Ann. Inst. Statist. Math.
Vol. 56, No. 1, 49–72 (2004)

NONPARAMETRIC REGRESSION WITH CURRENT STATUS DATA

TOSHIO HONDA

Institute of Social Sciences, University of Tsukuba, 1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8571, Japan, e-mail: honda@social.tsukuba.ac.jp

(Received January 10, 2002; revised February 7, 2003)

Abstract. We apply nonparametric regression to current status data, which often arises in survival analysis and reliability analysis. While no parametric assumption on the distributions has been imposed, most authors have employed parametric models like linear models to measure the covariate effects on failure times in regression analysis with current status data. We construct a nonparametric estimator of the regression function by modifying the maximum rank correlation (MRC) estimator. Our estimator can deal with the cases where the other estimators do not work. We present the asymptotic bias and the asymptotic distribution of the estimator by adapting a result on equicontinuity of degenerate U-processes to the setup of this paper.

Key words and phrases: Current status data, case 1 interval censoring, survival analysis, nonparametric regression, maximum rank correlation estimator, local linear regression, asymptotic properties, degenerate U-processes, VC-subgragh class, equicontinuity.

1. Introduction

Current status (or case 1 interval censored) data often arises in survival analysis and reliability analysis where we sometimes cannot observe a failure time directly and we only know whether or not it is in an interval given by an examination time.

Suppose that we observe (Y, C, X) and we are interested in $g(\cdot)$, where (Y, C, X) is defined by

$$\tilde{Y} = g(X) + \epsilon \quad \text{and} \quad Y = I\{\tilde{Y} \le C\}, \tag{1.1}$$

where $\tilde{Y}$ is an unobserved failure time, C is a random examination time, X is a random covariate, ϵ is a random error, and $I\{\}$ stands for an indicator function. We can also deal with monotone-transformed failure times, for example,

$$\log \tilde{Y} = g(X) + \epsilon \quad \text{and} \quad Y = I\{\log \tilde{Y} \le \log C\}. \tag{1.2}$$

No parametric assumption on the distributions is imposed on X, C, and ϵ. However, we assume that (C, X) and ϵ are independent as in assumption A1 below.

So far parametric forms have been assumed for $g(\cdot)$ in the literature of statistics, for example, a linear function. In this paper we do not specify any parametric forms for $g(\cdot)$ and estimate $g(x) - g(x_0)$ by assuming that the covariate X has a pivotal point or standard point x_0. Plotting the estimates of $g(x) - g(x_0)$ for various values of x with x_0 fixed will help us specify a parametric form of $g(\cdot)$ or inspect the goodness of fit of a parametric form of $g(\cdot)$. We state the reasons why we estimate $g(x) - g(x_0)$ later in this section. Hereafter we write 0 for the assumed pivotal point or standard point x_0.

Current status data can be thought of as binary response data in econometrics and a lot of work has been done in econometrics and statistics because of the importance.

In econometrics, a lot of authors have studied semiparametric binary choice models. They have considered parametric regression functions except for Horowitz (2001).

Han (1987) proposed the maximum rank correlation (MRC) estimator for linear models and proved its consistency. Sherman (1993) showed the asymptotic normality of the MRC estimator for linear models by using a result on equicontinuity of degenerate U-processes. According to Sherman (1993), the MRC estimator for linear models is not efficient. We construct our estimator from the MRC estimator. Cosslett (1987) calculated efficiency bounds for linear models. He showed that when $g(\cdot)$ is a linear function with a constant term, the information of the constant term vanishes under the assumption of zero-median error. But he also showed the information is positive when the error term is known to have a symmetric distribution. As for efficient estimators, Klein and Spady (1993) ingeniously constructed an efficient estimator for parametric regression functions in a general setup. Their estimator requires auxiliary bandwidth selection and trimming.

The estimator in Horowitz (2001) can be applied to the problem of estimating $g(x) - g(0)$ of this paper. The estimator is constructed by estimating the derivative of $g(\cdot)$ and numerically integrating the estimated derivative. However, it is impossible to apply the estimator if we have observed no covariate X on an interval between 0 and x.

Recently Chen (2000a, 2000b) considered binary choice models, where $g(\cdot)$ is a linear function with a constant term and the error is known to be symmetrically distributed. The estimator of Chen (2000a) attains the efficiency bound, but it requires smoothing parameter selection. The estimator of Chen (2000b) does not require smoothing parameter selection, but it does not attain the efficiency bound and needs a preliminary estimator of the slope parameter.

In statistics, several authors have considered linear regression with no constant term with current status data. For example, van der Laan *et al.* (1997) considered the case of discrete covariates and suggested a nonparametric regression procedure in the case of continuous covariates. Murphy *et al.* (1999) considered a penalized ML estimator and proved that it is efficient. Their penalized ML estimator requires smoothing parameter selection. Abrevaya (1999) studied the asymptotic properties of the estimator similar to the MRC estimator. Li and Zhang (1998) constructed an efficient estimator based on U-statistics. Their estimator also requires smoothing parameter selection. Shen (2000) studied linear regression with a constant term by assuming that the error term has zero mean. In Shen (2000) an efficient estimator is constructed based on the random sieve likelihood and the asymptotic properties are derived by using the results of Shen (1997) and so on. Andrews *et al.* (2001) also deals with linear regression with a constant term by introducing generalized location parameters.

Some authors studied current status data in other setups. For example, Huang (1996) studied the Cox model. Van der Laan and Robins (1998) considered the estimation of smooth functionals of the distribution function of failure times. See Huang and Wellner (1997) for other references. Some fundamental results on nonparametric ML estimation for current status data with no covariate are given in Groeneboom and Wellner (1992). They studied the estimation of the mean of the failure time $\tilde{Y}$ in the case of no covariate. They proved the $\sqrt{n}$-consistency and derived the asymptotic distribution under the compact support assumption in Section II.5.5 of Groeneboom and Wellner (1992). Note that the estimator of the mean is efficient. See also Section II.1.1 of Groeneboom and Wellner (1992), Example of 7.4.3 of van de Geer (2000), and Huang and Wellner (1995).

Van der Laan *et al.* (1997) suggested the following nonparametric regression estimator when the support of failure times is contained in that of examination times: We carry out nonparametric ML estimation of the distribution function of the failure time from the observations whose covariate is around x and calculate the mean of the ML estimate of the distribution function. However, they gave no theoretical analysis. Letting h be the bandwidth, from the results of Huang and Wellner (1995), we conjecture that this estimator has the variance of $O((nh)^{-1})$. The bias will be $O(h^2)$. The same kind of remarks will apply to the estimator in Andrews *et al.* (2001) when we apply it to nonparametric regression. The estimating procedure in Shen (2000) is another promising candidate to apply to nonparametric regression. However, we do not know how to analyze the procedure theoretically when it is applied to nonparametric regression.

In this paper we construct an estimator of $g(x) - g(0)$ by localizing Han's MRC estimator to the observations whose covariate is around 0 or x. By estimating $g(x)-g(0)$, we can alleviate the restriction that the support of failure times must be contained in that of examination times. For example, let us consider the case of log-transformed failure times. When assumption A4 below holds and the support of examination times is $(-\infty, a)(a > 0)$, we are able to carry out nonparametric inference for any smooth regression function and we do not have to care about the possible heavy tails of the distribution of failure times. See assumption A5 below.

We evaluate the asymptotic bias of the estimator in Section 2 and examine the asymptotic distribution by following Sherman (1993) in Section 3. Then we need to modify a result on equicontinuity of degenerate U-processes of de la Peña and Giné (1999) to apply it to the setup of this paper.

When the covariate X is discrete, our estimator will not reduce to an efficient estimator since the MRC estimator is not efficient. However, the estimator does not require any additional bandwidths. It requires only one bandwidth for the covariate X. Most of the other semiparametric estimators will need auxiliary bandwidths and parameters for trimming when they are applied to nonparametric regression.

Some other authors considered nonparametric regression in survival analysis, for example, Fan and Gijbels (1994) considered nonparametric regression with censored data.

We assume X is one-dimensional for simplicity of presentation. Before we define the estimator of $g(x) - g(0)$, we give some definitions and notations. We define Z_i and $\tilde{Z}_i$ by

$$Z_i = (C_i, X_i, \epsilon_i) \quad \text{and} \quad \tilde{Z}_i = (C_i, X_i), \tag{1.3}$$

respectively. We assume that $\{Z_i\}$ are i.i.d. in this paper. See also Y_i and $\tilde{Y}_i$ in (1.1).

We state assumptions A1–2. We do not allow ϵ to depend on X.

A1. (C_i, X_i) and ϵ_i are mutually independent. We allow only C_i to depend on X_i.

A2. The unknown regression function $g(\cdot)$ is twice continuously differentiable around 0 and x.

Writing h for the bandwidth tending to 0 as the sample size $n \to \infty$, we have from the Taylor expansion that

$$g(X_i) = g(0) - \beta_2^0 \frac{X_i}{h} + \frac{h^2}{2} g''(\theta_i X_i) \left(\frac{X_i}{h}\right)^2 \tag{1.4}$$

and

$$g(X_j) = g(x) - \beta_3^0 \frac{X_j - x}{h} + \frac{h^2}{2} g''(x + \eta_j (X_j - x)) \left(\frac{X_j - x}{h}\right)^2, \tag{1.5}$$

where $\beta_2^0 = -hg'(0)$ and $\beta_3^0 = -hg'(x)$. Putting

$$\beta_1^0 = g(x) - g(0),$$

we write β_0 for $(\beta_1^0, \beta_2^0, \beta_3^0)^T \in R^3$. Note that $\beta_0 \to (\beta_1^0, 0, 0)^T$ as $n \to \infty$.

By using the local linear approximation to $g(\cdot)$ in (1.4) and (1.5) and modifying the MRC estimator, we estimate β_0 by maximizing

$$\begin{aligned} \Gamma_n(\beta) = {} & \frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{h^2} K\left(\frac{X_i}{h}\right) K\left(\frac{X_j - x}{h}\right) \\ & \times \left[I\{Y_i > Y_j\} I\left\{ C_i + \beta_1 + \beta_2 \frac{X_i}{h} > C_j + \beta_3 \frac{X_j - x}{h} \right\} \right. \\ & \left. + I\{Y_i < Y_j\} I\left\{ C_i + \beta_1 + \beta_2 \frac{X_i}{h} < C_j + \beta_3 \frac{X_j - x}{h} \right\} \right], \end{aligned} \tag{1.6}$$

where $\beta = (\beta_1, \beta_2, \beta_3)^T$ and $K(\cdot)$ is a kernel function satisfying assumption A3. We do not care about $=$ in the indicator functions in (1.6) since the probability of equality is zero under technical assumptions.

A3. $K(\cdot)$ has compact support containing $[-1, 1]$ and $K(\cdot)$ is non-negative, bounded, and symmetric. In addition $\int K(t)dt = 1$.

Since $\Gamma_n(\cdot)$ in (1.6) usually has multiple maximum solutions, we should choose one among them by some rule. We write $\hat{\beta}$ for the chosen maximum solution. The results of this paper are independent of the rule for choosing $\hat{\beta}$. For example, the sample median is not uniquely determined when the sample size is even, and the asymptotic properties are independent of the definition of the sample median.

Theorems 2.1 and 3.1 in this paper show the well-known trade-off between the bias and the variance in nonparametric regression. We can define the asymptotically optimal bandwidth theoretically and the order is $O(n^{-1/5})$ when the covariate is one-dimensional. However, it seems to difficult to estimate it from the observations or to give an automatic selection rule like cross validation. Thus we give the result of a simulation study to see the effect of the bandwidth and how the estimator works. In the simulation study $\Gamma_n(\cdot)$ in (1.6) is maximized by using grid search.

Remark 1.1. We focus on local linear approximation in this paper. However, we can construct an estimator based on local constant approximation by considering the Taylor expansion of first order instead of (1.4) and (1.5). Then we do not have β_2 and β_3 in (1.6) and $\Gamma_n(\beta)$ is reduced to

$$\begin{aligned} \Gamma_n(\beta) = {} & \frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{h^2} K\left(\frac{X_i}{h}\right) K\left(\frac{X_j - x}{h}\right) \\ & \times [I\{Y_i > Y_j\} I\{C_i + \beta_1 > C_j\} + I\{Y_i < Y_j\} I\{C_i + \beta_1 < C_j\}]. \end{aligned}$$

The maximum solutions usually form an interval or intervals. The estimator is consistent and has the bias of $O(h^2)$ and the asymptotic normality as in Theorem 3.1. The proofs for the consistency and the asymptotic normality need no significant modification. As for the bias, we describe the necessary modifications in Remark 2.1 below. The local constant estimator may be better for computational purpose especially when we do not have so many observations.

We investigate the bias of the estimator and prove the consistency of the estimator in Section 2. The asymptotic normality is established in Section 3. The result of the simulation study is given in Section 4. Technical proofs are confined to Section 5.

2. Bias and consistency of the estimator

Putting

$$\Gamma(\beta) = \mathrm{E}\{\Gamma_n(\beta)\},$$

we define β_m by

$$\beta_m = \operatorname*{argmax}_{\beta \in R^3} \Gamma(\beta),$$

where $\beta_m = (\beta_1^m, \beta_2^m, \beta_3^m)^T$. $\Gamma(\cdot)$ and β_m depend on h. $\Gamma_n(\cdot)$ usually has multiple maximum solutions. However, β_m is deterministic and uniquely determined. In Theorem 2.1 we prove the uniqueness of β_m and evaluate $\beta_m - \beta_0$, which corresponds to the bias term since we show the asymptotic normality of $\hat{\beta} - \beta_m$ later in Section 3 by using theorems in Sherman (1993).

The consistency,

$$|\hat{\beta} - \beta_0| \to 0 \quad \text{in probability},$$

where $|\cdot|$ denotes the Euclidean norm, is established in Theorem 2.2. We write M_i for positive constants independent of h and n. The values of M_i differ from section to section. Note that arguments in this paper hold as n tends to ∞.

We introduce several assumptions and notations. As for the bandwidth, we only assume that h tends to 0. Suppose assumptions A1–7 hold throughout this paper.

A4. ϵ_1 has the bounded density function f_ϵ on R. f_ϵ is positive on R and continuous. F_ϵ stands for the distribution function.

A5. $C_2 - C_1$ has the positive density function around $\beta_1^0 = g(x) - g(0)$, where $C_1 \sim f_C(\cdot \mid 0)$, $C_2 \sim f_C(\cdot \mid x)$, and $f_C(\cdot \mid X)$ denotes the conditional density function of C on X.

Remark 2.1. $\tilde{Y}$ is usually nonnegative. When we deal with $\tilde{Y}$, not with $\log \tilde{Y}$, assumptions A4 and A5 should be replaced with A4' below. These assumptions are necessary for (5.2) in Section 5. We assume assumptions A4 and A5 for simplicity of presentation. When assumption A4 holds and the support of $C_2 - C_1$ is $(-\infty, \infty)$, assumption A5 holds automatically and we can deal with any smooth regression function. Assumption A4' below looks complicated. However, it just means that the intersection of the support of C_1 given $X_1 = 0$, that of Y_1 given $X_1 = 0$, that of $C_2 + g(0) - g(x)$ given $X_2 = x$, and that of $Y_2 + g(0) - g(x)$ given $X_2 = x$ is not empty.

A4'. The following relation holds for the support for the conditional distribution of C.

$$\{c \mid f_C(c \mid 0) > 0\} \cap \{c \mid f_C(c - g(0) + g(x) \mid x) > 0\} \cap \{\tilde{y} \mid f_\epsilon(\tilde{y} - g(0)) > 0\} \neq \phi.$$

A6. X_1 has the positive and continuous density function f_X around 0 and x.

A7. The conditional density function of C, $f_C(\cdot \mid X)$, and the derivative with respect to C, $f'_C(\cdot \mid X)$, are uniformly bounded from above around 0 and x. $|f_C(\cdot \mid X)|$ and $|f'_C(\cdot \mid X)|$ are uniformly bounded by an integrable function around 0 and x. $f'_C(\cdot \mid X)$ is uniformly Lipschitz continuous with respect to C around 0 and x. In addition we have that

$$\lim_{\tilde{x}\to 0} f_C(c \mid \tilde{x}) = f_C(c \mid 0) \quad \text{and} \quad \lim_{\tilde{x}\to x} f_C(c \mid \tilde{x}) = f_C(c \mid x) \quad \text{for any} \quad c.$$

We state Theorems 2.1 and 2.2. Then we prove Theorem 2.1, give a remark on the local constant estimator defined in Remark 1.1, and present an explicit expression of $\beta_m - \beta_0$. Theorem 2.2 is verified at the end of this section.

THEOREM 2.1. *If $h \to 0$, β_m is unique and the bias $\beta_m - \beta_0$ is written as*

$$\beta_m - \beta_0 = -\Lambda^{-1}\frac{\partial\Gamma}{\partial\beta}(\beta_0) + \mathrm{o}(h^2), \tag{2.1}$$

where Λ is a negative definite matrix and

$$\frac{\partial\Gamma}{\partial\beta}(\beta_0) = \mathrm{O}(h^2).$$

THEOREM 2.2. *If $h \to 0$ and $nh \to \infty$, we have*

$$|\hat{\beta} - \beta_0| \to 0 \quad \textit{in probability.}$$

PROOF OF THEOREM 2.1. We define two auxiliary random variables Y_i^* and $Y_j^\dagger$.

$$Y_i^* = \begin{cases} 1, & C_i - g(0) + \beta_2^0 \dfrac{X_i}{h} - \epsilon_i > 0 \\ 0, & \text{otherwise} \end{cases} \tag{2.2}$$

$$Y_j^\dagger = \begin{cases} 1, & C_j - g(x) + \beta_3^0 \dfrac{X_j - x}{h} - \epsilon_j > 0 \\ 0, & \text{otherwise.} \end{cases} \tag{2.3}$$

By replacing Y_i and Y_j in $\Gamma_n(\beta)$ with Y_i^* and $Y_j^\dagger$, we define $\tilde{\Gamma}_n(\beta)$ and $\tilde{\Gamma}(\beta)$.

$$\begin{aligned} \tilde{\Gamma}_n(\beta) = {} & \frac{1}{n(n-1)}\sum_{i\neq j}\frac{1}{h^2}K\left(\frac{X_i}{h}\right)K\left(\frac{X_j - x}{h}\right) \\ & \times\left[I\{Y_i^* > Y_j^\dagger\}I\left\{C_i + \beta_1 + \beta_2\frac{X_i}{h} > C_j + \beta_3\frac{X_j - x}{h}\right\}\right. \\ & \left. + I\{Y_i^* < Y_j^\dagger\}I\left\{C_i + \beta_1 + \beta_2\frac{X_i}{h} < C_j + \beta_3\frac{X_j - x}{h}\right\}\right], \end{aligned} \tag{2.4}$$

$$\tilde{\Gamma}(\beta) = \mathrm{E}\{\tilde{\Gamma}_n(\beta)\}. \tag{2.5}$$

We prove Theorem 2.1 by comparing $\Gamma(\cdot)$ and $\tilde{\Gamma}(\cdot)$. (1.1) and assumption **A4** imply

$$I\left\{K\left(\frac{X_i}{h}\right) > 0\right\} \mathrm{P}(Y_i^* \neq Y_i \mid \tilde{Z}_i) = \mathrm{O}(h^2) \quad \text{uniformly in } \tilde{Z}_i = (C_i, X_i) \tag{2.6}$$

and

$$I\left\{K\left(\frac{X_j - x}{h}\right) > 0\right\} \mathrm{P}(Y_j^\dagger \neq Y_j \mid \tilde{Z}_j) = \mathrm{O}(h^2) \quad \text{uniformly in } \tilde{Z}_j = (C_j, X_j). \tag{2.7}$$

By considering the conditional expectations of $\Gamma_n(\beta)$ and $\tilde{\Gamma}_n(\beta)$ on $\{\tilde{Z}_i\}$ and using (2.6) and (2.7), we can easily show that

$$\Gamma(\beta) = \tilde{\Gamma}(\beta) + \mathrm{O}(h^2) \quad \text{uniformly in } \beta. \tag{2.8}$$

The remainder of the proof consists of three steps, to prove that $\tilde{\Gamma}(\cdot)$ is uniquely maximized at β_0, to verify (2.14) below, and to establish (2.23) below which leads to the expression of $\beta_m - \beta_0$ in the theorem.

Step 1. We prove that $\tilde{\Gamma}(\cdot)$ is uniquely maximized at β_0 by noting that

$$\begin{aligned} &\mathrm{P}(Y_i^* > Y_j^\dagger \mid \tilde{Z}_i, \tilde{Z}_j) \gtrless \mathrm{P}(Y_i^* < Y_j^\dagger \mid \tilde{Z}_i, \tilde{Z}_j) \\ &\Leftrightarrow C_i + \beta_1^0 + \beta_2^0 \frac{X_i}{h} \gtrless C_j + \beta_3^0 \frac{X_j - x}{h}. \end{aligned} \tag{2.9}$$

We define $F(\tilde{Z}_1, \tilde{Z}_2; \beta)$ by

$$\begin{aligned} F(\tilde{Z}_1, \tilde{Z}_2; \beta) = {} & \frac{1}{h^2} K\left(\frac{X_1}{h}\right) K\left(\frac{X_2 - x}{h}\right) \\ & \times \Bigg[\mathrm{P}(Y_1^* > Y_2^\dagger \mid \tilde{Z}_1, \tilde{Z}_2) \\ & \quad \times \left\{I\left\{C_1 + \beta_1^0 + \beta_2^0 \frac{X_1}{h} > C_2 + \beta_3^0 \frac{X_2 - x}{h}\right\}\right. \\ & \quad \left. - I\left\{C_1 + \beta_1 + \beta_2 \frac{X_1}{h} > C_2 + \beta_3 \frac{X_2 - x}{h}\right\}\right\} \\ & + \mathrm{P}(Y_1^* < Y_2^\dagger \mid \tilde{Z}_1, \tilde{Z}_2) \\ & \quad \times \left\{I\left\{C_1 + \beta_1^0 + \beta_2^0 \frac{X_1}{h} < C_2 + \beta_3^0 \frac{X_2 - x}{h}\right\}\right. \\ & \quad \left. - I\left\{C_1 + \beta_1 + \beta_2 \frac{X_1}{h} < C_2 + \beta_3 \frac{X_2 - x}{h}\right\}\right\}\Bigg]. \end{aligned} \tag{2.10}$$

Then we can represent $\tilde{\Gamma}(\beta_0) - \tilde{\Gamma}(\beta)$ as

$$\tilde{\Gamma}(\beta_0) - \tilde{\Gamma}(\beta) = \mathrm{E}\{F(\tilde{Z}_1, \tilde{Z}_2; \beta)\}. \tag{2.11}$$

(2.9) implies

$$F(\tilde{Z}_1, \tilde{Z}_2; \beta) < 0 \quad \text{a.e.} \tag{2.12}$$

when

$$I\left\{C_1 + \beta_1^0 + \beta_2^0 \frac{X_1}{h} > C_2 + \beta_3^0 \frac{X_2 - x}{h}\right\} \neq I\left\{C_1 + \beta_1 + \beta_2 \frac{X_1}{h} > C_2 + \beta_3 \frac{X_2 - x}{h}\right\}.$$

Combining assumptions A5 and A6 and (2.11) and (2.12), we have

$$\tilde{\Gamma}(\beta_0) > \tilde{\Gamma}(\beta) \quad \text{for any} \quad \beta \neq \beta_0. \tag{2.13}$$

We have established that $\tilde{\Gamma}(\beta)$ is uniquely maximized at β_0.

Step 2. We show that

$$|\beta_0 - \beta_m| < M_1 h \quad \text{for some} \quad M_1. \tag{2.14}$$

(2.14) is proved if we show that for any M_2, there exists a positive number M_3 such that

$$\tilde{\Gamma}(\beta) < \tilde{\Gamma}(\beta_0) - M_2 h^2 \quad \text{for} \quad |\beta - \beta_0| > M_3 h. \tag{2.15}$$

Indeed (2.8) and (2.15) imply that

$$\Gamma(\beta_0) > \Gamma(\beta) + (M_2 - \mathrm{O}(1))h^2 \quad \text{for} \quad |\beta - \beta_0| > M_3 h.$$

We should choose a sufficiently large M_2. We postpone the proof of (2.15) to Section 5.

Step 3. We evaluate $\beta_m - \beta_0$ more accurately by considering the Taylor expansion of $\Gamma(\beta)$ at β_0 and the derivatives of $\tilde{\Gamma}(\beta)$ at β_0. Besides we prove the uniqueness of β_m and give an expression of $\beta_m - \beta_0$.

We rewrite $\Gamma(\beta)$ and $\tilde{\Gamma}(\beta)$ by writing T_1 and $\tilde{T}_2$ for X_1/h and $(X_2 - x)/h$, respectively.

$$\begin{aligned} \tilde{\Gamma}(\beta) = \int & K(T_1)K(\tilde{T}_2)f_X(T_1 h)f_X(x + \tilde{T}_2 h) \\ & \times \mathrm{E}\{F_\epsilon(C_1 - g(0) + \beta_2^0 T_1)(1 - F_\epsilon(C_2 - g(x) + \beta_3^0 \tilde{T}_2)) \\ & \quad \times I\{C_1 + \beta_1 + \beta_2 T_1 > C_2 + \beta_3 \tilde{T}_2\} \\ & \quad + (1 - F_\epsilon(C_1 - g(0) + \beta_2^0 T_1))F_\epsilon(C_2 - g(x) + \beta_3^0 \tilde{T}_2) \\ & \quad \times I\{C_1 + \beta_1 + \beta_2 T_1 < C_2 + \beta_3 \tilde{T}_2\} \mid T_1, \tilde{T}_2\} dT_1 d\tilde{T}_2 \end{aligned} \tag{2.16}$$

$$\begin{aligned} \Gamma(\beta) = \int & K(T_1)K(\tilde{T}_2)f_X(T_1 h)f_X(x + \tilde{T}_2 h) \\ & \times \mathrm{E}\{F_\epsilon(C_1 - g(T_1 h))(1 - F_\epsilon(C_2 - g(x + \tilde{T}_2 h))) \\ & \quad \times I\{C_1 + \beta_1 + \beta_2 T_1 > C_2 + \beta_3 \tilde{T}_2\} \\ & \quad + (1 - F_\epsilon(C_1 - g(T_1 h)))F_\epsilon(C_2 - g(x + \tilde{T}_2 h)) \\ & \quad \times I\{C_1 + \beta_1 + \beta_2 T_1 < C_2 + \beta_3 \tilde{T}_2\} \mid T_1, \tilde{T}_2\} dT_1 d\tilde{T}_2. \end{aligned} \tag{2.17}$$

$\tilde{\Gamma}(\beta)$ and $\Gamma(\beta)$ are twice continuously differentiable from the above expressions and assumption A7. In addition, (2.6), (2.7), assumption A7, and these expressions imply as in (2.8) that

$$\frac{\partial \tilde{\Gamma}}{\partial \beta}(\beta) = \frac{\partial \Gamma}{\partial \beta}(\beta) + \mathrm{O}(h^2) \quad \text{and} \quad \frac{\partial^2 \tilde{\Gamma}}{\partial \beta \partial \beta^T}(\beta) = \frac{\partial^2 \Gamma}{\partial \beta \partial \beta^T}(\beta) + \mathrm{O}(h^2) \tag{2.18}$$

uniformly in β.

We have from (2.13) and (2.18) that

$$\frac{\partial \tilde{\Gamma}}{\partial \beta}(\beta_0) = 0 \quad \text{and} \quad \frac{\partial \Gamma}{\partial \beta}(\beta_0) = \mathrm{O}(h^2). \tag{2.19}$$

We have from (2.16) and assumption A7 that

$$\lim_{n\to\infty} \frac{\partial^2 \tilde{\Gamma}}{\partial\beta\partial\beta^T}(\beta_0) = \Lambda, \tag{2.20}$$

where Λ is a 3×3 matrix and we give an expression of it later in this section. Besides (2.15) means

$$\Lambda \text{ is negative definite} \tag{2.21}$$

and (2.16), (2.18), and assumption A7 imply that

$$\frac{\partial^2 \Gamma}{\partial\beta\partial\beta^T}(\beta) = \Lambda + o(1) \quad \text{as} \quad |\beta - \beta_0| \to 0 \quad \text{and} \quad h \to 0. \tag{2.22}$$

$\Gamma(\cdot)$ is strictly concave around β_0 from (2.21) and (2.22) and we can represent $\Gamma(\beta)$ on $\{|\beta - \beta_0| < M_1 h\}$ as

$$\Gamma(\beta) = \Gamma(\beta_0) + (\beta - \beta_0)^T \frac{\partial\Gamma}{\partial\beta}(\beta_0) + \frac{1}{2}(\beta - \beta_0)^T(\Lambda + o(1))(\beta - \beta_0). \tag{2.23}$$

(2.23) is uniquely maximized by (2.1) on $\{|\beta - \beta_0| < M_1 h\}$. Hence the proof is complete. □

Remark 2.2. We can prove that the estimator based on the local constant approximation has the bias of $O(h^2)$ by assuming that f_ϵ has the bounded continuous derivative and that f_X is continuously differentiable around 0 and x. We can show that the bias is $O(h)$ as in the proof of Theorem 2.1. Then we use (2.23) on $\{|\beta - \beta_0| < Mh\}$ for sufficiently large M. If $\frac{d\Gamma}{d\beta_1}(\beta_0) = O(h^2)$, the bias is also $O(h^2)$. We can verify that $\frac{d\Gamma}{d\beta_1}(\beta_0) = O(h^2)$ by refining the evaluation in (2.26) and (2.27) below and using the symmetry of the kernel function.

We give an expression of the right-hand side of (2.1). We omit tedious calculations. At first we deal with Λ. Λ can be written as

$$\Lambda = \frac{\partial^2 \Gamma_\infty}{\partial\beta\partial\beta^T}((\beta_1^0, 0, 0)^T), \tag{2.24}$$

where

$$\begin{aligned}\Gamma_\infty(\beta) = {} & f_X(0) f_X(x) \\ & \times \mathrm{E}_\infty[F_\epsilon(C_1 - g(0))(1 - F_\epsilon(C_2 - g(x)))I\{C_1 + \beta_1 + \beta_2 T_1 > C_2 + \beta_3 T_2\} \\ & \quad + (1 - F_\epsilon(C_1 - g(0)))F_\epsilon(C_2 - g(x))I\{C_1 + \beta_1 + \beta_2 T_1 < C_2 + \beta_3 T_2\}]\end{aligned}$$

and $\mathrm{E}_\infty\{G(C_1, C_2, T_1, T_2)\}$ in the above expression is defined by

$$\begin{aligned}&\mathrm{E}_\infty\{G(C_1, C_2, T_1, T_2)\} \\ &\quad = \int G(C_1, C_2, T_1, T_2) f_C(C_1 \mid 0) f_C(C_2 \mid x) K(T_1) K(T_2) dC_1 dC_2 dT_1 dT_2.\end{aligned}$$

Next we deal with $\frac{1}{h^2}\frac{\partial\Gamma}{\partial\beta}(\beta_0)$ by exploiting the fact that

$$\frac{1}{h^2}\frac{\partial\Gamma}{\partial\beta}(\beta_0) = \frac{1}{h^2}\frac{\partial(\Gamma - \tilde{\Gamma})}{\partial\beta}(\beta_0). \tag{2.25}$$

To evaluate the difference between (2.16) and (2.17), we use the following expressions.

$$
\begin{aligned}
(2.26)\qquad & F_\epsilon(C_1 - g(T_1h))(1 - F_\epsilon(C_2 - g(x + \tilde{T}_2h))) \\
&\qquad - F_\epsilon(C_1 - g(0) + \beta_2^0 T_1)(1 - F_\epsilon(C_2 - g(x) + \beta_3^0\tilde{T}_2)) \\
&= -f_\epsilon(C_1 - g(0))(1 - F_\epsilon(C_2 - g(x)))\frac{1}{2}T_1^2h^2g''(0) \\
&\quad + f_\epsilon(C_2 - g(x))F_\epsilon(C_1 - g(0))\frac{1}{2}\tilde{T}_2^2h^2g''(x) + o(h^2)
\end{aligned}
$$

$$
\begin{aligned}
(2.27)\qquad & (1 - F_\epsilon(C_1 - g(T_1h)))F_\epsilon(C_2 - g(x + \tilde{T}_2h)) \\
&\qquad - (1 - F_\epsilon(C_1 - g(0) + \beta_2^0 T_1))F_\epsilon(C_2 - g(x) + \beta_3^0\tilde{T}_2) \\
&= f_\epsilon(C_1 - g(0))F_\epsilon(C_2 - g(x))\frac{1}{2}T_1^2h^2g''(0) \\
&\quad - f_\epsilon(C_2 - g(x))(1 - F_\epsilon(C_1 - g(0)))\frac{1}{2}\tilde{T}_2^2h^2g''(x) + o(h^2).
\end{aligned}
$$

Then from (2.16), (2.17), (2.25), (2.26), and (2.27), we obtain the following expression of $\frac{1}{h^2}\frac{\partial\Gamma}{\partial\beta}(\beta_0)$,

$$
(2.28)\qquad \frac{1}{h^2}\frac{\partial\Gamma}{\partial\beta}(\beta_0) = \frac{\partial\Omega}{\partial\beta}((\beta_1^0, 0, 0)^T) + o(1),
$$

where

$$
\begin{aligned}
\Omega(\beta) = {} & \frac{f_X(0)f_X(x)}{2} \\
& \times \mathrm{E}_\infty[\{-f_\epsilon(C_1 - g(0))(1 - F_\epsilon(C_2 - g(x)))T_1^2g''(0) \\
& \qquad + f_\epsilon(C_2 - g(x))F_\epsilon(C_1 - g(0))T_2^2g''(x)\} \\
& \qquad \times I\{C_1 + \beta_1 + \beta_2T_1 > C_2 + \beta_3T_2\} \\
& \qquad + \{f_\epsilon(C_1 - g(0))F_\epsilon(C_2 - g(x))T_1^2g''(0) \\
& \qquad\quad - f_\epsilon(C_2 - g(x))(1 - F_\epsilon(C_1 - g(0)))T_2^2g''(x)\} \\
& \qquad \times I\{C_1 + \beta_1 + \beta_2T_1 < C_2 + \beta_3T_2\}]
\end{aligned}
$$

and $\mathrm{E}_\infty\{\cdot\}$ is defined just below (2.24).

Finally we prove Theorem 2.2 which deals with the consistency of $\hat{\beta}$.

PROOF OF THEOREM 2.2. We can establish in the same way as in Lemma 5.1 below that if $nh \to \infty$,

$$
(2.29)\qquad \sup_{\beta\in R^3} |\Gamma_n(\beta) - \Gamma(\beta)| \to 0 \quad \text{in probability as} \quad n \to \infty.
$$

Combining (2.8) and (2.29), we have

$$
(2.30)\qquad \sup_{\beta\in R^3} |\Gamma_n(\beta) - \tilde{\Gamma}(\beta)| \to 0 \quad \text{in probability.}
$$

We can prove in the same way as (2.15) that for any $\delta_1 > 0$, there exists a positive number δ_2 such that

$$
(2.31)\qquad \tilde{\Gamma}(\beta_0) - \delta_1 > \tilde{\Gamma}(\beta) \quad \text{for} \quad |\beta - \beta_0| > \delta_2.
$$

Theorem 2.2 follows from (2.30) and (2.31). □

3. Asymptotic normality of the estimator

We establish the asymptotic normality of $\sqrt{nh}(\hat{\beta} - \beta_m)$ by applying Theorems 1 and 2 in Sherman (1993). Suppose that $nh \to \infty$ and $h \to 0$ and that assumptions A1–7 hold in this section, too. Then

THEOREM 3.1.
$$\sqrt{nh}(\hat{\beta} - \beta_m) \xrightarrow{D} \mathrm{N}(0, \Lambda^{-1}\Sigma\Lambda^{-1}),$$
where $\xrightarrow{D}$ *means convergence in law. The* 3×3 *matrix* Σ *is given by*

$$\begin{aligned}
\Sigma = {} & f_X(0) f_X(x) \\
& \times \mathrm{E}\left[\begin{pmatrix} \int K^2 dt & 0 & 0 \\ 0 & \int t^2 K^2 dt & 0 \\ 0 & 0 & 0 \end{pmatrix} f_X(x) f_C^2(C_1 + \beta_1^0 \mid x) \right. \\
& \times \{ I\{g(0) + \epsilon_1 < C_1\}(1 - F_\epsilon(C_1 - g(0)))^2 + I\{g(0) + \epsilon_1 > C_1\} F_\epsilon^2(C_1 - g(0)) \} \\
& + \begin{pmatrix} \int K^2 dt & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \int t^2 K^2 dt \end{pmatrix} f_X(0) f_C^2(C_2 - \beta_1^0 \mid 0) \\
& \left. \times \{ I\{g(x) + \epsilon_2 > C_2\} F_\epsilon^2(C_2 - g(x)) + I\{g(x) + \epsilon_2 < C_2\}(1 - F_\epsilon(C_2 - g(x)))^2 \} \right],
\end{aligned}$$

where $C_1 \sim f_C(\cdot \mid 0)$ *and* $C_2 \sim f_C(\cdot \mid x)$.

Theorems 2.1 and 3.1 imply that when X is one-dimensional, we can estimate $g(x) - g(0)$ in the order of $n^{-2/5}$ by taking $h = cn^{-1/5}$. The optimal c, which is denoted by c_0, depends on the unknown parameters in the theorems. It is possible to theoretically calculate c_0. However, it may be difficult to estimate the parameters at present. A simulation study is presented in Section 4 to see the effect of the bandwidth.

PROOF OF THEOREM 3.1. Theorem 3.1 follows from Theorems 1 and 2 in Sherman (1993) with n replaced with nh if all the conditions for the two theorems are satisfied. We will check all the conditions. In Section 3 we proved that Λ is negative definite and that $|\hat{\beta} - \beta_m| \to 0$ in probability. All we have to do is to establish (3.1), (3.8), and (3.9) below. (3.1) corresponds to (i) in Theorem 1, (3.8) corresponds to the asymptotic normality of W_n in Theorem 2, and (3.9) corresponds to (ii) in Theorem 1 and (4) in Theorem 2.

From the fact that $|\beta_m - \beta_0| \to 0$ and (2.22), we have

$$\Gamma(\beta) - \Gamma(\beta_m) \le -M_1 |\beta - \beta_m|^2 \tag{3.1}$$

when $|\beta - \beta_m|$ is sufficiently small.

We need two lemmas to establish (3.8) and (3.9). We introduce some notations and definitions for the lemmas.

Defining $T(i,j;\beta)$ by

$$T(i,j;\beta) = \frac{1}{h}K\left(\frac{X_i}{h}\right)K\left(\frac{X_j - x}{h}\right) \times \left[I\{Y_i > Y_j\}I\left\{C_i + \beta_1 + \beta_2\frac{X_i}{h} > C_j + \beta_3\frac{X_j - x}{h}\right\} + I\{Y_i < Y_j\}I\left\{C_i + \beta_1 + \beta_2\frac{X_i}{h} < C_j + \beta_3\frac{X_j - x}{h}\right\}\right], \tag{3.2}$$

we represent $\Gamma_n(\beta)$ as

$$\Gamma_n(\beta) = \frac{1}{n(n-1)}\sum_{i\neq j}\frac{1}{h}T(i,j;\beta).$$

We define $\Pi_n(\beta)$ and $U(i,j;\beta)$ by

$$\Pi_n(\beta) = h(\Gamma_n(\beta) - \Gamma_n(\beta_m)) \quad \text{and} \quad U(i,j;\beta) = T(i,j;\beta) - T(i,j;\beta_m), \tag{3.3}$$

respectively. Then the Hoeffding decomposition of $\Pi_n(\beta)$ is written as

$$\begin{aligned} \Pi_n(\beta) = {} & \mathrm{E}\{\Pi_n(\beta)\} \\ & + \frac{1}{n}\sum_{i=1}^{n}[\mathrm{E}\{U(i,j;\beta) \mid Z_i\} - \mathrm{E}\{\Pi_n(\beta)\}] \qquad (3.4) \\ & + \frac{1}{n}\sum_{j=1}^{n}[\mathrm{E}\{U(i,j;\beta) \mid Z_j\} - \mathrm{E}\{\Pi_n(\beta)\}] \\ & + \frac{1}{n(n-1)}\sum_{i\neq j}[U(i,j;\beta) - \mathrm{E}\{U(i,j;\beta) \mid Z_i\} \qquad (3.5) \\ & \qquad - \mathrm{E}\{U(i,j;\beta) \mid Z_j\} + \mathrm{E}\{\Pi_n(\beta)\}]. \end{aligned}$$

Putting (3.4) $= Q(\beta)$, i.e.,

$$Q(\beta) = \frac{1}{n}\sum_{i=1}^{n}[\mathrm{E}\{U(i,j;\beta) \mid Z_i\} - \mathrm{E}\{\Pi_n(\beta)\}] + \frac{1}{n}\sum_{j=1}^{n}[\mathrm{E}\{U(i,j;\beta) \mid Z_j\} - \mathrm{E}\{\Pi_n(\beta)\}], \tag{3.6}$$

we present Lemma 3.1.

LEMMA 3.1. *If $nh \to \infty$, $h \to 0$, and $|\beta - \beta_m| \to 0$, we have*

$$Q(\beta) = (\beta - \beta_m)^T\frac{\partial Q}{\partial \beta}(\beta_m) + \mathrm{o_p}(h|\beta - \beta_m|^2) \quad \textit{uniformly in } \beta \tag{3.7}$$

and

$$\sqrt{\frac{n}{h}}\frac{\partial Q}{\partial \beta}(\beta_m) \xrightarrow{D} \mathrm{N}(0,\Sigma). \tag{3.8}$$

We prove Lemma 3.1 later in this section. Lemma 3.2 deals with (3.5).

LEMMA 3.2. *If $nh \to \infty$, $h \to 0$, and $|\beta - \beta_m| \to 0$, we have*

$$\frac{1}{n(n-1)}\sum_{i\neq j}[U(i,j;\beta) - \mathrm{E}\{U(i,j;\beta) \mid Z_i\} - \mathrm{E}\{U(i,j;\beta) \mid Z_j\} + \mathrm{E}\{\Pi_n(\beta)\}] = \mathrm{o_p}\left(\frac{1}{n}\right)$$

uniformly in β.

We prove Lemma 3.2 in Section 5 by adapting the argument in the proof of Theorem 5.3.7 of de la Peña and Giné (1999) to the setup of this paper. Lemma 3.2 corresponds to Theorem 3 of Sherman (1993) and is related to equicontinuity of degenerate U-processes. What is crucial to the proof of Lemma 3.2 is that

$$\left\{K\left(\frac{X_1}{h}\right)K\left(\frac{X_2-x}{h}\right)\{Y_1 > Y_2\}I\left\{C_1+\beta_1+\beta_2\frac{X_1}{h} > C_2+\beta_3\frac{X_2-x}{h}\right\} \mid \beta\in R^3\right\}$$

and

$$\left\{K\left(\frac{X_1}{h}\right)K\left(\frac{X_2-x}{h}\right)\{Y_1 < Y_2\}I\left\{C_1+\beta_1+\beta_2\frac{X_1}{h} < C_2+\beta_3\frac{X_2-x}{h}\right\} \mid \beta\in R^3\right\}$$

are VC-subgraph classes and the VC-indices are independent of h.

Combining the Hoeffding decomposition of $\Pi_n(\beta)$, (2.22), (3.3), Lemmas 3.1 and 3.2, we have established that if $nh \to \infty$, $h \to 0$, and $|\beta - \beta_m| \to 0$,

$$\begin{aligned}\Gamma_n(\beta) - \Gamma_n(\beta_m) &= (\beta-\beta_m)^T(\Lambda + \mathrm{o_p}(1))(\beta-\beta_m)\\ &\quad + \frac{1}{h}(\beta-\beta_m)^T\frac{\partial Q}{\partial\beta}(\beta_m) + \mathrm{o_p}(|\beta-\beta_m|^2) + \mathrm{o_p}\left(\frac{1}{nh}\right)\end{aligned} \tag{3.9}$$

uniformly in β. Hence the proof is complete. □

We prove Lemma 3.1.

PROOF OF LEMMA 3.1. Consider the Taylor expansion of $Q(\beta)$ at β_m,

$$Q(\beta) = (\beta-\beta_m)^T\frac{\partial Q}{\partial\beta}(\beta_m) + \frac{1}{2}(\beta-\beta_m)^T\frac{\partial^2 Q}{\partial\beta\partial\beta^T}(\beta^*)(\beta-\beta_m). \tag{3.10}$$

We should show that

$$\sqrt{\frac{n}{h}}\frac{\partial Q}{\partial\beta}(\beta_m) \xrightarrow{D} \mathrm{N}(0,\Sigma) \tag{3.11}$$

and that

$$\frac{1}{h}\frac{\partial^2 Q}{\partial\beta\partial\beta^T}(\beta^*) = \mathrm{o}(1) \quad \text{uniformly in } \beta \quad \text{as} \quad |\beta-\beta_m| \to 0. \tag{3.12}$$

We consider (3.12) and then we go on to (3.11).

One of the conditional expectations in $Q(\beta)$ is written as

$$
\begin{aligned}
(3.13)\qquad & \mathrm{E}\{T(i,j;\beta) \mid Z_i\} \\
&= K\left(\frac{X_i}{h}\right) \\
&\quad \times \Bigg[I\{Y_i = 1\} \\
&\quad \times \int \left\{ \int_{-\infty}^{C_i+\beta_1+\beta_2 X_i/h-\beta_3(X_j-x)/h} (1 - F_\epsilon(C_j - g(X_j))) \right. \\
&\qquad \left. \times f_C(C_j \mid X_j) dC_j \right\} f_X(X_j) \frac{1}{h} K\left(\frac{X_j - x}{h}\right) dX_j \\
&\quad + I\{Y_i = 0\} \\
&\quad \times \int \left\{ \int_{C_i+\beta_1+\beta_2 X_i/h-\beta_3(X_j-x)/h}^{\infty} F_\epsilon(C_j - g(X_j)) \right. \\
&\qquad \left. \times f_C(C_j \mid X_j) dC_j \right\} f_X(X_j) \frac{1}{h} K\left(\frac{X_j - x}{h}\right) dX_j \Bigg].
\end{aligned}
$$

We deal with only (3.13) since the other term can be treated in the same way.

Assumptions A4 and A7, the definition of $U(i,j;\beta)$ in (3.3), and (3.13) imply

$$
(3.14)\qquad \left| \frac{1}{h} \frac{\partial^2}{\partial\beta\partial\beta^T} [\mathrm{E}\{T(i,j;\beta) \mid Z_i\} - \mathrm{E}\{T(i,j;\eta) \mid Z_i\}] \right|_\infty \le \frac{M_2}{h} K\left(\frac{X_i}{h}\right) |\beta - \eta|,
$$

where $|\cdot|_\infty$ denotes the sup norm in $R^{3\times 3}$.

Noting that there is no term related to $\mathrm{E}\{T(i,j;\beta_m)\}$ in (3.12), we have from (3.14) that

$$
\begin{aligned}
(3.15)\qquad & \left| \frac{1}{nh} \sum_{i=1}^n \frac{\partial^2}{\partial\beta\partial\beta^T} [\mathrm{E}\{U(i,j;\beta^*) \mid Z_i\} - \mathrm{E}\{\Pi_n(\beta^*)\}] \right|_\infty \\
&\le |\beta^* - (\beta_1^0, 0, 0)^T| \frac{M_3}{nh} \sum_{i=1}^n K\left(\frac{X_i}{h}\right) \\
&\quad + M_4 \Bigg| \frac{1}{nh} \sum_{i=1}^n \frac{\partial^2}{\partial\beta\partial\beta^T} [\mathrm{E}\{T(i,j;(\beta_1^0,0,0)^T) \mid Z_i\} \\
&\qquad\qquad - \mathrm{E}\{\Pi_n((\beta_1^0,0,0)^T)\}] \Bigg|.
\end{aligned}
$$

The first term on the right-hand side tends to 0 in probability uniformly in β since $|\beta^* - (\beta_1^0, 0, 0)^T|$ tends to 0 in probability uniformly as $|\beta - \beta_m| \to 0$. As for the second term, we can prove that it tends to 0 in probability by evaluating the second moment. The details are omitted since the calculation is standard one in the literature on nonparametric regression. Hence (3.12) is proved.

Next we deal with (3.11). Since

$$\frac{\partial}{\partial\beta}\mathrm{E}\{\Pi_n(\beta_m)\} = h\frac{\partial}{\partial\beta}\mathrm{E}\{\Gamma(\beta_m)\} = 0,$$

we can represent $\frac{1}{h}\frac{\partial Q}{\partial\beta}(\beta_m)$ as

$$\text{(3.16)} \qquad \frac{1}{h}\frac{\partial Q}{\partial\beta}(\beta_m) = \frac{1}{nh}\left[\sum_{i=1}^{n}\frac{\partial}{\partial\beta}\mathrm{E}\{T(i,j;\beta_m)\mid Z_i\} + \sum_{j=1}^{n}\frac{\partial}{\partial\beta}\mathrm{E}\{T(i,j;\beta_m)\mid Z_j\}\right].$$

We examine only the first term closely since the second term can be treated in the same way. The summand of the first term can be written as

$$\begin{aligned}
\text{(3.17)} \qquad & \left(\frac{\partial}{\partial\beta}\mathrm{E}\{T(i,j;\beta_m)\mid Z_i\}\right)^T \\
&= K\left(\frac{X_i}{h}\right) \\
&\times \Bigg[I\{Y_i=1\}\int\left(1,\frac{X_i}{h},-\frac{X_j-x}{h}\right) \\
&\times \left(1-F_\epsilon\left(C_i+\beta_1^m+\beta_2^m\frac{X_i}{h}-\beta_3^m\frac{X_j-x}{h}-g(X_j)\right)\right) \\
&\times f_C\left(C_i+\beta_1^m+\beta_2^m\frac{X_i}{h}-\beta_3^m\frac{X_j-x}{h}\mid X_j\right) \\
&\times f_X(X_j)\frac{1}{h}K\left(\frac{X_j-x}{h}\right)dX_j \\
&- I\{Y_i=0\}\int\left(1,\frac{X_i}{h},-\frac{X_j-x}{h}\right) \\
&\times F_\epsilon\left(C_i+\beta_1^m+\beta_2^m\frac{X_i}{h}-\beta_3^m\frac{X_j-x}{h}-g(X_j)\right) \\
&\times f_C\left(C_i+\beta_1^m+\beta_2^m\frac{X_i}{h}-\beta_3^m\frac{X_j-x}{h}\mid X_j\right) \\
&\times f_X(X_j)\frac{1}{h}K\left(\frac{X_j-x}{h}\right)dX_j\Bigg].
\end{aligned}$$

Assuming that $nh\to\infty$ and $h\to 0$, we can verify (3.11) by evaluating the second and third moments of (3.17) and appealing to the Lyapounov CLT, for example, on p. 362 of Billingsley (1994). We omit the details since this kind of argument is standard one in the literature on nonparametric regression. The covariance matrix Σ is defined in Theorem 3.1. □

4. Simulation study

We present the result of a simulation study to see the effect of the bandwidth and how the proposed procedure works. We have no automatic bandwidth selection rule like cross validation at present.

The data is generated by

$$\begin{aligned}
&X \sim \text{Unif}(-1.5, 1.5),\\
&C \sim \text{N}(0, 4),\\
&\epsilon \sim \text{N}(0, 0.3), \qquad \text{and}\\
&\tilde{Y} = x + x^2 + \epsilon.
\end{aligned}$$

We estimate $g(1) - g(0) = 2$ and $g(-1) - g(0) = 0$ by the local constant estimator in Remark 1.1 and the local linear estimator defined by (1.6). The kernel is the uniform kernel on $[-1, 1]$. The sample size is 500 and the tables are based on 200 repetitions for each entry.

In the above setup the covariate X, the censoring time C, and the error term ϵ are mutually independent. The censoring time C has a large variance compared to the error term. The censoring time and the error term take the values in $(-\infty, \infty)$. This means that we are considering cases like that of log-transformed failure times. Note that assumption A5 hold for any $g(\cdot)$. All the other assumptions are satisfied, too.

We maximize the objective functions by grid search. For the local constant estimator, the grid width is 0.05. For $\Gamma_n(\cdot)$ in (1.6), the grid search is carried out in two steps. At first the gird width is 0.1 for every β_i. Then we do the second step around the maximum solution of the first step. The gird width is 0.05 for every β_i.

As we mentioned in Section 1, the maximum solutions are not unique. We choose the average of the maximum solutions as the estimate of $g(x) - g(0)$.

We try 0.2, 0.25, and 0.3 for the bandwidth. SPLUS is used for the study on a personal computer. Tables 1 and 2 present the means and MSEs for the local constant estimator and the local linear estimator, respectively.

There is no significant difference between the local constant estimator and the local linear estimator. The maximization is complicated for the local linear estimator than for the local constant estimator. If the maximization and the choice of the estimate do not work for the local linear estimator, the performances of the local linear estimator

Table 1. Means and MSEs of the local constant estimator.

h		0.20	0.25	0.30
$x = 1.0$	Mean	1.99	1.98	1.93
	MSE	0.094	0.097	0.073
$x = -1.0$	Mean	−0.04	0.01	−0.02
	MSE	0.065	0.054	0.044

Table 2. Means and MSEs of the local linear estimator.

h		0.20	0.25	0.30
$x = 1.0$	Mean	2.00	2.02	2.00
	MSE	0.085	0.078	0.064
$x = -1.0$	Mean	−0.04	0.00	−0.02
	MSE	0.063	0.048	0.040

will be worse. Since there is no significant difference, maximization by grid search and the choice of the estimate seems to work well.

The effect of the variance is much more serious than that of the bias. Larger bandwidths may be recommended.

5. Technical proofs

In this section we prove (2.15) and Lemma 3.2.

PROOF OF (2.15). We establish that for any M_1, there exists a positive number M_2 such that

$$\tilde{\Gamma}(\beta) < \tilde{\Gamma}(\beta_0) - M_1 h^2 \qquad \text{for} \quad |\beta - \beta_0| > M_2 h.$$

We need to evaluate the difference between $P(Y_1^* > Y_2^\dagger \mid \tilde{Z}_1, \tilde{Z}_2)$ and $P(Y_1^* < Y_2^\dagger \mid \tilde{Z}_1, \tilde{Z}_2)$. From the definition of the conditional probability, we have

$$\begin{aligned}(5.1) \qquad & P(Y_1^* > Y_2^\dagger \mid \tilde{Z}_1, \tilde{Z}_2) - P(Y_1^* < Y_2^\dagger \mid \tilde{Z}_1, \tilde{Z}_2) \\ &= F_\epsilon\left(C_1 - g(0) + \beta_2^0 \frac{X_1}{h}\right)\left\{1 - F_\epsilon\left(C_2 - g(x) + \beta_3^0 \frac{X_2 - x}{h}\right)\right\} \\ &\quad - \left\{1 - F_\epsilon\left(C_1 - g(0) + \beta_2^0 \frac{X_1}{h}\right)\right\} F_\epsilon\left(C_2 - g(x) + \beta_3^0 \frac{X_2 - x}{h}\right).\end{aligned}$$

By the Taylor expansion, we have

$$(5.2) \qquad F_\epsilon(z)(1 - F_\epsilon(z+\delta)) - (1 - F_\epsilon(z))F_\epsilon(z+\delta) \approx -f_\epsilon(z)\delta \qquad \text{for small} \quad \delta.$$

Then (5.1), (5.2), and assumption A4 imply that there exist M_3, M_4, and M_5 which are independent of n, X_1/h, and $(X_2 - x)/h$ and satisfy

$$\begin{aligned}(5.3) \qquad & |P(Y_1^* > Y_2^\dagger \mid \tilde{Z}_1, \tilde{Z}_2) - P(Y_1^* < Y_2^\dagger \mid \tilde{Z}_1, \tilde{Z}_2)| \\ &\geq M_3 \left| C_1 - C_2 + \beta_1^0 + \beta_2^0 \frac{X_1}{h} - \beta_3^0 \frac{X_2 - x}{h} \right| \\ \text{on} \quad & \left\{ \left| C_1 - C_2 + \beta_1^0 + \beta_2^0 \frac{X_1}{h} - \beta_3^0 \frac{X_2 - x}{h} \right| < M_4 \quad \text{and} \quad |C_1| < M_5 \right\}\end{aligned}$$

(5.3) is crucial to the proof of (2.15).

We deal with two cases where $|\beta_1 - \beta_1^0| > M_2 h/9$ and where $|\beta_1 - \beta_1^0| \leq M_2 h/9$ separately.

When $\beta_1 - \beta_1^0 > M_2 h/9$, we have that

$$\begin{aligned}(5.4) \qquad & \left| \left(-\beta_1 - \beta_2 \frac{X_1}{h} + \beta_3 \frac{X_2 - x}{h} \right) - \left(-\beta_1^0 - \beta_2^0 \frac{X_1}{h} + \beta_3^0 \frac{X_2 - x}{h} \right) \right| \geq \frac{M_2 h}{9} \\ \text{on} \quad & (\{\beta_2 \geq \beta_2^0, X_1 \geq 0\} \cup \{\beta_2 \leq \beta_2^0, X_1 \leq 0\}) \\ & \cap (\{\beta_3 \geq \beta_3^0, X_2 \leq x\} \cup \{\beta_3 \leq \beta_3^0, X_2 \geq x\}).\end{aligned}$$

Combining (5.3) and (5.4), we obtain

$$(5.5) \qquad \tilde{\Gamma}(\beta) \leq \tilde{\Gamma}(\beta_0) - M_2 M_6 h^2 \qquad \text{for some} \quad M_6$$

by integrating $F(\tilde{Z}_1, \tilde{Z}_2; \beta)$ in (2.10) on

$$\left\{ C_1 - C_2 \mid -\beta_1 - \beta_2 \frac{X_1}{h} + \beta_3 \frac{X_2 - x}{h} < C_1 - C_2 < -\beta_1^0 - \beta_2^0 \frac{X_1}{h} + \beta_3^0 \frac{X_2 - x}{h} \right\}.$$

The case where $\beta_1 - \beta_1^0 < -M_2 h/9$ can be treated in the same way.

When $|\beta_1 - \beta_1^0| \leq M_2 h/9$ and $\beta_2 - \beta_2^0 > 6M_2 h/9$, we have that

$$\left| \left(-\beta_1 - \beta_2 \frac{X_1}{h} + \beta_3 \frac{X_2 - x}{h} \right) - \left(-\beta_1^0 - \beta_2^0 \frac{X_1}{h} + \beta_3^0 \frac{X_2 - x}{h} \right) \right| \geq \frac{2}{9} M_2 h \tag{5.6}$$

$$\text{on} \quad \left\{ \frac{X_1}{h} \geq \frac{1}{2} \right\} \cap (\{\beta_3 \geq \beta_3^0, X_2 \leq x\} \cup \{\beta_3 \leq \beta_3^0, X_2 \geq x\}).$$

Combining (5.3) and (5.6), we obtain

$$\tilde{\Gamma}(\beta) \leq \tilde{\Gamma}(\beta_0) - M_2 M_7 h^2 \quad \text{for some} \quad M_7 \tag{5.7}$$

by integrating $F(\tilde{Z}_1, \tilde{Z}_2; \beta)$ in (2.10) on

$$\left\{ C_1 - C_2 \mid -\beta_1 - \beta_2 \frac{X_1}{h} + \beta_3 \frac{X_2 - x}{h} < C_1 - C_2 < -\beta_1^0 - \beta_2^0 \frac{X_1}{h} + \beta_3^0 \frac{X_2 - x}{h} \right\}.$$

The other cases can be treated in the same way.

The proof is finished by taking a sufficiently large M_2 in (5.5), (5.7) and so on. □

Next we prove Lemma 3.2.

PROOF OF LEMMA 3.2. We should prove what is similar to (5.3.11)' in the proof of Theorem 5.3.7 of de la Peña and Giné (1999) with $k = 2$ (see (5.16) below) and that

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \sup_{|\beta - \beta_m| < \delta} \mathrm{E}\{(T(1,2;\beta) - T(1,2;\beta_m))^2\} = 0. \tag{5.8}$$

(5.8) is established from the definition of $T(i,j;\beta)$ in (3.2) and assumption A7 by considering the conditional expectations on $\{X_i\}$ of

$$\begin{aligned} &\left| I\left\{ C_1 + \beta_1 + \beta_2 \frac{X_1}{h} > C_2 + \beta_3 \frac{X_2 - x}{h} \right\} \right. \\ &\qquad \left. - I\left\{ C_1 + \beta_1^0 + \beta_2^0 \frac{X_1}{h} > C_2 + \beta_3^0 \frac{X_2 - x}{h} \right\} \right| \end{aligned} \tag{5.9}$$

and the other analogous terms. We omit the calculation.

Now we go on to the proof of (5.16) below. Recalling the definition of $U(i,j;\beta)$ in (3.3), we define $V(i,j;\beta)$ for symmetry with respect to Z_i and Z_j by

$$V(i,j;\beta) = \frac{1}{2}\{U(i,j;\beta) + U(j,i;\beta)\}. \tag{5.10}$$

We adopt the notations of de la Peña and Giné (1999) since we follow and adapt the proof of (5.3.11)' of de la Peña and Giné (1999), which addresses equicontinuity of degenerate

U-processes. In the proof of (5.3.11)' the class of kernels is fixed and the envelope function H satisfies

$$\lim_{M\to\infty} \mathrm{E}[H^2 I\{H > M\}] = 0.$$

However, neither of them holds in the setup of this paper. Thus we must modify the arguments of de la Peña and Giné (1999). The proof of (5.23) below is different from that of de la Peña and Giné (1999). The other parts of the proof of (5.16) below are essentially the same as on pp. 244–246 of de la Peña and Giné (1999). However, it is very important to make sure that the varying kernel classes do not affect the arguments.

We define π_2, π_1, and $U_n^{(2)}$ for $W(\cdot,\cdot)$ by

$$\begin{aligned}(5.11)\qquad \pi_2 W(Z_1,Z_2) &= W(Z_1,Z_2) - \mathrm{E}\{W(Z_1,Z_2) \mid Z_1\} \\ &\quad - \mathrm{E}\{W(Z_1,Z_2) \mid Z_2\} + \mathrm{E}\{W(Z_1,Z_2)\}\end{aligned}$$

$$(5.12)\qquad \pi_1 W(Z_1) = \mathrm{E}\{W(Z_1,Z_2) \mid Z_1\} - \mathrm{E}\{W(Z_1,Z_2)\}$$

and

$$(5.13)\qquad U_n^{(2)} W = \frac{1}{n(n-1)} \sum_{i\neq j} W(Z_i,Z_j)$$

and $U_n^{(1)} W$ for $W(\cdot)$ by

$$(5.14)\qquad U_n^{(1)} W = \frac{1}{n}\sum_{i=1}^{n} W(Z_i).$$

By using the above notations, the left-hand side of the expression of Lemma 3.2 is written as

$$(5.15)\qquad U_n^{(2)}\pi_2 V(\eta) \qquad \text{with } \beta \text{ replaced by } \eta.$$

We sometimes use η, η_1, and η_2 instead of β to avoid confusion. Then we have to prove

$$(5.16)\qquad \lim_{\delta\to 0}\limsup_{n\to\infty} \mathrm{E}\left[\sup_{d(\eta_1,\eta_2)<\delta} n|U_n^{(2)}(\pi_2(V(\eta_1)-V(\eta_2)))|\right] = 0,$$

where

$$d^2(\eta_1,\eta_2) = \mathrm{E}\{(V(Z_1,Z_2;\eta_1) - V(Z_1,Z_2;\eta_2))^2\}.$$

If (5.16) is established, (5.8) implies the lemma since $V(Z_1,Z_2;\beta_m) = 0$. What is essential to the proof is that the covering number of

$$\{V(Z_1,Z_2;\beta) \mid \beta \in R^3\}$$

has an upper bound independent of h. The bound is due to the following two facts.

1) $\{I\{C_1 + \beta_1 + \beta_2 \frac{X_1}{h} > C_2 + \beta_3 \frac{X_2 - x}{h}\} \mid \beta \in R^3\}$ is a VC-subgraph class with the VC-index no more than 6 from Example 3.7.4c of Geer (2000).

2) The proof of Lemma 2.6.18 (vi) of Vaart and Wellner (1996) implies that multiplying the class in 1) by $K(\frac{X_1}{h})K(\frac{X_2-x}{h})I\{Y_1 > Y_2\}$ increases the VC-index of the function class in 1) by at the most only one. It is because no subset of two points can be shattered on the set where all the functions vanish.

Refer to Vaart and Wellner (1996) for the definitions of covering number, envelope function, VC-subgraph class, and VC-index. The other three classes of functions appearing in the definition of $V(Z_1,Z_2;\beta)$ can be treated in the same way.

We put

$$\mathcal{G}_h = \{V(Z_1, Z_2; \beta) \mid \beta \in R^3\} \tag{5.17}$$

and the envelope function is given by

$$H(Z_1, Z_2) = \frac{1}{h}K\left(\frac{X_1}{h}\right)K\left(\frac{X_2 - x}{h}\right) + \frac{1}{h}K\left(\frac{X_1 - x}{h}\right)K\left(\frac{X_2}{h}\right). \tag{5.18}$$

By Applying Theorem 2.6.7 of Vaart and Wellner (1996) to the classes of functions appearing in the definition of $V(Z_1, Z_2; \beta)$ and combining the upper bounds from Theorem 2.6.7, we obtain the following upper bound of the covering number of $\mathcal{G}_h$, $N(\mathcal{G}_h, d_p, \epsilon\|H\|_p)$,

$$N(\mathcal{G}_h, d_p, \epsilon\|H\|_p) \leq M_8(\epsilon^{-M_9} \vee 1), \tag{5.19}$$

where P is an arbitrary probability measure and for $p \geq 1$,

$$d_p(f_1, f_2) = \left(\int |f_1 - f_2|^p dP\right)^{1/p} \quad \text{and} \quad \|H\|_p = \left(\int H^p dP\right)^{1/p}.$$

Equivalently we have

$$N(\mathcal{G}_h, d_p, \epsilon) \leq M_8\left(\frac{\|H\|_p}{\epsilon}\right)^{M_9} \vee 1. \tag{5.20}$$

We just give the necessary modifications and do not reproduce the arguments of de la Peña and Giné (1999) in this paper.

Defining $e^2_{n,2,2}(V, W)$ and $D_{n,2}(\delta)$ for $V(\cdot, \cdot)$ and $W(\cdot, \cdot)$ as in de la Peña and Giné (1999) by

$$e^2_{n,2,2}(V, W) = n^2\binom{n}{2}^{-2}\sum_{i<j}\{\pi_2(V - W)(Z_i, Z_j)\}^2 \tag{5.21}$$

and

$$D_{n,2}(\delta) = \left\{\sup_{\eta_1,\eta_2}\left|e^2_{n,2,2}(V(\eta_1), V(\eta_2)) - n^2\binom{n}{2}^{-1}\mathrm{E}\{(\pi_2(V(\eta_1) - V(\eta_2)))^2\}\right| > 2^2 2\delta^2\right\}, \tag{5.22}$$

we prove in Lemma 5.1 below that

$$\lim_{n\to\infty} \mathrm{P}(D_{n,2}(\delta)) = 0 \quad \text{for any} \quad \delta > 0. \tag{5.23}$$

The proof of (5.23), which corresponds to Lemma 5.3.6 of de la Peña and Giné (1999), is different from that of the lemma.

We also need to show that

$$U_n^{(2)}H^2 - \mathrm{E}\{H^2\} \to 0 \quad \text{in probability} \tag{5.24}$$

and

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\{H^2(Z_i, Z_{n+1}) \mid Z_i\} - \mathrm{E}\{H^2\} \to 0 \quad \text{in probability.} \tag{5.25}$$

These can be proved by appealing to Hoeffding's inequality on p. 165 of de la Peña and Giné (1999) if $nh \to \infty$ and $h \to 0$. Besides we have

$$\lim_{h\to 0} \mathrm{E}\{H^2\} = 2f_X(0)f_X(x)\left(\int K^2(t)dt\right)^2. \tag{5.26}$$

Since the arguments on pp. 244–246 of de la Peña and Giné (1999) are valid with (5.19), (5.20), (5.23)–(5.26), and minor modifications, the proof is complete. □

Lemma 5.1 corresponds to Lemma 5.3.6 of de la Peña and Giné (1999). We cannot apply the proof of the lemma to the setup of this paper since we do not have

$$\lim_{M\to\infty} \mathrm{E}[H^2 I\{H > M\}] = 0,$$

where H is the envelope function of $\mathcal{G}_h$. We have to go back to Theorem 5.2.2 and modify the proof of the theorem.

LEMMA 5.1. *For any $\delta > 0$, we have*

$$\mathrm{P}(D_{n,2}(\delta)) \to 0 \quad \text{as} \quad nh \to \infty \quad \text{and} \quad h \to 0. \tag{5.27}$$

PROOF. We prove that

$$\begin{aligned}\sup_{\eta_1,\eta_2} &|U_n^{(2)}(\pi_2(V(\eta_1) - V(\eta_2)))^2 - \mathrm{E}\{(\pi_2(V(\eta_1) - V(\eta_2)))^2\}| \\ &\to 0 \quad \text{in probability.}\end{aligned} \tag{5.28}$$

This is equivalent to Lemma 5.1. We define $\mathcal{H}_h$ by

$$\mathcal{H}_h = \{(\pi_2(V(Z_1, Z_2; \eta_1) - V(Z_1, Z_2; \eta_2)))^2 \mid \eta_1, \eta_2 \in R^3\}. \tag{5.29}$$

Letting $\{\tilde{\epsilon}_i\}$ be independent Rademacher variables, we have from the argument on pp. 226–227 of de la Peña and Giné (1999) that

$$\lim_{n\to\infty} \mathrm{E}\left\{\sup_{W\in\mathcal{H}_h}\left|\frac{(n-2)!}{n!}\sum_{i\neq j}\tilde{\epsilon}_i W(Z_i, Z_j)\right|\right\} = 0 \tag{5.30}$$

$$\Leftrightarrow \lim_{n\to\infty} \mathrm{E}\left\{\sup_{W\in\mathcal{H}_h}\left|\frac{(n-2)!}{n!}\sum_{i\neq j}\tilde{\epsilon}_i (W(Z_i, Z_j) - \mathrm{E}\{W\})\right|\right\} = 0 \tag{5.31}$$

$$\Leftrightarrow \lim_{n\to\infty} \mathrm{E}\left\{\sup_{W\in\mathcal{H}_h} |U_n^{(2)}W - \mathrm{E}\{W\}|\right\} = 0 \tag{5.32}$$

$$\Rightarrow \sup_{W\in\mathcal{H}_h} |U_n^{(2)}W - \mathrm{E}\{W\}| = 0 \quad \text{in probability.} \tag{5.33}$$

Therefore we prove (5.30). As in de la Peña and Giné (1999), we define $\tilde{e}_{n,1}$ and $e_{n,1}$ for $V(\cdot,\cdot)$ and $W(\cdot,\cdot)$ by

$$\tilde{e}_{n,1}(V,W) = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{1}{n-1}\sum_{j=1,j\neq i}^{n}(V-W)(Z_i,Z_j)\right| \tag{5.34}$$

and

$$e_{n,1}(V,W) = U_n^{(2)}(|V-W|). \tag{5.35}$$

Taking a τ-dense set $\mathcal{H}_{h,\tau}$ in $\mathcal{H}_h$ with respect to $\tilde{e}_{n,1}$, we have

$$\begin{aligned} &\mathrm{E}\left\{\sup_{W\in\mathcal{H}_h}\left|\frac{(n-2)!}{n!}\sum_{i\neq j}\tilde{\epsilon}_i W(Z_i,Z_j)\right| \mid \{Z_i\}\right\} \\ &\quad\leq \tau + \mathrm{E}\left\{\sup_{W\in\mathcal{H}_{h,n}}\left|\frac{(n-2)!}{n!}\sum_{i\neq j}\tilde{\epsilon}_i W(Z_i,Z_j)\right| \mid \{Z_i\}\right\}. \end{aligned} \tag{5.36}$$

Then (5.2.8) of de la Peña and Giné (1999) implies

$$\begin{aligned} &\mathrm{E}\left\{\sup_{W\in\mathcal{H}_{h,n}}\left|\frac{(n-2)!}{n!}\sum_{i\neq j}\tilde{\epsilon}_i W(Z_i,Z_j)\right| \mid \{Z_i\}\right\} \\ &\quad\leq M_{10}\frac{(n-2)!}{n!}[\log 2 + \log N(\mathcal{H}_h,\tilde{e}_{n,1},\tau)]^{1/2} \\ &\qquad\times \sup_{W\in\mathcal{H}_{h,n}}\left\{\sum_{i=1}^{n}\left(\sum_{j=1,j\neq i}^{n}W(Z_i,Z_j)\right)^2\right\}^{1/2}. \end{aligned} \tag{5.37}$$

Replacing W in the last term in (5.37) with the envelope function H, we obtain

$$\begin{aligned} &\mathrm{E}\left\{\sup_{W\in\mathcal{H}_{h,n}}\sum_{i=1}^{n}\left(\sum_{j=1,j\neq i}^{n}W(Z_i,Z_j)\right)^2\right\} \\ &\quad\leq M_{11}\mathrm{E}\left\{\sum_{i=1}\left\{\sum_{j=1,j\neq i}^{n}\left(\frac{1}{h^2}\left(K\left(\frac{X_i}{h}\right)K\left(\frac{X_j-x}{h}\right)\right.\right.\right.\right. \\ &\qquad\qquad\left. + K\left(\frac{X_i-x}{h}\right)K\left(\frac{X_j}{h}\right)\right)^2 \\ &\qquad + \left(K\left(\frac{X_i}{h}\right) + K\left(\frac{X_i-x}{h}\right)\right. \\ &\qquad\qquad\left.\left.\left.\left. + K\left(\frac{X_j}{h}\right) + K\left(\frac{X_j-x}{h}\right) + h\right)^2\right)\right\}^2\right\} \\ &\quad\leq \frac{M_{12}}{h^4}\sum_{i=1}^{n}h(nh+n^2h^2) \leq M_{13}\left(\frac{n^2}{h^2}+\frac{n^3}{h}\right). \end{aligned} \tag{5.38}$$

(5.30) follows from (5.36)–(5.38) if we establish that for any $\tau > 0$,

$$\lim_{n\to\infty} \frac{1}{(nh)^{1/2}} (\mathrm{E}\{\log N(\mathcal{H}_h, \tilde{e}_{n,1}, \tau)\})^{1/2} = 0. \tag{5.39}$$

Since $\tilde{e}_{n,1}(\cdot,\cdot) \le e_{n,1}(\cdot,\cdot)$, we have

$$N(\mathcal{H}_h, \tilde{e}_{n,1}, \tau) \le N(\mathcal{H}_h, e_{n,1}, \tau). \tag{5.40}$$

(5.40) implies that (5.39) follows if (5.41) below is verified.

$$\limsup_{n\to\infty} \mathrm{E}\{\log N(\mathcal{H}_h, e_{n,1}, \tau)\} < \infty \quad \text{for any} \quad \tau > 0. \tag{5.41}$$

We evaluate $e_{n,1}$ to prove (5.41). Since

$$\begin{aligned} &e_{n,1}((\pi_2(V(\eta_1) - V(\eta_2)))^2, (\pi_2(V(\theta_1) - V(\theta_2)))^2) \\ &\quad \le \frac{M_{14}}{h} U_n^{(2)}(|\pi_2(V(\eta_1) - V(\eta_2)) - \pi_2(V(\theta_1) - V(\theta_2))|) \\ &\quad \le \frac{M_{14}}{h}(e_{n,1}(\pi_2 V(\eta_1), \pi_2 V(\theta_1)) + e_{n,1}(\pi_2 V(\eta_2), \pi_2 V(\theta_2))), \end{aligned} \tag{5.42}$$

we have

$$N(\mathcal{H}_h, e_{n,1}, \tau) \le N^2\left(\tilde{\mathcal{G}}_h, e_{n,1}, \frac{\tau h}{2M_{14}}\right), \tag{5.43}$$

where

$$\tilde{\mathcal{G}}_h = \{\pi_2 V(\eta) \mid \eta \in R^3\}.$$

(5.20) implies that

$$\begin{aligned} &\log N^2\left(\tilde{\mathcal{G}}_h, e_{n,1}, \frac{\tau h}{2M_{14}}\right) \\ &\quad \le 8\log M_8 + 2M_9 \log\left(1 + \frac{2M_{14}}{h\tau} U_n^{(2)} H\right) + 2M_9 \log\left(1 + \frac{2M_{14}}{h\tau}\mathrm{E}\{H\}\right) \\ &\qquad + 4M_9 \log\left(1 + \frac{2M_{14}}{h\tau}\frac{1}{n}\sum_{i=1}^n \mathrm{E}\{H(Z_i, Z_{n+1}) \mid Z_i\}\right). \end{aligned} \tag{5.44}$$

We have from Jensen's inequality that

$$\mathrm{E}\left\{\log N^2\left(\mathcal{G}_h, e_{n,1}, \frac{\tau h}{2M_{14}}\right)\right\} \le 8\log M_8 + 8M_9 \log\left(1 + \frac{2M_{14}}{h\tau}\mathrm{E}\{H\}\right). \tag{5.45}$$

Since $h^{-1}\mathrm{E}\{H\}$ is bounded from above, (5.41) follows from (5.43)–(5.45). Hence the lemma is proved. □

Acknowledgements

The author appreciates helpful comments of two referees very much.

References

Abrevaya, J. (1999). Rank regression for current status data: asymptotic normality, *Statistics & Probability Letters*, **43**, 275–287.

Andrews, C., van der Laan, M. and Robins, J. (2001). Locally efficient estimation of regression parameters using current status data, Tech. Report, Department of Statistics, University of California, Berkeley (available at www.stat.berkeley.edu/~laan/).

Billingsley, P. (1994). *Probability and Measure*, 3rd ed., Wiley, New York.

Chen, S. (2000*a*). Efficient estimation of binary choice models under symmetry, *Journal of Econometrics*, **96**, 183–199.

Chen, S. (2000*b*). Rank estimation of a location parameter in the binary choice model, *Journal of Econometrics*, **98**, 317–334.

Cosslett, S. R. (1987). Efficiency bounds for distribution-free estimation of the binary choice and the censored regression models, *Econometrica*, **55**, 559–585.

de la Peña, V. H. and Giné, E. (1999). *Decoupling: From Dependence to Independence*, Springer, New York.

Fan, J. and Gijbels, I. (1994). Censored regression: Local linear approximations and their applications, *Journal of the American Statistical Association*, **89**, 560–570.

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

Han, A. K. (1987). Non-parametric analysis of a generalized regression model: The maximum rank correlation estimation, *Journal of Econometrics*, **35**, 303–316.

Horowitz, J. L. (2001). Nonparametric estimation of a generalized additive model with an unknown link function, *Econometrica*, **69**, 499–513.

Huang, J. (1996). Efficient estimation for the Cox model with interval censoring, *The Annals of Statistics*, **24**, 540–568.

Huang, J. and Wellner, J. A. (1995). Asymptotic normality of the NPMLE of linear functionals for interval censored data, case1, *Statistica Neerlandica*, **49**, 153–163.

Huang, J. and Wellner, J. A. (1997). Interval censoring survival data: A review of recent progress, *Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis* (eds. D. Y. Lin and T. R. Flemming), Lecture Notes in Statistics, No. 123, 123–170, Springer, New York.

Klein, R. W. and Spady, R. H. (1993). An efficient semiparametric estimator for binary response models, *Econometrica*, **61**, 387–421.

Li, G. and Zhang, C.-H. (1998). Linear regression with interval censored data, *The Annals of Statistics*, **26**, 1306–1327.

Murphy, S. A., van der Vaart, A. W. and Wellner, J. A. (1999). Current status regression, *Mathematical Methods of Statistics*, **8**, 407–425.

Shen, X. (1997). On methods of sieves and penalization, *The Annals of Statistics*, **25**, 2555–2591.

Shen, X. (2000). Linear regression with current status data, *Journal of the American Statistical Association*, **95**, 842–852.

Sherman, R. P. (1993). The limiting distribution of the maximum rank correlation estimator, *Econometrica*, **61**, 123–137.

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

van der Laan, M. and Robins, J. (1998). Locally efficient estimation with current status data and time-dependent covariates, *Journal of the American Statistical Association*, **93**, 693–701.

van der Laan, M., Bickel, P. J. and Jewell, N. P. (1997). Singly and doubly censored current status data: Estimation, asymptotics and regression, *Scandinavian Journal of Statistics*, **24**, 289–307.

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