}=1$ almost surely. (i) The density of x with respect to a reference measure $dx_1\times d\pi(x_{-1})$ conditioned on z exists, so that*

$$\mathbf{P}\{x\in A|z\} = \int_A p(x_1,x_{-1}|z)\,dx_1 d\pi(x_{-1})$$

for every Borel $A \subset \mathbb{R}^k$. *(ii) The density of* x_1 *conditioned on* (x_{-1}, z) *is everywhere positive and bounded on the real line, that is,* $p(x_1|x_{-1}, z) > 0$ *for every* $x \in \mathcal{X}$ *and*

$$\sup_{x \in \mathcal{X}} p(x_1|x_{-1}, z) < \infty. \tag{3.16}$$

(iii) The support of x *conditioned on* z *is not contained in any affine linear subspaces with positive probability. (iv) The density* $p(z)$ *is positive at every* $z \in \mathcal{Z}$.

(A.3) *The support* $\mathcal{X}_{-1}(\subset \mathbb{R}^{k-1})$ *of* x_{-1} *is bounded, that is, there exists* $M < \infty$ *such that* $\mathbf{P}\{\, |x_{-1}| \leq M \,|z\} = 1$.

(A.4) *For every distribution function* $H \in \mathcal{F}$, *let* $suppH$ *be the support of* H *defined by*

$$suppH := H^{-1}(0, 1) = \{u \in \mathbb{R} \,:\, 0 < H(u) < 1\}. \tag{3.17}$$

Then $suppF$ *is a bounded interval.*

(A.5) F *is twice continuously differentiable on* $suppF$ *with derivatives* $f = F'$ *and* $f' = F''$. *The derivatives are bounded on the support, too, and*

$$\inf_{u \in suppF} f(u) \geq c_f > 0$$

with a positive constant c_f.

(A.6) *The conditional density* $p(x|z)$ *is bounded and twice differentiable with respect to* x_1. *The derivative* $\partial_1 p(x_1, \cdots, x_k|z)$ *is continuous and bounded. The assumption implies that there is a positive constant* C_p *such that*

$$\sup_{x_{-1} \in \mathcal{X}_{-1}, z \in \mathcal{Z}} \big|p(x_1, x_{-1}|z) - p(\tilde{x}_1, x_{-1}|z)\big| \leq C_p|x_1 - \tilde{x}_1| \tag{3.18}$$

for any x_1 *and* $\tilde{x}_1$ *in* $\mathbb{R}$.

The joint density $p(x, z)$ *is twice continuously differentiable with respect to* z, *and*

$$\sup_{x,z} \left| \frac{\partial^2}{\partial z^2} p(x, z) \right| < \infty. \tag{3.19}$$

(A.7) *For two positive definite matrices* V_1 *and* V_2, *let* $V_1 > V_2$ *mean that* $V_1 - V_2$ *is also a positive definite matrix. Then,*

$$E(x_{-1}x'_{-1}|z) > E\left[\frac{f(x * \beta_z) \cdot f^*(x * \beta_z)}{F(x * \beta_z)(1 - F(x * \beta_z))} Var\,(x_{-1}|x * \beta_z, z) \,\middle|\, z \right],$$

where

$$f^*(u) := \int_{-\infty}^{u} \left(\int_{-\infty}^{\infty} v \, dF(v) - s \right) dF(s)$$

at every $u \in \mathbb{R}$.

(A.8) *As* $n \to \infty$, $h = h_n \downarrow 0$ *as well as* $n \cdot h_n \to \infty$.

(A.1) is placed to make optimizations well defined. **(A.2)** is standard identification assumption. Unbounded support of x_1 identifies F, and linear independence of x helps to identify β.

Lemma 5.17 will prove that the bounded support conditions **(A.3)** and **(A.4)** are sufficient for estimate $\hat{F}_b^*$ to have a bounded support for any $b \in \mathcal{B}$, too. Moreover, we can find a deterministic interval I such that

$$\mathbf{P}\left\{\bigcup_{b\in\mathcal{B}} supp\hat{F}_b \subset I\right\} = 1.$$

(A.5) and **(A.6)** will be used to compute the first derivative $\partial\mathcal{R}(b)$. In particular, **(A.5)** implies that F has the uniform second order expansion in the sense that there is a constant $C_F > 0$ such that

$$\sup_{u\in I} |F(u+\delta) - F(u) - f(u)\cdot\delta| \le C_F\delta^2 \tag{3.20}$$

as $\delta \to 0$ for any closed interval $I \subset suppF$. To see this, let $\theta = \theta(u,\delta)$ be a function such that

$$F(u+\delta) = F(u) + f(u)\cdot\delta + \frac{1}{2}f'(u+\theta(u,\delta)\cdot\delta)\cdot\delta^2$$

for every (u,δ), then

$$\sup_{u\in I} |F(u+\delta) - F(u) - f(u)\cdot\delta| \le \sup_{t\in suppF} |f'(t)|\cdot\delta^2$$

for every small δ. By the assumptions, it is also shown that f as well as F is Lipschitz continuous on $suppF$ because

$$|F(s) - F(t)| = \left|\int_t^s f(u)\,du\right| \le \sup_{u\in suppF} |f(u)\cdot|s-t|$$

and

$$|f(s) - f(t)| = \left|\int_t^s f'(u)\,du\right| \le \sup_{u\in suppF} |f'(u)|\cdot|s-t|$$

for any s and t in $suppF$.

(A.7) is used to evaluate the maximal eigenvalue of $\partial\mathcal{R}(\beta)$. **(A.7)** is satisfied if, for example, F is close to the normal distribution. To see this, suppose that β is constant and independent of z. Notice that $\sigma^2 E(x_{-1}x'_{-1})^{-1}$ with $\sigma^2 = Var(\varepsilon)$ is the efficient asymptotic variance of β-estimation in the case where $y = x*\beta - \epsilon$ were observable. In our case, where y is latent, the efficient asymptotic variance established by Chamberlain (1986) and Cosslett (1987) is

$$E\left[\frac{f(x*\beta)^2}{F(x*\beta)(1-F(x*\beta))}Var(x_{-1}|x*\beta)\right]^{-1}.$$

Therefore, for arbitrary F,

$$E(x_{-1}x'_{-1}|z) \ge E\left[\frac{f(x*\beta_z)^2/\sigma^2}{F(x*\beta_z)(1-F(x*\beta_z))}Var(x_{-1}|x*\beta_z, z)\,\middle|\, z\right]$$

holds in a positive semi-definite sense. If $F = N(\mu, \sigma^2)$, then $f^*(u) = \int^u (\mu - v)\, dF(v)$ satisfies

$$(f^*)'(u) = (\mu - u) f(u) = \sigma^2 \cdot \left(-\frac{u-\mu}{\sigma^2} f(u) \right) = \sigma^2 f'(u), \tag{3.21}$$

which implies $f^* = \sigma^2 f$, hence **(A.7)** is satisfied by $F = N(\mu, \sigma^2)$.

Even in a general case where $F \neq N(\mu, \sigma^2)$, we still might expect **(A.7)** to hold if $f = F'$ is unimodal with a bounded support.

Proposition 3.2 *For arbitrary density function f with mean $\mu = \int u f(v)\, dv$ and variance $\sigma^2 = \int (v - \mu)^2 f(v)\, dv$, let $f^*(u) = \int^u (\mu - v) f(v)\, dv$. Assume that f has a bounded support, then*

(i) *f^*/σ^2 is a density function,*

(ii) *f^*/σ^2 is increasing at $u < \mu$ and decreasing at $u > \mu$, hence f^*/σ^2 is unimodal with mode μ, and*

(iii) *if f is symmetric, so is f^*/σ^2. Furthermore, f^*/σ^2 is skewed to the same direction as f.*

Therefore, as long as f is unimodal, f^*/σ^2 will give a good approximation of f, and $F(u) = \int^u f$ satisfies **(A.7)**.

4 Simulations and An Empirical Example

4.1 Monte Carlo Simulations

We conduct a series of Monte Carlo experiments in order to ascertain the finite sample properties of our local WZ estimator. Three types of models are prepared as follows:

Model 1: $D = \{y > 0\}$, $y = x_1 + x_2 \beta_1(z) - \varepsilon$, where $x_1 \sim N[0, 2]$, $x_2 \sim N[0, 2]$, $\varepsilon \sim N[0, 1]$, and $\beta_1(z) = 1 + \cos(4z - 0.5)$

Model 2: $D = \{y > 0\}$, $y = x_1 + x_2 \beta_1(z) - \varepsilon$, where $x_1 \sim N[0, 2]$, $x_2 \sim N[0, 2]$, $\varepsilon \sim t[10]$, and $\beta_2(z) = 1 + \cos(4z - 0.5)$

Model 3: $D = \{y > 0\}$, $y = x_1 + x_2 \beta_1(z) - \varepsilon$, where $x_1 \sim N[0, 2]$, $x_2 \sim N[0, 2]$, $\varepsilon \sim N[-3, 1]$ or $N[3, 1]$ with equal probability, and $\beta_1(z) = 1 + \cos(4z - 0.5)$

where $t[10]$ denotes the t-distribution with degree of freedom 10. For these three set-ups, we set $z \in (0, 1]$ with equal interval, i.e., $z_i = i/n$. We generate two sizes of samples for Model 1 and Model 2, namely, $n = 150$ and $n = 500$; and one size of sample for Model 3, $n = 150$. For comparison, we also estimated the models in the case of $n = 150$ using the local iterative least squares estimator for the probit model with varying coefficients. Using the probit model, the first model is the fortunate case in which we correctly parametrically specified the error distribution, and the other two models are the cases with misspecification of error distributions. The estimation procedure for the varying coefficient probit model was explained in the last paragraph of Section 2.

For the kernel function, we used the Gaussian kernel considering the local identifiability of parameters. The bandwidth for each model was set arbitrarily to $h = 0.07$ in the case of $n = 150$, and $h = 0.05$ in the case of $n = 500$. In practice, the bandwidth should be chosen, for example, by the leave-one-out cross-validation method. However, since the cross-validation method is extremely computationally expensive, it is not realistic to conduct the cross-validation for each of Monte Carlo replications. The convergence criterion for the estimator was set as follows:

$$n^{-1} \sum_{i=1}^{n} \left| \hat{\beta}_1^{(t)}(z_i) - \hat{\beta}_1^{(t-1)}(z_i) \right| < 1/2n.$$

The maximum number of iterations was set to 100. We denote $\bar{t}$ as the minimum between 100 and the number of iterations when the criterion above is met.

The estimated varying coefficients are evaluated by the square root of mean squared error (RMSE) statistic:

$$RMSE_r = \left[n^{-1} \sum_{i=1}^{n} \left(\hat{\beta}_1^r(z_i) - \beta_1(z_i) \right)^2 \right]^{1/2},$$

and the squared bias (SB) stastic:

$$SB_i = \left[R^{-1} \sum_{r=1}^{R} \hat{\beta}_1^r(z_i) - \beta_1(z_i) \right]^2,$$

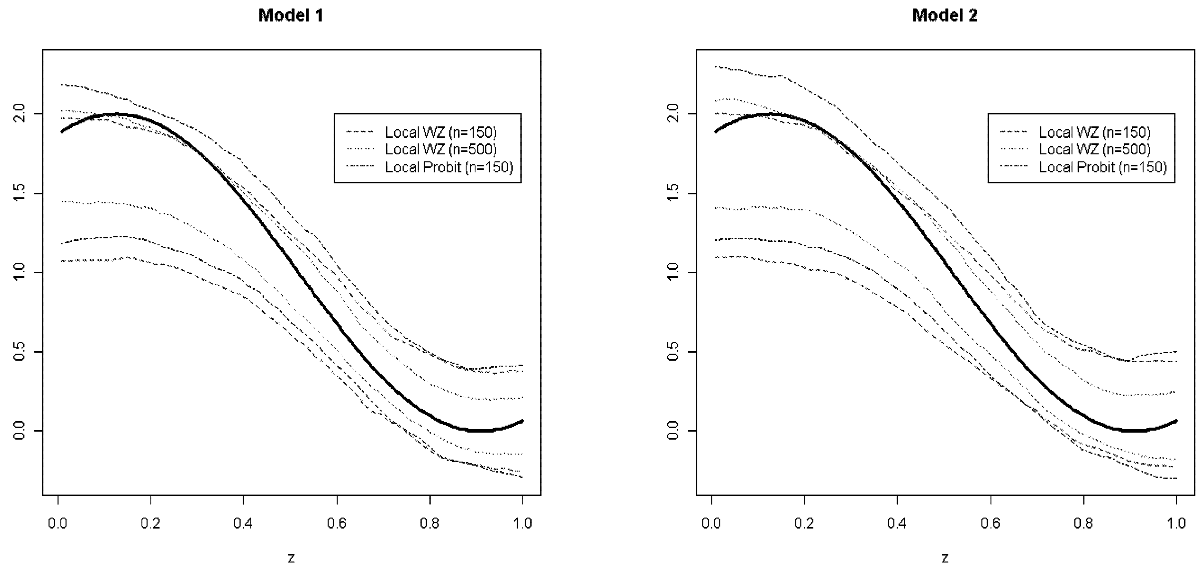
where $\hat{\beta}_1^r(z_i)$ is the estimate of $\beta_1(z_i)$ from the rth Monte Carlo replication, and R is the total number of the replications. The results reported below are based on 500 MC replications ($R = 500$). The simulations were coded in R version 2.8.1 (The R Foundation for Statistical Computing) and performed on a PC with Intel(R) Core(TM)2 Quad 2.66 GHz and 2 GB of RAM. The R codes used for the simulations are available from the authors. For all the simulations, the starting values were determined by the linear probability model with varying coefficients.

Table 1 summarizes the statistical information from the results of the simulations. The number of iterations required for convergence for the local WZ estimator is significantly larger than that for the local probit estimator in each model. As expected, in Model 1, the local WZ estimator was less accurate than the local probit estimator. In Model 2, although the error distribution is not normal, the local probit estimator performed well as compared to the local WZ estimator. This may be due to the shape similarity between the t-distribution and the normal distribution. In fact, when the error distribution was mixed normal, in Model 3, the mean of RMSE statistic for the local WZ estimator was significantly smaller than that for the local probit model. In Model 1 and Model 2, the increase in sample size from 150 to 500 remarkably improved the estimation accuracy.

The simulation results on the varying coefficients are graphically represented in Fig.4.1. In these figures, the true coefficient functions are drawn as a bold solid line. The dotted lines indicate the Monte Carlo 95 percent interval for each estimator. The first two of the figures clearly depict the importance of increasing sample size to improve the accuracy of the local WZ estimator. In the third figure, we can see that the local probit estimates are seriously biased upward due to the misspecification inerror distribution.

Table 1: Results of the Simulations

Model 1						
Estimator	Local WZ ($n = 150$)		Local Probit ($n = 150$)		Local WZ ($n = 500$)	
Mean of $\bar{t}$	36.580		20.670		58.340	
	RMSE	SB	RMSE	SB	RMSE	SB
Mean	0.3272	0.0857	0.2679	0.0441	0.1880	0.0276
Std. dev.	0.1100	0.0978	0.0942	0.0507	0.0664	0.0323
Min	0.0631	0.0000	0.0514	0.0000	0.0363	0.0000
Max	0.6306	0.2598	0.5265	0.1424	0.3994	0.0895
Model 2						
Estimator	Local WZ ($n = 150$)		Local Probit ($n = 150$)		Local WZ ($n = 500$)	
Mean of $\bar{t}$	33.244		20.606		56.266	
	RMSE	SB	RMSE	SB	RMSE	SB
Mean	0.3344	0.0898	0.2639	0.0356	0.1901	0.0261
Std. dev.	0.1135	0.0999	0.0987	0.0402	0.0714	0.0308
Min	0.0632	0.0000	0.0572	0.0000	0.0554	0.0000
Max	0.7112	0.2657	0.6182	0.1149	0.4434	0.0856
Model 3						
Estimator	Local WZ ($n = 150$)		Local Probit ($n = 150$)			
Mean of $\bar{t}$	66.822		13.022			
	RMSE	SB	RMSE	SB		
Mean	0.4319	0.0502	0.6033	0.0808		
Std. dev.	0.1524	0.0506	0.4685	0.1005		
Min	0.1196	0.0000	0.0974	0.0000		
Max	1.1720	0.1325	4.0266	0.3879		



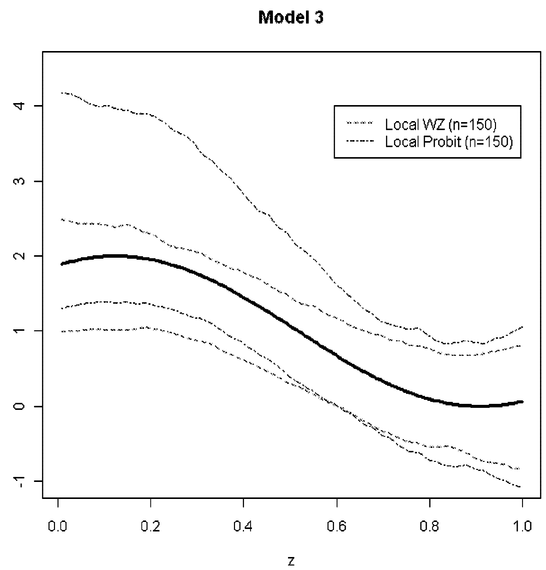


Figure 1: MC 95% Intervals for the Varying Coefficients

4.2 Dichotomous choice contingent valuation analysis

As a real data example, we conduct a dichotomous choice contingent valuation analysis using a varying-coefficient binary choice model. The contingent valuation method is one of the most popular and classic environmental evaluation techniques. This method of measuring the economic value of a good uses a questionnaire to directly elicit residents' willingness to pay for the good. The environmental good evaluated in this analysis is the landscape of the Rokko mountain in Hyogo prefecture, Japan. Since the rivers flowing through the Rokko mountain are torrential, and since the mountain primarily consists of weathered granite, there is a high risk of mudslides and debris flows due to heavy rainfall in the region. Therefore, a sand dam has been constructed around the mountain. The Japanese Ministry of Land, Infrastructure, and Transport is conducting a landscaping project that involves the renovation of the sand dam. The ministry performed a series of surveys to ascertain the extent to which the residents in the area were willing to pay for the renovation plan.[3] Respondents were randomly chosen from three cities located near the Rokko mountain: Kobe, Ashiya, and Nishinomiya. The questionnaire was mailed to the respondents in November 2008. Out of 1570 questionnaires sent out, the number of questionnaires collected was 706 (45 percent of the total number). One of the surveys was a double-bounded dichotomous choice contingent valuation survey, where respondents were first asked to state whether they were willing to pay a certain initial bid amount for a hypothetical renovation plan, and then asked a follow-up question with a higher (lower) bid if the response to the first bid was yes (no). In this analysis, we only use the initial responses as a single-bounded dichotomous choice model. The initial bid was randomly selected from among the following three values: JPY 250, JPY 1000, and JPY 4000.

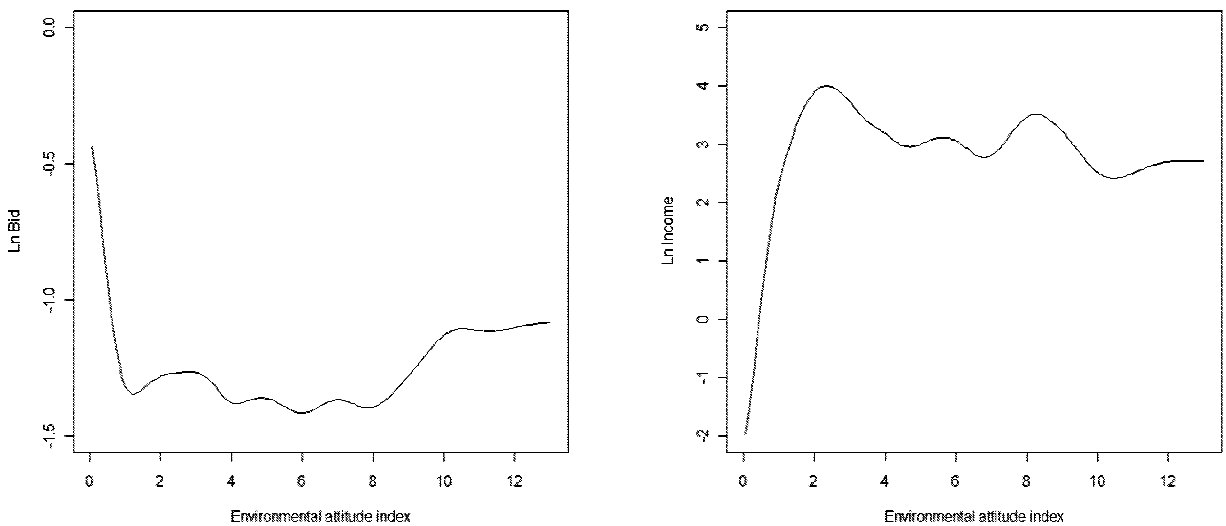
In this analysis, we consider the following model:

$$\begin{cases} D = \{y \geq 0\} \\ y = \text{Age} + \ln \text{Bid} \cdot \beta_{bid}(z) + \ln \text{Inc} \cdot \beta_{inc}(z) - \varepsilon \end{cases}$$

where y is the yes/no responses to the first bid amount, Age is the age of the respondent divided by 10, lnBid is the natural log of the bid amount, lnInc is the natural log of household annual income (in million Yen), and z is the one-dimensional index representing individual's environmental attitude. The coefficient to the age variable is normalized to one for identification. The environmental attitude index is created from a set of rating scale questions relating to various environmental problems. In total, the respondents were required to answer six importance rating questions: conservation of forests, river water quality improvement, preservation of the ecosystem, conservation of natural parks, promotion of eco-tourism, and prevention of climate change. The scale ranged from 1 (important) to 5 (unimportant), and 6 (undecided). We allocate 5 points if the answer is 1, 3 if it is 2, 1 if it is 3 or 6, 1/3 if it is 4, and 1/5 if it is 5. Then, U_i is calculated as the Euclidean norm of the points, i.e., the higher the total point score, the higher the level of pro-environmental attitude. In the data from the 706 respondents, we excluded data with missing values, and obtained a final sample size of 498 for analysis.

Fig.2 shows the the estimation results of $\beta_{bid}(\cdot)$, and $\beta_{inc}(\cdot)$. For the estimated coefficient function for the bid variable, $\hat{\beta}_{bid}(\cdot)$, we can see the tendency of people with stronger pro-environmental attitude and people with least pro-environmental attitude reacting less sensitively to the bid amount relative to moderately pro-environmental people. The estimated coefficient function for the income variable shows that the landscape of Rokko mountain is normal goods for most of the people. In contrast, for the least pro-environmental residents, the income effect can be negative. However, note that the accuracy of the estimates of functional coefficients in the neighborhood of boundary on the environmental attitude index should be questionable due to the property of localized estimator, and also note that these estimation results are crucially dependent on the specification of the environmental attitude index.

[3]The actual data collection was conducted by the Kinki division of the ministry called the Rokko Sabo Office (http://www.rokko.kkr.mlit.go.jp/). We are grateful to the Rokko Sabo Office and Professor Koichi Kuriyama of Kyoto University for their kind cooperation in allowing us to access this dataset.



Left: estimated coefficient function of ln Bid; Right: estimated coefficient function of ln Inc. (sample size: $n = 498$).

Figure 2: Estimated Coefficient Functions

5 Proof of Consistency

The proof consists of three steps. The first is to compute a linear approximation $\mathcal{R}(b(z_0)) \simeq \mathcal{R}(\beta(z_0)) + \partial\mathcal{R}(\beta(z_0))(b(z_0) - \beta(z_0))$. It is shown that $\partial\mathcal{R}(\beta(z_0))$ is a positive definite symmetric matrix and that $O < \partial\mathcal{R}(\beta(z_0)) < I_{k-1}$ in the sense of positive definite, hence $\mathcal{R}$ is locally contract in a neighborhood of β. This means that a population sequence $\{b^t(z_0)\}$ given by $b^t(z_0) = \mathcal{R}(b^{t-1}(z_0))$ converges to $\beta(z_0)$ as $t \to \infty$. The second step is to prove uniform consistency $\sup_b |\hat{\mathcal{R}}(b(z_0)) - \mathcal{R}(b(z_0))| \xrightarrow{p} 0$ as $n \to \infty$, which implies that $\sup_t \|\hat{b}^t(z_0) - b^t(z_0)\|_\infty \xrightarrow{p} 0$ as $n \to \infty$. In the third step, $\hat{\beta}(z_0) \xrightarrow{p} \beta(z_0)$ at every $z_0 \in \mathcal{Z}$ is obtained as a consequence of the first and second step.

5.1 Differentiability of $\mathcal{R}$

Lemma 5.1 *For every $b \in \mathcal{B}$, $F_b^* = \operatorname{argmin}_{H \in \mathcal{F}} E[D - H(x * b_z)]^2$ uniquely exists.*

(Proof) At every $b \in \mathcal{B}$ and $H \in \mathcal{F}$, let $H \circ b$ denote a mapping $(x, z) \mapsto H(x * b_z)$. Then, at every fixed $b \in \mathcal{B}$, $\{H \circ b : H \in \mathcal{F}\}$ is a convex, closed subset of $L_2(\mathbf{P})$. The Hilbert projection theorem proves the lemma.

□

Corollary 5.2 *At every b, F_b^* satisfies*

$$E\big[(D - F_b^*(x * b_z))(H(x * b_z) - F_b^*(x * b_z))\big] \leq 0 \tag{5.1}$$

for any $H \in \mathcal{F}$.

(Proof) Since $(1-t)\cdot F_b^*(x*b_z)+t\cdot H(x*b_z)\in\mathcal{F}\circ b$ for any $t\in[0,1)$ and $H\in\mathcal{F}$, a function

$$\rho(t)=E\big[D-(1-t)\cdot F_b^*(x*b_z)-t\cdot H(x*b_z)\big]^2$$

is minimized at $t=0$. Therefore, $\rho'(0)\geq 0$ or

$$E\big[\big(D-(1-t)\cdot F_b^*(x*b_z)-t\cdot H(x*b_z)\big)\big(F_b^*(x*b_z)-H(x*b_z)\big)\big]_{t=0}\geq 0.$$

□

Lemma 5.3 *$supp(F_b^*)$ is bounded for every $b\in\mathcal{B}$, and*

$$diam\left(supp(F_b^*)\backslash supp(F)\right)\to 0 \tag{5.2}$$

as $\|b-\beta\|_\infty\to 0$.

(Proof) Let M be a constant large enough to satisfy $diam(\mathcal{X}_{-1})\leq M$, $diam(B)\leq M$, and $supp(F)\subset(-M,M)$. Let $(\mathcal{X}\times\mathcal{Z})^\circ$ be a subset of $\mathcal{X}\times\mathcal{Z}$ such that $(x,z)\in(\mathcal{X}\times\mathcal{Z})^\circ$ implies that $\mathbf{P}\{D=1|x,z\}>0$ and $\mathbf{P}\{D=0|x,z\}>0$, or equivalently

$$(\mathcal{X}\times\mathcal{Z})^\circ=\{(x,z)\in\mathcal{X}\times\mathcal{Z}\,:\,0<F(x*\beta_z)<1\}.$$

By the assumptions,

$$(\mathcal{X}\times\mathcal{Z})^\circ\subset\{(x,z)\in\mathcal{X}\times\mathcal{Z}\,:\,|x*\beta_z|\leq M\}\subset\{(x,z)\in\mathcal{X}\times\mathcal{Z}\,:\,|x_1|\leq M(1+M)\}. \tag{5.3}$$

Hence closed hull $\overline{(\mathcal{X}\times\mathcal{Z})^\circ}$ is compact, and functionals

$$b\in\mathcal{B}\ \mapsto\ \begin{cases} u^*(b)=\max\left\{x*b_z\ :\ (x,z)\in\overline{(\mathcal{X}\times\mathcal{Z})^\circ}\right\} \\ \ell^*(b)=\min\left\{x*b_z\ :\ (x,z)\in\overline{(\mathcal{X}\times\mathcal{Z})^\circ}\right\}\end{cases}\quad(\in\mathbb{R}) \tag{5.4}$$

are well defined. Note that $supp(F_b^*)=(\ell^*(b),u^*(b))$ and, in particular $supp(F)=(\ell^*(\beta),u^*(\beta))$.

By (5.3), $u^*(b)\leq M(1+M)+diam(\mathcal{X}_{-1})\cdot diam(B)\leq M(1+2M)$ and $\ell^*(b)\geq -M(1+2M)$ for any $b\in\mathcal{B}$. Hence we have shown that $\bigcup_{b\in\mathcal{B}} supp(F_b)$ is bounded.

By the definition (5.4) of the functionals, there are $(\hat{x},\hat{z}),(\tilde{x},\tilde{z})\in\overline{(\mathcal{X}\times\mathcal{Z})^\circ}$ such that $u^*(b)=\hat{x}*b_{\hat{z}}$ and $u^*(\beta)=\tilde{x}*\beta_{\tilde{z}}$, hence

$$u^*(b)=\hat{x}*\beta_{\hat{z}}+\hat{x}_{-1}\cdot(b_{\hat{z}}-\beta_{\hat{z}})\leq u^*(\beta)+M\|b-\beta\|_\infty$$

and

$$u^*(\beta)=\tilde{x}*b_{\tilde{z}}+\tilde{x}_{-1}\cdot(\beta_{\tilde{z}}-b_{\tilde{z}})\leq u^*(b)+M\|b-\beta\|_\infty.$$

These imply that $|u^*(b) - u^*(\beta)| \leq M\|b - \beta\|_\infty$. In the same way, $|\ell^*(b) - \ell^*(\beta)| \leq M\|b - \beta\|_\infty$ is shown. Therefore,

$$diam\ (supp(F_b^*)\backslash supp(F)) \leq |u^*(b) - u^*(\beta)| + |\ell^*(b) - \ell^*(\beta)| \to 0$$

as $\|b - \beta\|_\infty \to 0$. □

Lemma 5.4 *Let $\mathcal{D}_b(u) = E(D\,|\,x * b_z = u)$, so that $\mathcal{D}_\beta(u) = F(u)$. Then there are intervals $I_b^1, I_b^2, \cdots$ indexed with $b \in \mathcal{B}$ satisfying the following conditions (See Fig.3):*

(i) *$I_b^j \subset supp(F_b^*)$ and $I_b^j \cap I_b^\ell = \emptyset$ for $j \neq \ell$,*

(ii) *$F_b^*(u) = E(\mathcal{D}_b(x * b_z)\,|\,x * b_z \in I_b^j)$ if $u \in I_b^j$ for some j, and*

(iii) *$\mathcal{D}_b(u) = F_b^*(u)$ if $u \notin I_b^j$ for any j.*

For later reference, let $I_b^0 = (\bigcup_{j\geq 1} I_b^j)^c$.

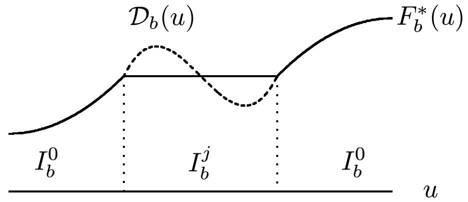


Figure 3: Projection F_b^* with Monotonicity Restriction and Conditional Expectation $\mathcal{D}_b$

(Proof) If $\mathcal{D}_b$ is nowhere decreasing on $supp(F_b^*)$, then $\mathcal{D}_b$ solves $\min_{H\in\mathcal{F}}\ E[D - H(x * b_z)]^2$ and is equal to F_b^*. If $\mathcal{D}_b$ is nowhere increasing, on the other extreme, the best approximation of $\mathcal{D}_b$ within $\mathcal{F}$ is constant $E[\mathcal{D}_b(x * b_z)]$. Combination of these insights proves the lemma. □

Lemma 5.5 *At every fixed $b \in \mathcal{B}$, let g_b be the density of random variable $x * b_z$. Then, there exists a positive constant C, which is independent of the choice of b, such that*

$$\sup_{u\in\mathbb{R}} |g_b(u) - g_\beta(u)| \leq C\|b - \beta\|_\infty \tag{5.5}$$

for any $b \in \mathcal{B}$.

(Proof) By the definition, $g_b(u) = \int p(u - x_{-1}'b_z, x_{-1}, z)d\pi(x_{-1}, z)$. The assumptions of the paper imply that there exists a constant C such that

$$|g_b(u) - g_\beta(u)| \leq \int \left|p(u - x_{-1}'b_z, x_{-1}, z) - p(u - x_{-1}'\beta_z, x_{-1}, z)\right| d\pi(x_{-1}, z) \leq C\|b - \beta\|_\infty.$$

□

Lemma 5.6 *Define function* $\zeta_b : \mathbb{R} \mapsto \mathbb{R}^{k-1}$ *indexed with* $b \in \mathcal{B}$ *by* $\zeta_b(u) = E(x_{-1}|x * b_z = u)$. *For any bounded interval* I, *there are positive constants* $C_\zeta(I)$ *and* $\delta(I)$ *such that* $\sup_{u\in I} |\zeta_b(u) - \zeta_\beta(u)| \leq C_\zeta(I)\|b-\beta\|_\infty$ *if* $\|b-\beta\|_\infty < \delta(I)$.

(Proof) For any bounded I, let $c(I) := \inf_{u\in I} g_\beta(u) > 0$ and $\delta(I) = \frac{c(I)}{2C}$, where C is a constant shown in the right hand side of inequality (5.5). Then, $\|b-\beta\|_\infty < \delta(I)$ implies

$$\inf_{u\in I} g_b(u) \geq \inf_{u\in I} g_\beta(u) - C\|b-\beta\|_\infty \geq \frac{c(I)}{2} > 0.$$

By the definition,

$$\zeta_b(u) = \frac{1}{g_b(u)} \int x_{-1} p(u - x'_{-1}b_z, x_{-1}, z) d\pi(x_{-1}, z),$$

hence there is a constant $C(I) > 0$, which depends on I, such that

$$\begin{aligned}
|\zeta_b(u) - \zeta_\beta(u)| &\leq \frac{\int |x_{-1}| \cdot \left|p(u - x'_{-1}b_z, x_{-1}, z) - p(u - x'_{-1}\beta_z, x_{-1}, z)\right| d\pi(x_{-1}, z)}{c(I)/2} \\
&\quad + \frac{|g_b(u) - g_\beta(u)|}{c(I)^2/4} \int |x_{-1}| p(u - x'_{-1}\beta_z, x_{-1}, z)\, d\pi(x_{-1}, z) \\
&\leq C(I)\|b-\beta\|_\infty
\end{aligned}$$

for any $u \in I$ and $b \in B$ such that $\|b-\beta\|_\infty < \delta(I)$. □

Lemma 5.7 *At every* $(x, z) \in \mathcal{X} \times \mathcal{Z}$, *consider a map* $\mathcal{D}_{x,z} : \mathcal{B} \mapsto \mathbb{R}$ *defined by*

$$\mathcal{D}_{x,z}(b) = \mathcal{D}_b(x * b_z) = E(D|x * b_z).$$

Then $\mathcal{D}_{x,z}$ *is almost surely differentiable at* $b = \beta$ *in the following sense: there exists a linear, continuous operator* $\partial\mathcal{D}_{x,z} : L_\infty(\mathcal{Z})^{k-1} \mapsto \mathbb{R}$ *defined by*

$$\partial\mathcal{D}_{x,z}(h) = f(x * \beta_z)\Big\{x'_{-1}h_z - E\big[x'_{-1}h_z \,|\, x * \beta_z\big]\Big\} \tag{5.6}$$

for every $h \in L_\infty(\mathcal{Z})^{k-1}$ *and* $\sup_{x,z} \Big|\mathcal{D}_{x,z}(\beta + h) - \mathcal{D}_{x,z}(\beta) - \partial\mathcal{D}_{x,z}(h)\Big| = o(\|h\|_\infty)$ *is satisfied.*

(Proof) Choose an arbitrary small $\delta > 0$. By the assumption (A.5),

$$\operatorname{ess.sup}_u \big|F(u+\delta) - \big\{F(u) + f(u)\cdot\delta\big\}\big| = o(\delta) \tag{5.7}$$

as $\delta \downarrow 0$. This implies that, for any $b \in \mathcal{B}$ such that $\|b-\beta\|_\infty < \delta$,

$$\begin{aligned}
\mathcal{D}_{x,z}(b) &= E[F(x * \beta_z) \,|\, x * b_z] \\
&= F(x * b_z) + E[F(x * \beta_z) - F(x * b_z) \,|\, x * b_z] \\
&= F(x * b_z) - E\big[f(x * b_z) \cdot x'_{-1}(b_z - \beta_z) \,|\, x * b_z\big] + E\big[o\big(x'_{-1}(b_z - \beta_z)\big) \,|\, x * b_z\big].
\end{aligned}$$

Let $M = diam(\mathcal{X}_{-1})$, and

$$E\big[o\big(x'_{-1}(b_z - \beta_z)\big)\,\big|\, x * b_z\big] \le M \cdot o(\|b-\beta\|_\infty) \le o(\delta)$$

uniformly in (x, z). Therefore,

$$\mathcal{D}_{x,z}(\beta + h) - F(x*\beta_z) = f(x*\beta_z)\big\{x'_{-1}h_z - E\left[x'_{-1}h_z \,\middle|\, x*b_z\right]\big\} + o(\delta)$$

holds for arbitrary $h : \mathcal{Z} \mapsto \mathbb{R}^{k-1}$ such that $\|h\|_\infty < \delta$. □

Lemma 5.8 *There exists constant C_F such that*

$$\sup_{(x,z)\in\mathcal{X}\times\mathcal{Z}} |F_b^*(x*b_z) - F(x*\beta_z)| \le C_F\|b-\beta\|_\infty \tag{5.8}$$

as well as $\sup_{(x,z)\in\mathcal{X}\times\mathcal{Z}} |\mathcal{D}_b(x*b_z) - F(x*\beta_z)| \le C_F\|b-\beta\|_\infty$.

(Proof) Let $\overline{\mathcal{D}}_b(u) = \sup_{v\le u}\mathcal{D}_b(v)$. By Lemma 5.4, $F_b^*(u) \le \overline{\mathcal{D}}_b(u)$, and

$$F_b^*(u) - F(u) \le \overline{\mathcal{D}}_b(u) - \mathcal{D}_\beta(u) \le \sup_v |\mathcal{D}_b(v) - \mathcal{D}_\beta(v)|\,.$$

By Assumption (A.5), there is constant $C > 0$ such that

$$|\mathcal{D}_b(u) - \mathcal{D}_\beta(u)| = \Big|E\Big[F(x*b_z) - F(x*\beta_z)\,|\, x*b_z = u\Big]\Big| \le C\|b-\beta\|_\infty.$$

In the same way, by using $\underline{\mathcal{E}}_b(u) = \inf_{v\ge u}\mathcal{E}_b(v)$, we can prove that $F_b(u) - F(u) \ge -C\|b-\beta\|_\infty$. Therefore,

$$\begin{aligned} |F_b^*(x*b_z) - F(x*\beta_z)| &\le |F_b^*(x*b_z) - F(x*b_z)| + |F(x*b_z) - F(x*\beta_z)| \\ &\le (2C + \sup_u f(u)\cdot diam(\mathcal{X}_{-1}))\cdot\|b-\beta\|_\infty. \end{aligned}$$

□

Lemma 5.9 *For any bounded interval I, there is a constant $C^*(I)$ and $\delta(I)$ such that*

$$\sup_{u\in I}\left|\frac{d\mathcal{D}_b}{du}(u) - f(u)\right| \le C^*(I)\cdot\|b-\beta\|_\infty \tag{5.9}$$

if $\|b-\beta\|_\infty < \delta(I)$.

(Proof) Fix an arbitrary bounded interval $I \subset \mathbb{R}$. By the definition of $\mathcal{D}_b$ and **(A.5)**, for arbitrary $s \to 0$ and $u \in I$, the Taylor expansion proves

$$\sup_{u\in I}\left|\frac{\mathcal{D}_b(u+s) - \mathcal{D}_b(u)}{s} - f(u)\right| \le \frac{\sup_{\|b-\beta\|_\infty<\delta(I)}\sup_{u\in I}|g_b'(u)|}{\inf_{\|b-\beta\|_\infty<\delta(I)}\inf_{u\in I} g_b(u)}\cdot\sup_{u\in I}|F(u) - \mathcal{D}_b(u)| + O(s).$$

By Lemma 5.8, the conclusion is shown. □

Lemma 5.10 *Let $A_b = a(x,z|b)$ be an arbitrary bounded, measurable function of (x,z) such that*

$$\sup_{(x,z)\in\mathcal{X}\times\mathcal{Z}} |a(x,z|b) - a(x,z|\beta)| \to 0$$

as $\|b-\beta\|_\infty \to 0$. Define map $\varphi : \mathcal{B} \mapsto L_\infty(\mathcal{Z})$ by $\varphi(b)(z) = E\left[\, A_b \cdot \left\{D - F_b^*(x * b_z)\right\} \,\middle|\, z \,\right]$, then φ is differentiable at $b = \beta$ in the following sense: there exists a linear, bounded operator $\partial\varphi : L_\infty(\mathcal{Z})^{k-1} \mapsto \mathbb{R}$ defined by

$$\partial\varphi(h) = -E\Big[\, A_\beta \cdot f(x*\beta_z)\Big\{x'_{-1}h_z - E[x'_{-1}h_z \,|\, x*\beta_z]\Big\} \,\Big|\, z \,\Big] \tag{5.10}$$

such that $\sup_{z\in\mathcal{Z}} |\varphi(\beta+h) - \varphi(\beta) - \partial\varphi(h)| = o(\|h\|_\infty)$ for every $h \in L_\infty(\mathcal{Z})^{k-1}$.

(Proof) By the definition, $\varphi(\beta) = 0$. At every b and z, let $\mathcal{X}_b^\circ(z) = \{x \in \mathcal{X} \,:\, 0 < F_b^*(x*b_z) < 1\}$, which is uniformly bounded for every b and z. For a small $\eta > 0$, define $\mathcal{X}_{\beta,\eta}^\circ(z)$ by

$$\mathcal{X}_{\beta,\eta}^\circ(z) = \{x \in \mathcal{X} \,:\, \eta < F(x*\beta_z) < 1-\eta\} \ (\subset \mathcal{X}_\beta^\circ(z)),$$

and we have

$$\begin{aligned}\varphi(b)(z) \;=\; & E\left[\, A_b \cdot (F(x*\beta_z) - F_b^*(x*b_z)) \cdot \{x \in \mathcal{X}_{\beta,\eta}^\circ(z)\} \,\middle|\, z \,\right] \\ & + E\left[\, A_b \cdot (F(x*\beta_z) - F_b^*(x*b_z)) \cdot \{x \in \mathcal{X}_\beta^\circ(z) \cup \mathcal{X}_b^\circ(z) \backslash \mathcal{X}_{\beta,\eta}^\circ(z)\} \,\middle|\, z \,\right]\end{aligned}$$

Since A_b is bounded, there exists a constant C such that

$$\begin{aligned}&\left|E\left[\, A_b \cdot (F(x*\beta_z) - F_b^*(x*b_z)) \cdot \{x \in \mathcal{X}_\beta^\circ(z) \cup \mathcal{X}_b^\circ(z) \backslash \mathcal{X}_{\beta,\eta}^\circ(z)\} \,\middle|\, z \,\right]\right| \\ &\le C \cdot \|b-\beta\|_\infty \cdot \mathbf{P}\{x \in \mathcal{X}_\beta^\circ(z) \cup \mathcal{X}_b^\circ(z) \backslash \mathcal{X}_{\beta,\eta}^\circ(z) \,|\, z\,\}.\end{aligned}$$

Lemma 5.3 implies that, as $b \to \beta$,

$$\mathbf{P}\{x \in \mathcal{X}_\beta^\circ(z) \cup \mathcal{X}_b^\circ(z) \backslash \mathcal{X}_{\beta,\eta}^\circ(z) \,|\, z\,\} \le \mathbf{P}\{x \in \mathcal{X}_b^\circ(z) \backslash \mathcal{X}_\beta^\circ(z) \,|\, z\,\} + \mathbf{P}\{x \in \mathcal{X}_\beta^\circ(z) \backslash \mathcal{X}_{\beta,\eta}^\circ(z) \,|\, z\,\} = o(1) + O(\eta)$$

uniformly in z.

When $b \to \beta$, on the other hand, Lemma 5.9 together with **(A.5)** implies that $\frac{d}{du}\mathcal{E}_b(x*b_z) > 0$ for any $x \in \mathcal{X}_{\beta,\eta}^\circ(z)$, and therefore that $F_b(x*b_z) \equiv \mathcal{E}_b(x*b_z)$ on $\mathcal{X}_{\beta,\eta}^\circ(z)$ when b is sufficiently close to β. By (5.6),

$$\begin{aligned}\varphi(\beta+h) \;=\; & -E\Big[\, A_\beta \cdot f(x*\beta_z)\Big\{x'_{-1}h_z - E(x'_{-1}h_z \,|\, x*\beta_z)\Big\} \cdot \{x \in \mathcal{X}_{\beta,\eta}^\circ(z)\} \,\Big|\, z \,\Big] \\ & + o(\|h\|_\infty) + O(\|h\|_\infty \cdot \eta)\end{aligned}$$

for any sufficiently small $h \in L_\infty(\mathcal{Z})$, and the result is obtained by $\eta \downarrow 0$. □

Proposition 5.11 *There is an open neighborhood $\mathcal{N}_\beta (\subset \mathcal{B})$ of β and constant $c \in [0,1)$ such that*

$$|\mathcal{R}(b(z)) - \mathcal{R}(\beta(z))| \le c \cdot |b_z - \beta_z| \tag{5.11}$$

for every $b \in \mathcal{N}_\beta$.

(Proof) Let $supp(F_b^*) = (\ell^*(b), u^*(b))$, where ℓ^* and u^* are defined by (5.4). By integrating by part,

$$(\mathcal{A}F_b^*)(x * b_z) = (x * b_z - \mu(F_b^*)) \cdot F_b^*(x * b_z) - \int_{\ell^*(b)}^{x*b_z} F_b^*(s)\, ds,$$

where $\mu(F_b^*) = (x * b_z) \cdot F_b^*(x * b_z) - \int_{\ell^*(b)}^{x*b_z} F_b^*(s)\, ds$. By Lemma 5.3 and Lemma 5.8,

$$\sup_{(x,z)\in\mathcal{X}\times\mathcal{Z}} \left| \frac{(\mathcal{A}F_b^*)(x * b_z)}{\kappa + F_b^*(x * b_z)(1 - F_b^*(x * b_z))} - \frac{(\mathcal{A}F)(x * \beta_z)}{\kappa + F(x * \beta_z)(1 - F(x * \beta_z))} \right| \to 0$$

as $\|b - \beta\|_\infty \to 0$. By Lemma 5.10, therefore,

$$\mathcal{R}(b(z)) = b_z - V_1(z)^{-1}V_2(z)(b_z - \beta_z) + o(\|b - \beta\|_\infty),$$

where $V_1(z)$ and $V_2(z)$ are $(k-1) \times (k-1)$ matrices defined by $V_1(z) = Var(x_{-1}|z)$ and

$$V_2(z) = E\left[\frac{f(x * \beta_z) \cdot (\mathcal{A}F)(x * \beta_z)}{\kappa + F(x * \beta_z)(1 - F(x * \beta_z))} Var(x_{-1}|x * \beta_z) \,\middle|\, z \right].$$

In a neighborhood $\mathcal{N}_\beta = \{b \in \mathcal{B} : \|b - \beta\|_\infty < \delta\}$ with sufficiently small $\delta > 0$, dynamics $b_z^{t+1} = \mathcal{R}(b_z^t)$ is approximated by

$$b_z^{t+1} - \beta_z = (I_{k-1} - V_1(z)^{-1}V_2(z))(b_z^t - \beta_z)$$

at every $z \in \mathcal{Z}$. Let $a_z^t := V_1(z)^{1/2}(b_z^t - \beta_z)$ for every t and $z \in \mathcal{Z}$ and let $A(z) = I_{k-1} - V_1(z)^{-1/2}V_2(z)V_1(z)^{-1/2}$, so that $a_z^{t+1} = A(z)a_z^t$. By the assumption, $V_1(z) > V_2(z)$ in the positive-definite sense, hence $O_{k-1} \le A(z) < I_{k-1}$. Let α_z be an eigenvector of $A(z)$ with the maximal eigenvalue λ_A. Then, $0 \le \alpha' A(z)\alpha = \lambda_A|\alpha|^2 < |\alpha|^2$ or $0 \le \lambda_A < 1$. This proves the proposition because

$$\begin{aligned} |\mathcal{R}(b(z)) - \mathcal{R}(\beta(z))| &\le |V_1(z)^{-1/2}AV_1(z)^{1/2}(b(z) - \beta(z))| + o(\|b - \beta\|_\infty) \\ &\le (\lambda_A + o(1))|b(z) - \beta(z)|. \end{aligned}$$

■

5.2 Uniform Consistency of $\hat{\mathcal{R}}_n$

To keep appearance of following proofs neat and clear, notations used in the literature of *empirical process theory* are very useful. The theory offers powerful tools to investigate asymptotic behaviors of a broad class of statistics. For details, see e.g. Pollard (1990), van der Vaart and Wellner (1996), van der Vaart (1998), van de Geer (2000) or Kosorok (2008). Let $\zeta_1, \zeta_2, \cdots$ be independent copies of $\zeta = (D, x, z, \epsilon)$. In the sequel $\mathbb{P}_n = n^{-1}\sum_{i=1}^n \delta_{\zeta_i}$ denotes the empirical measure based on $\{\zeta_1, \cdots, \zeta_n\}$. A feature of the empirical process theory is to treat $\mathbf{P}$ and $\mathbb{P}_n$ as integral operators defined by

$$\begin{cases} \mathbf{P}g = \int g(\zeta)\, \mathbf{P}(d\zeta) \\ \mathbb{P}_n g = \int g(\zeta)\, \mathbb{P}_n(d\zeta) = \dfrac{1}{n}\sum_{i=1}^n g(\zeta_i) \end{cases} \tag{5.12}$$

Particularly if g is nonstochastic, $\mathbf{P}g$ is equivalent to the usual expectation $Eg(\zeta)$.

In the following proofs, let $\|\cdot\|_r$ be a norm defined by

$$\|h\|_r = \left(\int |h(\zeta)|^r \, d\mathbf{P}(\zeta)\right)^{1/r}$$

for every measurable function h of ζ. Let $L_r(\mathbf{P})$ be a set of all functions h such that $\|h\|_r < \infty$.

We start our proof of consistency from the next lemma.

Lemma 5.12 *Define function $g(\,\cdot\,,(b,H))$ of $\zeta = (\delta, x, z) \in \{0,1\} \times \mathcal{X} \times \mathcal{Z}$ by*

$$g(\zeta,(b,H)) = 2(H(x*b_z) - F_b^*(x*b_z))\left(\delta - \frac{H(x*b_z)+F_b^*(x*b_z)}{2}\right) \tag{5.13}$$

for every $(b,H) \in \mathcal{B} \times \mathcal{F}$. Then,

$$\mathbf{P}\left|\hat{F}_b^*(x*b_z) - F_b^*(x*b_z)\right|^2 \le (\mathbb{P}_n - \mathbf{P})g(\zeta,(b,\hat{F}_b^*)) \tag{5.14}$$

at every $b \in \mathcal{B}$.

(Proof) By the definition, at every $b \in \mathcal{B}$, $\mathbb{P}_n(\delta - \hat{F}_b^*(x*b_z))^2 \le \mathbb{P}_n(\delta - F_b^*(x*b_z))^2$, hence

$$\mathbb{P}_n(\delta - \hat{F}_b^*(x*b_z))^2 - \mathbb{P}_n(\delta - F_b^*(x*b_z))^2 = -\mathbb{P}_n g((\delta,x,z),(b,\hat{F}_b^*)) \le 0$$

or $-\mathbf{P}g((\delta,x,z),(b,\hat{F}_b^*)) \le (\mathbb{P}_n - \mathbf{P})g((\delta,x,z),(b,\hat{F}_b^*))$. By (5.1),

$$\mathbf{P}\big(\delta - F_b^*(x*b_z)\big)\big(H(x*b_z) - F_b^*(x*b_z)\big) \le 0$$

for any $H \in \mathcal{F}$, which implies $\mathbf{P}g((\delta,x,z),(b,\hat{F}_b^*)) \le -\mathbf{P}\big(\hat{F}_b^*(x*b_z) - F_b^*(x*b_z)\big)^2$ or

$$\mathbf{P}\big(\hat{F}_b^*(x*b_z) - F_b^*(x*b_z)\big)^2 \le -\mathbf{P}g((\delta,x,z),(b,\hat{F}_b^*)) \le (\mathbb{P}_n - \mathbf{P})g((\delta,x,z),(b,\hat{F}_b^*)).$$

□

The inequality (5.14) means that the WZ estimator is consistent if class $\mathcal{G} = \{g(\,\cdot\,,(b,H)) : b \in \mathcal{B}, H \in \mathcal{F}\}$ is 'small' enough to satisfy the *uniform law of large numbers (ULLN)*,

$$\sup_{b\in\mathcal{B},H\in\mathcal{F}} |(\mathbb{P}_n - \mathbf{P})g(\zeta,(b,H))| \xrightarrow{p} 0 \quad (n \to \infty). \tag{5.15}$$

The ULLN is clearly satisfied if $\#\mathcal{G} < \infty$. Even in a general case where $\#\mathcal{G} = \infty$, the next proposition gives a sufficient condition for the ULLN.

Proposition 5.13 (Glivenko-Cantelli Theorem) *Let $\mathcal{G}$ be a subset of $L_r(\mathbf{P})$. For every small $\eta > 0$, the bracketing number $N_{[]}(\eta, \mathcal{G}, \|\cdot\|_r)$ of $\mathcal{G}$ is defined as the least number N of pairs $(\ell_j, u_j) \in L_r(\mathbf{P}) \times L_r(\mathbf{P})$ for $j = 1,2,\cdots,N$ such that (i) $\max_{1\le j\le N}\|u_j - \ell_j\|_r < \eta$ and that (ii) for any $g \in \mathcal{G}$, at least one pair (ℓ_j, u_j) satisfies $\ell_j \le g \le u_j$. The set $\{(\ell_1,u_1),\cdots,(\ell_N,u_N)\}$ satisfying conditions (i) and (ii) is called the $(L_r(\mathbf{P}),\eta)$-bracketing set of $\mathcal{G}$.*

If $N_{[]}(\eta, \mathcal{G}, \|\cdot\|_1) < \infty$ *for any* $\eta > 0$*, then*

$$\sup_{g \in \mathcal{G}} |(\mathbb{P}_n - \mathbf{P})g| \xrightarrow{p} 0 \tag{5.16}$$

as $n \to 0$*.*

For a proof of the GC theorem, see e.g. van der Vaart and Wellner (1996). In order to prove uniform consistency of $\hat{\mathcal{R}}_n$, therefore, we need to compute the bracketing number of the linear index family by the next lemma.

Lemma 5.14 *Let* Λ *be a set of single index functions* $\lambda : \mathcal{X} \times \mathcal{Z} \mapsto [0,1]$ *such that* $\lambda(x, z) = H(x * b_z)$ *with* $(b, H) \in \mathcal{B} \times \mathcal{F}$*. Then, there is a constant* $C_\Lambda > 0$ *such that*

$$\log N_{[]}(\eta, \Lambda, L_r(\mathbf{P})) \leq \frac{C_\Lambda}{\eta} \tag{5.17}$$

for every small $\eta > 0$*.*

A proof of the lemma is a straightforward extension of Example 7.4.3 of van de Geer (2000) or Theorem 2.7.5 of van der Vaart nad Wellner (1996). The result is generalized to a case of Lipschitz continuous functionals of Λ by the next lemma.

Corollary 5.15 *Let* $\mathcal{M}\,(\subset L_1(\mathbf{P}))$ *be a set of functions* $m(\,\cdot\,, (b, H))$ *such that*

$$\begin{aligned}&\left| m(\zeta, (b, H)) - m(\zeta, (\tilde{b}, \tilde{H}))\right| \\ &\qquad \leq C\left(|b_z - \tilde{b}_z| + |H(x * b_z) - \tilde{H}(x * \tilde{b}_z)| + \mathbf{P}|H(x * b_z) - \tilde{H}(x * \tilde{b}_z)|^2\right)\end{aligned}$$

for any (b, H) *and* $(\tilde{b}, \tilde{H})$ *in* $\mathcal{B} \times \mathcal{F}$*, where* $\zeta = (D, x, z)$ *and* C *is a positive constant. Then,*

$$\log N_{[\cdot]}(\eta, \mathcal{M}, \|\cdot\|_r) \leq \frac{C_\mathcal{M}}{\eta} \tag{5.18}$$

with nonessential constant $C_\mathcal{M} > 0$*.*

Thus we obtain the first consistency result of the estimator.

Lemma 5.16

$$\sup_{b \in \mathcal{B}} \mathbf{P}\left|\hat{F}_b^*(x * b_z) - F_b^*(x * b_z)\right|^2 \xrightarrow{p} 0 \tag{5.19}$$

as $n \to \infty$*.*

(Proof) From Lemma 5.12,

$$\sup_{b \in B} \mathbf{P}\left|\hat{F}_b^*(x * b_z) - F_b^*(x * b_z)\right|^2 \leq 2 \sup_{\lambda, \lambda' \in \Lambda} (\mathbb{P}_n - \mathbf{P}) m(\zeta, (\lambda, \lambda')),$$

where $m(\,\cdot\,,(\lambda,\lambda'))$ is a function of $\zeta\in\{0,1\}\times\mathcal{X}\times\mathcal{Z}$ defined by

$$m(\zeta,(\lambda,\lambda'))=(\lambda(x,z)-\lambda'(x,z))\left(\delta-\frac{\lambda(x,z)+\lambda'(x,z)}{2}\right)$$

with $(\lambda,\lambda')\in\Lambda\times\Lambda$. Let $\mathcal{M}=\{m(\,\cdot\,,(\lambda,\lambda'))\,:\,(\lambda,\lambda')\in\Lambda\times\Lambda\}$. By the definition,

$$\left|m(\zeta,(\lambda,\lambda'))-m(\zeta,(\tilde\lambda,\tilde\lambda'))\right|\le 3\big(|\lambda(x,z)-\tilde\lambda(x,z)|+|\lambda'(x,z)-\tilde\lambda'(x,z)|\big).$$

The proof of Lemma 5.15 implies that there is a constant $C_{\mathcal{M}}$ such that

$$\log N_{[\,]}(\eta,\mathcal{M},\|\cdot\|_1)\le\frac{C_{\mathcal{M}}}{\eta}\tag{5.20}$$

for every small $\eta>0$. Therefore, the Grivenko-Cantelli theorem proves uniform L_2-consistency of $\hat F_b^*$. □

In the following proof, we need to use uniform consistency of the integrals $\mu(\hat F_b^*)$ or $\mathcal{A}\hat F_b^*$ as well as consistency of $\hat F_b^*$ itself. For the purpose, the next lemma is very useful.

Lemma 5.17 *There exists a nonstochastic compact interval I such that*

$$\mathbf{P}\{I\supset\textstyle\bigcup_{b\in\mathcal{B}}supp\,\hat F_b^*\}=1.\tag{5.21}$$

(Proof) By (5.3), there exists $M>0$ such that

$$(\mathcal{X}\times\mathcal{Z})^\circ\subset\{(x,z)\in\mathcal{X}\times\mathcal{Z}\,|\,-M<x_1<M\}$$

as well as $M>\|b\|_\infty$ for any $b\in\mathcal{B}$ and $M>|x_{-1}|$ for any $x\in\mathcal{X}$. This implies that $\mathbf{P}\{D=1\,|\,x,z\}=1$ if $x_1>M$ and that $\mathbf{P}\{D=1\,|\,x,z\}=0$ if $x_1<-M$. For any $b\in\mathcal{B}$ and (x,z) such that $x*b_z>M(1+M)$, therefore,

$$x_1>M(1+M)-x'_{-1}b_z\ge M(1+M)+(M^2-|x_{-1}||b_z|)\ge M,$$

therefore, $D=1$ if $x*b_z>M(1+M)$ and $D=0$ if $x*b_z<-M(1+M)$.

By the definition, $\hat F_b^*$ minimizes a functional,

$$\begin{aligned}\sum_{i=1}^n(D_i-H(x_i*b(z_i)))^2 &= \sum_{i=1}^n(1-H(x_i*b(z_i)))^2\{x_i*b(z_i)>M(1+M)\}\\ &+\sum_{i=1}^n(0-H(x_i*b(z_i)))^2\{x_i*b(z_i)<-M(1+M)\}\\ &+\sum_{i=1}^n(D_i-H(x_i*b(z_i)))^2\{-M(1+M)\le x_i*b(z_i)\le M(1+M)\},\end{aligned}$$

with respect to $H\in\mathcal{F}$. Hence, with probability 1, $\hat F_b^*(u)\equiv 0$ if $u<-M(1+M)$ and $\hat F_b^*(u)\equiv 1$ if $u>M(1+M)$. Choose $I=(-M(1+M),M(1+M))$ and the lemma is proven. □

Lemma 5.18 *As $n\to\infty$, $\mu(\hat F_b^*)$ and $\mathcal{A}\hat F_b^*$ are uniformly consistent for $\mu(F_b^*)$ and $\mathcal{A}F_b^*$ over $b\in\mathcal{B}$.*

(Proof) We prove uniform consistency of $\mu(\hat{F}_b^*)$. Let I be a bounded interval such that $\cup_b supp(\hat{F}_b^*) \cup supp(F) \subset I$ with probability one. Since I is bounded, there exists a positive constant $c(I)$ such that

$$\left|\mu(\hat{F}_b^*) - \mu(F_b^*)\right| = \left|\int_I t\, d\hat{F}_b^*(t) - \int_I t\, dF_b^*(t)\right| \leq \frac{1}{c(I)}\sqrt{\mathbf{P}\left|\hat{F}_b^*(x * b_z) - F_b^*(x * b_z)\right|^2},$$

and Lemma 5.16 proves that $\sup_{b\in\mathcal{B}} \left|\mu(\hat{F}_b^*) - \mu(F_b^*)\right| \xrightarrow{p} 0$ as $n \to \infty$.

By the similar way,

$$\sup_{b\in\mathcal{B}} \mathbf{P}\left|(\mathcal{A}\hat{F}_b^*)(x * b_z) - (\mathcal{A}F_b^*)(x * b_z)\right|^2 \xrightarrow{p} 0$$

as well as $\sup_{b\in\mathcal{B}} \left|(\mathcal{A}\hat{F}_b^*)(x * b_z) - (\mathcal{A}F_b^*)(x * b_z)\right| \xrightarrow{p} 0$ at every (x, z) is also shown. □

Now the following proposition is immediately proven by combining the results of Lemma 5.16, 5.17 and 5.18.

Proposition 5.19 *As $n \to \infty$,*

$$\sup_{b\in\mathcal{B}} \left|\hat{\mathcal{R}}(b(z)) - \mathcal{R}(b(z))\right| \xrightarrow{p} 0 \tag{5.22}$$

at every $z \in \mathcal{Z}$.

5.3 Proof of Consistency

Let $\{b^t \in \mathcal{B} : t \geq 1\}$ be a sequence generated by the population updating rule, $b^t(\cdot) = \mathcal{R}(b^{t-1}(\cdot))$, started from b^0 in the neighborhood of β. If $\hat{\beta}^{t-1}$ and b^{t-1} are in the neighborhood of β, there is constant $0 \leq c < 1$ such that

$$|\hat{\beta}_z^t - b_z^t| \;\leq\; \sup{}_{b\in\mathcal{B}}\left|\hat{\mathcal{R}}_n(b) - \mathcal{R}(b)\right|(z) + c \cdot |\hat{\beta}_z^{t-1} - b_z^{t-1}| \leq \frac{1}{1-c}\sup{}_{b\in\mathcal{B}}\left|\hat{\mathcal{R}}_n(b) - \mathcal{R}(b)\right|(z) = o_p(1),$$

thus $\hat{\beta}^t$ is also in the neighborhood of β. Therefore,

$$|\hat{\beta}_z^t - \beta_z| \leq \frac{1}{1-c}\sup{}_b\left|\hat{\mathcal{R}}_n(b(z)) - R(b(z))\right| + |b_z^t - \beta_z| = o_p(1) + |b_z^t - \beta_z|,$$

which implies consistency of the local WZ estimator as $t \to \infty$ and $n \to \infty$. ■

References

[1] Ayer, M., Brunk, H.D., Ewing, G.M., Reid, W.T. and Silverman, E. (1955) An Empirical Distribution Function for Sampling with Incomplete Information. *Annals of Mathematical Statistics*, **26**, 641-647.

[2] Cai, Z., Fan, J. and Li, R. (2000) Efficient Estimation and Inferences for Varying-Coefficient Models. *Journal of the American Statistical Association*, **95**, 888-902.

[3] Chamberlain, G. (1986) Asymptotic Efficiency in Semi-parametric Models with Censoring. *Journal of Econometrics*, **32**, 189-218.

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

[5] Cosslett, S.R. (1987) Efficiency Bounds for Distribution-Free Estimators of the Binary Choice and the Censored Regression Models. *Econometrica*, **55**, 559-585.

[6] Dominitz, J. and Sherman, R.P. (2005) Some Convergence Theory for Iterative Estimation Procedures with an Application to Semiparametric Estimation. *Econometric Theory*, **21**, 838-863.

[7] Fan, J. and Zhang, W. (1999) Statistical Estimation in Varying Coefficient Models. *The Annals of Statistics*, **27**, 1491-1518.

[8] Gautier, E. and Kitamura, Y. (2009) Nonparametric Estimation in Random Coefficients Binary Choice Models. Cowles Foundation Discussion Paper No. 1721.

[9] Han, A.K. (1987) Non-parametric Analysis of a Generalized Regression Model: The Maximum Rank Correlation Estimator. *Journal of Econometrics*, **35**, 303-316.

[10] Hastie, T. and Tibshirani, R. (1993) Varying-Coefficient Models. *Journal of the Royal Statistical Society, Series B*, **55**, 757-796.

[11] Hensher, D.A. and Green, W.H. (2003) The Mixed Logit Model: The State of Practice. *Transportation*, **30**, 133-176.

[12] Horowitz, J. (1992) A Smoothed Maximum Score Estimator for the Binary Response Model. *Econometrica*, **60**, 505-531.

[13] Ichimura, H. (1993) Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single-Index Models. *Journal of Econometrics*, **58**, 71-120.

[14] Ichimura, H. and Thompson, T.S. (1998) Maximum Likelihood Estimation of a Binary Choice Model with Random Coefficients of Unknown Distribution. *Journal of Econometrics*, **86**, 269-295.

[15] Klein, R. and Spady, R.H. (1993) An Effcient Semiparametric Estimator for Binary Response Models. *Econometrica*, **61**, 387-421.

[16] Kosorok, M.R. (2008) *Introduction to Empirical Processes and Semiparametric Inference*. Springer.

[17] Louviere, J.J., Hensher, D.A. and Swait, J. (2000) *Stated Choice Methods: Analysis and Application*. Cambridge University Press.

[18] Maddala, G.S. (1983) *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge University Press.

[19] Manski, C.F. (1985) Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator. *Journal of Econometrics*, **27**, 313-333.

[20] McFadden, D. and Train, K. (2000) Mixed MNL Models of Discrete Response. *Journal of Applied Econometrics*, **15**, 447-470.

[21] Pagan, A. and Ullah, A. (1999) *Nonparametric Econometrics*. Cambridge University Press.

[22] Pollard, D. (1990) *Empirical Process: Theory and Applications*. NSF-CBMS Regional Conference Series in Probability and Statistics Volume 2.

[23] Revelt, D. and Train, K. (1999) Customer-Specific Taste Parameters and Mixed Logit. Working Paper, University of California at Berkeley.

[24] Sherman, R.P. (1993) The Limiting Distribution of the Maximum Rank Correlation Estimator. *Econometrica*, **61**, 123-137.

[25] Train, K. (2009) *Discrete Choice Models with Simulation*. Cambridge University Press.

[26] van de Geer, S. (2000) *Empirical Processes in M-Estimation*. Cambridge University Press.

[27] van der Vaart, A.W. (1998) *Asymptotic Statistics*. Cambridge University Press.

[28] van der Vaart, A.W. and Wellner, J.A. (1996) *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer.

[29] Wang, W. and Zhou, M. (1995) Iterative Least Squares Estimator of Binary Choice Models: A Semiparametric Approach. Working Paper no. E-180-95, The Center for Business and Economic Research, College of Business and Economics, University of Kentucky.