WINPEC Working Paper Series No.E1801
May 2018

Entropy Numbers of Semiparametric Linear Index Models

Hisatoshi Tanaka

Waseda INstitute of Political EConomy
Waseda University
Tokyo,Japan

Entropy Numbers of Semiparametric Linear Index Models

Hisatoshi Tanaka

May 6, 2018

Abstract

The goal of this paper is to evaluate the complexity of the linear index class by means of entropy numbers, which are the logarithms of the number of balls or intervals needed to cover the class. The results are applied to study asymptotic behavior of M-estimators for semiparametric regression models indexed with the linear index class.

JEL Classification: C14

Keywords: Uniform Entropy; Entropy with Bracketing; Semiparametric Least Squares Estimation; Semiparametric Least Absolute Distance Estimation

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

1 Introduction

Asymptotic properties of non/semiparametric estimation have been investigated by a number of authors from the perspective of the complexity of statistical model sets [*e.g.* van de Geer (1993), Wong and Shen (1995), Birgé et al. (1998), Gao and Wellner (2007)]. One way to measure complexity is to use entropy numbers, which are essentially the number of balls or intervals needed to cover the model set.

The goal of this paper is to compute the entropy numbers of the linear index class, which is a functional parameter set widely used in microeconometric modeling [*e.g.* Cosslett (1983), Manski (1985), Han (1987), Horowitz (1992), Ichimura (1993), Sherman (1993), Klein and Spady (1993)]. The main results of the paper are applied to show the consistency and rates of convergence of semiparametric estimators for models indexed with the linear index class.

The remainder of this paper is organized as follows. Section 2 defines the terms and concepts, and Section 3 presents the main theorems of the paper and provides their proofs. Section 4 provides examples of their application to semiparametric regressions. Section 5 concludes the paper.

2 Definitions

Let $(\mathcal{X}, \mathcal{A})$ be a k-dimensional Euclidean sample space and let $\mathcal{F}$ be a class of bounded nondecreasing functions defined on $\mathbb{R}$ and taking values in $[0,1]$. The linear index class is a set of functions $F \circ b : \mathcal{X} \mapsto [0,1]$ such that

$$F \circ b(x) = F(b \cdot x) \qquad (x \in \mathcal{X}), \tag{2.1}$$

where $F \in \mathcal{F}$ and $b \in B \subseteq \mathbb{R}^k$. When identification of $b = (b_1, \cdots, b_k)$ matters, the first element of b is set equal to one; the normalized coefficient set is $B_1 = \{(1, b_{-1}) : b_{-1} \in B_{-1}(\subseteq \mathbb{R}^{k-1})\}$, and the normalized linear index class is $\Theta_1 = \{F \circ b : F \in \mathcal{F}, b \in B_1\}$.

Let P be a probability measure on $(\mathcal{X}, \mathcal{A})$; for $1 \le r < \infty$, $\|\cdot\|_{P,r}$ is a semi norm defined by $\|g\|_{P,r} = \left(\int |g|^r \, dP\right)^{1/r}$; $L_r(P)$ is a set of functions g such that $\|g\|_{P,r} < \infty$; an open ball in $L_r(P)$ with center $\tau \in L_r(P)$ and radius $\delta > 0$ is a set of functions $g \in L_r(P)$ such that $\|g - \tau\|_{P,r} < \delta$.

Definition 2.1 Let $\mathcal{P}_n$ be a set of finite probability measures on $(\mathcal{X}, \mathcal{A})$ with n supporting points. Given a set $\mathcal{G}$ of functions on $\mathcal{X}$, the covering number $N(\delta, \mathcal{G}, \|\cdot\|_{P_n,r})$ is the minimum number of open balls in $L_r(P_n)$ needed to cover $\mathcal{G}$. The n-uniform entropy is a number given by

$$\mathcal{H}_{n,r}(\delta, \mathcal{G}) = \sup_{P_n \in \mathcal{P}_n} \log N(\delta, \mathcal{G}, \|\cdot\|_{P_n,r}). \tag{2.2}$$

Definition 2.2 Given two functions ℓ and u on $(\mathcal{X}, \mathcal{A}, P)$, the bracket $[\ell, u]$ is the set of all functions f with $\ell \le f \le u$. A δ-bracket is a bracket $[\ell, u]$ with $\|u - \ell\|_{P,r} < \delta$. The bracketing number $N_{[]}(\delta, \mathcal{G}, \|\cdot\|_{P,r})$ is the minimum number of δ-brackets needed to cover $\mathcal{G}$. The entropy of $\mathcal{G}$ with bracketing is a number given by

$$\mathcal{H}_{[]}(\delta, \mathcal{G}, \|\cdot\|_{P,r}) = \log N_{[]}(\delta, \mathcal{G}, \|\cdot\|_{P,r}). \tag{2.3}$$

Example 2.3 The class $\mathcal{F}$ satisfies

$$\mathcal{H}_{[]}(\delta, \mathcal{F}, \|\cdot\|_{P,r}) \leq \frac{K_r}{\delta} \tag{2.4}$$

for every probability measure P, every $r \geq 1$, and a constant K_r that depends only on r (see *e.g.*, van der Vaart and Wellner (1996), pp.159-162).

3 Theorems

Theorem 3.1 For arbitrary $n \geq 2$ and $\delta > 0$,

$\mathcal{H}_{n,r}(\delta, \Theta) \leq \lfloor 1/\delta \rfloor \log(n + \lfloor 1/\delta \rfloor) + 2k\log n.$

(Proof) Choose arbitrary $P_n \in \mathcal{P}_n$ and let $\mathcal{X}^n = \{x^1, \ldots, x^n\}$ be the support of P_n. For every $b \in B$, define the rank-order function $\rho_b : \mathcal{X}^n \mapsto \{1, \ldots, n\}$ by

$$\rho_b(x^i) = 1 + \left|\{\, \xi \in \mathcal{X}^n : b \cdot x^i > b \cdot \xi \,\}\right|, \tag{3.1}$$

where $|\cdot|$ denotes the cardinality of a set: $b \cdot x^i = \min\{b \cdot x^1, \ldots, b \cdot x^n\}$ if and only if $\rho_b(x^i) = 1$, while $b \cdot x^i = \max\{b \cdot x^1, \ldots, b \cdot x^n\}$ if $\rho_b(x^i) = n$.

Define an equivalent relation $\overset{n}{\sim}$ on B by

$$a \overset{n}{\sim} b \quad \Longleftrightarrow \quad \rho_a = \rho_b.$$

In words, a is equivalent to b if and only if a finite sequence $a \cdot x^1, \ldots, a \cdot x^n$ shares the same ordering as $b \cdot x^1, \ldots, b \cdot x^n$. If $b \cdot x^1 < \cdots < b \cdot x^n$, for example, $a \cdot x^1 < \cdots < a \cdot x^n$ must hold for all $a \overset{n}{\sim} b$.

Choose arbitrary small $\delta > 0$. Let $\mathcal{S}_n^\delta$ be a set of nondecreasing mappings from $\{1, 2, \cdots, n\}$ to $\{0,\ \delta,\ 2\delta,\ \cdots\ ,\ \lfloor 1/\delta \rfloor \cdot \delta\}$. Then, for every $\theta = F \circ b \in \Theta$, there exists $S \in \mathcal{S}_n^\delta$ such that $\|\theta - S \circ \rho_a\|_{P_n,r} < \delta$ for any $a \overset{n}{\sim} b$. Therefore, the covering number of Θ is bounded by $|\mathcal{S}_n^\delta| \times |\mathbb{R}^k / \overset{n}{\sim}|$. Since

$$|\mathcal{S}_n^\delta| \leq \begin{pmatrix} n + \lfloor 1/\delta \rfloor \\ n \end{pmatrix} \leq (n + \lfloor 1/\delta \rfloor)^{\lfloor 1/\delta \rfloor},$$

$|\mathbb{R}^k / \overset{n}{\sim}|$ is to be evaluated to complete the proof.

Suppose that $k = 1$ and that $x^1 \leq \cdots \leq x^n$. Then, $bx^1 \leq \cdots \leq bx^n$ if $b > 0$, $bx^1 \geq \cdots \geq bx^n$ if $b < 0$, and $bx^1 = \cdots = bx^n$ if $b = 0$. Therefore, $\mathbb{R}^1/\overset{n}{\sim} = \{(0, \infty),\ \{0\},\ (-\infty, 0)\}$, so that $|\mathbb{R}^1/\overset{n}{\sim}| = 3$. Now suppose that $k = 2$. Write $x^i = (x_1^i, x_2^i)$ for $i = 1, \ldots, n$ and $b = (b_1, b_2)$. Without any loss of generality, assume that $b_1 x_1^1 \leq \cdots \leq b_1 x_1^n$ and that $x_2^1 > \cdots > x_2^n$. If $b_2 \leq 0$, then $b \cdot x^1 \leq \cdots \leq b \cdot x^n$ holds. As $b_2 \uparrow \infty$, mutual exchanges among $b \cdot x^1, \ldots, b \cdot x^n$ take place until $b \cdot x^1 > \cdots > b \cdot x^n$ is attained. The mutual exchange takes place at most $(n-1) + (n-2) + \cdots + 1 = n(n-1)/2$ times. Therefore, $|\mathbb{R}^2/\overset{n}{\sim}| \leq 3 \times n(n-1)/2$. Recursively,

$$|\mathbb{R}^k/\overset{n}{\sim}| \leq 3\left(\frac{n(n-1)}{2}\right)^{k-1} \leq n^{2k}$$

is obtained for general k. ■

Theorem 3.2 Given $r \geq 1$ and P on $(\mathcal{X}, \mathcal{A})$, assume the following: (i) $E|x_{-1}|^r < \infty$, where $x_{-1} = (x_2, \cdots, x_k)$ for $x \in \mathcal{X}$. (ii) P possesses the density $p = dP/d\mu$ with respect to a reference measure μ, and the conditional density $p(x_1|x_{-1})$ of x_1 conditioned on x_{-1} is uniformly bounded on $\mathcal{X}$. (iii) B_1 is a bounded subset of $\{1\} \times \mathbb{R}^{k-1}$. Then, there is a constant K such that

$$\mathcal{H}_{[]}(\delta, \Theta_1, \|\cdot\|_{P,r}) \leq \frac{K}{\delta} \tag{3.2}$$

for any small $\delta > 0$.

(Proof) For every $a = (1, a_{-1})$ and $b = (1, b_{-1})$ in B_1 and for every $u \in \mathbb{R}$,

$$\begin{aligned} & P\{a \cdot x \leq u < b \cdot x\} \\ & = \int \left(\int_{u - b_{-1} \cdot x_{-1}}^{u - a_{-1} \cdot x_{-1}} p(x_1|x_{-1}) dx_1 \right) p(x_{-1})\, \mu(dx_{-1}) \\ & \leq C_0 |a - b|, \end{aligned} \tag{3.3}$$

where $C_0 = (\sup_x p(x_1|x_{-1}))\, E|x_{-1}| < \infty$ by Assumptions (i) and (ii).

To avoid nonessential complexities, we focus on the case of $k = 2$ in the following part of the proof. Moreover, we assume that $b \geq 0$, so that Assumption (iii) implies $B_1 \subset \{1\} \times [0, K)$ by sufficiently large $K > 0$. The proof can be extended to the general case of $k \geq 3$ and $B_{-1} \subset \mathbb{R}^{k-1}$ in a trivial manner.

Fix arbitrary $\delta_1 > 0$ and let $\gamma_j = (1, \delta_1 j) \in \{1\} \times [0, K + \delta_1)$ for $j = 0, 1, \cdots, N_B$, where $N_B = \lfloor K\delta_1^{-1} \rfloor + 1$. Define functions $\ell_j^\pm$ and $u_j^\pm$ by

$$\begin{cases} \ell_j^+(x) = (\gamma_{j-1} \cdot x)\mathbb{I}_{x_2 \geq 0}(x) \\ \ell_j^-(x) = (\gamma_j \cdot x)\mathbb{I}_{x_2 < 0}(x) \\ u_j^+(x) = (\gamma_j \cdot x)\mathbb{I}_{x_2 \geq 0}(x) \\ u_j^-(x) = (\gamma_{j-1} \cdot x)\mathbb{I}_{x_2 < 0}(x) \end{cases}$$

for $j = 1, \cdots, N_B$. Set $\ell_j = \ell_j^+ + \ell_j^-$ and $u_j = u_j^+ + u_j^-$, then $[\ell_j, u_j]$ is a $\delta_1 (E|x_{-1}|^r)^{1/r}$-bracket for the linear mapping class $\{x \mapsto b \cdot x : b \in B_1^+\}$ (see Fig 3.1).

Fix arbitrary $\delta_2 > 0$. For every $j = 0, 1, \ldots, N_B$, let P_j be the law of $\gamma_j \cdot x$, that is, $P_j A = P\{\gamma_j \cdot x \in A\}$; $M_j = N_{[]}(\delta_2, \mathcal{F}, \|\cdot\|_{P_j,r})$, and $[L_{ij}, U_{ij}]$ for $i = 1, \ldots, M_j$ is the corresponding δ_2-bracket covering $\mathcal{F}$. Without any loss of generality, all L_{ij} and U_{ij} are assumed to be nondecreasing. Example 2.3 shows that $\max_{1 \leq j \leq N_B} \log M_j \leq K_r \delta_2^{-1}$, where K_r depends only on r.

For every $\theta = F \circ b \in \Theta_1$, there exist (ℓ_j, u_j) and (L_{ij}, U_{ij}) such that $L_{ij} \circ \ell_j \leq \theta \leq U_{ij} \circ u_j$ and that

$$\|U_{ij} \circ u_j - L_{ij} \circ \ell_j\|_{P,r}$$

$$
\begin{aligned}
&\le \|(U_{ij} \circ u_j^+ - L_{ij} \circ u_j^+)\mathbb{I}_{x_2 \ge 0}\|_{P,r} \\
&\quad + \|(L_{ij} \circ u_j^+ - L_{ij} \circ \ell_j^+)\mathbb{I}_{x_2 \ge 0}\|_{P,r} \\
&\quad + \|(U_{ij} \circ u_j^- - L_{ij} \circ u_j^-)\mathbb{I}_{x_2 < 0}\|_{P,r} \\
&\quad + \|(L_{ij} \circ u_j^- - L_{ij} \circ \ell_j^-)\mathbb{I}_{x_2 < 0}\|_{P,r}.
\end{aligned}
$$

By the definition of P_j and P_{j-1},

$$
\begin{aligned}
&\|(U_{ij} \circ u_j^+ - L_{ij} \circ u_j^+)\mathbb{I}_{x_2 \ge 0}\|_{P,r} \\
&\quad + \|(U_{ij} \circ u_j^- - L_{ij} \circ u_j^-)\mathbb{I}_{x_2 < 0}\|_{P,r} \\
&\quad \le \|U_{ij} - L_{ij}\|_{P_j,r} + \|U_{i,j-1} - L_{i,j-1}\|_{P_{j-1},r} \\
&\quad \le 2\delta_2.
\end{aligned}
$$

Since L_{ij} is bounded and nondecreasing, there is a step function $S = \sum_m s_m \mathbb{I}_{(t_m,\infty)}$ such that $0 \le s_m \le 1$, that $\sum_m s_m \le 1$, and that $\sup_t |L_{ij}(t) - S(t)| < \delta_2$. Therefore,

$$
\begin{aligned}
&\|(L_{ij} \circ u_j^+ - L_{ij} \circ \ell_j^+)\mathbb{I}_{x_2 \ge 0}\|_{P,r} \\
&\le \|(L_{ij} \circ u_j^+ - S \circ u_j^+)\mathbb{I}_{x_2 \ge 0}\|_{P,r} \\
&\quad + \|(S \circ u_j^+ - S \circ \ell_j^+)\mathbb{I}_{x_2 \ge 0}\|_{P,r} \\
&\quad + \|(S \circ \ell_j^+ - L_{ij} \circ \ell_j^+)\mathbb{I}_{x_2 \ge 0}\|_{P,r} \\
&\le 2\delta_2 + \|(S \circ u_j^+ - S \circ \ell_j^+)\mathbb{I}_{x_2 \ge 0}\|_{P,r}.
\end{aligned}
$$

Inequality (3.3) together with $\sum_m s_m \le 1$ implies that

$$
\begin{aligned}
&\|(S \circ u_j^+ - S \circ \ell_j^+)\mathbb{I}_{x_2 \ge 0}\|_{P,r} \\
&\quad \le \sup_{j,t} \left\{ P \Big| \{\gamma_j \cdot x \le t\} - \{\gamma_{j-1} \cdot x \le t\} \Big| \right\}^{1/r} \\
&\quad \le (C_0\delta_1)^{1/r}.
\end{aligned}
$$

In the same way, $\|(L_{ij} \circ u_j^- - L_{ij} \circ \ell_j^-)\mathbb{I}_{x_2<0}\|_{P,r} \le (C_0\delta_1)^{1/r} + 2\delta_2$ is shown, hence

$$
\|U_{ij} \circ u_j - L_{ij} \circ \ell_j\|_{P,r} \le \delta := 2(C_0\delta_1)^{1/r} + 6\delta_2.
$$

Note that $\mathcal{H}_{[]}(\delta, \Theta_1, \|\cdot\|_{P,r}) \le \log \sum_{j=0}^{N_B} M_j$. In particular, by choosing $\delta_1 = \delta^r/4^r C_0$ and $\delta_2 = \delta/12$,

$$
\log \sum_{j=0}^{N_B} M_j \le \log\left(1 + \frac{K}{\delta_1}\right) + \frac{K_r}{\delta_2} \le \frac{2r + 12K_r}{\delta}
$$

holds for any sufficiently small $\delta > 0$. ∎

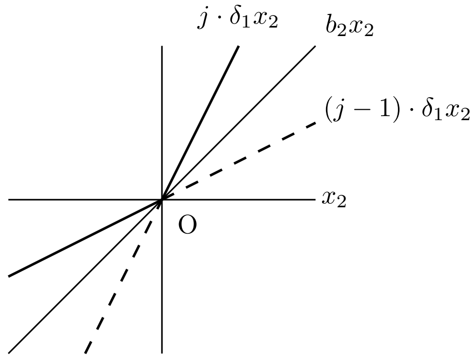

Figure 3.1: A bracket $[\ell_j, u_j]$ of a linear mapping $x \mapsto b \cdot x$.

4 Applications

4.1 Notations

Notations from empirical process theory are employed in this subsection. Let $(\mathcal{Z}, \mathcal{B}, P)$ be a probability space, and let $\mathcal{G} = \{g\}$ be a class of functions on $\mathcal{Z}$; $\mathbb{P}_n g = n^{-1}\sum_{i=1}^n g(z^i)$, where $z^1, \ldots, z^n$ are independently sampled from P; $Pg = \int g\,dP = E_P g(z)$; $\mathbb{G}_n$ maps $g \in \mathcal{G}$ into $\sqrt{n}(\mathbb{P}_n - P)g = n^{-1}\sum_{i=1}^n (g(z^i) - Eg(z^i))$ and is called the empirical process on $\mathcal{G}$; $\mathcal{G}$ is said to be P-Glivenko-Cantelli when $\sup_{g\in\mathcal{G}} |(\mathbb{P}_n - P)g| \to 0$ almost surely; $\mathcal{G}$ is P-Donsker if there exists a tight random element $\mathbb{Z}_g$ indexed with $g \in \mathcal{G}$ such that $\mathbb{G}_n g \overset{w}{\Rightarrow} \mathbb{Z}_g$ as $n \to \infty$, where "$\overset{w}{\Rightarrow}$" denotes weak convergence; for two functions ϕ, ψ of δ, $\phi \lesssim \psi$ means that there is a constant $C > 0$ such that $\phi(\delta) \le C\psi(\delta)$ for all δ [Andrews (1994), van der Vaart and Wellner (1996), van der Vaart (1998)].

4.2 Consistency

Theorem 3.1 implies that $n^{-1}\mathcal{H}_{n,1}(\delta, \Theta) \to 0$ as $n \to \infty$. This is a sufficient condition for Θ to be P-Glivenko-Cantelli [Vapnik and Červonenkis (1982)]. The P-Glovenko-Cantelli property of Θ plays an essential role in proving the consistency of M-estimators for linear index models.

Example 4.1 Let $\mathcal{Z} \subset \mathbb{R}^{k+1}$ and $z = (y, x) \in \mathcal{Z}$. Suppose a semiparametric regression model,

$$
y = \theta_0(x) + \epsilon, \quad \theta_0 \in \Theta, \tag{4.1}
$$

with a moment condition $E(\epsilon|x) = 0$. If $|y|$ is bounded by a positive constant, the model is consistently estimated by the Least Squares (LS) estimator $\hat{\theta}_n^{LS} = \text{argmin}\ \mathbb{P}_n g_\theta$ subject to $\theta \in \Theta$, where $g_\theta(z) = (y - \theta(x))^2$.

To see this, let $\mathcal{G} = \{g_\theta : \theta \in \Theta\}$. The definition of $\hat{\theta}_n^{LS}$ implies $\|\hat{\theta}_n^{LS} - \theta_0\|_{P,2}^2 \le \sup_{g\in\mathcal{G}} (\mathbb{P}_n - P)(g - g_{\theta_0})$. Since $|y|$ is bounded, there is a positive constant C such that $|g_\theta - g_{\theta_0}| \le C|\theta - \theta_0|$ for any θ. Therefore, $\mathcal{H}_{n,1}(\delta, \mathcal{G}) \le \mathcal{H}_{n,1}\left(C^{-1}\delta, \Theta\right) = o(n)$, and $\mathcal{G}$ also becomes P-Glivenko-Cantelli. Thus, $\|\hat{\theta}_n^{LS} - \theta_0\|_{P,2}^2 \to 0$ almost surely.

Example 4.2 If the moment condition for (4.1) is replaced with median$(\epsilon|x) = 0$, the Least Absolute Deviation (LAD) estimator $\hat{\theta}_n^{LAD} = \text{argmin}\ \mathbb{P}_n \lambda_\theta$ subject to $\theta \in \Theta$, where $\lambda_\theta(z) = y - \theta(x)$, becomes consistent for θ_0.

To see this, let $\mathcal{L} = \{\lambda_\theta : \theta \in \Theta\}$. Since $|\lambda_\theta - \lambda_{\theta_0}| \leq |\theta - \theta_0|$, $\mathcal{L}$ is also P-Glivenko-Cantelli. By the definition of $\hat{\theta}_n^{LAD}$, on the other hand, $0 \leq P(\lambda_{\hat{\theta}_n^{LAD}} - \lambda_{\theta_0}) \leq |(\mathbb{P}_n - P)(\lambda_{\hat{\theta}_n^{LAD}} - \lambda_{\theta_0})|$. Hence, $P(\lambda_{\hat{\theta}_n^{LAD}} - \lambda_{\theta_0}) \to 0$ almost surely. Now, let $p(\epsilon|x)$ be the conditional density of ϵ conditioned on x. Assume that $\pi_\epsilon := \inf_{\mathcal{X}} p(0|x) > 0$, then $\pi_\epsilon \|\theta - \theta_0\|_{P,2}^2 \leq P(\lambda_{\hat{\theta}} - \lambda_{\theta_0})$ for any θ in a neighborhood of θ_0. Thus, $\|\hat{\theta}_n^{LAD} - \theta_0\|_{P,2}^2 \to 0$ almost surely.

4.3 Rates of Convergence

Theorem 3.2 implies that $J_{[]}(\delta, \Theta_1, \|\cdot\|_{P,2}) := \int_0^\delta \sqrt{\mathcal{H}_{[]}(t, \Theta_1, \|\cdot\|_{P,2})}\, dt \lesssim \sqrt{\delta} < \infty$ for any small $\delta > 0$. The finite integrated entropy is a sufficient condition for Θ_1 to be a P-Donsker class [Donsker (1952)]. Moreover, it bounds a modulus of continuity of the empirical process $\mathbb{G}_n$ defined on the class. Suppose that $\mathcal{M}$ is a P-Donsker class with $J_{[]}(\delta, \mathcal{M}, \|\cdot\|_{P,2}) \lesssim \sqrt{\delta}$, and that $\mathcal{M}_\delta$ is a subset of $\mathcal{M}$ satisfying (i) $|\mu - \mu_0| \leq {}^\exists M < \infty$ and (ii) $\|\mu - \mu_0\| \lesssim \delta$ for any $\mu \in \mathcal{M}_\delta$, then,

$$E_P \sup{}_{\mu \in \mathcal{M}_\delta} |\mathbb{G}_n(\mu - \mu_0)| \lesssim \phi_n(\delta), \qquad (4.2)$$

where $\phi_n(\delta) = \sqrt{\delta} + 1/(\delta^2\sqrt{n})$ [Lemma 19.36 of van der Vaart (1998)].

One consequence of (4.2) is the cubic rate of convergence of M-estimators for linear index models. Let $\mathcal{M} = \{\mu_\tau : \tau \in \mathcal{T}\}$ be a P-Donsker class and let $\hat{\tau}_n$ be an M-estimator defined by $\hat{\tau}_n = \text{argmin}\, \mathbb{P}_n \mu_\tau$ subject to $\tau \in \mathcal{T}$. If $\|\tau - \tau\|_{P,2}^2 \lesssim P(\mu_\tau - \mu_{\tau_0})$, and if $E_p \sup\{|\mathbb{G}_n(\mu_\tau - \mu_{\tau_0})| : \|\tau - \tau_0\|_{P,2} < \delta\} \leq \phi_n(\delta)$, then $\|\hat{\tau}_n - \tau_0\|_{P,2} \leq O_p(n^{-1/3})$ [Theorem 3.2.5 of van der Vaart and Wellner (1996)].

Example 4.1 (Cont.) Assume (i)-(iii) of Theorem 3.2 and that $\theta_0 \in \Theta_1$. Normalize the LS estimator by solving $\hat{\theta}_n^{LS} = \text{argmin}\, \mathbb{P}_n g_\theta$ subject to $\theta \in \Theta_1$, then, $\|\hat{\theta}_n^{LS} - \theta_0\|_{P,2} \leq O_p(n^{-1/3})$. This rate is immediately obtained from the boundedness of $|y|$ and $P(g_\theta - g_{\theta_0}) = \|\theta - \theta_0\|_{P,2}^2$ for any $\theta \in \Theta_1$.

Example 4.2 (Cont.) Assume (i)-(iii) of Theorem 3.2 and that $\theta_0 \in \Theta_1$. Then, the normalized LAD estimator $\hat{\theta}_n^{LAD} = \text{argmin}\, \mathbb{P}_n \lambda_\theta$ subject to $\theta \in \Theta_1$ satisfies $\|\hat{\theta}_n^{LAD} - \theta_0\|_{P,2} \leq O_p(n^{-1/3})$ since $\|\theta - \theta_0\|_{P,2}^2 \leq \pi_\epsilon^{-1} P(\lambda_\theta - \lambda_{\theta_0})$ in a neighborhood of θ_0.

5 Conclusions

In this paper, two types of entropy numbers of the linear index class, n-uniform entropy and the entropy with bracketing, are computed. The results are applied to prove consistency and the $n^{1/3}$-rate of convergence of the Least Squares (LS) estimator and the Least Absolute Distance (LAD) estimator of a semiparametric regression model.

References

Andrews, D. W. K. (1994): "Asymptotics for Semiparametric Econometric Models via Stochastic Equicontinuity," *Econometrica*, 62, 43–72.

Birgé, L., P. Massart, et al. (1998): "Minimum Contrast Estimators on Sieves: Exponential Bounds and Rates of Convergence," *Bernoulli*, 4, 329–375.

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

Donsker, M. D. (1952): "Justification and Extension of Doob's Heuristic Approach to the Kolmogorov-Smirnov Theorems," *Annals of mathematical statistics*, 277–281.

Gao, F. and J. A. Wellner (2007): "Entropy Estimate for High-Dimensional Monotonic Functions," *Journal of Multivariate Analysis*, 98, 1751–1764.

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

Horowitz, J. L. (1992): "A Smoothed Maximum Score Estimator for the Binary Response Model," *Econometrica*, 60, 505–531.

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

Klein, R. W. and R. H. Spady (1993): "An Efficient Semiparametric Estimator of Binary Response Models," *Econometrica*, 61, 387–421.

Manski, C. F. (1985): "Semiparametric Analysis of Discrete Response : Asymptotic Properties of the Maximum Score Estimator," *Journal of Econometrics*, 27, 313–334.

Sherman, R. P. (1993): "The Limiting Distribution of the Maximum Rank Correlation Estimator," *Econometrica*, 61, 123–137.

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

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

van der Vaart, A. W. and J. A. Wellner (1996): *Weak Convergence and Empirical Processes: With Applications to Statistics*, Springer-Verlag, New York.

Vapnik, V. N. and A. J. Červonenkis (1982): "Necessary and Sufficient Conditions for the Uniform Convergence of Empirical Means to Their True Values," *Theory of Probability and its Applications*, 26, 532–553.

Wong, W. H. and X. Shen (1995): "Probability inequalities for likelihood ratios and convergence rates of sieve MLEs," *Annals of Statistics*, 339–362.