# SURVIVAL TIME REGRESSION INVOLVING COVARIATE MEASUREMENT ERROR

Truong, Young K.
Department of Biostatistics, University of North Carolina at Chapel Hill

# SURVIVAL TIME REGRESSION INVOLVING COVARIATE MEASUREMENT ERROR

By

Young K. Truong[*][†]

## Abstract

A class of deconvoluted kernel estimators is proposed to estimate the conditional survival, conditional cumulative hazard and the regression functions when the covariate is subject to measurement error. Under appropriate conditions, it is shown that the proposed estimators are consistent, and that they converge weakly to Gaussian processes.

*Key words: Conditional hazard function; Survival Function; Nonparametric regression; Covariate measurement errors; Deconvoluted kernel.*

## 1. Introduction

Prentice (1982) considered an important problem in survival analysis which deals with the estimation of the failure time regression model based on censored survival data when the covariate is subject to measurement error. The following study was described to support the approach. To relate thyroid cancer incidence to gamma radiation exposure level, survivors of the atomic bombs in Hiroshima and Nagasaki were assigned radiation dose level based on the location and the individual shielding characteristics. It was reported that individual radiation dose estimates may differ from actual exposure levels by as much as 30% in an appreciable fraction of study subjects.

The current paper is motivated by the study described above. Specifically, let $F(t|u)$ denote the conditional distribution function of a nonnegative random variable $T$ given $U = u$. Let $C$ be a random variable, set $Y = T \wedge C$ and $\delta = 1(T \leq C)$. In survival analysis, $T$ is usually the survival time (or failure time) or time to an event, and $U$ is the covariate; $C$ is the censoring time or withdrawal time and $Y$ is the observable time which may be the time to an event or time to withdraw. Also, the indicator variable $\delta$ equals 1 if the event occurs on or before the censoring time (if $T \leq C$) and it equals 0 otherwise.

Suppose the covariate $U$ is subject to measurement error and is not directly available; that is, it is only possible to observe $U$ through $X = U + Z$, where $Z$ is a random

error. Given a random sample from the distribution of $(X, Y, \delta)$, it is of interest to examine the covariate effect on the survival time by estimating the conditional cumulative hazard, the conditional survival function $S(t|u) = 1 - F(t|u)$ which is the probability that the survival time $T$ exceeds $t$ given $U = u$, the regression function $E(T|U = u)$, the conditional mean of survival time $T$ given $U = u$.

In this paper, our primary interest is on nonparametric estimation of $S(t|u)$ in the sense that there will be no parametric functional form imposed on the effect of $U$ on $T$. Moreover, the proposed procedure will be flexible enough to account for the censoring in the survival time and the measurement error in the covariate. To make the nonparametric problem identifiable, it is assumed (as in Fan and Truong (1993)) that the error $Z$ has a known distribution, and it is independent of $U$, $T$ and $C$. Also, $T$ and $C$ are conditionally independent given $U$.

In the absence of measurement errors, many approaches have been proposed for regression analyses with censored survival data. Parametric models are described in Elandt-Johnson and Johnson (1980), partly nonparametric (semi-parametric) methods are considered by Cox (1972), Miller (1976) and Buckley and James (1979). For further extensions and generalizations, see Kalbfleisch and Prentice (1980), Fleming and Harrington (1991), and Anderson et al. (1992). To explore structural relationships between the survival time and its covariates, especially for large data sets, Beran (1981) proposed a nonparametric approach for estimating the survival function. See also Doksum and Yandell (1982), Dabrowska (1987) and the references given therein for other nonparametric methods.

For uncensored observations, regression analysis involving measurement errors has been an important research topic. See Fuller (1987) for a survey on linear models; Carroll, et al. (1984), Stefanski and Carroll (1985) and Prentice (1986) for binary regression analyses; Armstrong (1985) for generalized linear models. To develop flexible procedures for exploratory data analyses, Fan and Truong (1993) proposed a nonparametric approach using the deconvoluted kernel method. A brief history of this development is given in Fan, Truong and Wang (1990, 1991).

For censored survival data in the presence of covariate measurement errors, Prentice (1982) proposed a method based on Cox's regression model. See also Pepe, Self and Prentice (1989). In view of the potential that bias may be introduced due to model misspecification in parametric or partly nonparametric estimation of the conditional survival function, the current approach generalizes Beran's nonparametric method by using the deconvoluted kernel estimator of Fan and Truong (1993) to account for the covariate measurement errors.

The rest of the paper is organized as follows. Section 2 describes the nonparametric methods of Beran (1981) and Fan and Truong (1993). Specifically, Section 2.1 describes the building block of nonparametric regression for measurement errors, namely the method of deconvoluted density estimation. Section 2.2 discusses the deconvoluted kernel regression estimation and gives a summary of theoretical properties of the procedure. Section 2.3 summarizes Beran's nonparametric method and his asymptotic results along with Dabrowska's contribution in this area. Our proposed procedure is given in

Section 2.4. Section 2.5 considers estimation of the regression function. Asymptotic properties of the proposed estimators are described in Section 3. Concluding remarks and open problems are given in Section 4. Proofs are presented in Section 5.

## 2.  Methods

### 2.1. Deconvoluted kernel density estimators

Given a random sample from the distribution of $X = U + Z$, one of the problems in errors-in-variables is to estimate the density function of $U$. To make this problem identifiable, it is assumed that the distribution of $Z$ is known and that $Z$ and $U$ are independent. Also, $Z$ has a non-vanishing characteristic function $\phi_Z(t)$.

Let $X_j = U_j + Z_j$, $j = 1, \ldots, n$ denote a random sample from the distribution of $X = U + Z$. Here $Z_1, \ldots, Z_n$ and $U_1, \ldots, U_n$ are independent and both $U_j$ and $Z_j$ are not observable. Denote the characteristic functions of $X$ and $U$ by $\phi_X(\cdot)$ and $\phi_U(\cdot)$, respectively. By Fourier inversion, the density function of $U$ is given by

$$f_U(u) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-itu)\phi_U(t)\, dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-itu)\frac{\phi_X(t)}{\phi_Z(t)}\, dt.$$

Thus the problem of estimating the density function of $U$ can be reduced to the estimation of the function

$$\phi_X(t) = \int_{-\infty}^{\infty} \exp(itx) f_X(x)\, dx,$$

which, in turn is a density estimation problem based on $X$. Let $K(\cdot)$ denote a kernel function; that is, $K(\cdot)$ satisfies conditions of a density function. Then the kernel estimate of the density function $f_X(\cdot)$ of $X$ is given by

$$\hat{f}_X(x) = \frac{1}{nb_n} \sum_{j=1}^{n} K\{b_n^{-1}(x - X_j)\} = K_{b_n} * F_n,$$

where '$*$' is the convolution of distribution functions, $K_{b_n}(x) = b_n^{-1}K(b_n^{-1}x)$, $F_n$ is the empirical distribution function of $X_1, \ldots, X_n$ and the bandwidth $b_n$ is a sequence of positive numbers that tend to zero as $n \to \infty$. See Silverman (1984) and Devroye and Györfi (1985).

Denote the Fourier transforms of $F_n(\cdot)$ and $K(\cdot)$ respectively by

$$\hat{\phi}_n(t) = \frac{1}{n} \sum_{j=1}^{n} \exp(itX_j) \quad \text{and} \quad \phi_K(t) = \int_{-\infty}^{\infty} \exp(itv)K(v)\, dv.$$

Then $\phi_X(\cdot)$ is estimated by

$$\hat{\phi}_X(t) = \int_{-\infty}^{\infty} \exp(itx)\hat{f}_X(x)\, dx = \hat{\phi}_n(t)\phi_K(tb_n).$$

Thus a natural estimator of the density function of $U$ is

$$\hat{f}_n(u) \quad = \quad \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-itu)\frac{\hat{\phi}_X(t)}{\phi_Z(t)}\, dt$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-itu)\phi_K(tb_n)\frac{\hat{\phi}_n(t)}{\phi_Z(t)}\, dt,$$

which can be rewritten in the kernel form:

$$\hat{f}_n(u) = \frac{1}{nb_n} \sum_{j=1}^{n} K_n\{b_n^{-1}(u - X_j)\}, \tag{2.1}$$

where

$$K_n(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-itx)\frac{\phi_K(t)}{\phi_Z(tb_n^{-1})}\, dt \tag{2.2}$$

is called the deconvoluted kernel. We give two examples of (2.2).

EXAMPLE 2.1. *Normal measurement errors*

Suppose $Z$ has a normal distribution with $\phi_Z(t) = \exp(-\frac{1}{2}\sigma_0^2 t^2)$. If the kernel $K(\cdot)$ is the inverse triangular density such that its Fourier transform is given by $\phi_K(t) = (1 - t^2)_+^3$. Then (2.2) yields,

$$K_n(x) = \frac{1}{\pi} \int_0^1 \cos(tx)(1 - t^2)^3 \exp\left(\frac{\sigma_0^2 t^2}{2b_n^2}\right) dt. \tag{2.3}$$

The optimal bandwidth is given by $b_n = c\sigma_0(\log n)^{-1/2}$ with $c > 1$. Graphs of (2.3) for different values of the constant factor $c$ are given in Fan, Truong and Wang (1991).

EXAMPLE 2.2. *Double exponential measurement errors*

Suppose $Z$ has a double exponential or Laplace distribution with $\phi_Z(t) = (1 + \sigma_0^2 t^2/4)^{-1}$. If $K(\cdot)$ is the Gaussian kernel $K(x) = (\sqrt{2\pi})^{-1}\exp(-x^2/2)$, then by (2.2),

$$K_n(x) = \frac{1}{\sqrt{2\pi}}\left(1 - \frac{\sigma_0^2}{4b_n^2}(x^2 - 1)\right)\exp\left(-\frac{1}{2}x^2\right). \tag{2.4}$$

The optimal bandwidth is given by $b_n \sim cn^{-1/9}$ with $c > 0$. Graphs of (2.4) for different values of the constant factor $c$ are given in Fan, Truong and Wang (1991).

Note that the deconvoluted kernel (2.2) depends on $n$ and $\phi_Z(\cdot)$ which is different from the usual kernel function. The deconvoluted kernel density estimator (2.1) was studied by Carroll and Hall (1988), Stefanski and Carroll (1990), and Zhang (1990) and Fan (1991a,b). Optimal properties will be summarized in the next section.

## 2.2. Deconvoluted kernel regression function estimators

We start with nonparametric regression for complete data. Suppose $(U_1, T_1), \ldots, (U_n, T_n)$ is a random sample from the distribution of $(U, T)$, then the kernel estimator of the regression function $m(u) = E(T|U = u)$ is given by

$$\hat{m}_n(u) = \frac{\sum_{j=1}^{n} T_j K\{b_n^{-1}(u - U_j)\}}{\sum_{j=1}^{n} K\{b_n^{-1}(u - U_j)\}}, \tag{2.5}$$

where $K(\cdot)$ is the kernel function and $b_n$ is the bandwidth. See Nadaraya (1964), Watson (1964) and Härdle (1990).

Next we consider nonparametric regression involving measurement errors. Suppose $X = U + Z$ and consider the problem of estimating the regression function $m(u) = E(T|U = u)$ from the random sample $(X_1, T_1), \ldots, (X_n, T_n)$. Here $X_j = U_j + Z_j$, $j = 1, \ldots, n$; $Z_1, \ldots, Z_n$ and $U_1, \ldots, U_n$ are independent. Since the deconvoluted kernel (2.2) has the effect of accounting for the measurement errors in $U_j$, a natural estimator of $m(u)$ is

$$\hat{m}_n(u) = \frac{\sum_{j=1}^n T_j K_n\{b_n^{-1}(u - X_j)\}}{\sum_{j=1}^n K_n\{b_n^{-1}(u - X_j)\}}, \tag{2.6}$$

where $K_n(\cdot)$ is given by (2.2).

Estimator (2.6) was introduced by Fan and Truong (1993). Under appropriate conditions, estimators (2.1) and (2.6) can be shown to possess optimal of rate of convergence. These results, depending upon the smoothness of the error distribution, are summarized as follows.

The distribution of $Z$ is said to be super smooth of order $\beta$: if the function $\phi_Z(\cdot)$ satisfies

$$d_0|t|^{\beta_0} \exp(-|t|^\beta/\gamma) \le |\phi_Z(t)| \le d_1|t|^{\beta_1} \exp(-|t|^\beta/\gamma) \quad \text{as } t \to \infty, \tag{2.7}$$

where $d_0, d_1, \beta, \gamma$ are positive constants and $\beta_0, \beta_1$ are constants.

On the other hand, $Z$ is said to be ordinary smooth of order $\beta$ if there are positive constants $d_0, d_1, \beta$ such that $\phi_Z(\cdot)$ satisfies

$$d_0|t|^{-\beta} \le |\phi_Z(t)| \le d_1|t|^{-\beta} \quad \text{as } t \to \infty. \tag{2.8}$$

For example, normal $N(0, 1)$ with $\beta = 2$ and Cauchy $1/\{\pi(1 + z^2)\}$ with $\beta = 1$ are super smooth distributions. While ordinary smooth distributions include Laplace or double exponential $2^{-1} \exp(-|z|)$ with $\beta = 2$ and Gamma $\alpha^p z^{p-1} \exp(-\alpha z)/\Gamma(p)$ with $\beta = p$.

The rates of convergence for deconvoluted kernel estimators (2.1) and (2.6) depend on $\beta$ — the order of smoothness of the error distribution. They also depend on the smoothness of the regression function and the marginal density function. For regression and density functions with bounded $k$-th derivatives, the optimal rates of convergence are given in the following table.

| Super smooth distribution | Rates of convergence | Ordinary smooth distribution | Rates of convergence |
|---|---|---|---|
| $N(0,1)$ | $(\log n)^{-k/2}$ | Gamma $(\alpha, p)$ | $n^{-k/(2k+2p+1)}$ |
| Cauchy $(0,1)$ | $(\log n)^{-k}$ | Laplace | $n^{-k/(2k+5)}$ |

A similar table for deconvoluted density estimation is given in Fan (1991a).

The optimal bandwidth $b_{\text{opt}}$ depends on the error distribution. For the super smooth error distribution of order $\beta$, $b_{\text{opt}} = c(\log n)^{-1/\beta}$ for some constant $c$ depending only on the error distribution and the kernel function. In the ordinary case, $b_{\text{opt}} = cn^{-1/(2k+2\beta+1)}$ for some constant $c$. Asymptotic normality of deconvoluted estimators (2.1) and (2.6) and confidence intervals are discussed in Fan (1991b) and Fan, Truong and Wang (1990). Based on simulation results reported in Fan and Truong (1993), it is important to know that the deconvoluted estimator (2.6) is robust to the misspecification of error distribution, and is significantly better than the ordinary (or naive) kernel method.

### 2.3. Kernel survival function estimators

In this section, we describe Beran's procedure for estimating the survival function $S(t|u) = P(T > t|U = u) = 1 - F(t|u)$, the probability that an individual's survival time before $t$, given that $U = u$. According to Tsiatis (1978), the observable random vector $(U, Y, \delta)$ does not identify $S(t|u)$ uniquely, unless the random variables $T$ and $C$ are conditionally independent given $U$. In our discussion, as well as in most survival analyses, the identifiability issue is resolved by imposing such conditional independence assumption.

Let $G(t|u)$ denote the conditional distribution function of the censoring time $C$ given $U = u$. Set $H_1(t|u) = P(Y > t, \delta = 1|U = u)$, $H_2(t|u) = P(Y > t|U = u)$. By the conditional independence assumption,

$$H_1(t|u) = \int_{s>t} \{1 - G(s|u)\}\, dF(s|u), \quad H_2(t|u) = \{1 - G(t|u)\}S(t|u).$$

Suppose $f(t|u)$ is the conditional density function of $T$ given $U = u$. That is, $f(t|u) = F'(t|u)$. Define the hazard and the cumulative hazard functions respectively by

$$\lambda(t|u) = \frac{f(t|u)}{S(t|u)}, \quad \Lambda(t|u) = \int_0^t \lambda(s|u)\, ds.$$

Then,

$$\Lambda(t|u) = \int_0^t \frac{f(s|u)}{S(s|u)}\, ds = -\log S(t|u), \tag{2.9}$$

and, under the conditional independence assumption,

$$\Lambda(t|u) = \int_0^t \frac{f(s|u)}{S(s|u)}\, ds = -\int_0^t \frac{H_1(ds|u)}{H_2(s|u)}. \tag{2.10}$$

The above relationships suggest that the problem of estimating $S(t|u)$ and $\Lambda(t|u)$ is now simplified to the estimation of two functions $H_1(t|u)$ and $H_2(t|u)$.

Let $(U_j, Y_j, \delta_j)$, $j = 1, \ldots, n$, denote a random sample ¿from the distribution of $(U, Y, \delta)$. Here $Y_j = \min(T_j, C_j)$; $T_j$ and $C_j$ are independent given $U_j$, and $(T_j, C_j)$, $j = 1, \ldots, n$ is a random sample from the distribution of $(T, C)$ which may or may not

be observable. Set $I_{1j} = I_{1j}(t) = 1(Y_j > t, \delta_j = 1)$ and $I_{2j} = I_{2j}(t) = 1(Y_j > t)$. ¿From (2.5), kernel estimators of $H_\nu(t|u), \nu = 1, 2$ are given respectively by

$$\hat{H}_{\nu n}(t|u) = \frac{\sum_{j=1}^n I_{\nu j} K\{b_n^{-1}(u - U_j)\}}{\sum_{j=1}^n K\{b_n^{-1}(u - U_j)\}}, \quad \nu = 1, 2. \tag{2.11}$$

Set $\Delta \hat{H}_{1n}(s|u) = \hat{H}_{1n}(s|u) - \hat{H}_{1n}(s-|u)$. By (2.9) and (2.10), the conditional cumulative hazard and the conditional survival functions are estimated respectively by

$$\hat{\Lambda}_n(t|u) = -\int_0^t \frac{d\hat{H}_{1n}(s|u)}{\hat{H}_{2n}(s|u)} = -\sum_{Y_i \leq t} \frac{\Delta \hat{H}_{1n}(Y_i|u)}{\hat{H}_{2n}(Y_i|u)} \tag{2.12}$$

and

$$\hat{S}_n(t|u) = \exp\left(-\hat{\Lambda}_n(t|u)\right). \tag{2.13}$$

Observe that (2.13) can be written as

$$\hat{S}_n(t|u) = \exp\left(-\sum_{Y_i \leq t} \frac{\Delta \hat{H}_{1n}(Y_i|u)}{\hat{H}_{2n}(Y_i|u)}\right)$$

$$= \prod_{Y_i \leq t} \exp\left(-\frac{\Delta \hat{H}_{1n}(Y_i|u)}{\hat{H}_{2n}(Y_i|u)}\right) \approx \prod_{Y_i \leq t} \left(1 - \frac{\Delta \hat{H}_{1n}(Y_i|u)}{\hat{H}_{2n}(Y_i|u)}\right),$$

which is the product estimator proposed by Beran (1981). Under reasonable assumptions, Beran established the uniform consistency of (2.11), (2.12) and (2.13). Dabrowska (1987) proved that these estimators converge weakly to Gaussian processes, and that these results also hold for nearest neighbour estimates. Note that (2.10) can be generalized to include discontinuous or jumps in $H_1$, see Beran (1981) and Dabrowska (1987) for the exposition.

### 2.4. Deconvoluted kernel estimator of the survival function

Suppose the process of obtaining the values of $U_j$ is subject to measurement errors and that it is only possible to observe $X_j = U_j + Z_j$, $j = 1, \ldots, n$ with $Z_j$ being independent random errors. According to (2.6), the deconvoluted kernel estimators of $H_\nu(t|u), \nu = 1, 2$, based on the random sample $(X_j, Y_j, \delta_j)$, $j = 1, \ldots, n$, are given respectively by

$$H_{\nu n}(t|u) = \frac{\sum_{j=1}^n I_{\nu, j} K_n\{b_n^{-1}(u - X_j)\}}{\sum_{j=1}^n K_n\{b_n^{-1}(u - X_j)\}}, \quad \nu = 1, 2. \tag{2.14}$$

Here $K_n(\cdot)$ is the deconvoluted kernel function (2.2). Consequently, the conditional cumulative hazard and the conditional survival functions are estimated respectively by

$$\Lambda_n(t|u) = -\int_0^t \frac{dH_{1n}(s|u)}{H_{2n}(s|u)} = -\sum_{Y_i \leq t} \frac{\Delta H_{1n}(Y_i|u)}{H_{2n}(Y_i|u)}, \tag{2.15}$$

and

$$S_n(t|u) = \exp\left(-\Lambda_n(t|u)\right). \tag{2.16}$$

Fan and Truong (1993) showed that estimators (2.14) possess optimal rates of convergence, and that they have asymptotic normal distributions (Fan, Truong and Wang, 1990). In the presence of censoring, it is however necessary to strengthen the above results to weak convergence to Gaussian processes. This will be discussed in Section 3.

## 2.5. Regression function estimation

The estimation of the conditional mean or regression functions is very important in practice. See Fan and Gijbels (1994). The regression function $m(u) = E(T|U = u)$ is given by

$$m(u) = \int_0^\infty S(t|u)\, dt.$$

To make $m(\cdot)$ identifiable, we truncate the upper bound of the integral, i.e.,

$$m(u; \tau_u) = \int_0^{\tau_u} S(t|u)\, dt, \tag{2.17}$$

where $\tau_u < \sup\{t : H_2(t|u) > 0\}$. See also Dabrowska (1987).

A natural estimate of the regression function is given by

$$m_n(u) = \int_0^\infty S_n(t|u)\, dt.$$

If the largest observation is censored, then the integral is infinite. In this case, we consider the estimate of the truncated mean regression function

$$m_n(u) = \int_0^{\tau_u} S_n(t|x)\, dt. \tag{2.18}$$

In the absence of covariate measurement error, Dabrowska (1987) proved that, under appropriate conditions, (2.18) is consistent, and that it converges weakly to a Gaussian process. In this paper, we will show that the same results hold in the context of measurement error.

## 3.  Asymptotic properties

In this section, we describe conditions under which the proposed estimators (2.14), (2.15), (2.16) and (2.18) will be shown to be consistent, and that they converge weakly to Gaussian processes.

The following condition is required for solving the identifiability problem mentioned earlier.

**Condition 1** *The random variables $T$ and $C$ are conditionally independent given $U$.*

The next condition is standard in nonparametric regression.

**Condition 2** (a) *Let $a < b$. The marginal density $f_U(\cdot)$ of the unobserved $U$ is bounded away from zero on the interval $[a, b]$, and has a bounded second derivative.*

(b) *For $t \geq 0$, the functions $H_1(t|u)$ and $H_2(t|u)$ have continuous second derivatives (as functions of $u$) on $[a, b]$.*

Recall that $\tau_u < \sup\{t : H_2(t|u) > 0\}$. Hence $H_1(\tau_u|u)$ and $H_2(\tau_u|u)$ are bounded away from zero.

Our asymptotic results depend on the smoothness condition of the error distribution, which will be described in the next two subsections.

## 3.1. Ordinary smooth error

This subsection considers the asymptotic properties of (2.14) – (2.16) and (2.18) involving ordinary smooth measurement errors.

The first part of the following condition ensures that the estimators (2.14) are well defined. The remaining deals with tail behaviors of the error distribution, which is satisfied by most ordinary smooth distributions such as double exponential and gamma distributions.

**Condition 3** *The characteristic function of the error distribution $\phi_Z(\cdot)$ does not vanish. Moreover, there are constants $\beta \geq 0, c \neq 0$ such that*

$$\phi_Z(s)s^\beta \to c, \quad \phi_Z'(s)s^{\beta+1} \to -\beta c \quad as \ s \to \infty.$$

In order for the bias term to achieve the desirable rate of convergence, the kernel function $K(\cdot)$ satisfies

**Condition 4** *The function $\phi_K(\cdot)$ is symmetric with $\int |\phi_K^{(4)}| < \infty$, $\phi_K(0) = 1$ and $\phi_K(v) = 1 + O(|v|^2)$ as $v \to 0$. Moreover,*

$$\int_{-\infty}^{\infty} \{|\phi_K(v)| + |\phi_K'(v)|\}|v|^{\beta+1} \, dv < \infty, \quad \int_{-\infty}^{\infty} |\phi_K(v)|^2 |v|^{2\beta+2} \, dv < \infty.$$

The above condition implies that the kernel function

$$K(x) = (2\pi)^{-1} \int \exp(ivx)\phi_K(v) \, dv$$

is a second order kernel; that is,

$$\int_{-\infty}^{\infty} K(x) \, dx = 1, \quad \int_{-\infty}^{\infty} xK(x) \, dx = 0, \quad \int_{-\infty}^{\infty} x^2 K(x) \, dx < \infty.$$

Moreover, there is a positive constant $D$ such that

$$|K(x)| \leq D(1 + |x|^4)^{-1}.$$

We start with the key result of the paper, which considers the weak convergence of the estimators (2.14).

Theorem 3.1. *Suppose Conditions 1–4 hold and that* $b_n \to 0$ *such that* $nb_n^{2\beta+1} \to$ $\infty$ *and* $nb_n^{2\beta+5} \to 0$ *as* $n \to \infty$. *Then the process* $\sqrt{nb_n^{2\beta+1}}(H_{1n}(t|u) - H_1(t|u), H_{2n}(t|u) - H_2(t|u))$ *converges weakly in* $D[0, \tau_u] \times D[0, \tau_u]$ *to a two-dimensional mean zero Gaussian process* $(W_1(s|u), W_2(t|u))$ *with covariance function given by*

$$\operatorname{cov}\{W_i(s|u), W_j(t|u)\}$$
$$= \frac{1}{2\pi|c|^2 f_U(u)} \int dv\, |v|^{2\beta} |\phi_K(v)|^2 \int dF_Z(z)\, f_U(u-z)$$
$$\times \{H_{\min\{i,j\}}(t|u-z) - H_i(s|u-z)H_j(t|u-z) \qquad (3.1)$$
$$+ (H_i(s|u-z) - H_i(s|u))(H_j(t|u-z) - H_j(t|u))\}, \quad i,j = 1, 2.$$

Observe that when there is no measurement error; that is, $Z = 0$ with probability 1, the above formula of the covariance is simplified to

$$\operatorname{cov}\{W_i(s|u), W_j(t|u)\} = \frac{1}{f_U(u)} \int K^2(v)\, dv \{H_{\min\{i,j\}}(t|u) - H_i(s|u)H_j(t|u)\}, \quad i,j = 1, 2,$$

which is identical to the one given in Dabrowska (1987).

It follows from the Skorokhod's construction that $\sqrt{nb_n^{2\beta+1}}(H_{1n}(s|u) - H_1(s|u),$ $H_{2n}(t|u) - H_2(t|u))$ converges almostly surely to $(W_1(s|u), W_2(t|u))$ in the supremum metric on $[0, \tau_u] \times [0, \tau_u]$. Hence, we conclude that

Corollary 3.1. *Suppose Conditions 1–4 hold and that* $b_n \to 0$ *such that* $nb_n^{2\beta+1} \to$ $\infty$ *and* $nb_n^{2\beta+5} \to 0$ *as* $n \to \infty$. *Then*

$$\sup_{s \le \tau_u} |H_{\nu n}(s|u) - H_\nu(s|u)| = O_p\left\{(nb_n^{2\beta+1})^{-1/2}\right\}, \quad a \le u \le b, \quad \nu = 1, 2.$$

Consequently, we obtain the consistency of $\Lambda_n$, $S_n$ and $m_n$ as defined by (2.15), (2.16) and (2.18), respectively.

Corollary 3.2. *Suppose Conditions 1–4 hold and that* $b_n \to 0$ *such that* $nb_n^{2\beta+1} \to$ $\infty$ *and* $nb_n^{2\beta+5} \to 0$ *as* $n \to \infty$. *Then, for* $a \le u \le b$,

$$\sup_{s \le \tau_u} |\Lambda_n(s|u) - \Lambda(s|u)| = O_p\left\{(nb_n^{2\beta+1})^{-1/2}\right\},$$

$$\sup_{s \le \tau_u} |S_n(s|u) - S(s|u)| = O_p\left\{(nb_n^{2\beta+1})^{-1/2}\right\}$$

*and*

$$|m_n(u; \tau_u) - m(u; \tau_u)| = O_p\left\{(nb_n^{2\beta+1})^{-1/2}\right\}.$$

In the absence of measurement errors, under fairly general conditions, Beran (1981) showed that $\sup_{s \le \tau_u} |H_{\nu n}(s|u) - H_\nu(s|u)| = o(1)$ almost surely. In proving this result, it is crucial that the kernel function $K(\cdot)$ is non-negative so that Glivenko-Cantelli theorem can be applied to yield the uniformity and the almost sure convergence. The situation becomes more complicated when there is measurement error, since the deconvoluted kernel weight may be negative. Nevertheless, these results are valid for convergence in probability by virtue of the Skorokhod construction.

The next result deals with the weak convergence of $\Lambda_n$ and $S_n$.

THEOREM 3.2. *Suppose Conditions 1–4 hold and that* $b_n \to 0$ *such that* $nb_n^{2\beta+1} \to \infty$ *and* $nb_n^{2\beta+5} \to 0$ *as* $n \to \infty$. *Then the processes* $\sqrt{nb_n^{2\beta+1}}(\Lambda_n(t|u) - \Lambda(t|u))$ *and* $\sqrt{nb_n^{2\beta+1}}(S_n(t|u) - S(t|u))$ *converge weakly in* $D[0, \tau_u]$ *to mean zero Gaussian processes with covariance structures given by* $\Gamma_1(s, t|u)$ *and* $S(s|u)S(t|u)\Gamma_1(s, t|u)$, *respectively.*

In the absence of measurement errors, it follows from integration by parts (see Breslow and Crowley (1974)) that the asymptotic covariance function of $\sqrt{nb_n^{2\beta+1}}(\Lambda_n(t|u) - \Lambda(t|u))$ is given by

$$\Gamma_1(s, t|u) = \frac{\int K^2(v)\, dv}{f_U(u)} \int_0^{s \wedge t} \frac{\Lambda(dz|u)}{H_2(z|u)}.$$

See Dabrowska (1987). The formula is quite complicated in the presence of measurement errors.

According to (2.17) and (2.18), the next result follows easily from the weak convergence of $\sqrt{nb_n^{2\beta+1}}(S_n(t|u) - S(t|u))$.

COROLLARY 3.3. *Suppose Conditions 1–4 hold and that* $b_n \to 0$ *such that* $nb_n^{2\beta+1} \to \infty$ *and* $nb_n^{2\beta+5} \to 0$ *as* $n \to \infty$. *Then* $\sqrt{nb_n^{2\beta+1}}(m_n(u) - m(u))$ *converges weakly to a mean zero Gaussian distribution with variance given by*

$$\int_0^{\tau_u} \int_0^{\tau_u} S(s|u)S(t|u)\Gamma_1(s, t|u)\, ds\, dt.$$

## 3.2. Normal error

This subsection considers the asymptotic properties of (2.14) – (2.16) and (2.18) when the covariate is subject to normal measurement error.

**Condition 5** $Z \sim N(0, \sigma_0^2)$ *and the deconvoluted kernel is given by (2.3).*

Set

$$\sigma_n \equiv \sigma_n(u) = b_n^{-1}\sqrt{E\left(K_n^2\{b_n^{-1}(u - X_1)\}\right)}.$$

THEOREM 3.3. *Suppose* $b_n = \sigma_0\{1 - 3(\log\log n)(\log n)^{-1}\}(\log n)^{-1/2}$ *and that Conditions 1, 2 and 5 hold. Then the process* $(\sqrt{n}/\sigma_n)(H_{1n}(t|u) - H_1(t|u), H_{2n}(t|u) - H_2(t|u))$ *converges weakly in* $D[0, \tau_u] \times D[0, \tau_u]$ *to a two-dimensional mean zero Gaussian process.*

In the case of ordinary smooth errors, the covariance function of the limiting Gaussian process is given by (3.1). There is no simple formula for the limiting covariance function of the above Gaussian process when the covariate measurement error is normal. This is due to the fact that, for normal errors, the variance of the deconvoluted kernel estimator converges to zero much slower than the bias term; see Fan and Truong (1993). It also explains why a different normalization is used in establishing the asymptotic distribution of (2.14).

The next result follows from the Skorokhod's construction.

COROLLARY 3.4. *Suppose $b_n = \sigma_0\{1 - 3(\log\log n)(\log n)^{-1}\}(\log n)^{-1/2}$ and that Conditions 1, 2 and 5 hold. Then*

$$\sup_{s \leq \tau_u} |H_{\nu n}(s|u) - H_\nu(s|u)| = O_p(\sigma_n/\sqrt{n}), \quad a \leq u \leq b, \quad \nu = 1, 2.$$

*Moreover,*

$$\sup_{s \leq \tau_u} |\Lambda_n(s|u) - \Lambda(s|u)| = O_p(\sigma_n/\sqrt{n}), \quad \sup_{s \leq \tau_u} |S_n(s|u) - S(s|u)| = O_p(\sigma_n/\sqrt{n})$$

*and*

$$|m_n(u; \tau_u) - m(u; \tau_u)| = O_p(\sigma_n/\sqrt{n}).$$

According to (4.11) of Fan, Truong and Wang (1990),

$$|K_n(\cdot)| = O\left(b_n^{7.4} \exp\{\sigma_0^2/(2b_n^2)\}\right).$$

It follows from $b_n = \sigma_0\{1 - 3(\log\log n)(\log n)^{-1}\}(\log n)^{-1/2}$ that $\sigma_n/\sqrt{n} \to 0$. Hence, the above results generalize the consistency results of Beran (1981) to the situation when the covariate is subject to normal error.

The next result considers the weak convergence of $\Lambda_n$ and $S_n$ for normal covariate measurement errors.

THEOREM 3.4. *Suppose $b_n = \sigma_0\{1 - 3(\log\log n)(\log n)^{-1}\}(\log n)^{-1/2}$ and that Conditions 1, 2 and 5 hold. Then the processes $(\sqrt{n}/\sigma_n)(\Lambda_n(t|u) - \Lambda(t|u))$ and $(\sqrt{n}/\sigma_n)$ $(S_n(t|u) - S(t|u))$ converge weakly in $D[0, \tau_u]$ to mean zero Gaussian processes.*

Consequently, Corollary 3.3 is also valid for normal errors as given in the next result.

COROLLARY 3.5. *Suppose $b_n = \sigma_0\{1 - 3(\log\log n)(\log n)^{-1}\}(\log n)^{-1/2}$ and that Conditions 1, 2 and 5 hold. Then $(\sqrt{n}/\sigma_n)(m_n(u) - m(u))$ converges weakly to a mean zero Gaussian distribution.*

Suppose the limiting covariance function of $(\sqrt{n}/\sigma_n)(\Lambda_n(t|u) - \Lambda(t|u))$ is denoted by $\Gamma_2(s, t|u)$. Then the limiting covariance function of $(\sqrt{n}/\sigma_n)(S_n(t|u) - S(t|u))$ is given by

$$S(s|u)S(t|u)\Gamma_2(s, t|u),$$

and the limiting covariance function of $(\sqrt{n}/\sigma_n)(m_n(u) - m(u))$ is

$$\int_0^{\tau_u} \int_0^{\tau_u} S(s|u)S(t|u)\Gamma_2(s, t|u)\, ds\, dt.$$

Proofs of these results are given in Section 5.

## 4.   Concluding remarks

Based on Beran (1981) and Fan and Truong (1993), the paper considers a new deconvoluted kernel method to estimate regression models involving censored survival data when the covariate is subject to measurement error. The approach is nonparametric, making no assumption about the functional form of the covariate effect, except for the smoothness conditions. Under appropriate conditions on the the marginal density and the conditional distribution functions, it is shown that the proposed estimators (2.14), (2.15), (2.16) and (2.18) are consistent, and they converge weakly to Gaussian processes. These results generalize the corresponding results of Beran (1981) and Dabrowska (1987) to covariates with measurement errors.

There are still many important issues to be addressed, a few of them is listed below.

*1. Variance estimates and confidence intervals.* To construct confidence intervals, it will be useful to have covariance function (3.1) estimated. One possible approach is to modify the method suggested by Fan, Truong and Wang (1990). Moreover, in the normal measurement error case, it will be important to compute and estimate the covariance function of the limiting Gaussian process. One can also consider using bootstrap to construct confidence intervals.

*2. Local linear estimators.* To prove Theorem 3.1, we impose a smoothness condition on the marginal density function as described in Condition 2 (a). This is unavoidable when using kernel estimator, see also Dabrowska (1987). However, in the absence of measurement errors, O'Grady (1994) considered a more flexible approach by using local linear method to estimate $H_1(t|u)$ and $H_2(t|u)$, and weak convergence results were also established. Local linear estimators have the advantage of being spatially adaptive; that is, their asymptotic properties depend only on the continuity of the marginal density, instead of the derivative as required by the usual kernel method. Methodologically, the extension to the case involving measurement error based on local linear methods is straightforward, see Fan, Truong and Wang (1991). However, it will be important to show that the local linear estimators of $(H_1(t|u), H_2(t|u))$ also converge weakly to a two dimensional Gaussian process. .

*3. Bandwidth selection.* The use of estimators (2.14), (2.15) and (2.16) depends on the choice of bandwidth $b_n$. Large value of bandwidth yields high bias in the estimate, while small value will inflate the variance term. It was discussed in Sections 2.1 and 2.2 that the optimal bandwidth depends on the error distribution and the smoothness of the underlying function. Hence, it is important to determine whether cross-validation or plug-in method still yields consistent estimates of the above functions.

*4. Quantile regression.* In the absence of measurement errors, Beran (1981), Dabrowska (1987) and Doksum and Yandell (1982) examined the problem of conditional quantile estimation. The latter paper also discussed simultaneous confidence bands for the mean and median regression, and demonstrated the usefulness of the nonparametric approach over the estimates derived from Cox's model and Buckley-James method on Stanford Heart transplant data. It will be of interest to extend our methodology to the estimation of quantile regression.

*5. Multivariate extensions.* For complete data, the area of nonparametric estima-

tion of the mean and median regression functions has been very active. However, it is known that kernel estimates are not efficient for handling covariates in high dimensional space, due to the *curse of dimensionality*; see Huber (1985) and Stone (1986). The procedures proposed in this paper have the same drawback. Recently, Kooperberg, Stone and Truong (1994a,b) considered the estimation of hazard regression using splines and their tensor products. It is of interest to extend this methodology to handle covariates involving measurement errors.

## 5. Proofs

The following notations will be used throughout this section. Set

$$K_{nj}(x) = b_n^{-1} K_n \{ b_n^{-1}(u - X_j) \}, \quad \tilde{f}(u) = E\{K_{n1}(u)\},$$

$$\tilde{H}_1(t, u) = E\{K_{n1}(u)I_{11}(t)\}, \quad \tilde{H}_2(t, u) = E\{K_{n1}(u)I_{21}(t)\},$$

where, we recall that $I_{11}(t) = 1(Y_1 > t, \delta_1 = 1)$ and $I_{21}(t) = 1(Y_1 > t)$.

### 5.1. Proof of Theorem 3.1

The argument is a refinement of Dabrowska (1987) and Fan, Truong and Wang (1990). Let $f_X$ and $f_Z$ denote the density functions of $X$ and $Z$, respectively. Note that

$$
\begin{aligned}
\tilde{f}(u) &= b_n^{-1} \int K_n \{ b_n^{-1}(u - x) \} f_X(x) \, dx \\
&= b_n^{-1} \int K_n \{ b_n^{-1}(u - x) \} \int f_U(x - z) \, dF_Z(z) \, dx \\
&= \frac{1}{2\pi b_n} \iint \exp \left\{ -it \left( \frac{u - x}{b_n} \right) \right\} \frac{\phi_K(t)}{\phi_Z(tb_n^{-1})} \, dt \int f_U(x - z) \, dF_Z(z) \, dx \\
&= \frac{1}{2\pi b_n} \int dt \frac{\phi_K(t)}{\phi_Z(tb_n^{-1})} \int dF_Z(z) \exp \left( \frac{itz}{b_n} \right) \int dv \exp \left\{ -it \left( \frac{u - v}{b_n} \right) \right\} f_U(v) \\
&= \frac{1}{2\pi b_n} \int dv f_U(v) \int dt \phi_K(t) \exp \left\{ -it \left( \frac{u - v}{b_n} \right) \right\} \\
&= b_n^{-1} \int dv f_U(v) K \{ b_n^{-1}(u - v) \} = f_U(u) + O(b_n^2).
\end{aligned}
\tag{5.1}
$$

The last equality follows from the property of the kernel function $K(\cdot)$ and Taylor expansion. Set $H_\nu(t, u) = H_\nu(t|u) f_U(u)$, $\nu = 1, 2$. Then a similar argument yields

$$\tilde{H}_\nu(t, u) = H_\nu(t, u) + O(b_n^2), \quad \nu = 1, 2. \tag{5.2}$$

Write

$$H_{\nu n}(t|u) - H_\nu(t|u) = \left\{ \frac{n^{-1} \sum_{j=1}^{n} K_{nj}(u)(\tilde{f}(u) I_{\nu j}(t) - \tilde{H}_\nu(t, u))}{\tilde{f}(u) \, n^{-1} \sum_{j=1}^{n} K_{nj}(u)} \right\}$$

$$+ \left\{ \frac{\tilde{H}_\nu(t,u)}{\tilde{f}(u)} - H_\nu(t|u) \right\}, \quad \nu = 1, 2. \quad (5.3)$$

In view of Lemma 2 of Fan and Truong (1993), (5.1), (5.2) and $\sqrt{nb_n^{2\beta+5}} \to 0$, it is sufficient to consider the asymptotic distribution of

$$W_{\nu n}(t) \equiv W_{\nu n}(t,u) = \sqrt{nb_n^{2\beta+1}} \left( n^{-1} \sum_{j=1}^n K_{nj}(u)\{\tilde{f}(u)I_{\nu j}(t) - \tilde{H}_\nu(t,u)\} \right), \quad \nu = 1, 2.$$

Observe that $E \sum K_{nj}(u)\{\tilde{f}(u)I_{\nu j}(t) - \tilde{H}_\nu(t,u)\} = 0$, so $W_{1n}(t)$ and $W_{2n}(t)$ are standardized sums of mean-zero random variables.

Let $t_1 < \cdots < t_p$ and let $c_{\nu,r}, r = 1, \ldots, p, \nu = 1, 2$, denote constants such that $\sum_r \sum_\nu c_{\nu,i}^2 \neq 0$. Then

$$E \left( K_{n1}(u) \sum_r \sum_\nu c_{\nu,r}\{\tilde{f}(u)I_{\nu 1}(t_r) - \tilde{H}_\nu(t_r,u)\} \right) = 0.$$

Set $\tilde{\sigma}^2(x) = E \left\{ \left( \sum_r \sum_\nu c_{\nu,r}\{\tilde{f}(u)I_{\nu 1}(t_r) - \tilde{H}_\nu(t_r,u)\} \right)^2 | X_1 = x \right\}$. Then

$$\mathrm{var} \left( K_{n1}(u) \sum_r \sum_\nu c_{\nu,r}\{\tilde{f}(u)I_{\nu 1}(t_r) - \tilde{H}_\nu(t_r,u)\} \right)$$

$$= E \left( K_{n1}^2(u) \, E \left\{ \left( \sum_r \sum_\nu c_{\nu,r}\{\tilde{f}(u)I_{\nu 1}(t_r) - \tilde{H}_\nu(t_r,u)\} \right)^2 | X_1 \right\} \right)$$

$$= b_n^{-2} \iint K_n^2 \left( \frac{u-x}{b_n} \right) \tilde{\sigma}^2(x) f_X(x) \, dx.$$

By Lemma 2.1 of Fan (1991b) and Parseval's identity,

$$\mathrm{var} \left( K_{n1}(u) \sum_r \sum_\nu c_{\nu,r}\{\tilde{f}(u)I_{\nu 1}(t_r) - \tilde{H}_\nu(t_r,u)\} \right) = O \left( \frac{b_n^{-2\beta-1}}{2\pi|c|^2} \int dv \, |v|^{2\beta} |\phi_K(v)|^2 \right). \quad (5.4)$$

Similarly, (see the argument in the proof of Theorem 3 of Fan and Truong (1993))

$$E \left| K_{n1}(u) \sum_r \sum_\nu c_{\nu,r}\{\tilde{f}(u)I_{\nu 1}(t_r) - \tilde{H}_\nu(t_r,u)\} \right|^{2+\delta} = O \left( b_n^{-(2+\delta)(1+\beta)+1} \right). \quad (5.5)$$

It follows from (5.4) and (5.5) and Lyapounov's theorem that $\sum_r \sum_\nu c_{\nu,r} W_{\nu n}(t_r)$ has an asymptotical normal distribution. Hence the finite dimensional distributions of $(W_{1n}(t), W_{2n}(t))$ are asymptotically normal.

Proof of tightness: For $t_1 < t < t_2$, we need to establish

$$E\{W_{1n}(t_1) - W_{1n}(t)\}^2 \{W_{1n}(t) - W_{1n}(t_2)\}^2 \leq c_1 \{\Psi_1(t_1) - \Psi_1(t)\}\{\Psi_1(t) - \Psi_1(t_2)\}, \quad (5.6)$$

where $c_1$ is a constant and $\Psi_1$ is a nondecreasing function. Write

$$W_{1n}(t_1) - W_{1n}(t) = \left(\frac{b_n^{2\beta+1}}{n}\right)^{1/2} \sum_{j=1}^{n} A_j,$$

$$W_{1n}(t) - W_{1n}(t_2) = \left(\frac{b_n^{2\beta+1}}{n}\right)^{1/2} \sum_{j=1}^{n} B_j,$$

where

$$\begin{aligned}
A_j &= K_{nj}(u)\{\tilde{f}(u)(I_1(t_1) - I_1(t)) - (\tilde{H}_1(t_1, u) - \tilde{H}_1(t, u))\}, \\
B_j &= K_{nj}(u)\{\tilde{f}(u)(I_1(t) - I_1(t_2)) - (\tilde{H}_1(t, u) - \tilde{H}_1(t_2, u))\}.
\end{aligned}$$

By simple algebra and note that $A_j$ and $B_j$ have mean zero,

$$E\{W_{1n}(t_1) - W_{1n}(t)\}^2\{W_{1n}(t) - W_{1n}(t_2)\}^2$$

$$= \left(\frac{b_n^{2\beta+1}}{n}\right)^2 E\left(\sum_{j=1}^{n} A_j\right)^2 \left(\sum_{j=1}^{n} B_j\right)^2$$

$$= \frac{b_n^{4\beta+2}}{n}\left\{EA_1^2 B_1^2 + 2(n-1)EA_1^2 EB_1^2 + 2(n-1)(EA_1 B_1)^2\right\}. \quad (5.7)$$

Let $f_{Y,X}$ and $f_{Y,U}$ denote joint density functions of $(Y, X)$ and $(Y, U)$, respectively. To establish an upper bound for (5.7), we start with

$$E(A_1^2)$$

$$= EK_{n1}^2(u)\{\tilde{f}(u)(I_1(t_1) - I_1(t)) - (\tilde{H}_1(t_1, u) - (\tilde{H}_1(t, u))\}^2$$

$$\leq b_n^{-2} \iint K_n^2\left(\frac{u-x}{b_n}\right)\{\tilde{f}(u)\}^2 1\{t_1 < y \leq t, \delta = 1\}f_{Y,X}(y, x)\, dy\, dx$$

$$= b_n^{-2} \iint K_n^2\left(\frac{u-x}{b_n}\right)\{\tilde{f}(u)\}^2 1\{t_1 < y \leq t, \delta = 1\}\int f_{Y,U}(y, x-z)\, dF_Z(z)\, dy\, dx$$

$$= b_n^{-2} \int dF_Z(z) \int dx K_n^2\left(\frac{u-x}{b_n}\right)\{\tilde{f}(u)\}^2 f_U(x-z)\{H_1(t_1|x-z) - H_1(t|x-z)\}.$$

It follows from Lemma 2.1 of Fan (1991b) and Parseval's identity that

$$E(A_1^2) \leq b_n^{-2\beta-1}f_U^2(u)\int dF_Z(z)\, f_U(u-z)\{H_1(t_1|u-z) - H_1(t|u-z)\}$$

$$\times \frac{1}{2\pi|c|^2}\int |v|^{2\beta}|\phi_K(v)|^2\, dv\,(1 + o(1)). \quad (5.8)$$

Similarly,

$$E(B_1^2) \leq b_n^{-2\beta-1}f_U^2(u)\int dF_Z(z)\, f_U(u-z)\{H_1(t|u-z) - H_1(t_2|u-z)\}$$

$$\times \frac{1}{2\pi|c|^2}\int |v|^{2\beta}|\phi_K(v)|^2\, dv\,(1 + o(1)). \quad (5.9)$$

The above argument also shows

$$E(A_1B_1) \leq b_n^{-2\beta-1}\frac{f_U^2(u)}{2\pi|c|^2}\int|v|^{2\beta}|\phi_K(v)|^2\,dv\int dF_Z(z)\,f_U(u-z)\tag{5.10}$$
$$\times\{H_1(t_1|u-z)-H_1(t|u-z)\}\{H_1(t|u-z)-H_1(t_2|u-z)\}(1+o(1)),$$

$$EA_1^2B_1^2 = O(b_n^{-2\beta-1}).\tag{5.11}$$

It follows from (5.7)–(5.11) that (5.6) is valid. Similarly, there is a nondecreasing function $\Psi_2$ and a constant $c_2$ such that

$$E\{W_{2n}(t_1)-W_{2n}(t)\}^2\{W_{2n}(t)-W_{2n}(t_2)\}^2\leq c_2\{\Psi_2(t_1)-\Psi_2(t)\}\{\Psi_2(t)-\Psi_2(t_2)\}.$$

According to (15.39) of Billingsley (1968), $(W_{1n}(\cdot),W_{2n}(\cdot))$ is tight. Hence, we conclude that the bivariate process $\sqrt{nb_n^{2\beta+1}}(H_{1n}(t|u)-H_1(t|u),H_{2n}(t|u)-H_2(t|u))$ converges weakly in $D[0,\tau_u]\times D[0,\tau_u]$ to a two dimensional Gaussian process $(W_1(t|u),W_2(t|u))$ with covariance function given by (3.1), which can be easily verified by invoking Lemma 2.1 of Fan (1991b) and the Parseval's identity. This completes the proof of Theorem 3.1.

### 5.2. Proof of Corollary 3.1

This follows easily from the Skorokhod's construction.

### 5.3. Proof of Corollary 3.2

Simple algebra and Corollary 3.1 yield

$$|\Lambda_n(t|u)-\Lambda(t|u)| = O\left(\sup_{s\leq t}|H_{2n}(s|u)-H_2(s|u)|\right)$$
$$+O\left(\sup_{s\leq t}|H_{1n}(s|u)-H_1(s|u)|\right).$$

The first result follows from Corollary 3.1.

The second result follows the first, (2.9) and (2.16). The third result is an easy consequence of the second result.

### 5.4. Proof of Theorem 3.2

Set $W_{\nu n}(t|u)=\sqrt{nb_n^{2\beta+1}}(H_{1n}(t|u)-H_1(t|u),H_{2n}(t|u)-H_2(t|u))$, $\nu=1,2$. By the approach analogous to Breslow and Crowsley (1974) and Gill (1983), we have

$$L_n(t|u)=\sqrt{nb_n^{2\beta+1}}\{\Lambda_n(t|u)-\Lambda(t|u)\}=\int_0^t\frac{W_{2n}(s|u)\,\Lambda(ds|u)}{H_2(s|u)}+\int_0^t\frac{W_{1n}(ds|u)}{H_2(s|u)}+o_p(1).$$
$$\tag{5.12}$$

According to Theorem 3.1 and the Skorokhod's construction, we may assume that the random vector $\sqrt{nb_n^{2\beta+1}}(H_{1n}(t|u)-H_1(t|u),H_{2n}(t|u)-H_2(t|u))$ converges almost surely

to $(W_1(t|u), W_2(t|u))$ in the sup norm on $[0, \tau_u] \times [0, \tau_u]$. That is,

$$\sup_{s \leq \tau_u} |W_{\nu n}(s|u) - W_\nu(s|u)| \to 0 \quad \text{with probability 1}, \quad \nu = 1, 2.$$

Then (5.12) converges almostly surely in the sup norm on $[0, \tau_u]$ to

$$L(t|u) = \int_0^t \frac{W_2(s|u) \, \Lambda(ds|u)}{H_2(s|u)} + \int_0^t \frac{W_1(ds|u)}{H_2(s|u)}, \quad 0 \leq t \leq \tau_u,$$

which is a mean-zero Gaussian process with covariance function given by

$$\Gamma_1(s, t|u) = \text{cov}\{L(s|u), L(t|u)\}.$$

This completes the first half of Theorem 3.2.

By Taylor expansion,

$$\exp(-\Lambda_n) = \exp(-\Lambda) - \exp(-\Lambda)(\Lambda_n - \Lambda) + \frac{1}{2}\exp(-\Lambda_n^*)(\Lambda_n - \Lambda)^2,$$

where $\Lambda_n^*$ is a random function having value between $\Lambda(t|u)$ and $\Lambda_n(t|u)$. Then

$$\sqrt{nb_n^{2\beta+1}}\{S_n(t|u) - S(t|u)\} = -S(t|u)L_n(t|u) + R_n(t, u),$$

where

$$R_n(t, u) = (nb_n^{2\beta+1})^{-1/2} \exp(-\Lambda_n^*)L_n^2(t|u).$$

By the weak convergence of $L_n(t|u)$ and the consistency of $\Lambda_n$, $R_n(t, u) \to 0$ in probability. Hence, $\sqrt{nb_n^{2\beta+1}}\{S_n(t|u) - S(t|u)\}$ converges weakly to a mean-zero Gaussian process $-S(t|u)L(t|u)$ with covariance function given by

$$S(s|u)S(t|u)\text{cov}\{L(s|u), L(t|u)\} = S(s|u)S(t|u)\Gamma_1(s, t|u).$$

This completes the proof of Theorem 3.2.

## 5.5. Proof of Corollary 3.3

This follows from the weak convergence of $\sqrt{nb_n^{2\beta+1}}\{S_n(t|u) - S(t|u)\}$.

## 5.6. Proofs of Theorems 3.3 and 3.4

According to (2.3) and (6.11) of Fan, Truong and Wang (1990), there is a positive constant $c_3$ such that

$$\sigma_n \geq c_3 b_n^7 \exp\left\{\sigma_0^2/(2b_n^2)\right\}. \tag{5.13}$$

Note that (5.1) and (5.2) are still valid for $Z \sim N(0, \sigma_0^2)$ and the deconvoluted kernel (2.3). It follows from (5.13), (5.1), (5.2) and $b_n = \sigma_0\{1 - 3(\log\log n)(\log n)^{-1}\}(\log n)^{-1/2}$ that

$$\frac{\sqrt{n}}{\sigma_n}\left\{\frac{\tilde{H}_\nu(t, u)}{\tilde{f}(u)} - H_\nu(t|u)\right\} = O\left(\frac{\sqrt{n}b_n^2}{b_n^7 \exp\{\sigma_0^2/(2b_n^2)\}}\right) = o(1).$$

Hence, in view of (5.3) and Lemma 4 of Fan and Truong (1993), the asymptotic normality of $(\sqrt{n}/\sigma_n)(H_{1n}(t|u) - H_1(t|u), H_{2n}(t|u) - H_2(t|u))$ follows from the asymptotic normality of

$$V_{\nu n}(t,u) = \frac{\sqrt{n}}{\sigma_n}\left(n^{-1}\sum_{j=1}^{n}K_{nj}(u)\left\{\tilde{f}(u)I_{\nu j}(t) - \tilde{H}_\nu(t,u)\right\}\right), \quad \nu = 1,2. \qquad (5.14)$$

Observe that $V_{\nu n}(t,u)$ is the standardized sum of mean-zero independent random variable It follows from the argument given in the proof of Theorem 3.1 that the finite dimensional distributions of $(V_{1n}(t,u), V_{2n}(t,u))$ are asymptotically normal.

The tightness of (5.14) can be established analogously as in the proof of Theorem 3.1. Thus $(V_{1n}(t,u), V_{2n}(t,u))$ converges weakly to a mean-zero Gaussian process with covariance function determined by

$$\lim_{n\to\infty}\text{cov}\left\{V_{1n}(t,u), V_{2n}(t,u)\right\}.$$

This completes the proof of Theorem 3.3.

The proof of Theorem 3.4 follows from the argument given in the proof of Theorem 3.2 and the weak convergence of $(\sqrt{n}/\sigma_n)(H_{1n}(t|u) - H_1(t|u), H_{2n}(t|u) - H_2(t|u))$.

## References

[ 1 ] Andersen, P. K., Borgan, Ø., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer-Verlag, New York.

[ 2 ] Armstrong, B. (1985). Measurement error in the generalized linear models. *Communications in Statistics — Simulation and Computation* 14 529–544.

[ 3 ] Beran, R. (1981). Nonparametric regression with randomly censored survival data. Manuscript.

[ 4 ] Billingsley, P. (1968). *Convergence of probability measures*. Wiley, New York.

[ 5 ] Breslow, N. and Crowly, J. (1974). A large sample study of the life table and product limit estimates under random censorship. *Ann. Statist.* 2 437–53.

[ 6 ] Buckley, J. (1984). Additive and multiplicative models for relative survival rates. *Biometrics* 40 51–62.

[ 7 ] Buckley, J. and James, I. (1979). Linear regression with censored data. *Biometrika* 66 429–36.

[ 8 ] Carroll, R. J., Spiegelman, C. H., Lan, K. K. G., Bailey, K. T., and Abbott, R. D. (1984). On errors-in-variables for binary regression models. *Biometrika* 70 19–25.

[ 9 ] Carroll, R. J. and Hall, P. (1988). Optimal rates of convergence for deconvoluting a density. *J. Amer. Statist. Assoc.* **83** 1184–1186.

[10] Fuller, W. A. (1987). *Measurement Error Models.* Wiley, New York.

[11] Cox, D. R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34** 187–220.

[12] Dabrowska, D. (1987). Nonparametric regression with censored survival time data. *Scandinavian Journal of Statistics* **14** 181–197.

[13] Devroye, L. P. and Györfi, L. (1985). *Nonparametric Density Estimation: The $L_1$ View.* Wiley, New York.

[14] Doksum, K. A. and Yandell, B. S. (1982). Properties of regression estimates based on censored survival data. *Festschrift for Erich Lehmann.* Wadsworth, Belmont.

[15] Elandt–Johnson, R. C. and Johnson, N. (1980). Survival models and data analysis. Wiley, New York.

[16] Fan, J. (1991a). On the optimal rates of convergence for nonparametric deconvolution problem. *Ann. Statist.* **19** 1257–1272.

[17] Fan, J. (1991b). Asymptotic normality for deconvolving kernel density estimators. *Sankhyā A* **53** 97–110.

[18] Fan, J. and Gijbels, I. (1994). Censored regression: Local linear approximations and their applications. *J. Amer. Statist. Assoc.* **89** 560–570.

[19] Fan, J. and Truong, Y. K. (1993). Nonparametric regression with errors-in-variables. *Ann. Statist.* **21** 1900–1925.

[20] Fan, J., Truong, Y. K. and Wang, Y. (1990). Measurement errors regression: A nonparametric approach. Technical report No. 2042, Department of Statistics, University of North Carolina at Chapel Hill.

[21] Fan, J., Truong, Y. K. and Wang, Y. (1991). Function estimation involving errors-in-variables. In *Nonparametric Functional Estimation and Related Topics*, ed. by G. G. Roussas.

[22] Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis.* Wiley, New York.

[23] Gill, R. D. (1983). Large sample behaviour of the product limit estimator on the whole line. *Ann. Statist.* **11** 49–58.

[24] Härdle, W. (1990). *Applied Nonparametric Regression.* Cambridge University Press, New York.

[25] Huber, P. (1985). Projection pursuit. *Ann. Statist.* **13** 435–525.

[26] Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data.* Wiley, New York.

[27] Kooperberg, C., Stone, C. J., and Truong, Y. K. (1994a). Hazard regression. Technical Report No. 389, Department of Statistics, University of California, Berkeley.

[28] Kooperberg, C., Stone, C. J., and Truong, Y. K. (1994b). The $L_2$ rate of convergence of hazard regression. Technical Report No. 390, Department of Statistics, University of California, Berkeley.

[29] Miller, R. G. (1976). Least squares regression with censored data. *Biometrika* **63** 449–64.

[30] Nadaraya, E. A. (1964). On estimating regression. *Theory Prob. Appl.* **10** 186–190.

[31] O'Grady, H. K. (1994). *Nonparametric Regression with Censored Data.* Institute of Statistics, Mimeo Series No. 2126T, University of North Carolina at Chapel Hill.

[32] Pepe, M. S., Self, S. G. and Prentice, R. L. (1989). Further results on covariate measurement errors in cohort studies with time to response data. *Statistics in Medicine* **8** 1167–1178.

[33] Prentice, R. L. (1982). Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika* **69** 331–342.

[34] Silverman, B. (1986). *Density Estimation.* Chapman and Hall, London.

[35] Stefanski, L. A. and Carroll, R. J. (1985). Covariate measurement error in logistic regression. *Ann. Statist.* **13** 1335–1351.

[36] Stefanski, L. A. and Carroll, R. J. (1987). Conditional scores and optimal scores for generalized linear measurement-error models. *Biometrika* **74** 703–716.

[37] Stefanski, L. A. and Carroll, R. J. (1990). Deconvoluting kernel density estimators. *Statistics* **21** 169–184.

[38] Stone, C. J. (1986). The dimensionality reduction principle for generalized additive models. *Ann. Statist.* **14** 590–606.

[39] Watson, G. S. (1964). Smooth regression analysis. *Sankhyā A* **26** 359–372.

[40] Zhang, C. H. (1990). Fourier methods for estimating mixing densities and distributions. *Ann. Statist.* **18** 806–830.