The Annals of Statistics
1995, Vol. 23, No. 3, 787–823

AN APPROACH TO NONPARAMETRIC REGRESSION FOR LIFE HISTORY DATA USING LOCAL LINEAR FITTING[1]

By Gang Li and Hani Doss

University of North Carolina, Charlotte and Ohio State University

Most hazard regression models in survival analysis specify a given functional form to describe the influence of the covariates on the hazard rate. For instance, Cox's model assumes that the covariates act multiplicatively on the hazard rate, and Aalen's additive risk model stipulates that the covariates have a linear additive effect on the hazard rate. In this paper we study a fully nonparametric model which makes no assumption on the association between the hazard rate and the covariates. We propose a class of estimators for the conditional hazard function, the conditional cumulative hazard function and the conditional survival function, and study their large sample properties. When the size of a data set is relatively large, this fully nonparametric approach may provide more accurate information than that acquired from more restrictive models. It may also be used to test whether a particular submodel gives a good fit to a given data set. Because our results are obtained under the multivariate counting process setting of Aalen, they apply to a number of models arising in survival analysis, including various censoring and random truncation models. Our estimators are related to the conditional Nelson–Aalen estimators proposed by Beran for the random censorship model.

1. Introduction and summary. Let T be the survival time of an individual with covariate vector $\mathbf{Z} = (Z_1, \dots, Z_p)$. To assess the influence of the covariate on T, by far the most commonly used model is Cox's proportional hazards model, which stipulates that

$$h(t|\mathbf{z}) \equiv \lim_{\Delta t \searrow 0} \frac{1}{\Delta t} P(T \le t + \Delta t | T > t; \mathbf{Z} = \mathbf{z}), \tag{1.1}$$

the hazard function for an individual with covariate $\mathbf{Z} = \mathbf{z}$, has the form $h(t|\mathbf{z}) = h_0(t)\exp(\boldsymbol{\beta}'\mathbf{z})$, where $\boldsymbol{\beta}$ is a vector of unknown regression coefficients and $h_0(t)$ is an unknown hazard function. This model has the major advantage that it is parsimonious and easy to understand: The effect of the covariates is neatly summarized by the vector $\boldsymbol{\beta}$. On the other hand, the assumed functional form of the hazard rate is extremely rigid.

Let $S(t|\mathbf{z}) = P(T > t|\mathbf{Z} = \mathbf{z})$ be the conditional survival function of T given that $\mathbf{Z} = \mathbf{z}$. Beran (1981) considered the more general model in which $\{S(\cdot|\mathbf{z})\}_{\mathbf{z}}$

Received April 1993; revised April 1994.
[1]Research supported by Air Force Office of Scientific Research Grant 90-0202.
AMS 1991 *subject classifications*. Primary 62G05, 62G07; secondary, 62E20 60F05.
Key words and phrases. Aalen model, censoring, counting process, hazard function, intensity, kernel, martingale, nearest neighbor.

is a completely arbitrary family of survival functions. Supposing densities exist, this is the model

$$h(t|\mathbf{z}) = \alpha(t,\mathbf{z}), \tag{1.2}$$

where we assume only that for each $\mathbf{z}$, $\alpha(\cdot,\mathbf{z})$ is a hazard function. Beran considered estimation of $S(t|\mathbf{z})$ and the corresponding cumulative hazard function $A(t,\mathbf{z}) = \int_0^t \alpha(s,\mathbf{z})\,ds$ under a random censorship model in which the survival time of each individual is observed until a censoring time. More specifically, the data consist of n i.i.d. triples $\{(X_i, \delta_i, \mathbf{Z}_i)\}_{i=1}^n$, where for each i, $X_i = \min(T_i, C_i)$ and $\delta_i = I(T_i \le C_i)$. Suppose that $\mathbf{Z}$ does not depend on time and that T is conditionally independent of C given $\mathbf{Z}$. Beran proposed a "local Nelson–Aalen estimator" of $A(t,\mathbf{z})$, which is described as follows. For each $\mathbf{z}$, let $K_1(t|\mathbf{z}) = P(X > t,\ \delta = 1|\mathbf{Z} = \mathbf{z})$ and $K(t|\mathbf{z}) = P(X > t|\mathbf{Z} = \mathbf{z})$. The representation

$$A(t,\mathbf{z}) = -\int_0^t \frac{dK_1(s|\mathbf{z})}{K(s-|\mathbf{z})}$$

[see Peterson (1977)] led Beran to propose the class of estimators of $A(t,\mathbf{z}_0)$ given by

$$\hat{A}\,(t,\mathbf{z}_0) = -\int_0^t \frac{d\hat{K}_1(s|\mathbf{z}_0)}{\hat{K}(s-|\mathbf{z}_0)}, \tag{1.3}$$

where $\hat{K}_1(t|\mathbf{z}_0) = \sum_{i=1}^n W_i(\mathbf{z}_0)I(X_i > t,\ \delta_i = 1)$ and $\hat{K}(t|\mathbf{z}_0) = \sum_{i=1}^n W_i(\mathbf{z}_0)I(X_i > t)$. Here $\{W_i(\mathbf{z}_0),\ i = 1,\dots,n\}$ is a set of nonnegative weights depending on the covariates only. For instance, the $W_i(\mathbf{z}_0)$'s can be taken as the nearest neighbor or kernel weights used in density estimation and nonparametric regression. When one uses the constant weights $W_i(\mathbf{z}_0) = 1/n$ for all i, $\hat{A}$ reduces to the ordinary Nelson–Aalen estimator [see Aalen (1978)]. Large sample properties of Beran-type estimators have been studied by many authors, for example, Beran (1981), Dabrowska (1987) and McKeague and Utikal (1990).

The purpose of this paper is to introduce and study a new class of estimators for the conditional cumulative hazard function, the conditional survival function and the conditional hazard function. We informally motivate our estimators as follows. Beran's estimators involve a local *averaging* in the $\mathbf{z}$ direction. It seems preferable to do a "local linear fit" in the $\mathbf{z}$ direction. It may be helpful to make an analogy with scatterplot smoothers in nonparametric regression, in which we are given data $\{(Y_i, Z_i)\}$ and we wish to estimate $E(Y|Z = z_0)$. Two standard estimators are the running average smoother and the running lines smoother. Let L_{z_0} denote the set of indices of all the z's that lie in a neighborhood of z_0. The running average smoother averages all the Y's whose indices lie in L_{z_0}. The running lines smoother involves doing a least squares fit using the points $\{(Y_l, Z_l);\ l \in L_{z_0}\}$. There are some problems with smoothers based on local averages, a notable one being large biases near the endpoints of the z region. In addition, they do not generally reproduce straight lines. For this reason, the running lines smoother is generally strongly preferred in practice. For example, the scatterplot smoother provided in the statistical programming language S is a running lines smoother ("loess"). See, for example, Chambers and Hastie [(1991), page 376].

We now wish to give a preliminary description of our estimators of the conditional cumulative hazard function in the context of the random censorship model described earlier. Fix a small neighborhood $\mathscr{N}_{\mathbf{z}_0}$. Within $\mathscr{N}_{\mathbf{z}_0}$ we have

$$\alpha(t,\mathbf{z}) \approx \beta_0(t,\mathbf{z}_0) + \beta_1(t,\mathbf{z}_0)z_1 + \cdots + \beta_p(t,\mathbf{z}_0)z_p. \tag{1.4}$$

The integrated version of this is

$$A(t,\mathbf{z}) \approx B_0(t,\mathbf{z}_0) + B_1(t,\mathbf{z}_0)z_1 + \cdots + B_p(t,\mathbf{z}_0)z_p, \tag{1.5}$$

where $B_j(t,\mathbf{z}_0) = \int_0^t \beta_j(s,\mathbf{z}_0)\,ds$ $(0 \le j \le p)$. Imagine now that we have equality in (1.5). We estimate $A(t,\mathbf{z}_0)$ by a function which has jumps only at the observed death times. Suppose there is an observed death at $X_{(i)}$. Let $\mathscr{I}_i = [X_{(i)}, X_{(i)} + dt)$, let $\mathscr{R}_i$ denote the set of individuals whose covariates are in $\mathscr{N}_{\mathbf{z}_0}$ and who are at risk of dying in the interval $\mathscr{I}_i$ and let n_i be the cardinality of $\mathscr{R}_i$. Each individual in $\mathscr{R}_i$ is observed to die in the interval $\mathscr{I}_i$ or not. Thus, we have a "response" vector of length n_i, consisting of $n_i - 1$ zeros and a single 1. If we formally regress this vector on the covariates z_l, $l \in \mathscr{R}_i$, we obtain an estimate of the increment of the functions $B_j(\cdot,\mathbf{z}_0)$, $j = 0,\dots,p$, at the point $t = X_{(i)}$. Summing up those increments gives an estimate of the $B_j(\cdot,\mathbf{z}_0)$'s, $j = 0,\dots,p$, and this gives an estimate of $A(\cdot,\mathbf{z}_0)$. We note that if we take $\mathscr{N}_{\mathbf{z}_0}$ to be the entire $\mathbf{z}$-space, this procedure gives Aalen's (1980) least squares estimator. Actually, we do not do an ordinary regression, but rather a weighted regression, in which individuals whose covariates are close to $\mathbf{z}_0$ are counted more heavily. A complete description of our estimators is given in Section 2.

It is not too difficult to see that if instead of taking a first-order Taylor expansion in (1.4) and (1.5), we take a zeroth-order Taylor expansion $\alpha(t,\mathbf{z}) \approx \alpha(t,\mathbf{z}_0)$, then we obtain the estimators proposed by Beran (1981). This is discussed further in Section 2.2 below.

We note that an estimate of the conditional cumulative hazard function naturally gives rise to an estimate of the conditional survival function $S(t|\mathbf{z})$ through the product–integral representation

$$\begin{aligned} S(t|\mathbf{z}) &= \mathscr{P}_{[0,t]}(1 - dA(s,\mathbf{z})) \\ &\equiv \exp\{A^c(t,\mathbf{z})\} \prod_{s \le t} (1 - \Delta A(s,\mathbf{z})), \end{aligned} \tag{1.6}$$

where A^c is the continuous component of A and $\Delta A(s,\mathbf{z}) \equiv A(s,\mathbf{z}) - A(s-,\mathbf{z})$ [see, e.g., Gill and Johansen (1990)]. Thus, an estimate $\tilde{A}(t|\mathbf{z})$ of $A(t|\mathbf{z})$ yields the estimate

$$\tilde{S}(t|\mathbf{z}) = \prod_{s \le t} \left(1 - \Delta\tilde{A}(s,\mathbf{z})\right),$$

and moreover the asymptotic distribution of $\tilde{S}(t|\mathbf{z})$ may be obtained from that of $\tilde{A}(t|\mathbf{z})$ via the functional version of the δ-method [see Gill (1989)]. Furthermore, estimates of functionals of $S(t|\mathbf{z})$ such as the mean life length and the median life length can in turn be obtained as functionals of $\tilde{S}(t|\mathbf{z})$, and in many cases (e.g., for the median life length) the asymptotics for these functionals can also be obtained via the δ-method.

The fully nonparametric approach provides useful alternatives to estimators derived from the Cox model. Also, our estimators may be used to determine whether a more restrictive model such as the Cox model gives an adequate fit to the data. This is done by comparing estimates under the fully nonparametric model with the estimates under the more restrictive model, and carrying out a formal test of goodness of fit (assuming that the distributions of the required test statistics can be obtained). See, for example, McKeague and Utikal (1991). They are also needed in the fitting of a "nonlinear additive risks model," which will be introduced and studied in a sequel to this paper. Our development in this paper proceeds within Aalen's (1978) framework of multivariate counting processes. This offers some important advantages. As is well known by now, Aalen's model encompasses a wide range of models arising in survival analysis, for example, very general forms of censoring (type I censoring, type II censoring and the important special case of random censoring) and random truncation models; see Chapter 3 of Andersen, Borgan, Gill and Keiding (1993) for a discussion of these and for additional examples.

In Section 2 we formally introduce our estimators of the conditional cumulative hazard function, the conditional survival function and the conditional hazard function, and discuss the choice of weight functions, giving emphasis to nearest neighbor and kernel weights. In Section 3 we state results which give sufficient conditions for weak convergence of our estimators. Specifically, these are Theorem 1, which pertains to weak convergence of the conditional cumulative hazard and the conditional survival functions, and Theorem 2, which deals with weak convergence of our estimates of the conditional hazard function. These two theorems are abstractly stated and pertain to an arbitrary family of weight functions. Section 4 gives results which state that the sufficient conditions of Theorems 1 and 2 are satisfied by the nearest neighbor and kernel weights. We also discuss briefly the rate of convergence of the estimators. In Section 5 we illustrate the procedures of this paper on a data set involving survival among diabetics. The survival times in this data set are both left truncated and right censored. Section 6 contains the proofs of our theorems. In that section we shall see that a by-product of our approach is a proof of weak convergence of Beran's estimators under the multivariate counting process setting described in Section 2 [cf. Theorems 3 and 4 and part (i) of the proof of Theorem 1].

2. The model and the estimators.

2.1. *Counting process formulation.* Let $\mathbf{N}^{(n)}(t) = (N_1^{(n)}(t), \dots, N_n^{(n)}(t))'$, $t \in [0, 1]$, be an n-component multivariate counting process with respect to the filtration $\mathscr{F}^{(n)} = \{\mathscr{F}^{(n)}(t)\colon t \in [0,1]\}$. Formally, this means that for each i the sample paths of $N_i^{(n)}$ are step functions, zero at time zero, with jumps of size $+1$ only, no two component processes can jump simultaneously and for each i, $N_i^{(n)}(t)$ is $\mathscr{F}^{(n)}(t)$-measurable. Intuitively, we think of $N_i^{(n)}(t)$ as a process that counts the number of failures for the ith subject during the time interval $[0,t]$ over the study period $[0,1]$. The σ-field $\mathscr{F}^{(n)}(t)$ is thought of as containing all the information that is available at time t.

For each i, let $\mathbf{Z}_i^{(n)}(t) = (Z_{i1}^{(n)}(t),\dots,Z_{ip}^{(n)}(t))'$, $t \in [0,1]$, be a predictable covariate process and let $Y_i^{(n)}(t)$ be a predictable $\{0,1\}$-valued process, indicating (by the value 1) that the ith subject is at risk at time t. Informally, "predictability" means that the values of $\mathbf{Z}_i^{(n)}(t)$ and $Y_i^{(n)}(t)$ are fixed given what has happened just before time t. Let $\boldsymbol{\lambda}^{(n)}(t) = (\lambda_1^{(n)}(t),\dots,\lambda_n^{(n)}(t))'$, $t \in [0,1]$, be the intensity process of $\mathbf{N}^{(n)}$. Thus, $M_i^{(n)}(t) = N_i^{(n)}(t) - \int_0^t \lambda_i^{(n)}(u)\,du$, $i = 1,\dots,n$, $t \in [0,1]$, are orthogonal locally square integrable martingales with respect to $\mathscr{F}^{(n)}$. A mathematically rigorous treatment of the theory of counting processes and martingales and related notions used in this paper are given in Chapter 2 of Andersen, Borgan, Gill and Keiding (1993).

We consider the nonparametric regression model

$$\lambda_i^{(n)}(t) = Y_i^{(n)}(t)\,\alpha\big(t,\mathbf{Z}_i^{(n)}(t)\big), \qquad i = 1,\dots,n, \tag{2.1}$$

where $\alpha(\cdot,\mathbf{z})$ is an arbitrary nonnegative deterministic hazard function, and our objective is to estimate $\alpha(\cdot,\cdot)$ and other related functions on the basis of the observations $(N_i^{(n)}, Y_i^{(n)}, \mathbf{Z}_i^{(n)})$, $i = 1,\dots,n$.

The model of random right censorship described in Section 1 is a special case of this setup. In the notation of Section 1, for each i define $N_i^{(n)}(t) = I(X_i \le t,\ \delta_i = 1)$ and $Y_i^{(n)}(t) = I(X_i \ge t)$. It is well known that $\mathbf{N}^{(n)}(t) = (N_1^{(n)}(t),\dots,N_n^{(n)}(t))'$, $t \in [0,1]$, is a multivariate counting process with each individual process $N_i^{(n)}$ having intensity process $\lambda_i^{(n)}(t) = Y_i^{(n)}(t)h(t|\mathbf{Z}_i)$, where $h(t|\mathbf{z})$ is given by (1.1); see, for example, Chapter 3 of Andersen, Borgan, Gill and Keiding (1993). In this case, model (2.1) corresponds to

$$h(t|\mathbf{Z}_i) = \alpha\big(t,\mathbf{Z}_i(t)\big), \qquad i = 1,\dots,n,$$

which is identical to model (1.2). As mentioned earlier, other important models in survival analysis fit into this counting process setting. In Section 5, we review how the random truncation model fits into this framework.

To ease the notation we shall suppress the superscript n in the rest of the paper.

2.2. *A class of nonparametric estimators.* Fix $\mathbf{z}_0 = (z_{01},\dots,z_{0p})' \in \mathbb{R}^p$ and define $A(t,\mathbf{z}_0) = \int_0^t \alpha(s,\mathbf{z}_0)\,ds$ and $S(\cdot|\mathbf{z}_0) = \mathscr{P}_{[0,\cdot]}(1 - dA)$ to be conditional cumulative hazard function and the conditional survival hazard function, respectively, under model (2.1). We wish to estimate $A(t,\mathbf{z}_0)$, $S(t|\mathbf{z}_0)$ and $\alpha(t,\mathbf{z}_0)$ as functions of t, $t \in [0,1]$. Then letting $\mathbf{z}_0$ range over the covariate space, we obtain estimates of A, S and α as functions of $(t,\mathbf{z})$.

Estimation of $A(\cdot, \mathbf{z}_0)$ *and* $S(\cdot|\mathbf{z}_0)$. We consider the first-order Taylor series expansion

$$\alpha(t, \mathbf{z}) = \beta_0(t, \mathbf{z}_0) + \beta_1(t, \mathbf{z}_0) z_1 + \cdots + \beta_p(t, \mathbf{z}_0) z_p + r(t, \mathbf{z}, \mathbf{z}_0). \tag{2.2}$$

If $r(t, \mathbf{z}, \mathbf{z}_0)$ was identically equal to zero over the whole $\mathbf{z}$-region, we would have

$$N_i(t) = \int_0^t Y_i(s)\big(\beta_0(s, \mathbf{z}_0) + \beta_1(s, \mathbf{z}_0) Z_{i1}(s) + \cdots + \beta_p(s, \mathbf{z}_0) Z_{ip}(s)\big)\, ds + M_i(t),$$

$t \in [0, 1]$, $i = 1, \dots, n$, or in a matrix form,

$$\mathbf{N}(t) = \int_0^t \mathbf{U}(s)\, d\mathbf{B}(s, \mathbf{z}_0) + \mathbf{M}(t), \qquad t \in [0, 1], \tag{2.3}$$

where $\mathbf{N}(t)$ is the multivariate counting process, $\mathbf{M}(t)$ is an $n \times 1$ vector of locally square integrable martingales, $\mathbf{B}(t, \mathbf{z}_0)$ is the $(p + 1) \times 1$ vector of the integrated regression functions $B_j(t, \mathbf{z}_0) = \int_0^t \beta_j(s, \mathbf{z}_0)\, ds$, $j = 0, \dots, p$, $\mathbf{U}(s) = \mathbf{Y}(s)(\mathbf{1}, \mathbf{Z}^*(s))$, $\mathbf{Y}(s) = \operatorname{diag}(Y_1(s), \dots, Y_n(s))$ and $\mathbf{Z}^*(s) = (\mathbf{Z}_1(s), \dots, \mathbf{Z}_n(s))'$. We can then estimate $\mathbf{B}(t, \mathbf{z}_0)$ by minimizing

$$(d\mathbf{N}(t) - \mathbf{U}(t)\, d\mathbf{B}(t, \mathbf{z}_0))'(d\mathbf{N}(t) - \mathbf{U}(t)\, d\mathbf{B}(t, \mathbf{z}_0))$$

for each $t \in [0, 1]$, which yields Aalen's (1980) least squares estimator

$$\mathbf{B}_n(t, \mathbf{z}_0) = \int_0^t (\mathbf{U}'(s)\mathbf{U}(s))^- \mathbf{U}'(s)\, d\mathbf{N}(s), \qquad t \in [0, 1].$$

Here we use the convention that for a square matrix A, A^- represents the inverse of A if A is invertible and the zero matrix otherwise. Thus, the estimator of $B_j(\cdot, \mathbf{z}_0)$ increases only at the points t at which one of the counting processes N_i has a jump.

In general, $r(t, \mathbf{z}, \mathbf{z}_0)$ is not equal to zero for all $\mathbf{z}$, and thus simply fitting a linear model is not appropriate. However, if we restrict ourselves to a small neighborhood of $\mathbf{z}_0$, $r(t, \mathbf{z}, \mathbf{z}_0)$ *is* close to zero. With this in mind we define, for each subject i, a predictable weight function $W_i(t, \mathbf{z}_0)$, which at time t assigns to subject i heavy weight if $\mathbf{Z}_i(t)$ is close to $\mathbf{z}_0$ and small weight otherwise. Minimizing

$$(d\mathbf{N}(t) - \mathbf{U}(t)\, d\mathbf{B}(t, \mathbf{z}_0))'\mathbf{W}(t, \mathbf{z}_0)(d\mathbf{N}(t) - \mathbf{U}(t)\, d\mathbf{B}(t, \mathbf{z}_0)),$$

for each $t \in [0, 1]$, gives the estimate

$$\mathbf{B}_n(t, \mathbf{z}_0) = \int_0^t J(s)(\mathbf{U}'(s)\mathbf{W}(s, \mathbf{z}_0)\mathbf{U}(s))^- \times \mathbf{U}'(s)\mathbf{W}(s, \mathbf{z}_0)\, d\mathbf{N}(s), \qquad t \in [0, 1], \tag{2.4}$$

where $J(s) \equiv I(\mathbf{U}'(s)\mathbf{W}(s, \mathbf{z}_0)\mathbf{U}(s)$ is invertible) and $\mathbf{W} = \operatorname{diag}(W_1, \dots, W_n)$. Finally, we define a locally weighted least squares estimator of $A(t, \mathbf{z}_0)$ by

$$A_n(t, \mathbf{z}_0) = (1, \mathbf{z}_0')\mathbf{B}_n(t, \mathbf{z}_0). \tag{2.5}$$

The conditional survival function $S(t|\mathbf{z}_0)$ in (1.6) is estimated by

$$S_n(t|\mathbf{z}_0) = \mathscr{P}_{[0,t]}(1 - dA_n) = \prod_{s \le t}(1 - \Delta A_n(s, \mathbf{z}_0)). \tag{2.6}$$

REMARK. Assume that one replaces (2.2) with a zeroth-order Taylor series expansion

$$\alpha(t, \mathbf{z}) = \alpha(t, \mathbf{z}_0) + r(t, \mathbf{z}, \mathbf{z}_0). \tag{2.7}$$

Then, by minimizing

$$(d\mathbf{N}(t) - \mathbf{Y}(t)\, dA(t, \mathbf{z}_0))'\mathbf{W}(t, \mathbf{z}_0)(d\mathbf{N}(t) - \mathbf{Y}(t)\, dA(t, \mathbf{z}_0)),$$

for every $t \in [0, 1]$, where $\mathbf{Y}(t) = (Y_1(t), \dots, Y_n(t))'$, we obtain the estimator

$$\hat{A}(t, \mathbf{z}_0) = \int_0^t I\left(\sum_{i=1}^n W_i(s, \mathbf{z}_0) Y_i(s) \neq 0\right) \times \frac{\sum_{i=1}^n W_i(s, \mathbf{z}_0) Y_i(s)\, dN_i(s)}{\sum_{i=1}^n W_i(s, \mathbf{z}_0) Y_i(s)}. \tag{2.8}$$

In the special case where the data consist of right censored observations and the covariates are independent of time, $\hat{A}$ is equal to (1.3), the estimator introduced by Beran (1981).

In Section 6 we show that under certain conditions A_n and $\hat{A}$ are asymptotically equivalent as $n \to \infty$. However, A_n and $\hat{A}$ behave quite differently for small or moderate size samples. Because (2.7) ignores the role played by the slope function $\alpha'_{\mathbf{z}}(t, \mathbf{z}_0)$, $\hat{A}$ tends to flatten out the covariate effects within the neighborhood of $\mathbf{z}_0$. This in general will cause a bias, which may be severe when $\mathbf{z}_0$ is close to the boundary of the $\mathbf{z}$-region. This effect was observed in simulation studies not reported here. In practical terms, the inclusion of the linear term in the estimation enables us to use larger neighborhoods of $\mathbf{z}_0$.

Estimation of the conditional hazard function. To estimate $\alpha(t, \mathbf{z}_0)$ for fixed $\mathbf{z}_0$, we shall smooth $A_n(t, \mathbf{z}_0)$ over t with a kernel function. Let K be a bounded density function supported on $[-1, 1]$ and satisfying

$$\int_{-1}^1 uK(u)\, du = 0 \tag{2.9}$$

and let $\{b_n\}$ be a sequence of positive numbers. Define

$$\alpha_n(t, \mathbf{z}_0) = \frac{1}{b_n}\int_0^1 K\left(\frac{t-s}{b_n}\right) A_n(ds, \mathbf{z}_0), \tag{2.10}$$

where $A_n(t, \mathbf{z}_0)$ is given by (2.5) and (2.4). In Section 3 we state theorems that assert that $\alpha_n(t, \mathbf{z}_0)$ is a consistent estimator of $\alpha(t, \mathbf{z}_0)$; the theorems also give the rate of convergence.

2.3. *Nearest neighbor estimates and kernel estimates.* Different choices of the weight functions $W_i(t, \mathbf{z}_0)$ yield different types of estimators. Following are some natural examples that may be used in practice.

Nearest neighbor estimates. Let $\{k_n\}$ be a sequence of positive integers. For k-nearest neighbor (k-NN) estimates the weights are defined by

$$W_i(t, \mathbf{z}_0) = w\left(\frac{\mathbf{z}_0 - \mathbf{Z}_i(t)}{R_n}\right) \Big/ \sum_{j=1}^{n} w\left(\frac{\mathbf{z}_0 - \mathbf{Z}_j(t)}{R_n}\right), \qquad 1 \le i \le n, \tag{2.11}$$

where $w(\cdot)$ is a density function in $\mathbb{R}^p$ that vanishes outside the unit ball $\{\mathbf{u} \in \mathbb{R}^p: |\mathbf{u}| \le 1\}$ and R_n is the Euclidean distance between $\mathbf{z}_0$ and the k_nth nearest of $\mathbf{Z}_1(t), \dots, \mathbf{Z}_n(t)$.

Kernel estimates. For kernel estimators, the weights are defined by

$$W_i(t, \mathbf{z}_0) = w\left(\frac{\mathbf{z}_0 - \mathbf{Z}_i(t)}{h_n}\right) \Big/ \sum_{j=1}^{n} w\left(\frac{\mathbf{z}_0 - \mathbf{Z}_j(t)}{h_n}\right), \qquad 1 \le i \le n, \tag{2.12}$$

where $w(\cdot)$ is a density function in $\mathbb{R}^p$ and $h_n > 0$ is the "bandwidth parameter."

An advantage of the k-NN estimates is that they are locally adaptive: when the covariates have small density at $\mathbf{z}_0$, observations around $\mathbf{z}_0$ are sparse, but R_n is then larger. For this reason, nearest neighbor estimates are usually preferred; for example, the S function "loess" mentioned in Section 1 is a k-NN estimate; see Hastie and Tibshirani [(1990), pages 29–30].

3. Weak convergence of the estimators. In this section we study large sample properties of the estimators defined in Section 2. Throughout the paper we shall assume that $\mathbf{Z}$ takes values only in $[0,1]^p$, and that $\sup_{(s,\mathbf{z}) \in [0,1]\times[0,1]^p} |\alpha(s, \mathbf{z})| = B < \infty$. We also adopt the convention that $0/0$ is 0.

3.1. *Notation and assumptions.* Fix $\mathbf{z}_0 = (z_{01}, \dots, z_{0p})' \in [0,1]^p$ and denote $\mathbf{Z}^* = (\mathbf{Z}_1, \dots, \mathbf{Z}_n)'$. Define

$$c_i(s, \mathbf{z}_0) = W_i(s, \mathbf{z}_0) Y_i(s) \Big/ \sum_{j=1}^{n} W_j(s, \mathbf{z}_0) Y_j(s), \qquad i = 1, \dots, n,$$

$$\begin{aligned} \mathbf{c}(s) &= (c_1(s, \mathbf{z}_0), \dots, c_n(s, \mathbf{z}_0))', \\ C(s) &= \operatorname{diag}(c_1(s, \mathbf{z}_0), \dots, c_n(s, \mathbf{z}_0)), \\ P(s) &= C(s) - \mathbf{c}(s)\mathbf{c}(s)', \\ \mathscr{T}_{\mathbf{z}_0} &= \Big\{ t \in [0,1]: \inf_{u \in [0,t]} \det(\mathbf{Z}^{*\prime} P \mathbf{Z}^*)(u) > 0 \Big\}, \\ J_1(s) &= I(s \in \mathscr{T}_{\mathbf{z}_0}), \end{aligned} \tag{3.1}$$

for all s and all i. The $c_i(s, \mathbf{z}_0)$'s are essentially the weights assigned to the n individuals, taking into account who is at risk at time s (those individuals who are not at risk are given a weight of zero). To be more precise, we need to look at the set $\mathscr{T}_{\mathbf{z}_0}$, and when thinking about the definition of this set it is very helpful to first consider the case where the covariate is one dimensional. In this case, the condition $\det(\mathbf{Z}^{*\prime} P \mathbf{Z}^*)(u) > 0$ is the condition that $\sum_{i=1}^n c_i(s, z_0)(Z_i(s) - \sum_{j=1}^n c_j(s, z_0) Z_j(s))^2 > 0$. Note that if $\sum_{j=1}^n W_j(s, \mathbf{z}_0) Y_j(s) = 0$, then $c_i(s, \mathbf{z}_0) = 0$ for all i (by our convention that $0/0 = 0$), which means that s cannot be in $\mathscr{T}_{\mathbf{z}_0}$. Thus, if $s \in \mathscr{T}_{\mathbf{z}_0}$, we have $\sum_{j=1}^n W_j(s, \mathbf{z}_0) Y_j(s) > 0$, so that $\sum_{i=1}^n c_i(s, \mathbf{z}_0) = 1$. Therefore, if $s \in \mathscr{T}_{\mathbf{z}_0}$, then $\sum_{i=1}^n c_i(s, z_0) Z_i(s)$ and $\sum_{i=1}^n c_i(s, z_0)(Z_i(s) - \sum_{j=1}^n c_j(s, z_0) Z_j(s))^2$ represent the weighted mean and weighted variance of the $Z_i(s)$'s. For the case where the covariate is multidimensional, if $s \in \mathscr{T}_{\mathbf{z}_0}$, then

$$\mathbf{Z}^*(s)'\mathbf{c}(s) = \sum_{i=1}^n c_i(s, \mathbf{z}_0) \mathbf{Z}_i(s)$$

and

$$\mathbf{Z}^*(s)' P(s) \mathbf{Z}^*(s) = \sum_{i=1}^n c_i(s, \mathbf{z}_0) \mathbf{Z}_i(s) \mathbf{Z}_i(s)' - \left(\sum_{i=1}^n c_i(s, \mathbf{z}_0) \mathbf{Z}_i(s) \right) \left(\sum_{i=1}^n c_i(s, \mathbf{z}_0) \mathbf{Z}_i(s) \right)'$$

are the weighted mean and weighted covariance matrix of $\mathbf{Z}_1(s), \ldots, \mathbf{Z}_n(s)$.

The main results of this section are proved under the following two sets of conditions. Condition A is needed for the results pertaining to the estimators of the conditional cumulative hazard function and the conditional survival function, and Condition B is needed to obtain the asymptotics for estimators of the conditional hazard function. Although it would seem at first sight that these conditions are forbidding and unintuitive, in fact this is not the case, and in Section 4 we give relatively straightforward verification that the conditions are satisfied by the k-NN and kernel estimates. The limits in Conditions A and B are taken as $n \to \infty$.

CONDITION A. Let $\{a_n\}_1^\infty$ be a sequence of positive numbers. These will be connected with the smoothing parameter of the estimators; for example, when kernel estimators are used, a_n will be taken to be the bandwidth and when k-NN estimators are used, a_n will be taken to be k_n/n.

(A1) $P(\mathscr{T}_{\mathbf{z}_0} = [0,1]) \to 1$.

(A2) There exists a nonnegative measurable function $g_\delta(s, \mathbf{z})$ indexed by $\delta \geq 0$ and defined on $[0,1] \times [0,1]^p$, such that for $\delta = 0$ and for some $\delta > 0$,

$$\int_0^1 \left| (na_n)^{1+\delta} \sum_{i=1}^n c_i^{2+\delta}(s, \mathbf{z}_0) - g_\delta(s, \mathbf{z}_0) \right| ds \to_P 0.$$

(A3) $\sqrt{na_n} \int_0^1 |\sum_{i=1}^n c_i(s, \mathbf{z}_0)(\alpha(s, \mathbf{Z}_i(s)) - \alpha(s, \mathbf{z}_0))| \, ds \to_P 0$.

(A4) $\sqrt{na_n} \int_0^1 \sum_{i=1}^n c_i(s, \mathbf{z}_0)(\alpha(s, \mathbf{Z}_i(s)) - \alpha(s, \mathbf{z}_0))^2 \, ds \to_P 0$.

(A5) $na_n c_i(s, \mathbf{z}_0) = O_p(1)$ uniformly in s and i.
(A6) $\sqrt{na_n} \int_0^1 J_1(s)(\mathbf{Z}^*(s)'\mathbf{c}(s) - \mathbf{z}_0)'(\mathbf{Z}^*(s)'P(s)\mathbf{Z}^*(s))^{-1}(\mathbf{Z}^*(s)'\mathbf{c}(s) - \mathbf{z}_0)\, ds \to_P 0$.

CONDITION B. Let $\{b_n\}_1^\infty$ be the sequence of positive numbers used in (2.10) and assume that $b_n \to 0$. Let $\{a_n\}_1^\infty$ be a sequence of positive numbers such that $a_n \to 0$.

(B1) $P(\mathscr{T}_{\mathbf{z}_0} = [0,1]) \to 1$.
(B2) There exists a nonnegative measurable function $g_\delta(s, \mathbf{z})$ indexed by $\delta \geq 0$ and defined on $[0,1] \times [0,1]$, such that for each $t \in [0,1]$,

$$\int_{t-b_n}^{t+b_n} \left| (na_n)^{1+\delta} \sum_{i=1}^n c_i^{2+\delta}(s, \mathbf{z}_0) - g_\delta(s, \mathbf{z}_0) \right| ds = o_p(b_n)$$

for $\delta = 0$ and some $\delta > 0$.
(B3) $\sqrt{na_n/b_n} \int_{t-b_n}^{t+b_n} |\sum_{i=1}^n c_i(s, \mathbf{z}_0)(\alpha(s, \mathbf{Z}_i(s)) - \alpha(s, \mathbf{z}_0))|\, ds \to_P 0$ for each $t \in [0,1]$.
(B4) $\int_{t-b_n}^{t+b_n} \sum_{i=1}^n c_i(s, \mathbf{z}_0)(\alpha(s, \mathbf{Z}_i(s)) - \alpha(s, \mathbf{z}_0))^2\, ds = O_p(b_n^3)$ for each $t \in [0,1]$.
(B5) $na_n c_i(s, \mathbf{z}_0) = O_p(1)$ uniformly in s and i.
(B6) $na_n b_n^2 \int_{t-b_n}^{t+b_n} J_1(s)(\mathbf{Z}^*(s)'\mathbf{c}(s) - \mathbf{z}_0)'(\mathbf{Z}^*(s)'P(s)\mathbf{Z}^*(s))^{-1}(\mathbf{Z}^*(s)'\mathbf{c}(s) - z_0) \times ds \to_P 0$.

3.2. *Main theorems.* We now state our main results. The proofs are given in Section 6. Let $D[0,1]$ be the standard Skorohod space on $[0,1]$.

THEOREM 1. *Let $A_n(t, \mathbf{z}_0)$ and $S_n(t|\mathbf{z}_0)$ be defined by* (2.5) *and* (2.6), *and let $\{a_n\}_1^\infty$ be a sequence of positive numbers such that $na_n \to \infty$. Then, under conditions* (A1)–(A6),

$$\sqrt{na_n}\,(A_n(\cdot, \mathbf{z}_0) - A(\cdot, \mathbf{z}_0)) \to_d U(\cdot, \mathbf{z}_0) \quad \text{in } D[0,1] \tag{3.2}$$

and

$$\sqrt{na_n}\,(S_n(\cdot|\mathbf{z}_0) - S(\cdot|\mathbf{z}_0)) \to_d S(\cdot|\mathbf{z}_0)U(\cdot, \mathbf{z}_0) \quad \text{in } D[0,1], \tag{3.3}$$

where $U(\cdot, \mathbf{z}_0)$ is a continuous Gaussian martingale with mean zero and variance function

$$v(t) = \int_0^t g_0(s, \mathbf{z}_0)\,\alpha(s, \mathbf{z}_0)\, ds.$$

REMARK. It is worth noting that Theorem 1 does not assume $a_n \to 0$. This provides a unified approach for establishing weak convergence of both conditional and unconditional empiricals. See the remark following the proof of Theorem 1 in Section 6.

THEOREM 2. *Let $\alpha_n(t, \mathbf{z}_0)$ be the estimator of the hazard function $\alpha(t, \mathbf{z}_0)$ defined by* (2.10). *Assume the sequence $\{b_n\}_1^\infty$ of positive numbers appearing in*

(2.10) *satisfies* $nb_n^{p+3} \to \infty$ *and* $nb_n^{p+5} \to 0$. *Let* $\{a_n\}_1^\infty$ *be a sequence of positive numbers such that* $a_n \sim b_n^p$. *Suppose also that* $\alpha(t, \mathbf{z}_0)$ *is twice differentiable with respect to* t, *that* $D = \sup_{t \in [0,1]} |\alpha_t''(t, \mathbf{z}_0)| < \infty$ *and that* $g_0(t, \mathbf{z}_0)$ *is continuous in* $t \in (0,1)$. *Then, under conditions* (B1)–(B6),

(3.4) $\sqrt{na_n b_n}\,(\alpha_n(t, \mathbf{z}_0) - \alpha(t, \mathbf{z}_0)) \to_d \mathcal{N}(0, \sigma_t^2)$ *for every* $t \in (0,1)$,

where

$$\sigma_t^2 = g_0(t, \mathbf{z}_0)\,\alpha(t, \mathbf{z}_0) \int_{-1}^{1} K^2(u)\,du. \tag{3.5}$$

4. Asymptotics for *k*-NN and kernel estimates. In this section, we study large sample properties of the k-NN estimates and kernel estimates defined in Section 2 for the important special case where:

1. The observations $(N_1, Y_1, \mathbf{Z}_1), \ldots, (N_n, Y_n, \mathbf{Z}_n)$ are i.i.d.
2. The covariates $\mathbf{Z}_i$'s are time independent and have a common density function $f(\mathbf{z})$.
3. Each predictable indicator process Y_i has paths which are left continuous and of bounded variation.

Theorems 3 and 4 state that the sufficient Conditions A and B are satisfied when we use k-NN and kernel weights, respectively. The proofs of these two theorems are similar, with the proof of Theorem 3 being somewhat harder. In addition, as we discussed earlier, we feel that the result for k-NN estimates is more practically relevant. For these reasons, we have given only the proof of Theorem 3 in this paper; the proof of Theorem 4 appears in the technical report by Li and Doss (1994).

Because the weight functions used in (2.4) no longer depend on time, we shall use the notation $W_i(\mathbf{z})$ instead of $W_i(t, \mathbf{z})$. We note the trivial fact that if Y_i has finitely many jumps (as will be the case in all the situations of interest to us), then Y_i is of bounded variation. The following regularity conditions are also assumed throughout this section.

(R1) For each $s \in [0,1]$, the partial derivative $\partial\alpha/\partial\mathbf{z}$ exists at $\mathbf{z}_0$ and is bounded in $s \in [0,1]$. In addition, there exists a constant K_1 that is independent of s such that

$$\left| \alpha(s, \mathbf{z}) - \alpha(s, \mathbf{z}_0) - \left(\frac{\partial\alpha}{\partial\mathbf{z}}(s, \mathbf{z}_0) \right)' (\mathbf{z} - \mathbf{z}_0) \right| \le K_1 \|\mathbf{z} - \mathbf{z}_0\|^2.$$

(R2) For each s, the subdensity function

$$f(\mathbf{z}, s) = \frac{\partial^p}{\partial z_1 \cdots \partial z_p} P(\mathbf{Z}_1 \le \mathbf{z}, Y_1(s) = 1) \tag{4.1}$$

of the subdistribution function $P(\mathbf{Z}_1 \le \mathbf{z}, Y_1(s) = 1)$ exists. In addition, there is a constant M that is independent of s such that for all $(\mathbf{z}_1, s), (\mathbf{z}_2, s) \in [0,1]^p \times [0,1]$,

$$|f(\mathbf{z}_2, s) - f(\mathbf{z}_1, s)| \le M \|\mathbf{z}_2 - \mathbf{z}_1\|. \tag{4.2}$$

(R3) The density function $f(\mathbf{z})$ of $\mathbf{Z}_i$ is positive and continuous at $\mathbf{z}_0$. Suppose further that the function $H(s, \mathbf{z}) \equiv P(Y_1(s) = 1 | \mathbf{Z}_1 = \mathbf{z})$ is continuous in $\mathbf{z}$ at $\mathbf{z}_0$ for every fixed s and satisfies $\inf_{s \in [0,1]} H(s, \mathbf{z}_0) > 0$.

Because $f(\mathbf{z}, s) = 0$ for $\mathbf{z} \notin [0,1]^p$, (4.2) implies that $f(\mathbf{z}, s)$ is bounded on $\mathbb{R}^p \times [0,1]$.

4.1. *Asymptotics for k-NN estimates.*

THEOREM 3. *Assume the setup described in the beginning of this section. Let $W_i(\mathbf{z})$ be defined by* (2.11), *where $w(\cdot)$ is a bounded radially symmetric density function satisfying*

$$w(\mathbf{u}) = 0 \quad \text{for all } \|\mathbf{u}\| > 1, \tag{4.3}$$

$$w(c\mathbf{u}) \geq w(\mathbf{u}) \quad \text{for any } 0 \leq c \leq 1 \text{ and } \mathbf{u} \in \mathbb{R}^p. \tag{4.4}$$

Let $a_n = k_n/n$, $b_n = (k_n/n)^{1/p}$ and

$$g_\delta(s, \mathbf{z}_0) = \left(\frac{2}{H(s, \mathbf{z}_0)} \right)^{1+\delta} \int w^{2+\delta}(\mathbf{u})\, d\mathbf{u}.$$

(i) *If $k_n \to \infty$ and $k_n^{p+4}/n^4 \to 0$, then $na_n \to \infty$ and* (A1)–(A6) *hold.*
(ii) *If $k_n^{p+3}/n^3 \to \infty$ and $k_n^{p+5}/n^5 \to 0$, then $nb_n^{p+3} \to \infty$, $nb_n^{p+5} \to 0$ and* (B1)–(B6) *hold.*

4.2. *Asymptotics for kernel estimates.*

THEOREM 4. *Assume the setup described in the beginning of this section. Let $W_i(\mathbf{z})$ be defined by* (2.12), *where $w(\cdot)$ is a bounded radially symmetric density function satisfying*

$$c_1 I_{(\|\mathbf{u}\| \leq r)} \leq w(\mathbf{u}) \leq c_2 I_{(\|\mathbf{u}\| \leq r)} \quad \text{for some positive constants } r,\ c_1 \text{ and } c_2. \tag{4.5}$$

Let $a_n = h_n^p$, $b_n = h_n$ and

$$g_\delta(s, \mathbf{z}) = \frac{1}{\{f(\mathbf{z}) H(s, \mathbf{z})\}^{1+\delta}} \int w^{2+\delta}(\mathbf{u})\, d\mathbf{u}.$$

(i) *If $nh_n^p \to \infty$ and $nh_n^{p+4} \to 0$, then $na_n \to \infty$ and* (A1)–(A6) *hold.*
(ii) *If $nh_n^{p+3} \to \infty$ and $nh_n^{p+5} \to 0$, then $nb_n^{p+3} \to \infty$, $nb_n^{p+5} \to 0$ and* (B1)–(B6) *hold.*

REMARK. Our results are close to optimal in tems of rate of convergence. As an example, we give a brief discussion on this point for the kernel estimates. Details on optimal local smoothing in the multivariate counting process model will be discussed in a sequel to this paper. By "rate of convergence" we shall mean the normalizing factors $(na_n)^{1/2}$ appearing in (3.2) and (3.3), and $(na_n b_n)^{1/2}$ appearing in (3.4), respectively. Clearly we wish to take a_n and b_n to be as big as possible. For simplicity, let us assume $p = 1$.

The assumptions in part (i) of Theorem 4 require that $h_n = o(n^{-1/5})$, giving a rate of convergence just slightly under $n^{2/5}$ for the conditional cumulative hazard and conditional survival estimates A_n and S_n. From the point view of the smoothing literature on density and regression function estimation, this is the best rate that one can expect to achieve in weak convergence results of the form (3.2) and (3.3) without asymptotic bias. In fact, it can be shown by modifying our method that when $h_n = O(n^{-1/5})$, both A_n and S_n have an $n^{2/5}$ rate convergence with an asymptotic bias term. The asymptotic bias is a small price to pay for the faster rate of convergence (therefore substantial gain in the reduction of mean square error).

Similar remarks can be made on the rate of convergence of the conditional hazard estimate $\alpha_n(t, z)$. Without asymptotic bias, our assumption $h_n = o(n^{-1/6})$ in part (ii) of Theorem 4 leads to the rate $(na_n b_n)^{1/2} = o(n^{1/3})$ for the convergence of $\alpha_n(t, z)$, which is arbitrarily close to the optimal rate for estimating a $\mathscr{C}^2$ density in two dimensions (one dimension for z and one for t). One can show that our method gives the expected optimal $n^{1/3}$ rate of convergence for $\alpha_n(t, z)$ with an asymptotic bias when $h_n = O(n^{-1/6})$. It is worth noting that an optimal rate estimator of $\alpha(t, z)$ can be obtained by smoothing any optimal rate estimator of $A(t, z)$ in the t direction with appropriately selected bandwidths $b_n = O(h_n)$.

5. Illustration on survival in diabetics data. We apply the nonparametric regression method described in Section 2 to study survival among insulin-dependent diabetics in Fyn County, Denmark, using data collected by Dr. Anders Green from Odense, Denmark. This data set consists of 1499 patients who suffered from insulin-dependent diabetes mellitus ("diabetes" for short) on July 1, 1973. The data were obtained by recording all insulin prescriptions in the National Health Service files for this county during a five month period covering the above date, and subsequent check of each patient's medical record. Each patient was then followed from July 1, 1973 until death, emigration or January 1, 1982, whichever came first. On January 1, 1982, there were 254 observed deaths among 783 male diabetics and 237 observed deaths among 716 female patients. Of interest is the mortality of diabetics, taking into account potential risk factors. Here we shall focus on the effect of age at diabetes onset on the duration of disease. The date of diabetes onset is defined to be the first time the physician established the diagnosis.

Clearly this data set is *right censored* since some patients either were still alive on January 1, 1982 or had early emigration. It is also *left truncated* because a diabetic may be included in the followup study only if he or she was alive on July 1, 1973. More precisely, for patient i, let X_i = survival time (the period from diabetes onset to death), C_i = time elapsed from diabetes onset to emigration or January 1, 1982, $\delta_i = I(X_i \le C_i)$ and T_i = length of the period from diabetes onset to July 1, 1973. Then a triple $(T_i, \min\{X_i, C_i\}, \delta_i)$ is observed only if patient i is included in the study and $T_i < X_i$. Nothing is observable for patient i if $T_i \geq X_i$, that is, if patient i died before July 1, 1973. In addition, the age Z_i at onset of diabetes was recorded for each case i.

Because the chance of survival may vary with sex, we do separate analyses for the male and female groups. Let I_f denote the index set for the female group. Assume that $(T_i, \min\{X_i, C_i\}, \delta_i, Z_i)$, $i \in I_f$, are independent and identically distributed. Assume further that the truncation time T_i and the censoring time C_i are conditionally independent of the survival time X_i and that $T_i < C_i$ with conditional probability 1, given that $Z_i = z$. For each $i \in I_f$, define $N_i^{(n)}(t) = I(T_i < X_i \leq t,\ \delta_i = 1)$ and $Y_i^{(n)}(t) = I(T_i < t \leq \min\{X_i, C_i\})$. Then $(N_i^{(n)}(t),\ i \in I_f)$, $t \in [0, \infty)$, is a multivariate counting process with each $N_i^{(n)}$ having intensity process $\lambda_i^{(n)}(t) = Y_i^{(n)}(t)h(t|Z_i)$, where $h(t|z)$ is the conditional hazard function of the survival time X given that $Z = z$. [See, e.g., Section III.3 of Andersen, Borgan, Gill and Keiding (1993).] Therefore this model falls into the counting process framework described in Section 2. We assume the same probability model for the male group.

The Fyn county diabetes data have been studied by many authors [see, e.g., Green, Borch-Johnsen, Andersen, Hougaard, Keiding, Kreiner and Deckert (1985) (hereafter, G85) and Andersen, Borch-Johnsen, Deckert, Green, Hougaard, Keiding and Kreiner (1985) (hereafter A85)]. A85 analyzed a subset of this data set that included those who had diagnosis established before age 31 years. It is intuitive that the effect of the covariate Z may depend on the time variable. For instance, a patient who is diagnosed as having diabetes at age 30 is more likely to die after 40 years than after 20 years. However, for a person who has diagnosis established at age 10, the chance of dying after 40 years may not be very much different from that after 20 years. So on intuitive grounds, one can question the appropriateness of the classical Cox (1972) model. The data analysis in A85 confirms that this model does not give a good fit to the data. They also showed that the hazards for female and male diabetics are not proportional. A85 used models of multiplicative hazards type with both individual characteristics and changing trends in mortality included in the baseline hazard function. The completely nonparametric regression method proposed in this paper provides a natural alternative inference method for the Fyn county diabetes data.

For each group, we computed the estimate $S_n(t|z)$ [see (2.6)] of the conditional survival function $S(t|z) = P(X > t|Z = z)$ using the weight function given by (2.11) with $w(u) = \frac{1}{2}I(-1 \leq u \leq 1)$ and $k = 30\% \times$ sample size. (We actually tried different values for k and there were no substantial differences among the conclusions being reached.) Figures 1 and 2 give plots of $S_n(t|z)$ as a function of t for $z = 5, 10, 15, \dots, 80$ for female and male groups respectively. Figure 3 compares 95% simultaneous confidence bands of the conditional survival function $S(t|z)$ between female and male patients for $z = 10$, 25, 40 and 70. Figure 4 compares the plots of the estimated median survival time versus the covariate z between the two sex groups.

As mentioned earlier, one expects to see the general trend that $S(t|z)$ decreases as z increases. Figures 1 and 2 reveal this trend for both sex groups, and also show that this trend does not behave in a uniform way. The influence of z is much more significant over some z intervals than over others. For instance, the survival probabilities for female diabetics drop dramatically as z goes from 30 to 45, but the changes that occur as z varies from 5 to 30 are less significant. A similar conclusion can be drawn for male patients.

Figures 1 and 2 also indicate that there is an interaction between the influence of age z at diagnosis and duration t of disease. For the female group, Figure 1(a) shows that for $z \leq 30$, the influence of z is more significant over the range $20 \leq t \leq 30$ than it is over the range $t \leq 20$ or $30 \leq t \leq 38$. [This effect was also mentioned on page 925 of A85 in which a slightly different time variable was used.] We do not draw any conclusion for the range $t \geq 38$ since the nonparametric estimator $S_n(t|z)$ is not stable in its right tail. For $30 \leq z \leq 45$ [Figure 1(b)], the influence of z is very significant and the magnitude of this influence goes up dramatically as t increases. For example, $(S_n(10|30))/(S_n(10|45)) = 1.18$ compared to $(S_n(30|30))/(S_n(30|45)) = 4.0$. For $45 \leq z \leq 65$ [Figure 1(c)], the influence of z is also significant, but the interaction between z and t is more difficult to discern. When $z \geq 65$ [Figure 1(d)], the influence of z is essentially insignificant. Similar effects of z are found (Figure 2) for male diabetics except that the interaction between the effects of z and t is less serious.

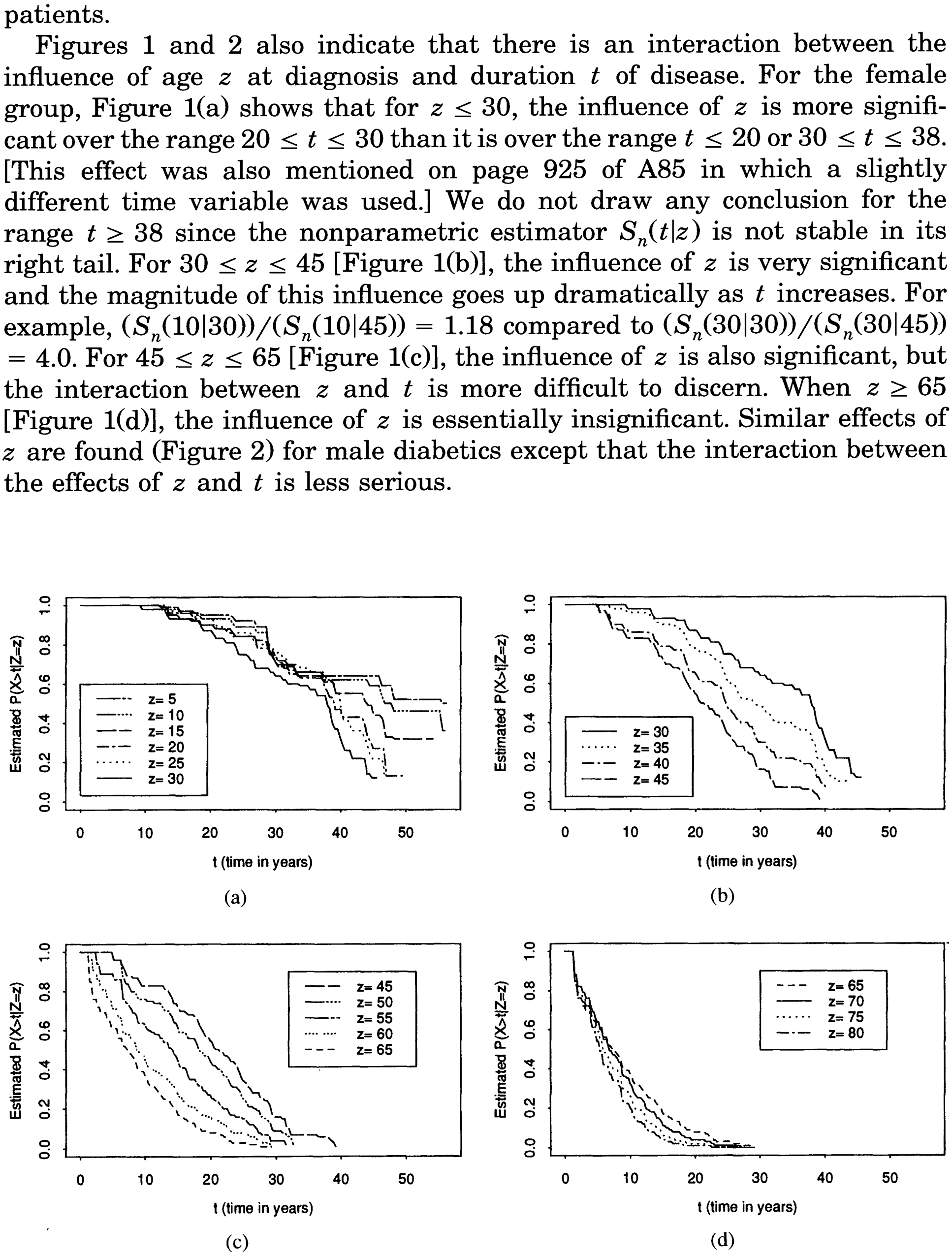

FIG. 1. *Estimated conditional survival functions for female diabetics* (z = *age at diagnosis of diabetes*).

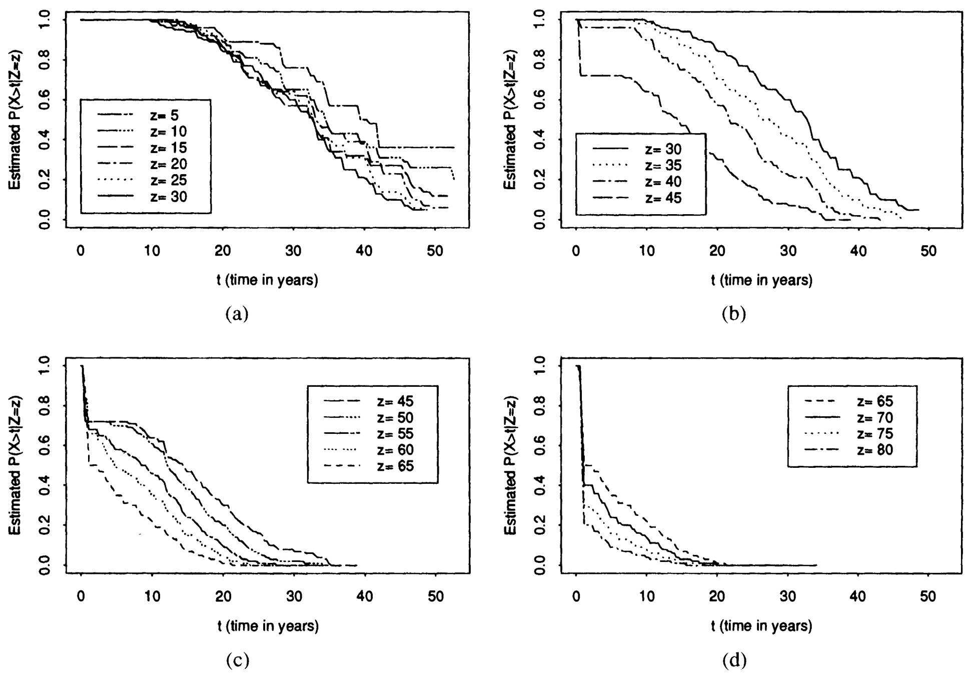

FIG. 2. *Estimated conditional survival functions for male diabetics* (z = *age at diagnosis of diabetes*).

Comparison of 95% confidence bands of conditional survival function $S(\cdot|z)$ between female and male patients (Figure 3) shows that the survival probabilities of female diabetics are consistently higher than those of male diabetics for different levels of z. The same trend is also shown in the estimated median plots (Figure 4) for the different sexes.

6. Proofs of the theorems.

PROOF OF THEOREM 1. It is helpful at this point to review the definitions given in (3.1). Recall that $\hat{A}(t,\mathbf{z}_0)$ is defined by (2.8). Let us write

$$\hat{A}\,(t,\mathbf{z}_0) = \int_0^t \tilde{J}(s)\mathbf{c}(s)'\,d\mathbf{N}(s), \tag{6.1}$$

where $\tilde{J}(s) = I(\sum_{i=1}^n c_i(s,\mathbf{z}_0) \neq 0)$. Note that $\tilde{J}(s)$ is the indicator that is required to be 1 when doing a weighted average in the definition of the Beran-type estimator (6.1), and $J(s)$ [defined right after (2.4)] is the indicator that is required to be 1 when doing a weighted least squares fit. Define $\mathbf{h}(s) = (h_1(s,\mathbf{z}_0),\dots,h_n(s,\mathbf{z}_0))'$ by

$$\mathbf{h}(s) = P(s)\mathbf{Z}^*(s)(\mathbf{Z}^*(s)'P(s)\mathbf{Z}^*(s))^{-1}(\mathbf{z}_0 - \mathbf{Z}^*(s)'\mathbf{c}(s)) \tag{6.2}$$

and also define $R^{(n)}(t,\mathbf{z}_0) = \int_0^t J(s)\mathbf{h}(s)'\,d\mathbf{N}(s)$.

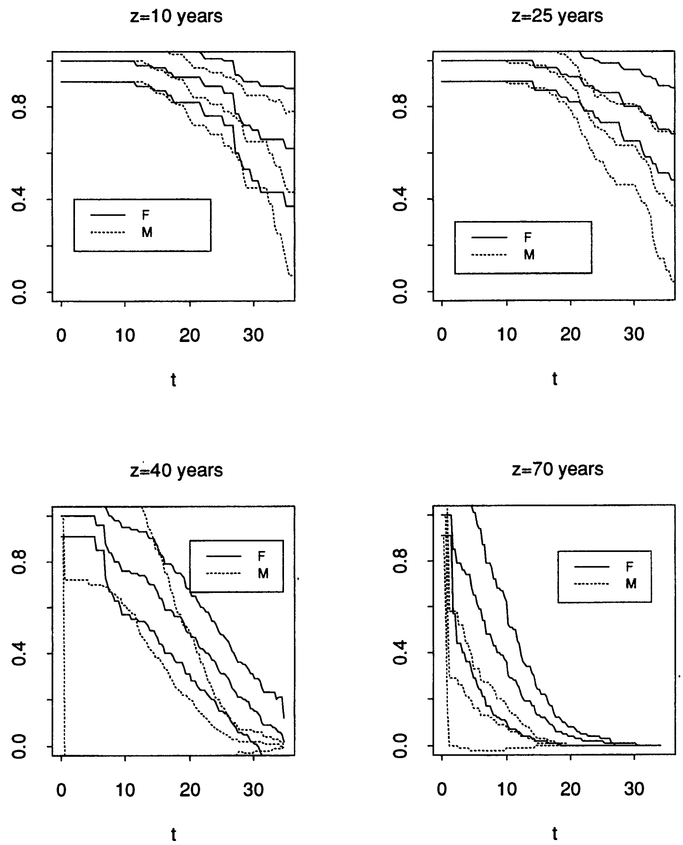

FIG. 3. *95% confidence bands for conditional survival functions $S(t|z)$ (z = age at diagnosis of diabetes, F = female, M = male).*

The key to proving Theorem 1 is to first establish the decomposition

$$
\begin{aligned}
A_n(t,\mathbf{z}_0) &= (1,\mathbf{z}_0')\mathbf{B}_n(t,\mathbf{z}_0) \\
&= \int_0^t J(s)(1,\mathbf{z}_0')\big(\mathbf{U}(s)'\mathbf{W}(s,\mathbf{z}_0)\mathbf{U}(s)\big)^{-1} \\
&\qquad \times \mathbf{U}(s)'\mathbf{W}(s,\mathbf{z}_0)\,d\mathbf{N}(s) \\
&= \int_0^t J(s)(1,\mathbf{z}_0')\begin{pmatrix} 1 & \mathbf{c}(s)'\mathbf{Z}^*(s) \\ \mathbf{Z}^*(s)'\mathbf{c}(s) & \mathbf{Z}^*(s)'C(s)\mathbf{Z}^*(s)\end{pmatrix}^{-1} \\
&\qquad \times \begin{pmatrix} \mathbf{c}(s)' \\ \mathbf{Z}^*(s)'C(s)\end{pmatrix} d\mathbf{N}(s) \\
&= \int_0^t J(s)\mathbf{c}(s)'\,d\mathbf{N}(s) + \int_0^t J(s)\mathbf{h}(s)'\,d\mathbf{N}(s) \\
&= \hat{A}\,(t,\mathbf{z}_0) + R^{(n)}(t,\mathbf{z}_0) + \int_0^t \big(J(s) - \tilde{J}(s)\big)\mathbf{c}(s)'\,d\mathbf{N}(s).
\end{aligned}
\tag{6.3}
$$

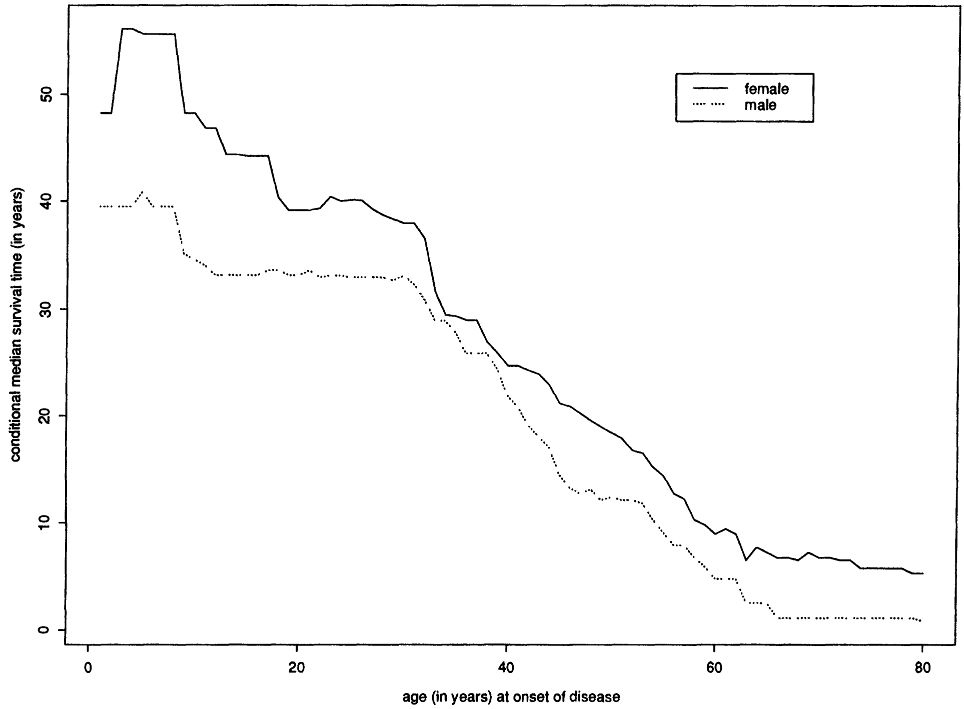

FIG. 4. *Conditional median survival time plot for female and male diabetics.*

In the fourth equality in (6.3), we use the formula [see Problem 2.7 of Rao (1973), page 33]

$$\begin{pmatrix} 1 & \mathbf{b}' \\ \mathbf{b} & D \end{pmatrix}^{-1} = \begin{pmatrix} 1 + \mathbf{b}'E^{-1}\mathbf{b} & -\mathbf{b}'E^{-1} \\ -E^{-1}\mathbf{b} & E^{-1} \end{pmatrix},$$

where $E = D - \mathbf{b}\mathbf{b}'$. Also, the fact that E^{-1} exists if and only if the inverse on the left-hand side of the above equation exists [see, e.g., Problem 2.4 of Rao (1973), page 32] implies that $J(s) = I(\mathbf{Z}^*(s)'P(s)\mathbf{Z}^*(s)$ is invertible). The term $R^{(n)}(t, \mathbf{z}_0)$ in (6.3) represents the effect of doing a "local linear fit" instead of doing a "local average."

Under Assumption (A1), the probability that the third term in (6.3) equals 0 for all $t \in [0, 1]$ tends to 1. Thus it suffices to show that

(a) $\sqrt{na_n}(\hat{A}(\cdot, \mathbf{z}_0) - A(\cdot, \mathbf{z}_0)) \to_d U$ under (A1)–(A5), where U is defined in Theorem 1.

(b) $\sqrt{na_n} \sup_{t \in [0,1]} |R^{(n)}(t, \mathbf{z}_0)| \to_P 0$ under (A1)–(A6).

We first prove (a). Define

$$X_1^{(n)}(t, \mathbf{z}_0) = \sqrt{na_n} \int_0^t J(s) \sum_{i=1}^n c_i(s, \mathbf{z}_0)\, dM_i(s),$$

$$X_2^{(n)}(t, \mathbf{z}_0) = \sqrt{na_n} \int_0^t J(s) \sum_{i=1}^n c_i(s, \mathbf{z}_0)\big(\alpha(s, \mathbf{Z}_i(s)) - \alpha(s, \mathbf{z}_0)\big)\, ds,$$

where $M_i(\cdot) = N_i(\cdot) - \int_0^{\cdot} Y_i(s)\alpha(s, \mathbf{Z}_i(s))\, ds$, $1 \le i \le n$ are orthogonal locally square integrable martingales. Then

$$
\begin{aligned}
&\sqrt{na_n}\left(\hat{A}\,(\cdot, \mathbf{z}_0) - A(t, \mathbf{z}_0)\right) \\
(6.4)\qquad &= X_1^{(n)}(t, \mathbf{z}_0) + X_2^{(n)}(t, \mathbf{z}_0) + \sqrt{na_n}\int_0^t (1 - J(s))\,\alpha(s, \mathbf{z}_0)\, ds.
\end{aligned}
$$

By the version of Rebolledo's martingale central limit theorem stated as Theorem I.2 in Andersen and Gill (1982), we have

$$
\begin{aligned}
(6.5)\qquad X_1^{(n)}(\cdot, \mathbf{z}_0) &= \int_0^{\cdot} J(s) \sum_{i=1}^n \sqrt{na_n}\, c_i(s, \mathbf{z}_0)\, dM_i(s) \\
&\to_d U(\cdot, \mathbf{z}_0) \quad \text{in } D[0,1]
\end{aligned}
$$

if the following two conditions hold:

(i) For each $t \in [0, 1]$,

$$
\langle X_1^{(n)}, X_1^{(n)} \rangle(t) \to_P \int_0^t g_0(s, \mathbf{z}_0)\,\alpha(s, \mathbf{z}_0)\, ds,
$$

where

$$
\langle X_1^{(n)}, X_1^{(n)} \rangle(t) = \int_0^t \sum_{i=1}^n \left(J(s)\sqrt{na_n}\, c_i(s, \mathbf{z}_0)\right)^2 Y_i(s)\,\alpha(s, \mathbf{Z}_i(s))\, ds.
$$

(ii) (Lindeberg condition.) For each $\varepsilon > 0$,

$$
\begin{aligned}
&\int_0^1 \sum_{i=1}^n \left(J(s)\sqrt{na_n}\, c_i(s, \mathbf{z}_0)\right)^2 Y_i(s)\,\alpha(s, \mathbf{Z}_i(s)) \\
&\qquad \times I\left(J(s)\sqrt{na_n}\, c_i(s, \mathbf{z}_0) > \varepsilon\right) ds \to_P 0.
\end{aligned}
$$

Before going further, we note that for any three sets of functions $d_i(s)$, $x_i(s)$, $y_i(s)$, $i = 1, \dots, n$, we have (the argument s is omitted)

$$
\begin{aligned}
(6.6)\qquad \int \sum_{i=1}^n d_i |x_i y_i| &\le \int \left(\sum_{i=1}^n d_i x_i^2\right)^{1/2} \left(\sum_{i=1}^n d_i y_i^2\right)^{1/2} \\
&\le \left(\int \sum_{i=1}^n d_i x_i^2\right)^{1/2} \left(\int \sum_{i=1}^n d_i y_i^2\right)^{1/2}.
\end{aligned}
$$

We now proceed to verify (i) and (ii). To check (i) we note that

$$
\begin{aligned}
&\left| \langle X_1^{(n)}, X_1^{(n)} \rangle(t) - \int_0^t J(s) g_0(s, \mathbf{z}_0)\,\alpha(s, \mathbf{z}_0)\, ds \right| \\
&\quad \le na_n \int_0^1 J(s) \sum_{i=1}^n c_i^2(s, \mathbf{z}_0) Y_i(s) |\alpha(s, \mathbf{Z}_i(s)) - \alpha(s, \mathbf{z}_0)|\, ds \\
&\qquad + \int_0^1 J(s) \left| \sum_{i=1}^n na_n c_i^2(s, \mathbf{z}_0) Y_i(s)\,\alpha(s, \mathbf{z}_0) - g_0(s, \mathbf{z}_0)\,\alpha(s, \mathbf{z}_0) \right| ds
\end{aligned}
$$

$$\le na_n\Big(\sup_s \max_i c_i(s,\mathbf{z}_0)\Big)\int_0^1 \sum_{i=1}^n c_i(s,\mathbf{z}_0)|\alpha(s,\mathbf{Z}_i(s)) - \alpha(s,\mathbf{z}_0)|\,ds$$

$$(6.7)\qquad + B\int_0^1 \left|\sum_{i=1}^n na_n c_i^2(s,\mathbf{z}_0) - g_0(s,\mathbf{z}_0)\right| ds$$

$$\le na_n\Big(\sup_s \max_i c_i(s,\mathbf{z}_0)\Big)$$

$$\times\left\{\int_0^1 \sum_{i=1}^n c_i(s,\mathbf{z}_0)\big(\alpha(s,\mathbf{Z}_i(s)) - \alpha(s,\mathbf{z}_0)\big)^2\,ds\right\}^{1/2}$$

$$+ B\int_0^1 \left|\sum_{i=1}^n na_n c_i^2(s,\mathbf{z}_0) - g_0(s,\mathbf{z}_0)\right| ds$$

$$\to_P 0$$

from (A5), (A4) and (A2), where the last inequality follows from (6.6). Hence (i) follows from (A1) and (6.7). Furthermore, for any $\varepsilon > 0$,

$$\int_0^1 \sum_{i=1}^n \left(J(s)\sqrt{na_n}\,c_i(s,\mathbf{z}_0)\right)^2 Y_i(s)\,\alpha(s,\mathbf{Z}_i(s)) I\left(J(s)\sqrt{na_n}\,c_i(s,\mathbf{z}_0) > \varepsilon\right) ds$$

$$\le \int_0^1 na_n \sum_{i=1}^n c_i^2(s,\mathbf{z}_0)\,\alpha(s,\mathbf{Z}_i(s))\left(\frac{J(s)\sqrt{na_n}\,c_i(s,\mathbf{z}_0)}{\varepsilon}\right)^{\delta} ds$$

$$\le \frac{B}{\varepsilon^{\delta}}(na_n)^{-\delta/2}\int_0^1 (na_n)^{1+\delta}\sum_{i=1}^n c_i^{2+\delta}(s,\mathbf{z}_0)\,ds$$

$$\to_P 0$$

from (A2) and the assumption that $na_n \to \infty$. Hence (ii) holds. This proves (6.5).

It is easy to see that (A3) implies

$$(6.8)\qquad \sup_{t\in[0,1]} |X_2^{(n)}(t,\mathbf{z}_0)| \to_P 0.$$

Moreover, the probability that the third term in (6.4) equals 0 for all $t \in [0,1]$ tends to 1 under Assumption (A1). This, together with (6.4), (6.5) and (6.8), implies that (a) holds.

We now prove (b). Let $\boldsymbol{\alpha}_n(s) = (\alpha(s,\mathbf{Z}_1(s)),\dots,\alpha(s,\mathbf{Z}_n(s)))'$. Using the fact that $h_i(s,\mathbf{z}_0)Y_i(s) = h_i(s,\mathbf{z}_0)$ for each i, we see that

$$\sqrt{na_n}\,R^{(n)}(t,\mathbf{z}_0) = \sqrt{na_n}\int_0^t J(s)\mathbf{h}(s)'\,d\mathbf{N}(s)$$

$$= R_1^{(n)}(t,\mathbf{z}_0) + R_2^{(n)}(t,\mathbf{z}_0) + R_3^{(n)}(t,\mathbf{z}_0),$$

where

$$R_1^{(n)}(t,\mathbf{z}_0) = \sqrt{na_n}\int_0^t J_1(s)\mathbf{h}(s)'\boldsymbol{\alpha}_n(s)\,ds,$$

$$R_2^{(n)}(t,\mathbf{z}_0) = \sqrt{na_n}\int_0^t J_1(s)\mathbf{h}(s)'\,d\mathbf{M}(s),$$

$$R_3^{(n)}(t,\mathbf{z}_0) = \sqrt{na_n}\int_0^t (J(s) - J_1(s))\mathbf{h}(s)'\,d\mathbf{N}(s).$$

We shall show that $\sup_{t\in[0,1]}|R_j^{(n)}(t,\mathbf{z}_0)| \to_P 0$ for $j = 1, 2$ and 3. Consider $R_1^{(n)}(t,\mathbf{z}_0)$. For every s, we have $P(s)\mathbf{1} = (C(s) - \mathbf{c}(s)\mathbf{c}(s)')\mathbf{1} = \mathbf{0}$, which implies that $\mathbf{h}(s)'\mathbf{1} = \mathbf{0}$. Thus,

$$\begin{aligned}
|R_1^{(n)}(t,\mathbf{z}_0)| &= \left|\sqrt{na_n}\int_0^t J_1(s)\mathbf{h}(s)'(\boldsymbol{\alpha}_n(s) - \mathbf{1}\alpha(s,\mathbf{z}_0))\,ds\right| \\
&\le \sqrt{na_n}\int_0^1 \Big|\big\{J_1(s)(\mathbf{z}_0 - \mathbf{Z}^*(s)'\mathbf{c}(s))' \\
&\qquad \times(\mathbf{Z}^*(s)'P(s)\mathbf{Z}^*(s))^{-1}\mathbf{Z}^*(s)'P(s)^{1/2}\big\} \\
&\qquad \times\big\{P(s)^{1/2}(\boldsymbol{\alpha}_n(s) - \mathbf{1}\alpha(s,\mathbf{z}_0))\big\}\Big|\,ds \\
&\le \sqrt{na_n}\bigg(\int_0^1 \|J_1(s)P(s)^{1/2}\mathbf{Z}^*(s)(\mathbf{Z}^*(s)'P(s)\mathbf{Z}^*(s))^{-1} \\
&\qquad (\mathbf{z}_0 - \mathbf{Z}^*(s)'\mathbf{c}(s))\|^2\,ds\bigg)^{1/2} \qquad (6.9)\\
&\qquad \times\left(\int_0^1 \|P^{1/2}(s)(\boldsymbol{\alpha}_n(s) - \mathbf{1}\alpha(s,\mathbf{z}_0))\|^2\,ds\right)^{1/2} \\
&\le \bigg(\sqrt{na_n}\int_0^1 J_1(s)(\mathbf{z}_0 - \mathbf{Z}^*(s)'\mathbf{c}(s))'(\mathbf{Z}^*(s)'P(s)\mathbf{Z}^*(s))^{-1} \\
&\qquad \times(\mathbf{z}_0 - \mathbf{Z}^*(s)'\mathbf{c}(s))\,ds\bigg)^{1/2} \\
&\qquad \times\bigg(\sqrt{na_n}\int_0^1 (\boldsymbol{\alpha}_n(s) - \mathbf{1}\alpha(s,\mathbf{z}_0))'C(s) \\
&\qquad \times(\boldsymbol{\alpha}_n(s) - \mathbf{1}\alpha(s,\mathbf{z}_0))\,ds\bigg)^{1/2},
\end{aligned}$$

where the second inequality follows from (6.6), with $d_i \equiv 1$. Together with (A6) and (A4), this implies that $\sup_{t\in[0,1]}|R_1^{(n)}(t,\mathbf{z}_0)| \to_P 0$.

Now consider $R_2^{(n)}(t,\mathbf{z}_0)$. Because each $\sqrt{na_n}J_1(t)h_i(t)$ is a bounded predictable process, $R_2^{(n)}(t,\mathbf{z}_0)$ is a locally square integrable martingale. By the version of Lenglart's inequality stated as Theorem I.1 in Andersen and Gill (1982), for each $\eta > 0$ and $\varepsilon > 0$, we have

$$P\Big(\sup_{t\in[0,1]}|R_2^{(n)}(t,\mathbf{z}_0)| > \eta\Big) \le \frac{\varepsilon}{\eta^2} + P\big(\langle R_2^{(n)}, R_2^{(n)}\rangle(1) > \varepsilon\big).$$

Now

$$
\begin{aligned}
&\langle R_2^{(n)}, R_2^{(n)}\rangle(1)\\
&\quad = na_n \int_0^1 J_1(s) \sum_{i=1}^n h_i^2(s,\mathbf{z}_0) Y_i(s)\,\alpha(s, Z_i(s))\, ds\\
&\quad \le Bna_n \int_0^1 J_1(s)\mathbf{h}(s)'\mathbf{h}(s)\, ds\\
&\quad = Bna_n \int_0^1 J_1(s)(\mathbf{z}_0 - \mathbf{Z}^*(s)'\mathbf{c}(s))'(\mathbf{Z}^*(s)'P(s)\mathbf{Z}^*(s))^{-1}\mathbf{Z}^*(s)'\\
&\qquad\quad \times P(s)^2\mathbf{Z}^*(s)(\mathbf{Z}^*(s)'P(s)\mathbf{Z}^*(s))^{-1}(\mathbf{z}_0 - \mathbf{Z}^*(s)'\mathbf{c}(s))\, ds \qquad (6.10)\\
&\quad \le B\Big(na_n \max_i \sup_{s\in[0,1]} c_i(s,\mathbf{z}_0)\Big)\int_0^1 J_1(s)(\mathbf{z}_0 - \mathbf{Z}^*(s)'\mathbf{c}(s))'\\
&\qquad\quad \times (\mathbf{Z}^*(s)'P(s)\mathbf{Z}^*(s))^{-1}(\mathbf{z}_0 - \mathbf{Z}^*(s)'\mathbf{c}(s))\, ds,
\end{aligned}
$$

where in the last step we use the fact that for any $\mathbf{x}$ (the arguments s and $\mathbf{z}_0$ are omitted)

$$
\mathbf{x}'P^2\mathbf{x} = (\mathbf{x}'P^{1/2})P(P^{1/2}\mathbf{x}) \le (\mathbf{x}'P^{1/2})C(P^{1/2}\mathbf{x}) \le \Big(\max_i \sup_s c_i\Big)\mathbf{x}'P\mathbf{x}.
$$

Hence (A5) and (A6) imply that $\sup_{t\in[0,1]}|R_2^{(n)}(t,\mathbf{z}_0)| \to_P 0$.

Finally, we see that $\sup_{t\in[0,1]}|R_3^{(n)}(t,\mathbf{z}_0)| \to_P 0$ from (A1) and the fact that $J_1(s) \le J(s)$. This proves (b) and completes the proof of the first part of Theorem 1.

The second part of Theorem 1 is a direct consequence of part (1) together with the compact differentiability of the product integral [see Theorem 8 of Gill and Johansen (1990)] and a functional version of the δ-method [see Theorem 3 of Gill (1989)]. □

REMARK. Part (a) of the proof gives weak convergence of the Beran-type estimators under (A2)–(A5) and the assumptions that $na_n \to \infty$ and that $P(\Sigma_i W_i(\mathbf{z}_0)Y_i(s) > 0$ for all $s \in [0,1]) \to 1$. An interesting observation is that this also yields weak convergence of the unconditional Nelson–Aalen estimator: Assume $\alpha(t,\mathbf{z}) \equiv \alpha(t)$ and let $a_n \equiv 1$ and $W_i(\mathbf{z}) \equiv 1/n$. Then these sufficient conditions are satisfied if $(1/n)\sum Y_i(s)$ converges in probability uniformly to a function that is bounded away from 0 on the interval $[0,1]$.

PROOF OF THEOREM 2. Consider the identity (6.3) and recall that under (B1), the probability that the third term on the right side of (6.3) is identically 0 over $[0,1]$ tends to 1. Fix t in $(0,1)$, define

$$
\begin{aligned}
\hat{\alpha}(t,\mathbf{z}_0) &= \frac{1}{b_n}\int_0^1 K\left(\frac{t-s}{b_n}\right)\hat{A}\,(ds,\mathbf{z}_0),\\
r^{(n)}(t,\mathbf{z}_0) &= \frac{1}{b_n}\int_0^1 K\left(\frac{t-s}{b_n}\right)R^{(n)}(ds,\mathbf{z}_0)
\end{aligned}
\qquad (6.11)
$$

and write

$$\alpha_n(t,\mathbf{z}_0) = \frac{1}{b_n}\int_0^1 K\left(\frac{t-s}{b_n}\right) A_n(ds,\mathbf{z}_0) = \hat{\alpha}(t,\mathbf{z}_0) + r^{(n)}(t,\mathbf{z}_0) + \zeta_n(t,\mathbf{z}_0),$$

where the probability that $\sup_{u\in[0,1]}|\zeta_n(u,\mathbf{z}_0)|$ equals 0 tends to 1. We will show that:

(i) $\sqrt{na_nb_n}\,(\hat{\alpha}(t,\mathbf{z}_0) - \alpha(t,\mathbf{z}_0)) \to_d N(0,\sigma_t^2)$, where σ_t^2 is defined by (3.5).
(ii) $\sqrt{na_nb_n}\,r^{(n)}(t,\mathbf{z}_0) \to_P 0$.

Theorem 2 then follows immediately.

We first prove (i). We have

$$\begin{aligned}\hat{\alpha}(t,\mathbf{z}_0) = {} & \frac{1}{b_n}\int_0^1 J(s)K\left(\frac{t-s}{b_n}\right)\sum_{i=1}^n c_i(s,\mathbf{z}_0)\,dM_i(s) - \zeta_n(t,\mathbf{z}_0)\\ & + \frac{1}{b_n}\int_0^1 J(s)K\left(\frac{t-s}{b_n}\right)\sum_{i=1}^n c_i(s,\mathbf{z}_0)\alpha(s,\mathbf{Z}_i(s))\,ds.\end{aligned}$$

For large n we have

$$\begin{aligned}\alpha(t,\mathbf{z}_0) &= \left\{\frac{1}{b_n}\int_0^1 K\left(\frac{t-s}{b_n}\right)ds\right\}\alpha(t,\mathbf{z}_0)\\ &= \frac{1}{b_n}\int_0^1 J(s)K\left(\frac{t-s}{b_n}\right)\sum_{i=1}^n c_i(s,\mathbf{z}_0)\alpha(t,\mathbf{z}_0)\,ds\\ &\quad + \frac{1}{b_n}\int_0^1 (1-J(s))K\left(\frac{t-s}{b_n}\right)\alpha(t,\mathbf{z}_0)\,ds,\end{aligned}$$

so that

$$\sqrt{na_nb_n}\,(\hat{\alpha}(t,\mathbf{z}_0) - \alpha(t,\mathbf{z}_0)) = X^{(n)}(1) + I_{n1} - I_{n2} - \sqrt{na_nb_n}\,\zeta_n(t,\mathbf{z}_0),$$

where

$$\begin{aligned}X^{(n)}(\tau) &= \left(\frac{na_n}{b_n}\right)^{1/2}\int_0^\tau J(s)K\left(\frac{t-s}{b_n}\right)\sum_{i=1}^n c_i(s,\mathbf{z}_0)\,dM_i(s), \qquad \tau\in[0,1],\\ I_{n1} &= \left(\frac{na_n}{b_n}\right)^{1/2}\int_0^1 J(s)K\left(\frac{t-s}{b_n}\right)\\ &\quad\times\sum_{i=1}^n c_i(s,\mathbf{z}_0)\big(\alpha(s,\mathbf{Z}_i(s)) - \alpha(t,\mathbf{z}_0)\big)\,ds,\\ I_{n2} &= \left(\frac{na_n}{b_n}\right)^{1/2}\left\{\int_0^1(1-J(s))K\left(\frac{t-s}{b_n}\right)ds\right\}\alpha(t,\mathbf{z}_0).\end{aligned}$$

Now it suffices to show that $I_{nj} \to_P 0$ for $j=1,2$ and that $X^{(n)}(1) \to_d$

$N(0, \sigma_t^2)$. Consider I_{n1}. Let

$$V_{n1} = \left(\frac{na_n}{b_n}\right)^{1/2} \int_0^1 J(s) K\left(\frac{t-s}{b_n}\right) \left| \sum_{i=1}^n c_i(s, \mathbf{z}_0)(\alpha(s, \mathbf{Z}_i(s)) - \alpha(s, \mathbf{z}_0)) \right| ds,$$

$$V_{n2} = \left(\frac{na_n}{b_n}\right)^{1/2} \left| \int_0^1 J(s) K\left(\frac{t-s}{b_n}\right)(\alpha(s, \mathbf{z}_0) - \alpha(t, \mathbf{z}_0))\, ds \right|$$

and note that

$$|I_{n1}| \le V_{n1} + V_{n2}. \tag{6.12}$$

Since $K(u)$ vanishes outside the interval $[-1, 1]$ and $M \equiv \sup_u K(u) < \infty$, we have

$$\int_0^1 K\left(\frac{t-s}{b_n}\right) f(s)\, ds \le \int_{t-b_n}^{t+b_n} K\left(\frac{t-s}{b_n}\right) f(s)\, ds \le M \int_{t-b_n}^{t+b_n} f(s)\, ds$$

for every nonnegative function f. Hence, in (6.12), $V_{n1} \to_P 0$ by (B3). Recalling that $D = \sup_{s \in [0,1]} |\alpha_s''(s, \mathbf{z}_0)|$, we have for large n

$$\begin{aligned} V_{n2} &\le \left(\frac{na_n}{b_n}\right)^{1/2} \left| \alpha_t'(t, \mathbf{z}_0) \int_0^1 K\left(\frac{t-s}{b_n}\right)(s-t)\, ds \right| \\ &\quad + D\left(\frac{na_n}{b_n}\right)^{1/2} \int_0^1 K\left(\frac{t-s}{b_n}\right)(s-t)^2\, ds \\ &\le \left(\frac{na_n}{b_n}\right)^{1/2} \left| \alpha_t'(t, \mathbf{z}_0) b_n^2 \int_{-1}^1 uK(u)\, du \right| \\ &\quad + MD\left(\frac{na_n}{b_n}\right)^{1/2} \int_{t-b_n}^{t+b_n} (s-t)^2\, ds \\ &= 0 + O\left(\sqrt{nb_n^{p+5}}\right) \to 0. \end{aligned}$$

Therefore, $I_{n1} \to_P 0$. We also have $I_{n2} \to_P 0$ from (B1).

To prove $X^{(n)}(1) \to_d N(0, \sigma_t^2)$, we apply Theorem II.5.1 of Anderson, Borgan, Gill and Keiding (1993) with $\mathscr{T}_0 = \{1\}$ to $X^{(n)}(\tau) = \int_0^\tau \sum_{i=1}^n H_i(s, \mathbf{z}_0)\, dM_i(s)$, $\tau \in [0, 1]$, where

$$H_i(s, \mathbf{z}_0) = J(s)\left(\frac{na_n}{b_n}\right)^{1/2} K\left(\frac{t-s}{b_n}\right) c_i(s, \mathbf{z}_0), \qquad i = 1, \dots, n.$$

Thus we need to check the following conditions.

(a) $\langle X^{(n)}, X^{(n)} \rangle(1) \to_P \sigma_t^2$.

(b) For each $\varepsilon > 0$,

$$\int_0^1 \sum_{i=1}^n H_i^2(s, \mathbf{z}_0) I(H_i(s, \mathbf{z}_0) > \varepsilon) Y_i(s) \alpha(s, \mathbf{Z}_i(s))\, ds \to_P 0.$$

To verify (a), we write

$$
\begin{aligned}
\langle X^{(n)}, X^{(n)}\rangle(1) &= \int_0^1 J(s)\frac{na_n}{b_n}K^2\left(\frac{t-s}{b_n}\right)\sum_{i=1}^n c_i^2(s,\mathbf{z}_0)\alpha(s,\mathbf{Z}_i(s))\,ds \\
&= \int_0^1 J(s)\frac{na_n}{b_n}K^2\left(\frac{t-s}{b_n}\right) \\
&\qquad\times \sum_{i=1}^n c_i^2(s,\mathbf{z}_0)\big(\alpha(s,\mathbf{Z}_i(s)) - \alpha(s,\mathbf{z}_0)\big)\,ds \\
&\qquad + \int_0^1 J(s)\frac{1}{b_n}K^2\left(\frac{t-s}{b_n}\right) \\
&\qquad\quad\times\left(na_n\sum_{i=1}^n c_i^2(s,\mathbf{z}_0) - g_0(s,\mathbf{z}_0)\right)\alpha(s,\mathbf{z}_0)\,ds \\
&\qquad\quad + \int_0^1 J(s)\frac{1}{b_n}K^2\left(\frac{t-s}{b_n}\right)g_0(s,\mathbf{z}_0)\alpha(s,\mathbf{z}_0)\,ds \\
&\equiv I_1 + I_2 + I_3.
\end{aligned}
$$

In the above expression,

$$
\begin{aligned}
|I_1| &\le \int_0^1 \frac{na_n}{b_n}K^2\left(\frac{t-s}{b_n}\right)\sum_{i=1}^n c_i^2(s,\mathbf{z}_0)|\alpha(s,\mathbf{Z}_i(s)) - \alpha(s,\mathbf{z}_0)|\,ds \\
&\le M^2\frac{na_n}{b_n}\int_{t-b_n}^{t+b_n}\sum_{i=1}^n c_i^2(s,\mathbf{z}_0)|\alpha(s,\mathbf{Z}_i(s)) - \alpha(s,\mathbf{z}_0)|\,ds \\
&\le \frac{M^2}{b_n}\left(na_n\max_i\sup_s c_i(s,\mathbf{z}_0)\right) \\
&\qquad\times\int_{t-b_n}^{t+b_n}\sum_{i=1}^n c_i(s,\mathbf{z}_0)|\alpha(s,\mathbf{Z}_i(s)) - \alpha(s,\mathbf{z}_0)|\,ds \\
&= O_p\left(\frac{1}{b_n}\right)\left\{\int_{t-b_n}^{t+b_n}\sum_{i=1}^n c_i(s,\mathbf{z}_0)\big(\alpha(s,\mathbf{Z}_i(s)) - \alpha(s,\mathbf{z}_0)\big)^2\,ds\right\}^{1/2} \\
&\qquad\times\left\{\int_{t-b_n}^{t+b_n}ds\right\}^{1/2} \\
&= O_p\left(\frac{1}{b_n}\right)\big(O_p(b_n^3)\big)^{1/2}(2b_n)^{1/2} = O_p(b_n),
\end{aligned}
$$

where the first equality follows from (B5) and (6.6) and the second equality from (B4). Also,

$$
|I_2| \le \frac{BM^2}{b_n}\int_{t-b_n}^{t+b_n}\left|na_n\sum_{i=1}^n c_i^2(s,\mathbf{z}_0) - g(s,\mathbf{z}_0)\right|ds = O_p\left(\frac{1}{b_n}\right)o_p(b_n) = o_p(1)
$$

[recall that $B = \sup|\alpha(s, \mathbf{z})| < \infty$] and $I_3 \to_P \sigma_t^2$ by (B1) and the assumption that $g_0(s, \mathbf{z}_0)$ and $\alpha(s, \mathbf{z}_0)$ are continuous in s. Therefore,

$$\langle X^{(n)}, X^{(n)} \rangle(1) = I_1 + I_2 + I_3 \to_P g_0(t, \mathbf{z}_0)\alpha(t, \mathbf{z}_0) \int K^2(u)\, du = \sigma_t^2.$$

We now prove (b). For the $\delta > 0$ in (B2), we have

$$\begin{aligned}
&\int_0^1 \sum_{i=1}^n H_i^2(s, \mathbf{z}_0) I(H_i(s, \mathbf{z}_0) > \varepsilon) Y_i(s)\alpha(s, \mathbf{Z}_i(s))\, ds \\
&\quad \le \frac{1}{\varepsilon^\delta} \int_0^1 \sum_{i=1}^n H_i^{2+\delta}(s, \mathbf{z}_0) Y_i(s)\alpha(s, \mathbf{Z}_i(s))\, ds \\
&\quad = \frac{1}{\varepsilon^\delta} \int_0^1 J(s) \left(\left(\frac{n a_n}{b_n} \right)^{1/2} \right)^{2+\delta} K^{2+\delta}\left(\frac{t-s}{b_n} \right) \\
&\qquad \times \sum_{i=1}^n c_i^{2+\delta}(s, \mathbf{z}_0) Y_i(s)\alpha(s, \mathbf{Z}_i(s))\, ds \\
&\quad \le \frac{BM^{2+\delta}}{\varepsilon^\delta} \left((n a_n)^{-\delta/2} b_n^{-1-\delta/2} \right) (n a_n)^{1+\delta} \int_{t-b_n}^{t+b_n} \sum_{i=1}^n c_i^{2+\delta}(s, \mathbf{z}_0)\, ds \\
&\quad = O_p\left((n a_n)^{-\delta/2} b_n^{-1-\delta/2} \right) O_p(b_n) = O_p\left(\left(n b_n^{p+1} \right)^{-\delta/2} \right),
\end{aligned}$$

where the second equality follows from (B2) and the assumption that $g(s, \mathbf{z}_0)$ is continuous in s. The last term converges to 0 in probability since $n b_n^{p+1} \to \infty$.

Therefore we have proved part (i).

We now prove (ii). For $r^{(n)}(t, \mathbf{z}_0)$ defined by (6.11), we have

$$\begin{aligned}
\sqrt{n a_n b_n}\, r^{(n)}(t, \mathbf{z}_0) &= \left(\frac{n a_n}{b_n} \right)^{1/2} \int_0^1 J(s) K\left(\frac{t-s}{b_n} \right) \mathbf{h}(s)'\, d\mathbf{N}(s) \\
&= r_1^{(n)} + r_2^{(n)}(1) + r_3^{(n)},
\end{aligned}$$

where

$$r_1^{(n)} = \left(\frac{n a_n}{b_n} \right)^{1/2} \int_0^1 J_1(s) K\left(\frac{t-s}{b_n} \right) \mathbf{h}(s)' \boldsymbol{\alpha}_n(s)\, ds,$$

$$r_2^{(n)}(\tau) = \left(\frac{n a_n}{b_n} \right)^{1/2} \int_0^\tau J_1(s) K\left(\frac{t-s}{b_n} \right) \mathbf{h}(s)'\, d\mathbf{M}(s), \qquad \tau \in [0, 1],$$

$$r_3^{(n)} = \left(\frac{n a_n}{b_n} \right)^{1/2} \int_0^1 (J(s) - J_1(s)) K\left(\frac{t-s}{b_n} \right) \mathbf{h}(s)'\, d\mathbf{N}(s),$$

and we recall that $\boldsymbol{\alpha}_n(s) = (\alpha(s, \mathbf{Z}_1(s)), \dots, \alpha(s, \mathbf{Z}_n(s)))'$ and $\mathbf{h}(s)$ is defined by (6.2). We shall show that $r_1^{(n)}$, $r_2^{(n)}(1)$ and $r_3^{(n)}$ all converge to zero in probability. Using the fact that for $s \in \mathscr{T}_{\mathbf{z}_0}$ we have $\mathbf{h}(s)'\mathbf{1} = 0$, we see that

$$
\begin{aligned}
|r_1^{(n)}| &= \left| \sqrt{\frac{na_n}{b_n}} \int_0^1 J_1(s) K\left(\frac{t-s}{b_n}\right) \mathbf{h}(s)'(\boldsymbol{\alpha}_n(s) - \mathbf{1}\alpha(s,\mathbf{z}_0))\, ds \right| \\
&\le M \sqrt{\frac{na_n}{b_n}} \int_{t-b_n}^{t+b_n} J_1(s) |\mathbf{h}(s)'(\boldsymbol{\alpha}_n(s) - \mathbf{1}\alpha(s,\mathbf{z}_0))|\, ds \\
&\le M \sqrt{\frac{na_n}{b_n}} \left(\int_{t-b_n}^{t+b_n} J_1(s)(\mathbf{z}_0 - \mathbf{Z}^*(s)'\mathbf{c}(s))'(\mathbf{Z}^*(s)'P(s)\mathbf{Z}^*(s))^{-1} \right. \\
&\qquad \left. \times (\mathbf{z}_0 - \mathbf{Z}^*(s)'\mathbf{c}(s))\, ds \right)^{1/2} \\
&\qquad \times \left(\int_{t-b_n}^{t+b_n} (\boldsymbol{\alpha}_n(s) - \mathbf{1}\alpha(s,\mathbf{z}_0))'C(s)(\boldsymbol{\alpha}_n(s) - \mathbf{1}\alpha(s,\mathbf{z}_0))\, ds \right)^{1/2} \\
&= O_p\left(\sqrt{\frac{na_n}{b_n}}\right) \left(o_p\left(\frac{1}{na_n b_n^2}\right)\right)^{1/2} \left(O_p(b_n^3)\right)^{1/2} = o_p(1),
\end{aligned}
$$

where to obtain the second inequality we reason as we did to obtain (6.9), and to obtain the second equality we use (B6) and (B4).

To prove that $r_2^{(n)}(1) \to_P 0$, we apply Lenglart's inequality to $r_2^{(n)}$: For any η, $\varepsilon > 0$

$$
P\left(\sup_\tau |r_2^{(n)}(\tau)| > \eta \right) \le \frac{\varepsilon}{\eta^2} + P\left(\langle r_2^{(n)}, r_2^{(n)} \rangle (1) > \varepsilon \right). \tag{6.13}
$$

We have

$$
\begin{aligned}
\langle r_2^{(n)}, r_2^{(n)} \rangle (1) &= \frac{na_n}{b_n} \int_0^1 J_1(s) K^2\left(\frac{t-s}{b_n}\right) \sum_{i=1}^n h_i^2(s,\mathbf{z}_0) Y_i(s) \alpha(s, Z_i(s))\, ds \\
&\le \frac{CM^2}{b_n} na_n \int_{t-b_n}^{t+b_n} J_1(s) \mathbf{h}(s)'\mathbf{h}(s)\, ds \\
&\le \frac{CM^2}{b_n} \left(na_n \max_i \sup_{s\in[0,1]} c_i(s,\mathbf{z}_0) \right) \\
&\qquad \times \int_{t-b_n}^{t+b_n} J_1(s)(\mathbf{z}_0 - \mathbf{Z}^*(s)'\mathbf{c}(s))' \\
&\qquad \times (\mathbf{Z}^*(s)'P(s)\mathbf{Z}^*(s))^{-1}(\mathbf{z}_0 - \mathbf{Z}^*(s)'\mathbf{c}(s))\, ds \\
&= O_p\left(\frac{1}{b_n}\right) O_p(1) o_p\left(\frac{1}{na_n b_n^2}\right) \\
&= o_p\left(\frac{1}{nb_n^{p+3}}\right) \to_P 0.
\end{aligned} \tag{6.14}
$$

To obtain the second inequality in (6.14) we reason as we did to obtain (6.10). The second equality in (6.14) follows from (B5) and (B6), and the last assertion follows since $nb_n^{p+3} \to \infty$ by assumption. Therefore, $r_2^{(n)}(1) \to_P 0$ by (6.13) and (6.14). Finally, $r_3^{(n)} \to_P 0$ from (B1). Therefore,

$$\sqrt{na_n b_n}\, r^{(n)} = r_1^{(n)} + r_2^{(n)}(1) + r_3^{(n)} \to_P 0. \qquad \square$$

To prove Theorem 3, we shall need to use some known results concerning k-NN estimators in nonparametric regression and density estimation. These are stated as the next two propositions.

PROPOSITION 6.1. *Assume that* $(Y, \mathbf{X}), (Y, \mathbf{X}_1), \dots, (Y, \mathbf{X}_n)$ *are i.i.d. random vectors taking values in* $\mathbb{R} \times \mathbb{R}^p$ *and that* $\mathbf{X}$ *has a continuous density function* $g(\mathbf{x})$. *For each* $\mathbf{x} \in \mathbb{R}^p$, *define* $m(\mathbf{x}) = E(Y|\mathbf{X} = \mathbf{x})$ *and* $m_n(\mathbf{x}) = \sum_{i=1}^n W_i(\mathbf{x}) Y_i$, *where*

$$W_i(\mathbf{x}) = w\left(\frac{\mathbf{x} - \mathbf{X}_i}{R_n}\right) \Big/ \sum_{j=1}^n w\left(\frac{\mathbf{x} - \mathbf{X}_j}{R_n}\right),$$

R_n *is the Euclidean distance from* $\mathbf{x}$ *to the kth closest of* $\mathbf{X}_1, \dots, \mathbf{X}_n$ *and* $w(\cdot)$ *is a bounded density function on* $\mathbb{R}^p$ *satisfying* $\|\mathbf{u}\|^p w(\mathbf{u}) \to 0$ *as* $\|\mathbf{u}\| \to \infty$ *and* (4.4). *Let* $\mathbf{x}_0 \in \mathbb{R}^p$ *such that* $g(\mathbf{x}_0) > 0$. *Assume that* $m(\mathbf{x})$ *and* $\operatorname{Var}(Y|\mathbf{X} = \mathbf{x})$ *exist in a neighborhood of* $\mathbf{x}_0$. *Assume further that* $m(\mathbf{x})$ *is continuous at* $\mathbf{x}_0$ *and* $\operatorname{Var}(Y|\mathbf{X} = \mathbf{x})$ *is bounded in a neighborhood of* $\mathbf{x}_0$. *Then, if* $k_n \to \infty$ *and* $k_n/n \to 0$,

$$m_n(\mathbf{x}_0) \to_P m(\mathbf{x}_0).$$

PROOF. This is Proposition 1 of Collomb (1980). $\square$

PROPOSITION 6.2. *Let* $\mathbf{X}_1, \dots, \mathbf{X}_n$ *be i.i.d.* $\mathbb{R}^p$*-valued random vectors with bounded density function* $g(\mathbf{x})$. *For each* $\mathbf{x} \in \mathbb{R}^p$, *define*

$$\hat{g}_n(\mathbf{x}) = \frac{1}{nR_n^p} \sum_{i=1}^n w\left(\frac{\mathbf{x} - \mathbf{X}_i}{R_n}\right),$$

where R_n *is defined as in Proposition* 6.1 *and* $w(\cdot)$ *is a bounded density function on* $\mathbb{R}^p$ *satisfying* (4.3), (4.4) *and* $w(-\mathbf{u}) = w(\mathbf{u})$ *for all* $\mathbf{u} \in \mathbb{R}^p$. *Then, if* $k_n \to \infty$ *and* $k_n/n \to 0$, *we have*

$$\hat{g}_n(\mathbf{x}) \to_P g(\mathbf{x}) \quad \text{at every continuity point } \mathbf{x} \text{ of } g. \tag{6.15}$$

When $w(\mathbf{u}) = (1/\gamma(p)) I(\|\mathbf{u}\| \le 1)$ *with* $\gamma(p) = \int_{\|\mathbf{u}\| \le 1} d\mathbf{u} = 2\pi^{p/2}/(p\Gamma(p/2))$, (6.15) *is the statement*

$$\frac{k_n}{\gamma(p) n R_n^p} \to_P g(\mathbf{x}) \quad \text{at every continuity point } \mathbf{x} \text{ of } g. \tag{6.16}$$

PROOF. Let $\mathbf{x} \in \mathbb{R}^p$ and $\varepsilon > 0$. By Theorem 1.1 of Moore and Yackel (1977), there exist $\eta > 0$ and a finite set of positive numbers $\alpha_1, \dots, \alpha_M$

such that $|\hat{g}_n(\mathbf{x}) - g(\mathbf{x})| > \varepsilon$ implies that either $|f_n(\mathbf{x}, \alpha_j) - g(\mathbf{x})| > \eta$ or $|g_n(\mathbf{x}, \alpha_j) - g(\mathbf{x})| > \eta$ for at least one j in $\{1, \dots, M\}$, where

$$f_n(\mathbf{x}, \alpha) = \frac{1}{nh_n(\alpha)^p} \sum_{i=1}^{n} w\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n(\alpha)}\right),$$

$$g_n(\mathbf{x}, \alpha) = \frac{1}{nh_n(\alpha)^p} \sum_{i=1}^{n} \frac{1}{2} I\left(\left|\frac{\mathbf{x} - \mathbf{X}_i}{h_n(\alpha)}\right| \le 1\right),$$

$h_n(\alpha)$ is determined by $k_n = \alpha n h_n(\alpha)^p$ and the choice of η and $\alpha_1, \dots, \alpha_M$ is uniform in $n, \mathbf{x}$ and the sample point ω. This implies

$$\begin{aligned} &P\big(|\hat{g}_n(\mathbf{x}) - g(\mathbf{x})| > \varepsilon\big) \\ &\quad \le \sum_{j=1}^{M} P\big(|f_n(\mathbf{x}, \alpha_j) - g(\mathbf{x})| > \eta\big) + \sum_{j=1}^{M} P\big(|g_n(\mathbf{x}, \alpha_j) - g(\mathbf{x})| > \eta\big) \to 0, \end{aligned}$$

where the convergence statement is a consequence of Theorem 3.1.2 of Prakasa Rao (1983). Thus, $\hat{g}_n(\mathbf{x}) \to_P g(\mathbf{x})$. □

The following result is needed to prove Theorem 3.

LEMMA 6.1. *Under the conditions of Theorem* 3, *we have*:

(i) $(n(k_n/n)^{(p-4)/p})^{1/4} \sum_{i=1}^n c_i(s, \mathbf{z}_0)(\mathbf{Z}_i - \mathbf{z}_0) = o_p(1)$ *uniformly in* $s \in [0, 1]$, *if* $k_n \to \infty$ *and* $k_n^{p+4}/n^4 \to 0$.

(ii) $(n(k_n/n)^{(p+1)/p}) \sum_{i=1}^n c_i(s, \mathbf{z}_0)(\mathbf{Z}_i - \mathbf{z}_0) = o_p(1)$ *uniformly in* $s \in [0, 1]$, *if* $k_n \to \infty$ *and* $k_n^{p+5}/n^5 \to 0$.

PROOF. We only prove Part (i) of the lemma since the second part is proved in an identical way. Denote $\alpha_n = (n(k_n/n)^{(p-4)/p})^{1/4}$. Then

$$\begin{aligned} &\alpha_n \sum_{i=1}^{n} c_i(s, \mathbf{z}_0)(\mathbf{Z}_i - \mathbf{z}_0) \\ &\quad = \alpha_n \frac{\sum_{i=1}^n W_i(\mathbf{z}_0) Y_i(s)(\mathbf{Z}_i - \mathbf{z}_0)}{\sum_{j=1}^n W_j(\mathbf{z}_0) Y_j(s)} \qquad (6.17) \\ &\quad = \frac{(\alpha_n/nR_n^{p-1}) \sum_{i=1}^n w((\mathbf{Z}_i - \mathbf{z}_0)/R_n) Y_i(s)((\mathbf{Z}_i - \mathbf{z}_0)/R_n)}{(1/nR_n^p) \sum_{j=1}^n w((\mathbf{Z}_j - \mathbf{z}_0)/R_n) \sum_{j=1}^n W_j(\mathbf{z}_0) Y_j(s)}. \end{aligned}$$

We first show that if $k_n \to \infty$ and $k_n/n \to 0$, then

$$\sup_{s \in [0,1]} \left| \sum_{i=1}^{n} W_i(\mathbf{z}_0) Y_i(s) - H(s, \mathbf{z}_0) \right| \to_P 0, \qquad (6.18)$$

where $H(s, \mathbf{z}_0)$ is defined in (R3). For each $i = 1, \dots, n$, write $Y_i(s) = Y_{ia}(s) - Y_{ib}(s)$, where $Y_{ia}(s)$ and $Y_{ib}(s)$ are left-continuous nondecreasing random functions. By Proposition 6.1,

$$\left| \sum_{i=1}^{n} W_i(\mathbf{z}_0) Y_{ia}(s) - H_a(s, \mathbf{z}_0) \right| \to_P 0, \tag{6.19}$$

$$\left| \sum_{i=1}^{n} W_i(\mathbf{z}_0) Y_{ia}(s+) - H_a(s+, \mathbf{z}_0) \right| \to_P 0,$$

for every s, where $H_a(s, \mathbf{z}_0) = E(Y_{1a}(s) | \mathbf{Z}_1 = \mathbf{z}_0)$ and we have used the fact that $E(Y_{1a}(s+) | \mathbf{Z}_1 = \mathbf{z}_0) = E(\lim_{k \to \infty} Y(s + 1/k) | \mathbf{Z}_1 = \mathbf{z}_0) = \lim_{k \to \infty} E(Y(s + 1/k) | \mathbf{Z}_1 = \mathbf{z}_0) = H_a(s+, \mathbf{z}_0)$ by the bounded convergence theorem. Using the standard arguments similar to those in the proof of Theorem 5.5.1 of Chung (1974), it is shown that (6.19) implies

$$\sup_{s \in [0,1]} \left| \sum_{i=1}^{n} W_i(\mathbf{z}_0) Y_{ia}(s) - H_a(s, \mathbf{z}_0) \right| \to_P 0.$$

Similarly, we have

$$\sup_{s \in [0,1]} \left| \sum_{i=1}^{n} W_i(\mathbf{z}_0) Y_{ib}(s) - H_b(s, \mathbf{z}_0) \right| \to_P 0$$

with $H_b(s, \mathbf{z}_0) = E(Y_{1b}(s) | \mathbf{Z}_1 = \mathbf{z}_0)$. Therefore, (6.18) holds since $H(s, \mathbf{z}_0) = H_a(s, \mathbf{z}_0) - H_b(s, \mathbf{z}_0)$.

Note also that $(1/nR_n^p) \sum_{j=1}^n w((\mathbf{Z}_j - \mathbf{z}_0)/R_n)$ converges in probability to $f(\mathbf{z}_0)$, by Proposition 6.2.

Recall that R_n is the Euclidean distance from $\mathbf{z}_0$ to the k_nth closest of $\mathbf{Z}_1, \dots, \mathbf{Z}_n$. Let $(\tilde{\mathbf{Z}}_1, \tilde{Y}_1(s)), \dots, (\tilde{\mathbf{Z}}_{k_n}, \tilde{Y}_{k_n}(s))$ be the k_n points among $(\mathbf{Z}_1, Y_1(s)), \dots, (\mathbf{Z}_n, Y_n(s))$ such that $\mathbf{Z}_i$ lies in the ball centered at $\mathbf{z}_0$ and of radius R_n. Then, for $A_i \subset \{\mathbf{z}\colon \|\mathbf{z} - \mathbf{z}_0\| \le r\}$, $i = 1, \dots, k_n$, and $y_1, \dots, y_{k_n} \in \{0, 1\}$, a direct calculation shows that the joint conditional distribution function $P(\tilde{\mathbf{Z}}_1 \in A_1, \tilde{Y}_1(s) = y_1, \dots, \tilde{\mathbf{Z}}_{k_n} \in A_{k_n}, \tilde{Y}_{k_n}(s) = y_{k_n} | R_n = r)$ is given by

$$\prod_{i=1}^{k_n} \frac{P(\mathbf{Z}_i \in A_i, Y_i(s) = y_i)}{G(r)},$$

where $G(r) \equiv P(\|\mathbf{Z}_1 - \mathbf{z}_0\| \le r)$. So given $R_n = r$, $(\tilde{\mathbf{Z}}_1, \tilde{Y}_1(s)), \dots, (\tilde{\mathbf{Z}}_{k_n}, \tilde{Y}_{k_n}(s))$ are conditionally independent and identically distributed. This also implies that the conditional subdensity function of $P(\tilde{\mathbf{Z}}_1 \le \mathbf{z}, \tilde{Y}_1(s) = 1 | R_n = r)$ is given by $(f(\mathbf{z}, s))/G(r))$, where $f(\mathbf{z}, s)$ is defined by (4.1).

Fix $l \in \{1, \dots, p\}$. For each $s \in [0, 1]$, denote by $\eta_n(s)$ the lth component of the numerator of (6.17). Then since w vanishes outside the unit ball,

$$\eta_n(s) = \frac{\alpha_n}{nR_n^{p-1}} \sum_{i=1}^{k_n} w\left(\frac{\tilde{\mathbf{Z}}_i - \mathbf{z}_0}{R_n} \right) \tilde{Y}_i(s) \left(\frac{\tilde{Z}_{il} - z_{0l}}{R_n} \right).$$

We note that $|w((\tilde{\mathbf{Z}}_i - \mathbf{z}_0)/R_n)\tilde{Y}_i(s)((\tilde{Z}_{il} - z_{0l})/R_n)|$ is bounded by a constant, say C_1, which is independent of s. By Theorem 2 of Hoeffding (1963),

$$\begin{aligned}P(\eta_n(s) - E(\eta_n(s)|R_n) > \varepsilon) &= E(P(\eta_n(s) - E(\eta_n(s)|R_n) > \varepsilon|R_n))\\ &\le \exp\left\{-2k_n\left(\frac{\varepsilon}{(\alpha_n/nR_n^{p-1})(k_n)C_1}\right)^2\right\}\\ &= \exp\left\{-\left(\frac{2\varepsilon^2}{C_1^2}\right)k_n^{1/2}\left(\frac{k_n}{nR_n^p}\right)^{(2(1-p))/p}\right\}\\ &\to 0 \quad \text{uniformly in } s \in [0,1],\end{aligned}$$

where the convergence follows from (6.16) and the assumption that $k_n \to \infty$. Similarly

$$P(\eta_n(s) - E(\eta_n(s)|R_n) < -\varepsilon) = P(-\eta_n(s) - E(-\eta_n(s)|R_n) > \varepsilon) \to 0$$

uniformly in $s \in [0, 1]$. Thus

$$\eta_n(s) - E(\eta_n(s)|R_n) \to_P 0 \quad \text{uniformly in } s \in [0,1]. \tag{6.20}$$

Moreover,

$$\begin{aligned}|E(\eta_n(s)|R_n)| &= k_n\left(\frac{\alpha_n}{nR_n^{p-1}}\right)\left|E\left(w\left(\frac{\tilde{\mathbf{Z}}_i - \mathbf{z}_0}{R_n}\right)\tilde{Y}_i(s)\left(\frac{\tilde{Z}_{il} - z_{0l}}{R_n}\right)\middle|R_n\right)\right|\\ &= \frac{k_n\alpha_n}{nR_n^{p-1}}\left|\int w\left(\frac{\mathbf{z} - \mathbf{z}_0}{R_n}\right)\left(\frac{z_l - z_{0l}}{R_n}\right)\frac{f(\mathbf{z}, s)}{G(R_n)}\,d\mathbf{z}\right|\\ &= \frac{k_n\alpha_n R_n}{nG(R_n)}\left|\int w(\mathbf{u})u_l(f(\mathbf{z}_0 + R_n\mathbf{u}, s) - f(\mathbf{z}_0, s))\,d\mathbf{u}\right|\\ &\le M\frac{k_n\alpha_n R_n^2}{nG(R_n)}\int w(\mathbf{u})|u_l|\cdot\|\mathbf{u}\|\,d\mathbf{u} \quad [\text{by (4.2)}]\\ &= O_p\left(\left(\frac{k_n^{p+4}}{n^4}\right)^{1/(4p)}\right) \quad \text{uniformly in } s \in [0,1],\end{aligned}$$

where the fifth line uses (6.16) and the fact that as $r \to 0$,

$$\begin{aligned}G(r) &= \gamma(p)f(\mathbf{z}_0)r^p + \int_{\|z - \mathbf{z}_0\| \le r}(f(z) - f(\mathbf{z}_0))\,dz\\ &= \gamma(p)f(\mathbf{z}_0)r^p + o(r^p).\end{aligned} \tag{6.21}$$

Hence, if $k_n^{p+4}/n^4 \to 0$,

$$\begin{aligned}\eta_n(s) &= (\eta_n(s) - E(\eta_n(s)|R_n)) + E(\eta_n(s)|R_n)\\ &\to_P 0 \quad \text{uniformly in } s \in [0,1].\end{aligned}$$

Therefore,

$$\left(n(k_n/n)^{(p-4)/p}\right)^{1/4} \sum_{i=1}^{n} c_i(s, \mathbf{z}_0)(\mathbf{Z}_i - \mathbf{z}_0) = o_p(1). \qquad \square$$

PROOF OF THEOREM 3. Before checking Conditions A and B, we first show that

$$\frac{1}{R_n^2} \sum_i c_i(s, \mathbf{z}_0)(\mathbf{Z}_i - \mathbf{z}_0)(\mathbf{Z}_i - \mathbf{z}_0)' \to_P \operatorname{diag}\left(\int u_1^2 w(\mathbf{u})\, d\mathbf{u}, \ldots, \int u_p^2 w(\mathbf{u})\, d\mathbf{u}\right) \tag{6.22}$$

uniformly in $s \in [0,1]$ if $k_n \to \infty$ and $k_n/n \to 0$. For each l, $1 \le l \le p$,

$$\begin{aligned} &\frac{1}{R_n^2} \sum_{i=1}^{n} c_i(s, \mathbf{z}_0)(Z_{il} - z_{0l})^2 \\ &= \sum_{i=1}^{n} W_i(\mathbf{z}_0) Y_i(s) \left(\frac{Z_{il} - z_{0l}}{R_n}\right)^2 \Big/ \sum_{j=1}^{n} W_j(\mathbf{z}_0) Y_j(s) \\ &= \frac{(1/nR_n^p) \sum_{i=1}^{n} w((\mathbf{Z}_i - \mathbf{z}_0)/R_n)((Z_{il} - z_{0l})/R_n)^2 Y_i(s)}{(1/nR_n^p) \sum_{j=1}^{n} w\left((\mathbf{Z}_j - \mathbf{z}_0)/R_n\right) \sum_{j=1}^{n} W_j(\mathbf{z}_0) Y_j(s)}. \end{aligned} \tag{6.23}$$

It follows from Proposition 6.2 and (6.18) that the denominator in (6.23) converges in probability to $f(\mathbf{z}_0)H(s, \mathbf{z}_0) = f(\mathbf{z}_0, s)$ uniformly in $s \in [0,1]$. We next show that the numerator in (6.23) converges in probability to $f(\mathbf{z}_0, s) \int w(\mathbf{u}) u_l^2\, d\mathbf{u}$. It then follows that

$$\frac{1}{R_n^2} \sum_{i=1}^{n} c_i(s, \mathbf{z}_0)(Z_{il} - z_{0l})^2 \to_P \int u_l^2 w(\mathbf{u})\, d\mathbf{u} \quad \text{uniformly in } s \in [0,1].$$

Denote by $\xi_n(s)$ the numerator in (6.23). Then since w vanishes outside the unit ball,

$$\xi_n(s) = \frac{1}{nR_n^p} \sum_{i=1}^{k_n} w\left(\frac{\tilde{\mathbf{Z}}_i - \mathbf{z}_0}{R_n}\right) \tilde{Y}_i(s) \left(\frac{\tilde{Z}_{il} - z_{0l}}{R_n}\right)^2,$$

where $(\tilde{\mathbf{Z}}_1, \tilde{Y}_1(s)), \ldots, (\tilde{\mathbf{Z}}_{k_n}, \tilde{Y}_{k_n}(s))$ are the k_n points among $(\mathbf{Z}_1, Y_1(s)), \ldots, (\mathbf{Z}_n, Y_n(s))$ such that $\mathbf{Z}_i$ lies in the ball centered at $\mathbf{z}_0$ and of radius R_n. Recall that given $R_n = r$, $(\tilde{\mathbf{Z}}_1, \tilde{Y}_1(s)), \ldots, (\tilde{\mathbf{Z}}_{k_n}, \tilde{Y}_{k_n}(s))$ are conditionally independent and identically distributed with the conditional subdensity function $P(\tilde{\mathbf{Z}}_1 \le \mathbf{z}, \tilde{Y}_1(s) = 1 | R_n = r)$ given by $(f(\mathbf{z}, s))/(G(r))$ (see the third paragraph in the proof of Lemma 6.1). Hence,

$$\begin{aligned} E(\xi_n(s)|R_n) &= \frac{k_n}{nR_n^p} E\left(w\left(\frac{\tilde{\mathbf{Z}}_i - \mathbf{z}_0}{R_n} \right) \tilde{Y}_i(s) \left(\frac{\tilde{Z}_{il} - z_{0l}}{R_n} \right)^2 \middle| R_n \right) \\ &= \frac{k_n}{nR_n^p} \int w\left(\frac{\mathbf{z} - \mathbf{z}_0}{R_n} \right) \left(\frac{z_l - z_{0l}}{R_n} \right)^2 \frac{f(\mathbf{z}, s)}{G(R_n)}\, d\mathbf{z} \\ &= \frac{k_n}{nG(R_n)} \int w(\mathbf{u}) u_l^2 f(\mathbf{z}_0 + R_n \mathbf{u}, s)\, d\mathbf{u} \\ &\to_P f(\mathbf{z}_0, s) \int w(\mathbf{u}) u_l^2\, d\mathbf{u} \quad \text{uniformly in } s \in [0,1], \end{aligned}$$

where the convergence statement follows from (6.21), (6.16) and (4.2). Moreover, as in (6.20), we have

$$\xi_n(s) - E(\xi_n(s)|R_n) \to_P 0 \quad \text{uniformly in } s \in [0,1].$$

Therefore,

$$\begin{aligned} \xi_n(s) &= (\xi_n(s) - E(\xi_n(s)|R_n)) + E(\xi_n(s)|R_n) \\ &\to_P f(\mathbf{z}_0, s) \int w(\mathbf{u}) u_l^2\, d\mathbf{u} \quad \text{uniformly in } s \in [0,1]. \end{aligned} \tag{6.24}$$

For $l \ne m$ $(1 \le l, m \le p)$, we have

$$\begin{aligned} &\frac{1}{R_n^2} \sum_{i=1}^n c_i(s, \mathbf{z}_0)(Z_{il} - z_{0l})(Z_{im} - z_{0m}) \\ &\quad = (1/nR_n^p) \textstyle\sum_{i=1}^n w((Z_i - \mathbf{z}_0)/R_n) Y_i(s) \\ &\qquad \times ((\mathbf{Z}_{il} - z_{0l})/R_n)((Z_{im} - z_{0m})/R_n) \\ &\qquad \times \left[(1/nR_n^p) \textstyle\sum_{j=1}^n w((\mathbf{Z}_j - \mathbf{z}_0)/R_n) \sum_{j=1}^n W_j(\mathbf{z}_0) Y_j(s) \right]^{-1}. \end{aligned} \tag{6.25}$$

It follows from Proposition 6.2 and (6.18) that the denominator of (6.25) converges in probability to $f(\mathbf{z}_0)H(s, \mathbf{z}_0)$ uniformly in $s \in [0,1]$. We need to show that the numerator of (6.25) converges to zero in probability uniformly in $s \in [0,1]$, and this is accomplished using the technique we used to prove the convergence of the numerator of (6.17) in the proof of Lemma 6.1 [cf. also the proof of (6.24)]. Thus the left-hand side of (6.25) converges to zero in probability uniformly in $s \in [0,1]$. Therefore, (6.22) holds.

Now we prove part (i) of the theorem. We have

$$
\begin{aligned}
\frac{1}{R_n^2}\mathbf{Z}^{*\prime}P\mathbf{Z}^*(s) &= \frac{1}{R_n^2}\sum_{i=1}^n c_i(s,\mathbf{z}_0)(\mathbf{Z}_i-\mathbf{z}_0)(\mathbf{Z}_i-\mathbf{z}_0)' \\
&\quad -\frac{1}{R_n^2}\left(\sum_{i=1}^n c_i(s,\mathbf{z}_0)(\mathbf{Z}_i-\mathbf{z}_0)\right)\left(\sum_{i=1}^n c_i(s,\mathbf{z}_0)(\mathbf{Z}_i-\mathbf{z}_0)\right)' \\
&= \frac{1}{R_n^2}\sum_{i=1}^n c_i(s,\mathbf{z}_0)(\mathbf{Z}_i-\mathbf{z}_0)(\mathbf{Z}_i-\mathbf{z}_0)' \\
&\quad -\frac{1}{R_n^2}\left(o_P\left(\left(n\left(\frac{k_n}{n}\right)^{(p-4)/p}\right)^{-1/4}\right)\right)^2 \qquad (6.26)\\
&= \frac{1}{R_n^2}\sum_{i=1}^n c_i(s,\mathbf{z}_0)(\mathbf{Z}_i-\mathbf{z}_0)(\mathbf{Z}_i-\mathbf{z}_0)' - o_p\left((k_n)^{-1/2}\right) \\
&\to_P \operatorname{diag}\left(\int u_1^2 w(\mathbf{u})\,d\mathbf{u},\dots,\int u_p^2 w(\mathbf{u})\,d\mathbf{u}\right)
\end{aligned}
$$

uniformly in $s \in [0,1]$,

and each integral in the last line of (6.26) is positive. Here, the second equality follows from part (i) of Lemma 6.1, the third equality from (6.16) and the convergence statement from (6.22). This implies (A1).

To verify (A2), it suffices to show that for $a_n = k_n/n$ and for every $\delta \geq 0$,

$$
\sup_{s\in[0,1]}\left|(na_n)^{1+\delta}\sum_{i=1}^n c_i^{2+\delta}(s,\mathbf{z}_0) - g_\delta(s,\mathbf{z}_0)\right| \to_P 0. \qquad (6.27)
$$

Let $\tilde{w}(\cdot) = w^{2+\delta}(\cdot)/\int w^{2+\delta}(\mathbf{u})\,d\mathbf{u}$ and

$$
\tilde{W}_i(\mathbf{z}_0) \equiv \tilde{w}\left(\frac{\mathbf{Z}_i-\mathbf{z}_0}{h_n}\right)\Big/\sum_{j=1}^n \tilde{w}\left(\frac{\mathbf{Z}_j-\mathbf{z}_0}{h_n}\right).
$$

Then since $\tilde{w}(\cdot)$ is a density function that satisfies the regularity conditions needed to apply Propositions 6.1 and 6.2, we have

$$
\begin{aligned}
&k_n^{1+\delta}\sum_{i=1}^n W_i^{2+\delta}(\mathbf{z}_0)Y_i(s) \\
&\quad = \left(\frac{k_n}{nR_n^p}\right)^{1+\delta}\left[\sum_{i=1}^n \tilde{W}_i(\mathbf{z}_0)Y_i(s)\right] \qquad (6.28)\\
&\qquad \times\left[\frac{(1/nR_n^p)\sum_{j=1}^n\tilde{w}\left((\mathbf{Z}_j-\mathbf{z}_0)/R_n\right)}{\left((1/nR_n^p)\sum_{j=1}^n w\left((\mathbf{Z}_j-\mathbf{z}_0)/R_n\right)\right)^{2+\delta}}\right]\int w^{2+\delta}(\mathbf{u})\,d\mathbf{u} \\
&\quad \to_P (\gamma(p))^{1+\delta}H(s,\mathbf{z}_0)\int w^{2+\delta}(\mathbf{u})\,d\mathbf{u} \quad \text{uniformly in } s\in[0,1],
\end{aligned}
$$

where the convergence of the factor in the first set of brackets in the second line of (6.28) follows from the arguments leading to (6.18) and the convergence of the numerator and denominator in the second set of brackets follows by Proposition 6.2. Thus,

$$
\begin{aligned}
&k_n^{1+\delta} \sum_{i=1}^{n} c_i^{2+\delta}(s, \mathbf{z}_0) \\
&\quad = k_n^{1+\delta} \sum_{i=1}^{n} W_i^{2+\delta}(\mathbf{z}_0) Y_i(s) \Bigg/ \left(\sum_{i=1}^{n} W_i(\mathbf{z}_0) Y_i(s) \right)^{2+\delta} \\
&\quad \to_P \left(\frac{\gamma(p)}{H(s, \mathbf{z}_0)} \right)^{1+\delta} \int w^{2+\delta}(\mathbf{u})\, d\mathbf{u} \quad \text{uniformly in } s \in [0,1]
\end{aligned}
$$

and hence (6.27) holds.

By regularity condition (R1),

$$
\begin{aligned}
&\left| \sum_{i=1}^{n} c_i(s, \mathbf{z}_0)(\alpha(s, \mathbf{Z}_i) - \alpha(s, \mathbf{z}_0)) \right| \\
&\quad \le \left| \left(\frac{\partial \alpha}{\partial \mathbf{z}}(s, \mathbf{z}_0) \right)' \sum_{i=1}^{n} c_i(s, \mathbf{z}_0)(\mathbf{Z}_i - \mathbf{z}_0) \right| + O_p(R_n^2) \\
&\quad = o_p\left(\left(n \left(\frac{k_n}{n} \right)^{(p-4)/p} \right)^{-1/4} \right) + O_p\left(\left(\frac{k_n}{n} \right)^{2/p} \right) \quad \text{uniformly in } s \in [0,1],
\end{aligned}
$$

where the last statement follows from part (i) of Lemma 6.1 and (6.16). Thus

$$
\begin{aligned}
&\sqrt{k_n} \int_0^1 \left| \sum_{i=1}^{n} c_i(s, \mathbf{z}_0)(\alpha(s, \mathbf{Z}_i) - \alpha(s, \mathbf{z}_0)) \right| ds \\
&\quad = o_p\left((k_n^{p+4}/n^4)^{1/(4p)} \right) + O_p\left((k_n^{p+4}/n^4)^{1/(2p)} \right) \to_P 0
\end{aligned}
$$

when $k_n^{p+4}/n^4 \to 0$. Therefore, (A3) holds.

Since (R1) implies that $|\alpha(s, \mathbf{z}) - \alpha(s, \mathbf{z}_0)| \le M_1 \|\mathbf{z} - \mathbf{z}_0\|$ for some constant $M_1 > 0$,

$$
\begin{aligned}
&\sqrt{k_n} \left| \sum_{i=1}^{n} c_i(s, \mathbf{z}_0)(\alpha(s, \mathbf{Z}_i) - \alpha(s, \mathbf{z}_0))^2 \right| \\
&\quad \le \sqrt{k_n} M_1 \sum_{i=1}^{n} c_i(s, \mathbf{z}_0) \|\mathbf{Z}_i - \mathbf{z}_0\|^2 \\
&\quad = \sqrt{k_n} O_p(R_n^2) = O_p\left((k_n^{p+4}/n^4)^{1/(2p)} \right)
\end{aligned}
$$

uniformly in $s \in [0,1]$, where the first equality follows from (6.22) and the second equality from (6.16). Together with the assumption that $k_n^{p+4}/n^4 \to 0$, this implies (A4).

Now we check (A5). Note that for every i and s,

$$
\begin{aligned}
k_n c_i(s, \mathbf{z}_0) &= k_n \frac{W_i(\mathbf{z}_0) Y_i(s)}{\sum_{j=1}^n W_j(\mathbf{z}_0) Y_j(s)} \\
&\le \frac{k_n}{nR_n^p} \frac{\sup_{\mathbf{u}} w(\mathbf{u})}{(1/nR_n^p)\sum_{j=1}^n w\big((\mathbf{Z}_j - \mathbf{z}_0)/R_n\big) \sum_{j=1}^n W_j(\mathbf{z}_0) Y_j(s)}.
\end{aligned} \tag{6.29}
$$

Furthermore, by Proposition 6.2 and (6.18), the right-hand side of (6.29) converges in probability uniformly in $s \in [0,1]$ to $\gamma(p)\sup_{\mathbf{u}} w(\mathbf{u})/H(s, \mathbf{z}_0)$, which is bounded because $\inf_{s \in [0,1]} H(s, \mathbf{z}_0) > 0$ by (R3). Therefore,

$$k_n c_i(s, \mathbf{z}_0) = O_p(1) \quad \text{uniformly in } i \text{ and } s \in [0,1].$$

Finally, because

$$
\begin{aligned}
&\sqrt{k_n}\, J_1(s)(\mathbf{z}^*(s)'\mathbf{c}(s) - \mathbf{z}_0)'(\mathbf{Z}^*(s)'P(s)\mathbf{Z}^*(s))^{-1}(\mathbf{Z}^*(s)'\mathbf{c}(s) - \mathbf{z}_0) \\
&= J_1(s)\left(\frac{k_n}{nR_n^p}\right)^{2/p}\left(\left(n\left(\frac{k_n}{n}\right)^{(p-4)/p}\right)^{1/4} \sum_{i=1}^n c(s, \mathbf{z}_0)(\mathbf{Z}_i - \mathbf{z}_0)\right)' \\
&\quad \times \left(\frac{1}{R_n^2}\mathbf{Z}^*(s)'P(s)\mathbf{Z}^*(s)\right)^{-1} \\
&\quad \times \left(\left(n\left(\frac{k_n}{n}\right)^{(p-4)/p}\right)^{1/4} \sum_{i=1}^n c(s, \mathbf{z}_0)(\mathbf{Z}_i - \mathbf{z}_0)\right),
\end{aligned}
$$

(A6) follows immediately from (6.16), part (i) of Lemma 6.1 and (6.26).

The proof of part (ii) of the theorem uses part (ii) of Lemma 6.1 and is completely parallel to the proof of part (i). □

Acknowledgments. We wish to thank the Associate Editor and referee for their insightful comments that led to an improved version of this paper. We are grateful to Dr. Anders Green for providing the Fyn diabetes data. We also thank Per Kragh Andersen for facilitating the transfer of the data.

REFERENCES

AALEN, O. O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.* **6** 701–726.

AALEN, O. O. (1980). *A Model for Nonparametric Regression Analysis of Counting Processes. Lecture Notes in Statist.* **2** 1–25. Springer, New York.

ANDERSEN, P. K., BORCH-JOHNSEN, K., DECKERT, T., GREEN, A., HOUGAARD, P., KEIDING, N. and KREINER, S. (1985). A Cox regression model for the relative mortality and its applications to diabetes mellitus survival data. *Biometrics* **41** 921–932.

ANDERSEN, P. K., BORGAN, O., GILL, R. D. and KEIDING, N. (1993). *Statistical Models Based on Counting Processes*. Springer, New York.

ANDERSEN, P. K. and GILL, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.* **10** 1100–1120.

BERAN, R. (1981). Nonparametric regression with randomly censored survival data. Technical Report, Dept. Statistics, Univ. California, Berkeley.

CHAMBERS, J. and HASTIE, T. (1991). *Statistical Models in S.* Wadsworth and Brooks/Cole, Pacific Grove, CA.

CHUNG, K. L. (1974). *A Course in Probability Theory*, 2nd ed. Academic Press, New York.

COLLOMB, G. (1980). *Estimation de la Regression par la Méthode des k points les Plus Proches avec Noyau: Quelques Propriétés de Convergence Ponctuelle. Lecture Notes in Math.* **821** 159–175. Springer, New York.

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

DABROWSKA, D. M. (1987). Nonparametric regression with censored survival time data. *Scand. J. Statist.* **14** 181–197.

GILL, R. D. (1989). Non- and semi-parametric maximum likelihood estimators and the von Mises method (Part 1). *Scand. J. Statist.* **16** 97–128.

GILL, R. D. and JOHANSEN, S. (1990). A survey of product-integration with a view towards application in survival analysis. *Ann. Statist.* **18** 1501–1555.

GREEN, A., BORCH-JOHNSEN, K., ANDERSEN, P. K., HOUGAARD, P., KEIDING, N., KREINER, S. and DECKERT, T. (1985). The relative mortality of IDDM in Denmark 1933–1982. *Diabetologia* **28** 339–342.

HASTIE, T. and TIBSHIRANI, R. (1990). *Generalized Additive Models.* Chapman and Hall, New York.

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

LI, G. and DOSS, H. (1994). An approach to nonparametric regression for life history data using local linear fitting. Technical Report 94-1, Dept. Mathematics, Univ. North Carolina, Charlotte.

MCKEAGUE, I. W. and UTIKAL, K. J. (1990). Inference for a nonlinear counting process regression model. *Ann. Statist.* **18** 1172–1187.

MCKEAGUE, I. W. and UTIKAL, K. J. (1991). Goodness-of-fit tests for additive hazards and proportional hazards models. *Scand. J. Statist.* **18** 177–195.

MOORE, D. S. and YACKEL, J. W. (1977). Consistency properties of nearest neighbor density function estimators. *Ann. Statist.* **5** 143–154.

PETERSON, A. V., JR. (1977). Expressing the Kaplan–Meier estimator as a function of empirical survival functions. *J. Amer. Statist. Assoc.* **72** 854–858.

PRAKASA RAO, B. L. S. (1983). *Nonparametric Functional Estimation.* Academic Press, New York.

RAO, C. R. (1973). *Linear Statistical Inference and Its Applications.* Wiley, New York.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF NORTH CAROLINA AT CHARLOTTE
CHARLOTTE, NORTH CAROLINA 28223

DEPARTMENT OF STATISTICS
OHIO STATE UNIVERSITY
1958 NEIL AVENUE
COLUMBUS, OHIO 43210-1247