第45卷第11期 中国科学技术大学学报 Vol. 45, No. 11
2015年11月 JOURNAL OF UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA Nov. 2015

Article ID: 0253-2778(2015)11-0881-12

Random weighting method for smoothed binary response model

YUAN Min[1], WU Xiaoyan[2]

(1. *Department of Probability and Statistics, School of Mathematical Sciences, University of Science and Technology of China, Hefei* 230026, *China*;
2. *Department of Mathematics, Electronic Engineering Institute of the People's Liberation Army, Hefei* 230037, *China*)

Abstract: The smoothed score estimating method has the nice properties of $\sqrt{n}$-consistency and asymptotic normality under some regular assumptions. However, the asymptotic variance involves quantities related to the unknown error distribution which is hard to be accurately estimated. A random weighting method for estimating the asymptotic variance of the maximum smoothed score estimators was proposed. The random weighting estimator is shown to be consistent and asymptotically normal. Statistical inference is thus possible with the variance estimates.

Key words: random weighting method; maximum smoothed score function; binary response model

CLC number: O212.1 **Document code**: A doi: 10.3969/j.issn.0253-2778.2015.11.001

2010 Mathematics Subject Classification: Primary 62F12; Secondary 62F40

Citation: Yuan Min, Wu Xiaoyan. Random weighting method for smoothed binary response model[J]. Journal of University of Science and Technology of China, 2015, 45(11): 881-892.

光滑两值响应模型的随机加权方法

袁 敏[1], 吴小燕[2]

(1. 中国科学技术大学数学科学学院概率统计系, 安徽合肥 230026; 2. 中国人民解放军电子工程学院数学系, 安徽合肥 230037)

摘要: 光滑计分估计在一定的正则条件下具有$\sqrt{n}$相合性及渐近正态性. 然而由于渐近方差包含了与未知误差分布有关的未知函数参数而很难被精确估计. 这里提出了用随机加权的方法估计极大光滑计分估计的渐近方差, 并从理论证明和计算机模拟两方面证明了所提出的估计具有相合性和渐近正态性.

关键词: 随机加权方法; 极大光滑计分函数; 两值响应模型

0 Introduction

The binary response model has been applied extensively in economics, sociology, medicine and many other fields. This model assumes that the binary response variable is governed by a latent

Received: 2014-06-27; **Revised**: 2015-04-25

Foundation item: Supported by National Natural Science Foundation of China (11201452).

Biography: YUAN Min (corresponding author), female, born in 1981, PhD/associate Prof. Research field: statistics. E-mail: myuan@ustc.edu.cn

variable which is linearly related to some explanatory variables, but the link function that connects the probability function of the response and the latent variable is assumed to be completely unknown. In this sense, different from the (parametric) generalized linear models for binary response variables, the binary response model is semi-parametric.

The binary response model assumes that the response variable y depends on the explanatory variables x through

$$y=\begin{cases}1, & \text{if } \beta' x+e\geqslant 0;\\ 0, & \text{otherwise}\end{cases} \qquad (1)$$

where e is the unobserved error variable with completely unknown distribution except that median$(e|x)=0$, the $p\times 1$ vector x stands for the explanatory variables and β is the $p\times 1$ regression coefficients. In model (1), only x and y are observable. The expectation of y given x is

$$E(y\mid x)=P(y=1\mid x)=$$
$$P(e\geqslant -\beta' x\mid x)=\overline{F}(-\beta' x),$$

where $F=1-\overline{F}$ is the cumulative distribution function of e, which is completely unspecified except that $F(0)=1/2$. Therefore the link function, the inverse function $\overline{F}^{-1}$, is unspecified. Obviously the model is not identifiable to scale transform on the parameter β. Without loss of generality, we assume that at least one of the regression coefficients is not zero and $\beta_1=1$ (if the effect of x_1 is negative, one can change the sign of x_1). In what follows, for a vector

$$b=(b_1,b_2,\cdots,b_p)'\in R^p,$$

we denote $\tilde{b}=(b_2,\cdots,b_p)'$ for the vector composed by components of b except the first coordinate. Therefore for the true regression coefficients β of the binary response model, $\tilde{\beta}=(\beta_2,\cdots,\beta_p)'$.

Generally, in the binary response model the error distribution (or the link function) is assumed to be unknown and the model is therefore robust in the link specifications. Many well-known parametric regression models such as the logistic regression model and the probit model are sub-models of the binary response model when the error distribution is taken to be the logistic and normal distributions.

Let (x_i,y_i), $i=1,\cdots,n$, be the samples of (x,y). Manski[3] defined the score function

$$S_n^*(b)=\frac{1}{n}\sum_{i=1}^{n}(2y_i-1)\,I(b' x_i\geqslant 0) \qquad (2)$$

where $I(A)$ is the indicator function of event A. He proved that the maximum score estimator of $\tilde{\beta}$, the maximizer of the score function subject to the constraint of $b_1=1$, is consistent. Pollard[13] showed that, at the rate of $n^{\frac{1}{3}}$, the maximum score estimator converges in distribution to the random variable that maximizes a certain Gaussian process.

Obviously, it is difficult to carry out statistical inference based on such complex distribution. Horowitz[1] modified the score function by replacing the indicator function by a smooth function

$$S_n(b)=\frac{1}{n}\sum_{i=1}^{n}(2y_i-1)\,K\left(\frac{b' x_i}{\sigma_n}\right) \qquad (3)$$

where $K(\cdot)$ is a certain smooth function, and the scale is chosen such that $\sigma_n\to 0$ as $n\to\infty$. The maximum smoothed score estimator (MSSE) of $\tilde{\beta}$ is defined as the maximizer of the score function $S_n(b)$ subject to the constraint of $b_1=1$. Horowitz showed that the maximum smoothed score estimators of $\tilde{\beta}$ are consistent and asymptotically normally distributed. However statistical inference is still difficult to implement since the asymptotic variance in the limit distribution involves parameters related to the nonparametric density function of the error e, which may not be accurately estimated in practice.

In this article we propose using the random weighting method to estimate the variance of the MSSE. The random weighting method has been investigated and studied as an alternative method to the bootstrap method and is also referred to as the Bayesian bootstrap method[17,23]. Zheng[19] and Dudley[9] called it the random weighting method. Rao and Zhao[12] applied the method to M-estimation in linear models.

For the maximum smoothed score estimation problem in the binary response model, we can randomly weight the summand of the smoothed score function repeatedly. The randomly weighted version of the estimators can be used to estimate the asymptotic variance of the original maximum smoothed score estimators. The random weighting method put an extra weight, w_i, into the summation of the smoothed score function,

$$S_n^w(b) = \frac{1}{n}\sum_{i=1}^{n} w_i(2y_i - 1)K\left(\frac{b'x_i}{\sigma_n}\right) \tag{4}$$

where

$$\left.\begin{array}{l} w_1, w_2, \cdots \text{i. i. d. } P(w_1 \geqslant 0) = 1, \\ Ew_1 = 1,\ Ew_1^2 = \tau \geqslant 1,\ Ew_1^4 < \infty \end{array}\right\} \tag{5}$$

and the sequence $\{w_i\}$ and $\{(x_i, y_i)\}$ are independent. Then the random weighting maximum smoothed score estimator (WMSSE) $\tilde{\beta}_n^w$ is defined as the maximizer of the smoothed score function.

In this article, we will show that under some conditions the random weighting version of MSSE, $\tilde{\beta}_n^w$, is consistent and the conditional distribution of $\tilde{\beta}_n^w - \tilde{\beta}_n$ given the data $\{(x_i, y_i), i=1,2,\cdots,n\}$ is the same as that of $\tilde{\beta}_n - \tilde{\beta}$ asymptotically.

1 Main results

Firstly, we assume the following assumptions hold as they are necessary for both the parameter's identification and asymptotic normality.

(Ⅰ) (regularity)

(a) The support of the distribution of x is not contained in any proper linear subspace of R^p.

(b) $0 < P(y=1|x) < 1$ for almost every x.

(c) For almost every $\tilde{x} = (x_2, \cdots, x_p)'$, the distribution of x_1 conditional on $\tilde{x}$ has everywhere a positive density with respect to Lebesgue measure.

(d) median$(e|x) = 0$ for almost every x.

Secondly, we need the following conditions on the smooth function K, as given by Horowitz[1].

(K1) $K(v)$ is a continuous function of the real line into itself such that: $|K(v)| < M$ for some finite M and all v in $(-\infty, \infty)$, $\lim_{v\to-\infty} K(v) = 0$ and $\lim_{v\to\infty} K(v) = 1$.

(K2) K is twice differentiable everywhere, $K'(\cdot)$, $K''(\cdot)$ are uniformly bounded, $K'(\cdot)$ is symmetrical about 0 and each of the following integrals over $(-\infty, +\infty)$, $\int (K'(v))^4 \mathrm{d}v$, $\int (K'(v))^2 \mathrm{d}v$, $\int v^2 K''(v)\mathrm{d}v$, $\int v^3 K''(v)\mathrm{d}v$ are all finite.

(K3) For some integer $h \geqslant 2$ and each integer $1 \leqslant i \leqslant h$, $\int |v^i K'(v)\mathrm{d}v| < \infty$ and

$$\int v^i K'(v)\mathrm{d}v = \begin{cases} 0, & \text{if } i \leqslant h-1; \\ d \neq 0, & \text{if } i = h. \end{cases}$$

(K4) For any integer i between 0 and h, $\eta > 0$, and any sequence $\{\sigma_n\}$ converging to zero,

$$\lim_{v\to\infty} \sigma_n^{i-h} \int_{|\sigma_n v| > \eta} |v^i K'(v)|\,\mathrm{d}v = 0,$$

$$\lim_{v\to\infty} \sigma_n^{-1} \int_{|\sigma_n v| > \eta} |K''(v)|\,\mathrm{d}v = 0.$$

Since K is twice differentiable everywhere, $S_n(b)$, $S_n^w(b)$ are also twice differentiable with respect to $\tilde{b} \equiv (b_2, \cdots, b_p)'$, define

$$U_n(b, \sigma_n) = \frac{\partial S_n(b)}{\partial \tilde{b}},\ U_n^w(b, \sigma_n) = \frac{\partial S_n^w(b)}{\partial \tilde{b}}$$

and

$$Q_n(b, \sigma_n) = \frac{\partial^2 S_n(b)}{\partial \tilde{b}\partial \tilde{b}'},\ Q_n^w(b, \sigma_n) = \frac{\partial^2 S_n^w(b)}{\partial \tilde{b}\partial \tilde{b}'}.$$

Let $\beta_n^w \equiv (\beta_{n1}^w, \tilde{\beta}_n^{w'})'$ denote the maximizer of (4), and assume that $\tilde{\beta}$ is an interior point of $\tilde{B}$, then with probability approaching 1 as $n \to \infty$, $\tilde{\beta}_n^w$ is an interior point of $\tilde{B}$, $\beta_{n1}^w = \beta_1 = 1$ and $U_n^w(\beta_n^w, \sigma_n) = 0$. A Taylor expansion of $U_n^w(\beta_n^w, \sigma_n)$ at true parameter β yields

$$U_n^w(\beta_n^w, \sigma_n) = U_n^w(\beta, \sigma_n) + Q_n^w(\beta_n^*, \sigma_n)(\tilde{\beta}_n^w - \tilde{\beta}) = 0,$$

where β_n^* is between β_n^w and β. We will show that there is a real function $\rho(n)$ such that $\rho(n)U_n^w(\beta, \sigma_n)$ is asymptotically normally distributed if σ_n, the bandwidth, is suitably chosen so that $\lg n/(n\sigma_n^4) \to 0$ and $n\sigma_n^{2h+1} \to \lambda$ as $n \to \infty$, where h is a positive integer selected according to some criteria. In addition, under some mild conditions $Q_n^w(\beta_n^*, \sigma_n)$ converges in probability to nonsingular, non stochastic matrix Q, then

$$\rho(n)(\tilde{\beta}_n^w - \tilde{\beta}) = -Q^{-1}\rho(n)U_n^w(\beta,\sigma_n) + o_p(1)$$

Thus, the standard Taylor series methods of asymptotic distribution theory yield the result that $\rho(n)(\tilde{\beta}_n^w - \tilde{\beta})$ is distributed asymptotically as $-Q^{-1}\rho(n)U_n^w(\beta,\sigma_n)$.

To formalize these ideas, let $z=\beta' x$. By assumption (Ⅰ), the distribution of z conditional on $\tilde{x}$ has everywhere positive density with respect to Lebesgue measure, for almost every $\tilde{x}$. Let $p(z|\tilde{x})$ denote this density. For each positive integer i, define $p^{(i)}(z|\tilde{x}) = \partial^i p(z|\tilde{x})/\partial z^i$, whenever the derivative exists and $p^{(0)}(z|\tilde{x}) = p(z|\tilde{x})$. Let $F(\cdot|z,\tilde{x})$ denote the cumulative distribution of e conditional on $\tilde{x}$ and z, define $F^{(i)}(-z \mid z,\tilde{x}) = \partial^i F(-z \mid z,\tilde{x})/\partial z^i$ whenever these derivatives exist. Define the scalar constants α_A and α_D by

$$\alpha_A = \int_{-\infty}^{+\infty} v^h K'(v)\mathrm{d}v,\ \alpha_D = \int_{-\infty}^{+\infty} (K'(v))^2 \mathrm{d}v$$

whenever these quantities exist. For each integer $h \geqslant 2$, define the $p-1$ dimensional vector A and two $(p-1)\times(p-1)$ matrices D and Q by

$$A = -\alpha_A \sum_{i=1}^{h} \frac{1}{i!(h-i)!} E[F^{(i)}(0 \mid 0,\tilde{x}) p^{(h-i)}(0 \mid \tilde{x})\tilde{x}],$$

$$D = \alpha_D E[\tilde{x}\tilde{x}' p(0 \mid \tilde{x})],$$

$$Q = 2E[\tilde{x}\tilde{x}' F^{(1)}(0 \mid 0,\tilde{x}) p(0 \mid \tilde{x})].$$

In order to prove the consistency and asymptotical normality, we need to make the following additional assumptions (Ⅱ)~(Ⅵ):

(Ⅱ) $E\|\tilde{x}\|^4 < \infty$.

(Ⅲ) (a) For each integer $i(1 \leqslant i \leqslant h-1)$, all z's in a neighborhood of 0, almost every $\tilde{x}$, and some $M<\infty$, $p^{(i)}(z|\tilde{x})$ exists and is a continuous function of z satisfying $|p^{(i)}(z|\tilde{x})| \leqslant M$, in addition, $|p(z|\tilde{x})| \leqslant M$ for all z's and almost every $\tilde{x}$.

(b) For each integer $i(1 \leqslant i \leqslant h)$, all z's in a neighborhood of 0, almost every $\tilde{x}$, and some $M < \infty$, $F^{(i)}(-z \mid z,\tilde{x})$ exists and is a continuous function of z satisfying $|F^{(i)}(-z|z,\tilde{x})| \leqslant M$.

(Ⅳ) $\frac{\lg n}{n\sigma_n{}^4} \to 0, \sigma_n \to 0$ as $n \to \infty$.

(Ⅴ) The true regression coefficient $\beta = (\beta_1, \beta_2, \cdots, \beta_p)'$, $\beta_1 = 1$, $\tilde{\beta} = (\beta_2, \cdots, \beta_p)'$ is contained in a compact subset $\tilde{B} \in R^{p-1}$.

(Ⅵ) Q is negative definite.

Assumptions (Ⅲ) and (K1)~(K4) insure the existence of A, D, Q as well as the convergence of certain sequences of the integrals that arise in the proof of asymptotic normality. (Ⅳ) are analogous to assumptions made in kernel density estimation. $K'(\cdot)$ is the kernel function, σ_n is the bandwidth in the kernel estimation. (Ⅴ) is the standard in asymptotic distribution theory for Taylor series expansion.

In the text bellow, c is constant and may vary from place to place, and may be different in even one formula.

In this article, notation such as $\mathscr{L}^*$, P^*, E^*, Var^*, F_n^* refer to probability calculation under the condition that $(x_1, y_1), \cdots, (x_n, y_n)$ are given.

Theorem 1 (consistency) Under the assumption (Ⅰ), the random weighting maximum smoothed score estimator $\tilde{\beta}_n^w$ defined as the maximizer of (4) converges almost surely to the true parameter $\tilde{\beta}$.

Theorem 2 (asymptotic normality) Assume that under model (1), Conditions (Ⅰ)~(Ⅵ) and (K1)~(K4) are held with $\tau=2$ in the conditions imposed on the weights (5), and $\tilde{\beta}_n, \tilde{\beta}_n^w$ are defined as the maximizer of (3) and (4), respectively. When $n\sigma_n^{2h+1} \to \lambda$, as $n \to \infty$, for $h \geqslant 2$, we have

$$\sup_t | P^*(\sqrt{n\sigma_n}(\tilde{\beta}_n^w - \tilde{\beta}_n) \leqslant t) - P(\sqrt{n\sigma_n}(\tilde{\beta}_n - \tilde{\beta}) \leqslant t) | \xrightarrow{P} 0, \text{ as } n \to \infty.$$

2 Simulation studies

This section presents results of simulation studies. To compute $\tilde{\beta}_n^w$ or $\tilde{\beta}_n$ in application, it is necessary to assign a numerical value to the bandwidth parameter σ_n. Since $\tilde{\beta}_n^w$ or $\tilde{\beta}_n$ can be quite sensitive to the choice of σ_n, it is necessary to have a good choice for the bandwidth. The bandwidth selection is common in nonparametric and semiparametric estimation, however, there is no satisfactory solution. Horowitz[1] discussed the problem of bandwidth choice. It is a data-driven method which is optimal in the sense of minimizing the mean square error (MSE). In this article we also choose the optimal bandwidth which minimizes the mean square error of $\widetilde{\beta}_n$ (or $\widetilde{\beta}_n^w$). Unlike Horowitz's method, we directly search for the optimal bandwidth over a discrete set of [0.1, 0.9] which contains all the possible values of σ_n in our simulation.

The model we used to generate data is the same as that used by Horowitz[1]

$$y = \begin{cases} 1, & \text{if } x_1 + \beta x_2 + e \geqslant 0; \\ 0, & \text{otherwise.} \end{cases}$$

We use the same parameter settings as Ref. [1] in order to compare the behavior of maximum random weighting smoothed estimates with the maximum smoothed estimates. The true value of β is 1, $x_1 \sim N(0, 1)$, and $x_2 \sim N(1, 1)$. There are four different distributions of e, that is, distribution L: $e \sim$ logistic with median 0 and variance 1; distribution U: $e \sim$ uniform with median 0 and variance 1; distribution T3: $e \sim$ student's t distribution with 3 degrees of freedom normalized to have variance 1 and distribution H: $e = 0.25(1 + 2z^2 + z^4)v$, where $z = x_1 + x_2$, v is a logistic random variable with mean 0 and variance 1. With distribution H, e is heteroscedastic.

We used two smoothing functions in our simulations. The first one is $K_2(x) = \Phi(x)$ corresponding to $h = 2$ where $\Phi(x)$ is the cumulative standard normal distribution function. The second smooth function is defined as follows ($h=4$):

$$K_4(x) = \begin{cases} 0 & \text{if } x < -5, \\ \frac{1}{2} + \frac{105}{64}\left\{\frac{x}{5} - \frac{5}{3}\left(\frac{x}{5}\right)^3 + \frac{7}{5}\left(\frac{x}{5}\right)^5 - \frac{3}{7}\left(\frac{x}{5}\right)^7\right\} & \\ \quad \text{if } -5 \leqslant x \leqslant 5, & \\ 1 & \text{if } x > -5. \end{cases}$$

The weighting variables are generated from the exponential distribution (i.e. $w_1, w_2, \cdots, w_n$ iid$\sim$exp(1)) and Poisson distribution (i.e. $w_1, w_2, \cdots, w_n$ iid$\sim$Poisson(1)). The sample size are $n=250$, 500 and 1 000 in the simulation. There are 1 000 replications per experiment.

Both the smoothed score function and the random weighted smoothed score function have many local maximizers, a global optimization algorithm is necessary to compute β_n and β_n^w. Although such methods, for example, tunneling and simulated annealing, are available, we search for the optima over a discrete set of β values, which is convenient in our simulation as β is one-dimensional.

Tabs. 1 and 2 report the mean, mean square error (MSE) of MSSE and WMSSE estimators and the optimal bandwidths under different scenarios. From Tab. 1, we can see that all estimators are consistent and robust with respective to different error distributions. The bias of WMSSE is larger than MSSE when the sample size is 250. However, the difference becomes negligible when sample size becomes larger. The mean square error of WMSSE is always larger than MSSE. It is within our expectation as the introduction of random weights brings more variation when estimating model parameters. Similar properties are observed with the kernel function $K_4(x)$ (see Tabs. 3 and 4 for details). In conclusion, estimators from both smoothed score function and random weighted smoothed score function are good choices in the binary response model.

Figs. 1 and 2 are the Q-Q plots of $\sqrt{n\sigma_n}\beta_n$ and $\sqrt{n\sigma_n}\beta_n^w$ with sample size $n=250$, $n=1\ 000$ and Poisson weights. σ_n is the optimal bandwidth selected according to the minimum mean square error criteria. We also have the Q-Q plots for $n=500$ and exponential weights under different error distributions. These figures are not shown here as they have similar patterns. Although distribution approximation of MSSE by WMSSE is not very satisfactory when sample size is small, the Q-Q plots become more diagonal when the sample size becomes larger, indicating that the random weighting method is desirable in distribution approximation for moderately large sample size.

Tab. 1 Monte Carlo simulation results of MSSE and WMSSE for $h=2$, smoothed function $K(x)=K_2(x)$, exponential weight, sample size $n=250, 500, 1\ 000$, error distributions L, U, T3 and H

sample size	error distribution	MSSE			WMSSE		
		Bias	MSE	bandwidth	Bias	MSE	bandwidth
250	L	0.096 2	0.099 5	0.15	0.153 1	0.278 0	0.12
	U	0.199 4	0.216 3	0.28	0.200 7	0.488 4	0.12
	T3	0.081 0	0.039 3	0.23	0.105 6	0.091 5	0.21
	H	0.026 4	0.010 6	0.13	0.044 9	0.032 6	0.14
500	L	0.093 0	0.039 8	0.27	0.074 0	0.081 2	0.18
	U	0.108 0	0.085 7	0.25	0.156 1	0.282 4	0.14
	T3	0.052 9	0.018 4	0.20	0.082 9	0.044 4	0.22
	H	0.020 2	0.004 3	0.14	0.027 5	0.009 7	0.14
1 000	L	0.052 8	0.017 6	0.19	0.072 9	0.033 8	0.23
	U	0.088 8	0.038 9	0.24	0.082 8	0.083 3	0.18
	T3	0.044 6	0.009 6	0.19	0.057 8	0.021 4	0.20
	H	0.015 2	0.002 2	0.11	0.019 4	0.004 3	0.12

Tab. 2 Monte Carlo simulation results of MSSE and WMSSE for $h=2$, smoothed function $K(x)=K_2(x)$, Poisson weight, sample size $n=250, 500, 1\ 000$, error distributions L, U, T3 and H

sample size	error distribution	MSSE			WMSSE		
		Bias	MSE	bandwidth	Bias	MSE	bandwidth
250	L	0.081 9	0.082 7	0.17	0.155 3	0.246 2	0.14
	U	0.189 4	0.305 4	0.23	0.261 0	0.664 4	0.10
	T3	0.078 4	0.045 2	0.23	0.110 0	0.131 5	0.19
	H	0.028 3	0.010 3	0.12	0.050 8	0.059 3	0.11
500	L	0.067 3	0.038 6	0.20	0.100 3	0.094 4	0.21
	U	0.159 5	0.094 7	0.32	0.163 1	0.228 9	0.23
	T3	0.052 8	0.018 7	0.20	0.064 7	0.043 7	0.19
	H	0.018 2	0.004 3	0.13	0.029 9	0.009 7	0.14
1 000	L	0.051 7	0.015 7	0.20	0.081 8	0.036 0	0.23
	U	0.072 4	0.034 7	0.22	0.094 0	0.087 0	0.20
	T3	0.038 7	0.008 4	0.19	0.058 0	0.018 7	0.22
	H	0.013 0	0.002 0	0.11	0.021 5	0.004 4	0.13

Tab. 3 Monte Carlo simulation results of MSSE and WMSSE for $h=4$, smoothed function $K(x)=K_4(x)$, exponential weight, sample size $n=250, 500, 1\ 000$, error distributions L, U, T3 and H

sample size	error distribution	MSSE			WMSSE		
		Bias	MSE	bandwidth	Bias	MSE	bandwidth
250	L	0.088 6	0.048 0	0.33	0.130 1	0.194 6	0.27
	U	0.169 2	0.158 6	0.47	0.284 4	0.526 0	0.47
	T3	0.077 5	0.033 6	0.30	0.119 3	0.087 2	0.32
	H	0.017 5	0.008 0	0.14	0.029 8	0.028 9	0.13
500	L	0.068 1	0.024 7	0.32	0.081 2	0.055 3	0.32
	U	0.120 1	0.047 8	0.45	0.202 5	0.193 3	0.48
	T3	0.040 9	0.013 5	0.25	0.054 5	0.034 8	0.26
	H	0.011 7	0.003 4	0.13	0.013 7	0.007 0	0.15
1 000	L	0.037 5	0.010 4	0.27	0.050 8	0.023 3	0.28
	U	0.066 6	0.020 6	0.40	0.079 6	0.042 8	0.40
	T3	0.032 4	0.007 0	0.24	0.046 6	0.015 0	0.26
	H	0.010 7	0.001 9	0.13	0.012 9	0.003 5	0.14

Tab. 4 Monte Carlo simulation results of MSSE and WMSSE for $h=4$, smoothed function $K(x)=K_4(x)$, Poisson weight, sample size $n=250$, 500, 1 000, error distributions L, U, T3 and H

sample size	error distribution	MSSE			WMSSE		
		Bias	MSE	bandwidth	Bias	MSE	bandwidth
250	L	0.075 2	0.054 1	0.31	0.117 3	0.157 5	0.29
	U	0.139 6	0.167 2	0.41	0.242 9	0.477 2	0.41
	T3	0.059 5	0.030 7	0.27	0.071 7	0.090 8	0.24
	H	0.024 8	0.008 6	0.15	0.041 7	0.024 8	0.16
500	L	0.067 4	0.025 1	0.32	0.079 2	0.072 0	0.30
	U	0.138 1	0.051 1	0.47	0.151 1	0.146 7	0.39
	T3	0.037 0	0.013 5	0.23	0.047 0	0.032 7	0.24
	H	0.017 5	0.004 0	0.13	0.017 4	0.008 7	0.10
1 000	L	0.035 2	0.010 9	0.27	0.084 1	0.027 0	0.35
	U	0.063 1	0.023 5	0.36	0.092 9	0.052 9	0.40
	T3	0.024 9	0.006 7	0.21	0.046 2	0.014 5	0.25
	H	0.010 0	0.001 9	0.12	0.014 7	0.003 9	0.13

3 Conclusion

This paper has described a random weighting version of Horowitz's smoothed score estimator for the binary response model. The smoothed estimator converges to a normal distribution with the variance parameter related to an unknown nonparametric density function. It is thus difficult to carry out statistical inference based on such limit distribution. The randomly weighted version of the estimators can be used to estimate the asymptotic variance of the original maximum smoothed score estimators. We proved that the conditional distribution of the random weighed maximum smoothed score estimator given data could accurately approximate the distribution of Horowitz's smoothed score estimator whether it is normal or not.

A popularly used method for approximating distribution, bootstrap could be viewed as the special case of random weighting method by setting the weights to be distributed as multinomial distribution, i. e $(w_1, \cdots, w_n) \sim$ multinomial$(n, (1/n, 1/n, \cdots, 1/n))$. Random weighting method has advantages over the bootstrap method in the sense that it is more flexible to choose weights. Simulations have indicated that the approximations of asymptotic distribution are likely to be accurate with moderate large samples. In conclusion, the random weighting method provides a useful approximation to the maximum smoothed score estimator in the binary response model and makes statistical inference possible.

Appendix

This appendix presents the proof of three main theorems and several lemmas. These lemmas establish properties of the functions $S_n(b)$, $S_n^*(b)$ and $S_n^w(b)$ that are used in proving theorems. For $b \in R^p$, define,

$$S(b) = 2P(y=1, b'x \geqslant 0) - P(b'x > 0).$$

Define sets B by $B \equiv \{1\} \times \widetilde{B}$ and B^* by $B^* \equiv \{1\} \times R^{p-1}$.

Results from the empirical process are needed to prove the uniform convergency summarized in Lemma A. 1 and Lemma A. 2 (Theorem 4. 8 and Theorem 8. 3 in Ref. [13].

Lemma A. 1 Let $\mathscr{F}$ be a bounded subset of R^n with envelop F and pseudo-dimension at most V. Then there exist constants A and W, depending only on V, such that $D_1(\varepsilon \mid \alpha \odot F \mid, \alpha \odot \mathscr{F}) \leqslant A\left(\frac{1}{\varepsilon}\right)^W$, for $0 < \varepsilon \leqslant 1$, for every rescaling vector α of non-negative constants.

D_1 is the package number under L_1 distance, the definitions of package number and pseudo-

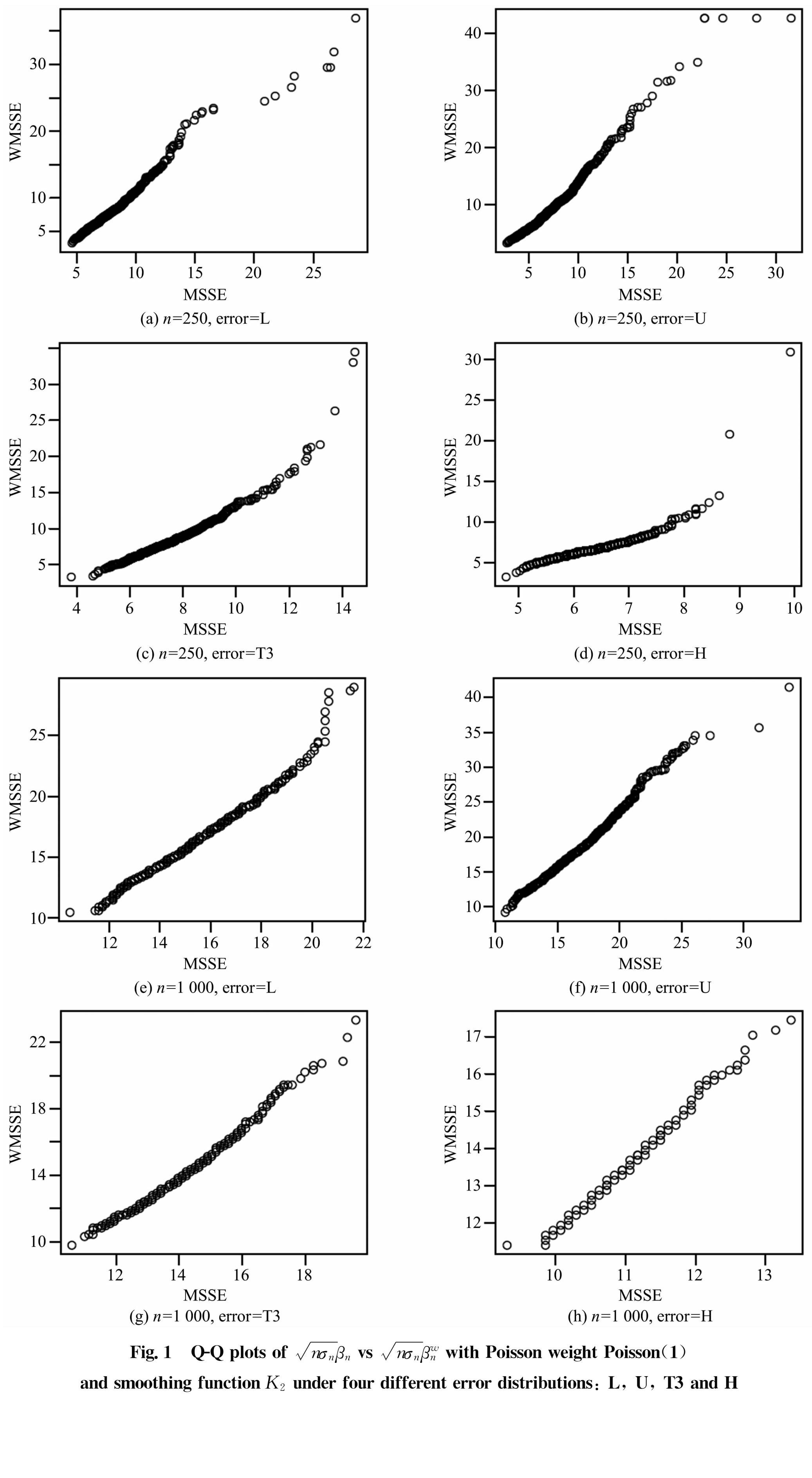

Fig. 1 Q-Q plots of $\sqrt{n\sigma_n}\beta_n$ vs $\sqrt{n\sigma_n}\beta_n^w$ with Poisson weight Poisson(1) and smoothing function K_2 under four different error distributions: L, U, T3 and H

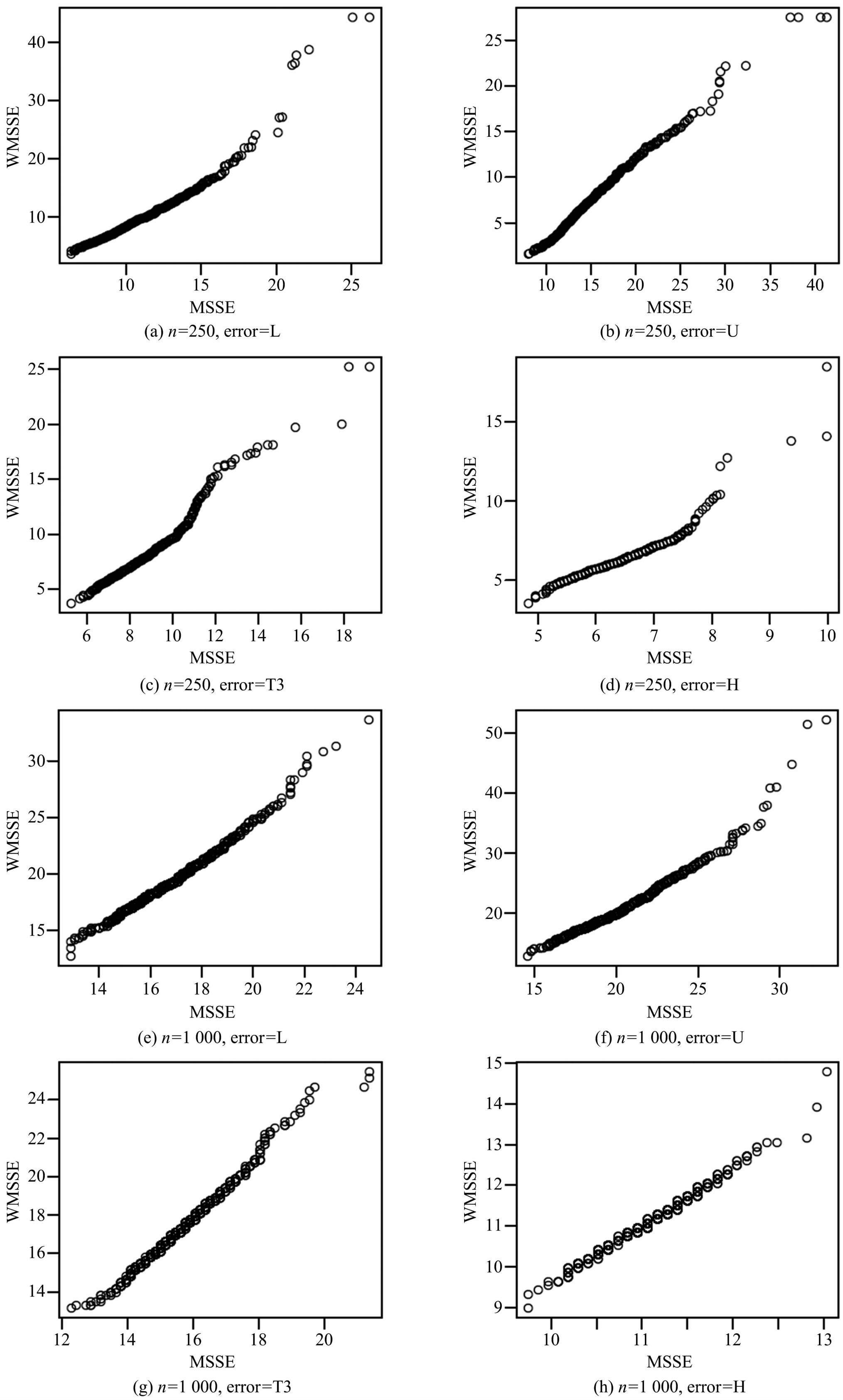

Fig. 2　Q-Q plots of $\sqrt{n\sigma_n}\beta_n$ vs $\sqrt{n\sigma_n}\beta_n^w$ with Poisson weight Poisson(1)
and smoothing function K_4 under four different error distributions：L，U，T3 and H

dimension refer to Ref.[13, Definitions (3.3) and (4.3)]. $\odot$ is inner product operation.

Lemma A. 2 $\{f_i(\omega, t): t \in T\}$ are independent, manageable series with envelopes $F_i(\omega)$, if $\sum_{i=1}^{n} \frac{EF_i^2}{i^2} < \infty$, then as $n \to \infty$, we have

$$\frac{1}{n} \sup_{t \in T} | S_n(\omega, t) - M_n(t) | \xrightarrow{a.s.} 0 \quad (A1)$$

where $S_n(\omega, t) = \sum_{i=1}^{n} f_i(\omega, t)$, $M_n(t)$ represents the expectation of $S_n(\omega, t)$.

Definition of "manageable random functions" can be found in Ref.[13, Definition (7.9)].

Lemma A. 3 Under assumption (Ⅰ)~(Ⅳ), (Ⅵ) and (K1)~(K4),

$$\lim_{n \to \infty} Q_n^w(\beta_n^w, \sigma_n) = Q \text{ in prob.}.$$

Proof It is easy to show that $EQ_n^w = EQ_n$ and $\mathrm{Var}Q_n^w = (\tau - 1)\mathrm{Var}Q_n$. These two equations together with Ref. [1, Lemma 9] prove Lemma A. 3. □

Proof of Theorem 1 Ref.[3, Lemma 3 and Lemma 5] proved $S(b)$ is continuous and achieved its unique maxima at true value β. Thus, we only need to prove $|S_n^w(b) - S(b)| \to 0$ uniformly hold for $b \in B$ as $n \to \infty$. Denote $M(b) = E(S_n^w(b))$, we prove $\lim_{v \to \infty} \sup_{b \in B^*} | S_n^w(b) - M(b) | = 0$ and $\lim_{v \to \infty} \sup_{b \in B^*} |S(b) - M(b)| = 0$ respectively.

First of all, for some $\alpha > 0$,

$$| S(b) - M(b) | = \left| E(2y_i - 1)\left(K\left(\frac{b'x_i}{\sigma_n}\right) - I(b'x_i \geqslant 0)\right)\right| \leqslant$$
$$E\left|K\left(\frac{b'x_i}{\sigma_n}\right) - I(b'x_i \geqslant 0)\right| \leqslant$$
$$E\left|K\left(\frac{b'x_i}{\sigma_n}\right) - I(b'x_i \geqslant 0)\right| I(|b'x_i| \geqslant \alpha) +$$
$$E\left|K\left(\frac{b'x_i}{\sigma_n}\right) - I(b'x_i \geqslant 0)\right| I(|b'x_i| < \alpha).$$

Since $\lim_{v \to \infty} K(v) = 0$, $\lim_{v \to \infty} K(v) = 1$, therefore, for any given $\varepsilon > 0$, there exists a constant A, when $|v| > A$, $|I(v \geqslant 0) - K(v)| < \varepsilon$. Then, when $\alpha/\sigma_n > A$,

$$| K(b'x_i/\sigma_n) - I(b'x_i \geqslant 0) | I(|b'x_i| \geqslant \alpha) \leqslant$$
$$| K(b'x_i/\sigma_n) - I(b'x_i \geqslant 0) | \leqslant \varepsilon$$

uniformly hold over $b \in B^*$. By the dominated convergence theorem,

$$\lim_{v \to \infty} \sup_{b \in B^*} E\left|K\left(\frac{b'x_i}{\sigma_n}\right) - I(b'x_i \geqslant 0)\right| \cdot I(|b'x_i| \geqslant \alpha) = 0 \quad (A2)$$

The bounding property of function $K(\cdot)$ implies that there exists a constant c such that

$$| I(b'x_i \geqslant 0) - K(b'x_i/\sigma_n) | \leqslant c,$$

therefore, we have

$$E\left|K\left(\frac{b'x_i}{\sigma_n}\right) - I(b'x_i \geqslant 0)\right| I(|b'x_i| < \alpha) \leqslant$$
$$cE\, I(|b'x_i| < \alpha) = cP(|b'x_i| < \alpha).$$

Ref.[1, Lemma 4] proved that $P(|b'x_i| < \alpha)$ converges to 0 uniformly over B^* as α goes to 0. Thus,

$$\lim_{v \to \infty} \sup_{b \in B^*} E\left|K\left(\frac{b'x_i}{\sigma_n}\right) - I(b'x_i \geqslant 0)\right| \cdot I(|b'x_i| \leqslant \alpha) = 0 \text{ as } \alpha \to 0 \quad (A3)$$

(A2) and (A3) imply $\lim_{v \to \infty} \sup_{b \in B^*} |S(b) - M(b)| = 0$.

Now we use Lemma A. 2 to prove $\lim_{v \to \infty} \sup_{b \in B^*} |S_n^w(b) - M(b)| = 0$. Taking $f_i(\omega, b) = w_i K(b'x_i/\sigma_n)$ in Lemma A. 2 we obtain $\{f_i\}$ with envelop $\{F_i\}$, $F_i = c|w_i|$, and furthermore $f_i(\omega, b)$ are independent with respect to their envelops, manageable and

$$\sum_{i=1}^{\infty} \frac{EF_i^2}{i^2} = c^2\tau^2 \sum_{i=1}^{\infty} \frac{1}{i^2} < \infty$$

So, by applying Lemma A. 2 we obtain

$$\lim_{n \to \infty} \sup_{b \in B^*} | S_n^w(b) - M(b) | = 0.$$

This completes the proof. □

Proof of Theorem 2 The Taylor expansion of $U_n^w(\beta, \sigma_n)$ at β_n^w yield

$$\sqrt{n\sigma_n} U_n^w(\beta, \sigma_n) + Q_n^w(b_n^*, \sigma_n)\sqrt{n\sigma_n}(\tilde{\beta}_n^w - \tilde{\beta}) = 0,$$

where b_n^* is between β and β_n^w. By applying Lemma A. 3 we have

$$\sqrt{n\sigma_n}(\tilde{\beta}_n^w - \tilde{\beta}) = -Q^{-1}\sqrt{n\sigma_n} U_n^w(\beta, \sigma_n) + o_p(1) =$$
$$-Q^{-1} \frac{1}{\sqrt{n\sigma_n}} \sum_{i=1}^{n} (w_i - 1)(2y_i - 1)\tilde{x}_i K'\left(\frac{\beta' x_i}{\sigma_n}\right) + o_p(1),$$

Let

$$z_n = \frac{1}{\sqrt{n\sigma_n}} \sum_{i=1}^{n} (w_i - 1)(2y_i - 1)\tilde{x}_i K'\left(\frac{\beta' x_i}{\sigma_n}\right),$$

then

$$\left.\begin{array}{l}\sqrt{n\sigma_n}(\tilde{\beta}_n^w-\tilde{\beta}_n)=-Q^{-1}(z_n+\xi_n),\\ E^*\|\xi_n^2\wedge 1\|\to 0 \text{ in pr. as } n\to\infty\end{array}\right\}\quad (A4)$$

$$E^*z_nz'_n=E^*\frac{1}{n\sigma_n}\sum_{i=1}^n(w_i-1)^2(2y_i-1)^2\cdot\left(K'\left(\frac{\beta'x_i}{\sigma_n}\right)^2\tilde{x}_i\tilde{x}'_i\right)=\frac{1}{n\sigma_n}\sum_{i=1}^n\tilde{x}_i\tilde{x}'_i\left(K'\left(\frac{\beta'x_i}{\sigma_n}\right)\right)^2.$$

Since $E\frac{1}{\sigma_n}\tilde{x}\tilde{x}'\left(K'\left(\frac{\beta'x}{\sigma_n}\right)\right)^2\to D$ exists, so when $n\to\infty$,

$$Ez_nz'_n=E\frac{1}{\sigma_n}\tilde{x}\tilde{x}'\left(K'\left(\frac{\beta'x}{\sigma_n}\right)\right)^2\xrightarrow{\text{a.s.}}D\quad (A5)$$

Let U_0 be a be denumerable dense subset in the unit sphere, $U=\{\gamma\in R^{p-1}:\|\gamma\|=1\}$. It is necessary to prove

$$P^*(\gamma'_0Z_n\leqslant v)\to\Phi\left(\frac{v}{\gamma'_0D\gamma_0}\right)\text{ in pr. as } n\to\infty\quad (A6)$$

where $\gamma_0\in U_0$, $v\in R^1$, Φ is the standard normal distribution.

Let

$$\eta_{ni}=\frac{1}{\sqrt{n\sigma_n}}(w_i-1)(2y_i-1)\gamma'\tilde{x}_iK'\left(\frac{\beta'x_i}{\sigma_n}\right),$$

then

$$\gamma'z_n=\sum_{i=1}^n\eta_{ni},\ E^*\eta_{ni}=0,$$

$$E^*\eta_{ni}^2=\frac{1}{n\sigma_n}K'\left(\frac{\beta'x_i}{\sigma_n}\right)^2\gamma'\tilde{x}_i\tilde{x}'_i\gamma\triangleq\delta_i^2.$$

Let $B_n=\sum_{i=1}^n\delta_i^2$, we need to verify the Lindeberg condition

$$\frac{1}{B_n}\sum_{i=1}^n\int_{|\eta_{ni}|\geqslant\varepsilon\sqrt{B_n}}\frac{1}{n\sigma_n}(w_i-1)^2(2y_i-1)^2\cdot\left(K'\left(\frac{\beta'x_i}{\sigma_n}\right)\right)^2\gamma'\tilde{x}_i\tilde{x}'_i\gamma\,\mathrm{d}F_i^*(w_i)\to 0\text{ in pr.}\quad (A7)$$

for any given $\varepsilon>0$.

That is to say

$$L_n(\varepsilon):=\frac{1}{B_n}\sum_{i=1}^nE^*\eta_{ni}^2I(|\eta_{ni}|\geqslant\varepsilon\sqrt{B_n})\xrightarrow{P.}0\quad (A8)$$

Firstly,

$$B_n\xrightarrow{p.}\gamma'D\gamma\text{ as } n\to\infty\quad (A9)$$

It is because

$$E(B_n-EB_n)^2=\frac{1}{n}\mathrm{Var}\left[\frac{1}{\sigma_n}\left(K'\left(\frac{\beta'x_1}{\sigma_n}\right)\right)^2\gamma'\tilde{x}_1\tilde{x}'_1\gamma\right]\leqslant E\left[\frac{1}{\sigma_n^2}\left(K'\left(\frac{\beta'x_1}{\sigma_n}\right)\right)^4(\gamma'\tilde{x}_1\tilde{x}'_1\gamma)^2\right]=$$

$$\frac{1}{n}E\left[E\left(\frac{1}{\sigma_n^2}\left(K'\left(\frac{\beta'x_1}{\sigma_n}\right)\right)^4(\gamma'\tilde{x}_1\tilde{x}'_1\gamma)^2\mid\tilde{x}_1\right)\right]=\frac{1}{n}E\int_{-\infty}^{\infty}\frac{1}{\sigma_n^2}\left(K'\left(\frac{z}{\sigma_n}\right)\right)^4p(z\mid\tilde{x}_1)\mathrm{d}z(\gamma'\tilde{x}_1\tilde{x}'_1\gamma)^2\overset{z/\sigma_n=u}{=\!=\!=}$$

$$\frac{1}{n\sigma_n}E\int_{-\infty}^{\infty}(K'(u))^4p(\sigma_nu\mid\tilde{x}_1)\mathrm{d}u(\gamma'\tilde{x}_1\tilde{x}'_1\gamma)^2.$$

By dominated convergence theorem, as $n\to\infty$,

$$E\int_{-\infty}^{\infty}(K'(u))^4p(\sigma_nu\mid\tilde{x}_1)\mathrm{d}u\,(\gamma'\tilde{x}_1\tilde{x}'_1\gamma)^2\to E\int(K'(u))4p(0\mid\tilde{x}_1)\mathrm{d}u(\gamma'\tilde{x}_1\tilde{x}'_1\gamma)^2=\int_{-\infty}^{\infty}(K'(u))^4\mathrm{d}uEp(0\mid\tilde{x}_1)(\gamma'\tilde{x}_1\tilde{x}'_1\gamma)^2.$$

(Ⅱ) and (K2) imply the right side above is bounded and noticing that $n\sigma_n\to\infty$, so

$$E(B_n-E(B_n))^2\to 0,$$

as $n\to\infty$. Similarly,

$$EB_n=\frac{1}{\sigma_n}E\left(K'\left(\frac{\beta'x_1}{\sigma_n}\right)\right)^2\gamma'\tilde{x}_1\tilde{x}'_1\gamma=\frac{1}{\sigma_n}E\left[E\left[\left(K'\left(\frac{\beta'x_1}{\sigma_n}\right)\right)^2\mid\tilde{x}_1\right]\gamma'\tilde{x}_1\tilde{x}'_1\gamma\right]=$$

$$\frac{1}{\sigma_n}E\int\left(K'\left(\frac{z}{\sigma_n}\right)\right)^2p(z\mid\tilde{x}_1)\mathrm{d}z\gamma'\tilde{x}_1\tilde{x}'_1\gamma\overset{z/\sigma_n=u}{=\!=\!=}E\int(K'(u))^2p(\sigma_nu\mid\tilde{x}_1)\mathrm{d}u\gamma'\tilde{x}_1\tilde{x}'_1\gamma.$$

By dominated convergence theorem, when $n\to\infty$,

$$EB_n\to\int(K'(u)^2)E[p(0\mid\tilde{x}_1)\gamma'\tilde{x}_1\tilde{x}'_1\gamma]=\gamma'D\gamma\quad (A10)$$

$E(B_n-EB_n)^2\to 0$ and (A10) imply (A9) as $n\to\infty$.

By (A9), in order to prove (A8) it is necessary to prove

$$E\left[\frac{1}{\sigma_n}(w_1-1)^2\left(K'\left(\frac{\beta'x_1}{\sigma_n}\right)\right)^2\gamma'\tilde{x}_1\tilde{x}'_1\gamma\,I\cdot\left(\left|\frac{1}{n\sigma_n}(w_1-1)^2\left(K'\left(\frac{\beta'x_1}{\sigma_n}\right)\right)^2\gamma'\tilde{x}_1\tilde{x}'_1\gamma\right|\geqslant\varepsilon^2\right)\to 0\text{, in pr.}$$

Let

$$T_n = \frac{1}{n\sigma_n}\left(K'\left(\frac{\beta' x_1}{\sigma_n}\right)\right)^2 \gamma' \tilde{x}_1 \tilde{x}'_1 \gamma.$$

Since

$$E\left[\frac{1}{n\sigma_n}\left(K'\left(\frac{\beta' x_1}{\sigma_n}\right)\right)^2 \gamma' \tilde{x}_1 \tilde{x}'_1 \gamma\right] =$$

$$E\left[E\frac{1}{n\sigma_n}\left(K'\left(\frac{\beta' x_1}{\sigma_n}\right)\right)^2 \gamma' \tilde{x}_1 \tilde{x}'_1 \gamma \mid \tilde{x}_1\right] =$$

$$\frac{1}{n\sigma_n}E\left(K'\left(\frac{z}{\sigma_n}\right)\right)^2 p(z \mid \tilde{x}_1)\mathrm{d}z\gamma' \tilde{x}_1 \tilde{X}'_1 \gamma \xlongequal{z/\sigma_n = u}$$

$$\frac{1}{n}E\int (K'(u))^2 p(\sigma_n u \mid \tilde{x}_1)\mathrm{d}u\gamma' \tilde{x}_1 \tilde{x}'_1 \gamma.$$

By dominated convergence theorem, as $n \to \infty$, we have

$$E\int (K'(u))^2 p(\sigma_n u \mid \tilde{x}_1)\mathrm{d}u\gamma' \tilde{x}_1 \tilde{x}'_1 \gamma \to \gamma' D\gamma,$$

thus $ET_n \to 0$ as $n \to \infty$.

Therefore,

$$T_n \to 0, \text{ in pr. as } n \to \infty \tag{A11}$$

Besides,

$$E\left[\frac{1}{\sigma_n}(w_1 - 1)^2\left(K'\left(\frac{\beta' x_1}{\sigma_n}\right)\right)^2 \gamma' \tilde{x}_1 \tilde{x}'_1 \gamma\right] =$$

$$EB_n \to \gamma' D\gamma \tag{A12}$$

by (A11) and (A12), the Lindeberg condition (A7) is established. This completes the proof. □

References

[1] Horowitz J L. A smoothed maximum score estimator for the binary response model[J]. Econometrica, 1992, 60: 505-531.

[2] Manski C. Semiparametric analysis of binary response from respnonse-based samples [J]. Journal of Econometrics, 1986, 31: 31-40.

[3] Manski C. Semiparametric analysis of discrete response [J]. Journal of Econometrics, 1985, 27: 313-333.

[4] Manski C. Identification in binary response model[J]. Journal of the American Statistical Association, 1988, 83:403, 729-738.

[5] Matzkin R L. Nonparametric identification and estimation of poly-chotomous choice model[J]. Journal of Econometrics, 1993, 58: 137-168.

[6] Klein R W, Spady R H. An efficient semiparametric estimator for binary response model [J]. Econometricas, 1993, 61(2): 387-421.

[7] Manski C. Maximum score estimation of the stochastic model of choice[J]. Journal of Econometrics, 1975, 3: 205-228.

[8] Vapnik V N, Chervonenkis A Y. On the uniform convergence of relative frequencies of events to their probabilities [J]. Theory of Probability and Its Applications, 1971, 16: 264-280.

[9] Dudley R M. Universal Donsker classes and metric entropy[J]. The Annals of Probability, 1987, 15: 1 306-1 326.

[10] Wenocur R S, Dudley R M. Some special vapnik-chervonenkis classes[J]. Discrete Mathematics, 1981, 33: 313-318.

[11] Kim J, Pollard D. Cube root asymptotics[J]. Ann Statist, 1990, 18: 191-219.

[12] Rao C R, Zhao L C. Approximation to the distribution of M-estimates in linear models by randomly weighted bootstrap [J]. Sankhya: The Indian Journal of Statistics, 1992, 54: 323-331.

[13] Pollard D. Empirical processes: Theory and applications [C]// NSF-CBMS Regional Conference Series in Probability and Statistics, Vol. 2. Hayward, CA: Institute of Mathematical Statistics, 1990.

[14] Horowitz J L. Semiparametic Methods in Econometrics [M]. New York: Springer, 1998.

[15] Rao C R. Approximation to the distribution of M-estimates in linear models by randomly weighted bootstrap [J]. Sankhya: The Indian Journal of Statistics, 1992, 54: 323-331.

[16] Freedman D A. Bootstrapping regression models[J]. Ann Statist, 1981, 9: 1 218-1 228.

[17] Rubin D B. The Bayesian bootstrap[J]. Ann Statist, 1981, 9:130-134.

[18] Stout W F. Almost Sure Convergence [M]. New York: Academic Press, 1974.

[19] Zheng Z G. Random weighting method[J]. Acta Math Appl Sinica, 1987, 10: 247-253. (in Chinese)

[20] Rao R R. Relations between weak and uniform convergence of measures with applications[J]. Annals of Mathematical Statistics, 1962, 33: 659-680.

[21] Ameniya T. Advanced Economitrics[M]. Cambridge: Harvard University Press, 1985.

[22] vander Vaart A W. Asymptotic Statistics [M]. Cambridge: Cambridge University Press, 1998.

[23] Shao Jun, Tu Dongsheng. The Jackknife and Bootstrap [M]. New York: Springer-Verlag, 1995.