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

by

Gang Li and Hani Doss
Purdue University — Florida State University

Technical Report #93-16C

March 1993

Research supported by Air Force Office of Scientific Research Grant 90-0202.

*

An Approach to Nonparametric Regression for Life History Data Using Local Linear Fitting

Gang Li
Department of Statistics
Purdue University

Hani Doss*
Department of Statistics
Florida State University

March 1993

Abstract

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 (1972) model assumes that the covariates act multiplicatively on the hazard rate, and Aalen's (1980) 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 good fit to a given data set. Because our results are obtained under the multivariate counting process setting of Aalen (1978), 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 (1981) for the random censorship model.

Keywords and phrases: Aalen model; censoring; counting process; hazard function; intensity; kernel; martingale; nearest neighbor.

*Research supported by Air Force Office of Scientific Research Grant 90-0202.

1 Introduction and Summary

Let T be the survival time of an individual with covariate vector $\boldsymbol{Z} = (Z_1, \ldots, Z_p)$. 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

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

the hazard function for an individual with covariate $\boldsymbol{Z} = \boldsymbol{z}$, has the form

$$h(t \mid \boldsymbol{z}) = h_0(t) \exp(\beta' \boldsymbol{z}), \tag{1.2}$$

where β is a vector of unknown regression coefficients and $h_0(t)$ is an unknown and unspecified nonnegative 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 β. On the other hand, the form (1.2) of the hazard rate is extremely rigid.

Let $S(t \mid \boldsymbol{z}) = P(T > t \mid \boldsymbol{Z} = \boldsymbol{z})$ denote the distribution of the survival time for an individual with covariate vector $\boldsymbol{z}$. Beran (1981) considered the more general model in which $\{S(\cdot \mid \boldsymbol{z})\}_{\boldsymbol{z}}$ is a completely arbitrary family of distribution functions. Supposing densities exist, this is the model

$$h(t \mid \boldsymbol{z}) = \alpha(t, \boldsymbol{z}) \tag{1.3}$$

where we assume only that for each $\boldsymbol{z}$, $\alpha(\cdot, \boldsymbol{z})$ is a hazard function. Beran considered estimation of $S(t \mid \boldsymbol{z})$ and the corresponding cumulative hazard function $A(t, \boldsymbol{z}) = \int_0^t \alpha(s, \boldsymbol{z}) ds$ under a random censorship model in which for each of n individuals, the survival time T_i of an individual is observed until a censoring time C_i. More specifically, suppose that the data consists of n i.i.d. triples $(X_1, \delta_1, \boldsymbol{Z}_1), \ldots, (X_n, \delta_n, \boldsymbol{Z}_n)$ where $X_i = \min(T_i, C_i)$ and $\delta_i = I(T_i \leq C_i)$, $i = 1, \ldots, n$. Suppose further that $\boldsymbol{Z}$ does not depend on time, and that T is conditionally independent of C given $\boldsymbol{Z}$. Beran proposed a "local Nelson-Aalen estimator" of $A(t, \boldsymbol{z})$, which is described as follows. For each $\boldsymbol{z}$, let $K_1(t \mid \boldsymbol{z}) = P(X > t,\ \delta = 1 \mid \boldsymbol{Z} = \boldsymbol{z})$ and $K(t \mid \boldsymbol{z}) = P(X > t \mid \boldsymbol{Z} = \boldsymbol{z})$. The representation

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

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

$$\hat{A}(t, \boldsymbol{z}_0) = -\int_0^t \frac{d\hat{K}_1(s \mid \boldsymbol{z}_0)}{\hat{K}(s- \mid \boldsymbol{z}_0)} \tag{1.4}$$

where $\hat{K}_1(t \mid \boldsymbol{z}_0) = \sum_{i=1}^n W_i(\boldsymbol{z}_0) I(X_i > t,\ \delta_i = 1)$ and $\hat{K}(t \mid \boldsymbol{z}_0) = \sum_{i=1}^n W_i(\boldsymbol{z}_0) I(X_i > t)$. Here $\{W_i(\boldsymbol{z}_0),\ i = 1, \ldots, n\}$ is a set of nonnegative weights depending on the covariates only. For instance the $W_i(\boldsymbol{z}_0)$'s can be taken as the nearest neighbor or kernel weights used in density estimation and nonparametric regression. When one uses the constant weights $W_i(\boldsymbol{z}_0) = 1/n$ for all individuals, $\hat{A}$ reduces to the ordinary Nelson-Aalen estimator (see

Aalen (1978)). Beran (1981) showed that $\hat{A}$ is strongly consistent if one uses nearest neighbor or kernel weights. Weak convergence results for these estimators were later obtained by Dabrowska (1987). McKeague and Utikal (1990) obtained weak convergence results for estimators of the type proposed by Beran, except that the local averaging is done by "binning" the data. Their study was carried out in the framework of Aalen's (1978) multiplicative intensity model.

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 $\boldsymbol{z}$ direction. It seems preferable to do a "local linear fit" in the $\boldsymbol{z}$ 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 $E(Y \mid Z = z_0)$. 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 (if the data lie exactly along a straight line, but the z_i's are not equally spaced, the smoother need not yield a straight line). 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"). For more on this, see the discussion and references on p. 376 of Chambers and Hastie (1991).

Our estimators are developed in the context of Aalen's (1978) multiplicative intensity model. We now wish to give a preliminary description of our estimators of the conditional cumulative hazard function, and this is easiest to do in the simpler context of the random censorship model described earlier. Fix a small neighborhood $\mathcal{N}_{\boldsymbol{z}_0}$. Within $\mathcal{N}_{\boldsymbol{z}_0}$ we have

$$\alpha(t, \boldsymbol{z}) \approx \beta_0(t, \boldsymbol{z}_0) + \beta_1(t, \boldsymbol{z}_0) z_1 + \cdots + \beta_p(t, \boldsymbol{z}_0) z_p. \tag{1.5}$$

The integrated version of this is

$$A(t, \boldsymbol{z}) \approx B_0(t, \boldsymbol{z}_0) + B_1(t, \boldsymbol{z}_0) z_1 + \cdots + B_p(t, \boldsymbol{z}_0) z_p, \tag{1.6}$$

where $B_j(t, \boldsymbol{z}_0) = \int_0^t \beta_j(s, \boldsymbol{z}_0) ds$ $(0 \le j \le p)$. Imagine now that we have equality in (1.6). We estimate $A(t, \boldsymbol{z}_0)$ by a function which has jumps only at the observed death times. Suppose there is an observed death at $X_{(i)}$. Let $\mathcal{I}_i$ denote the interval $[X_{(i)}, X_{(i)} + dt)$, let $\mathcal{R}_i$ denote the set of individuals whose covariates are in $\mathcal{N}_{\boldsymbol{z}_0}$ and who are at risk of dying in the interval $\mathcal{I}_i$, and let n_i denote the cardinality of $\mathcal{R}_i$. Each individual in $\mathcal{R}_i$ is observed to die in the interval $\mathcal{I}_i$ or not. Thus, we have a "response" vector of length n_i, consisting of $n_i - 1$ zeros and a single one. If we formally regress this vector on the covariates z_l, $l \in \mathcal{R}_i$, we obtain an estimate of the increment of the functions $B_j(\cdot, \boldsymbol{z}_0)$, $j = 1, \ldots, p$, at the point $t = X_{(i)}$. Summing up those increments gives an estimate of the $B_j(\cdot, \boldsymbol{z}_0)$'s, $j = 1, \ldots, p$, and this gives an estimate of $A(\cdot, \boldsymbol{z}_0)$. We note that if we take the neighborhood $\mathcal{N}_{\boldsymbol{z}_0}$ to be the entire $\boldsymbol{z}$-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 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 (1.5) and (1.6), we take a zeroth-order Taylor expansion $\alpha(t, z) \approx \alpha(t, z_0)$, 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 $S(t \mid z)$ through the product-integral representation

$$S(t \mid z) = \mathcal{P}_{[0,t]}(1 - dA(s, z)) \equiv \exp\{A^c(t, z)\} \prod_{s \leq t} \left(1 - \Delta A(s, z)\right), \tag{1.7}$$

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

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

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

As mentioned earlier, our development proceeds within Aalen's (1978) framework of multivariate counting processes. This approach 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 (censoring by fixed constants, 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. In that section we also discuss some technical points relating to the rate of convergence of the estimators. Since these points are important we give a very brief summary of our discussion here. In "standard" density estimation and nonparametric regression the mean integrated squared error of nearest neighbor and kernel estimators with an optimally chosen smoothing parameter is of the order of $n^{-4/5}$. On the other hand, for histogram-type estimators with an optimally chosen bin width, the mean integrated squared error is of order $n^{-2/3}$. For the more difficult problem of regression in a counting process setting, the best available rate for histogram-type estimators was obtained by McKeague and Utikal (1990), whose estimators have an integrated squared error which is almost $\mathcal{O}_p(n^{-1/2})$. As one would expect from the results in the "standard" setting, one can achieve a better rate using nearest neighbor or kernel smoothers, and indeed the estimators we propose have an integrated squared error which is almost $\mathcal{O}_p(n^{-4/5})$. Further remarks on the rate of convergence appear in Section 4.3.

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).

There are several motivations for the development of our estimators. First, we believe that our estimators will provide useful alternatives to estimators derived from the Cox model. Second, 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).

The initial motivation for the development of our estimators was that they were needed to carry out the steps in the fitting of an "additive risks model" that will be studied in a sequel. A fundamental problem with estimators based on local fitting is that they do not work well in high dimensions. This is because of the well-known "curse of dimensionality": neighborhoods with a fixed number of points become less local as the dimension increases. To overcome this problem one can consider the additive risks model

$$h(t \mid z) = \alpha_0(t) + \alpha_1(t, z_1) + \cdots + \alpha_p(t, z_p), \tag{1.8}$$

which is more restrictive than the fully nonparametric model (1.3), but has the advantage that it is easier to understand when the dimension is high. To ensure the identifibility of model (1.8) we assume that $E[\alpha_j(t, Z_j)] = 0$ for $j = 1, \ldots, p$. Estimation of the cumulative functions $A_j(t, z_j) = \int_0^t \alpha_j(s, z_j)ds$, $j = 1, \ldots, p$, is done via a "backfitting algorithm" that is similar to the one described in the framework of ordinary regression by Hastie and Tibshirani (1990). Here we give a brief description of the algorithm for $p = 2$. Let $\hat{A}_0$ be the ordinary Nelson-Aalen estimator of $A_0(t) = \int_0^t \alpha_0(s)ds$ that is obtained under the model $h(t \mid z) = \alpha_0(t)$. Let $\hat{A}_j^{(0)}$ be initial estimates of A_j, $j = 1, 2$. We fit the model (1.8) by estimating A_1, acting as if A_2 and A_0 are known to be equal to $\hat{A}_2^{(0)}$ and $\hat{A}_0$ respectively. This gives an estimate $\hat{A}_1^{(1)}$. The next step is to estimate A_2 in model (1.8) acting as if A_1 and A_0 are known to be equal to $\hat{A}_1^{(1)}$ and $\hat{A}_0$ respectively. This gives an estimate $\hat{A}_2^{(1)}$. Now we repeat this procedure, using $\hat{A}_1^{(1)}$ and $\hat{A}_2^{(1)}$ instead of $\hat{A}_1^{(0)}$ and $\hat{A}_2^{(0)}$, and iterate "until convergence". The point is that each step of the algorithm is carried out using the fully nonparametric approach based on model (1.3) with $p = 1$.

2 The Model and the Estimators

2.1 Counting Process Formulation

Let $\boldsymbol{N}^{(n)}(t) = (N_1^{(n)}(t), \ldots, N_n^{(n)}(t))'$, $t \in [0,1]$, be an n-component multivariate counting process with respect to the filtration $\mathcal{F}^{(n)} = \{\mathcal{F}^{(n)}(t) : t \in [0,1]\}$. 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 $\mathcal{F}^{(n)}(t)$-measurable. Intuitively, we think of $N_i^{(n)}(t)$ as a process that counts the number of failures for the i^{th} subject during the time interval $[0,t]$ over the study period $[0,1]$. The σ-field $\mathcal{F}^{(n)}(t)$ is thought of as containing all the information that is available at time t.

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

We consider the nonparametric regression model

$$\lambda_i^{(n)}(t) = Y_i^{(n)}(t)\alpha(t, \boldsymbol{Z}_i^{(n)}(t)), \quad i = 1, \ldots, n \tag{2.1}$$

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

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 $\boldsymbol{N}^{(n)}(t) = (N_1^{(n)}(t), \ldots, N_n^{(n)}(t))'$, $t \in [0,1]$ is a multivariate counting process with each individual process $N_i^{(n)}$ having intensity process $\lambda_i^{(n)}(t) = Y_i^{(n)}(t)h(t \mid \boldsymbol{Z}_i)$ where $h(t \mid \boldsymbol{z})$ is given by (1.1); see e.g. Chapter III of Andersen et al. (1993). In this case, model (2.1) corresponds to

$$h(t \mid \boldsymbol{Z}_i) = \alpha(t, \boldsymbol{Z}_i(t)), \quad i = 1, \ldots, n$$

which is identical to model (1.3). 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 $\boldsymbol{z}_0 = (z_{01}, \ldots, z_{0p})' \in \mathbb{R}^p$, and define $A(t, \boldsymbol{z}_0) = \int_0^t \alpha(s, \boldsymbol{z}_0)ds$ and $S(\cdot \mid \boldsymbol{z}_0) = \mathcal{P}_{[0,\cdot]}(1 - dA)$ to be conditional cumulative hazard function and the conditional survival hazard function, respectively, under model (2.1). We wish to estimate $A(t, \boldsymbol{z}_0)$, $S(t \mid \boldsymbol{z}_0)$, and $\alpha(t, \boldsymbol{z}_0)$ as functions of t, $t \in [0,1]$. Then letting $\boldsymbol{z}_0$ range over the covariate space, we obtain estimates of A, S and α as functions of $(t, \boldsymbol{z})$.

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

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

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

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

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

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

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

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

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

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

Here $J(t) = I(\text{rank}(\boldsymbol{U}(t)) = p+1)$ is the indicator that $\boldsymbol{U}(t)$ has full rank and 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 $B_j(\cdot, \boldsymbol{z}_0)$ increases only at the points t at which one of the counting processes N_i has a jump.

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

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

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

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

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

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

The conditional survival function $S(t \mid \boldsymbol{z}_0)$ in (1.7) is estimated by

$$S_n(t \mid \boldsymbol{z}_0) = \mathcal{P}_{[0,t]}(1 - dA_n) = \prod_{s \leq t} \Big(1 - \Delta A_n(s, \boldsymbol{z}_0)\Big) \tag{2.6}$$

where the second equality follows from the fact that $A_n(\cdot, \boldsymbol{z}_0)$ is a step function.

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

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

Then, by minimizing

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

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

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

In the special case where the data consists of right censored observations and the covariates are independent of time, $\hat{A}$ is equal to (1.4), the estimator introduced by Beran (1981). In Section 6 we show that 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 equation (2.7) ignores the role played by the slope function $\alpha'_{\boldsymbol{z}}(t, \boldsymbol{z}_0)$, $\hat{A}$ tends to flatten out the covariate effects within the neighborhood of $\boldsymbol{z}_0$. This in general will cause a bias, which may be severe when $\boldsymbol{z}_0$ is close to the boundary of the $\boldsymbol{z}$-region. This effect was observed in simulation studies not reported here. In practical terms, the inclusion of the linear term in the estimation procedure in effect enabled us to use larger neighborhoods of $\boldsymbol{z}_0$.

Estimation of The Conditional Hazard Function To estimate $\alpha(t, \boldsymbol{z}_0)$ for fixed $\boldsymbol{z}_0$, we shall smooth $A_n(t, \boldsymbol{z}_0)$ 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

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

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

2.3 Nearest Neighbor Estimates and Kernel Estimates

Different choices of the weight functions $W_i(t, \boldsymbol{z}_0)$ 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

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

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

Kernel Estimates For kernel estimators, the weights are defined by

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

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 $\boldsymbol{z}_0$, observations around $\boldsymbol{z}_0$ 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 pp. 29–30 of Hastie and Tibshirani (1990).

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 $\boldsymbol{Z}$ takes values only in $[0,1]^p$, and that $\sup_{(s,\boldsymbol{z})\in[0,1]\times[0,1]^p} |\alpha(s, \boldsymbol{z})| = B < \infty$. We also adopt the convention that $0/0$ is 0.

3.1 Notation and Assumptions

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

$$\begin{aligned}
c_i(s, \boldsymbol{z}_0) &= W_i(s, \boldsymbol{z}_0)Y_i(s) \Big/ \textstyle\sum_{j=1}^n W_j(s, \boldsymbol{z}_0)Y_j(s), \quad i = 1, \dots, n \\
\boldsymbol{c}(s) &= (c_1(s, \boldsymbol{z}_0), \dots, c_n(s, \boldsymbol{z}_0))' \\
C(s) &= \operatorname{diag}(c_1(s, \boldsymbol{z}_0), \dots, c_n(s, \boldsymbol{z}_0)) \\
P(s) &= C(s) - \boldsymbol{c}(s)\boldsymbol{c}(s)' \\
\mathcal{T}_{\boldsymbol{z}_0} &= \{t \in [0,1] : \inf_{u\in[0,t]} \det(\boldsymbol{Z}^{*\prime} P \boldsymbol{Z}^*)(u) > 0\} \\
J_1(s) &= I(s \in \mathcal{T}_{\boldsymbol{z}_0})
\end{aligned} \tag{3.1}$$

for all s and all i. The $c_i(s, \boldsymbol{z}_0)$'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 $\mathcal{T}_{\boldsymbol{z}_0}$, 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 $\det(\boldsymbol{Z}^{*'}P\boldsymbol{Z}^*)(u) > 0$ 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 $\sum_{j=1}^n W_j(s, \boldsymbol{z}_0)Y_j(s) = 0$, then $c_i(s, \boldsymbol{z}_0) = 0$ for all i (by our convention that $0/0 = 0$), which means that s cannot be in $\mathcal{T}_{\boldsymbol{z}_0}$. Thus, if $s \in \mathcal{T}_{\boldsymbol{z}_0}$, we have $\sum_{j=1}^n W_j(s, \boldsymbol{z}_0)Y_j(s) > 0$, so that $\sum_{i=1}^n c_i(s, \boldsymbol{z}_0) = 1$. Therefore, if $s \in \mathcal{T}_{\boldsymbol{z}_0}$, 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 $s \in \mathcal{T}_{\boldsymbol{z}_0}$, then $\boldsymbol{Z}^*(s)'\boldsymbol{c}(s) = \sum_{i=1}^n c_i(s, \boldsymbol{z}_0)\boldsymbol{Z}_i(s)$ and $\boldsymbol{Z}^*(s)'P(s)\boldsymbol{Z}^*(s) = \sum_{i=1}^n c_i(s, \boldsymbol{z}_0)\boldsymbol{Z}_i(s)\boldsymbol{Z}_i(s)' - (\sum_{i=1}^n c_i(s, \boldsymbol{z}_0)\boldsymbol{Z}_i(s))(\sum_{i=1}^n c_i(s, \boldsymbol{z}_0)\boldsymbol{Z}_i(s))'$ are the weighted mean and weighted covariance matrix of $\boldsymbol{Z}_1(s), \ldots, \boldsymbol{Z}_n(s)$.

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 equal to k_n/n.

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

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

$$\int_0^1 \Big|(na_n)^{1+\delta} \sum_{i=1}^n c_i^{2+\delta}(s, \boldsymbol{z}_0) - g_\delta(s, \boldsymbol{z}_0)\Big| ds \xrightarrow{P} 0$$

for $\delta = 0$ and for some $\delta > 0$.

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

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

(A5) $na_n c_i(s, \boldsymbol{z}_0) = O_p(1)$ uniformly in s and i.

(A6) $\sqrt{na_n} \int_0^1 J_1(s)\big(\boldsymbol{Z}^*(s)'\boldsymbol{c}(s) - \boldsymbol{z}_0\big)'\big(\boldsymbol{Z}^*(s)'P(s)\boldsymbol{Z}^*(s)\big)^{-1}\big(\boldsymbol{Z}^*(s)'\boldsymbol{c}(s) - \boldsymbol{z}_0\big) ds \xrightarrow{P} 0.$

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) $P(\mathcal{T}_{z_0} = [0,1]) \to 1$.

(B2) There exists a nonnegative measurable function $g_\delta(s,z)$ indexed by $\delta \geq 0$, and defined on $[0,1]\times[0,1]$, such that for each $t \in [0,1]$

$$\int_{t-b_n}^{t+b_n} \Big|(na_n)^{1+\delta} \sum_{i=1}^n c_i^{2+\delta}(s,z_0) - g_\delta(s,z_0)\Big| ds = o_p(b_n)$$

for $\delta = 0$ and some $\delta > 0$.

(B3) $\sqrt{\frac{na_n}{b_n}} \int_{t-b_n}^{t+b_n} \Big|\sum_{i=1}^n c_i(s,z_0)\Big(\alpha(s,Z_i(s)) - \alpha(s,z_0)\Big)\Big| ds \xrightarrow{P} 0$ for each $t \in [0,1]$.

(B4) $\int_{t-b_n}^{t+b_n} \sum_{i=1}^n c_i(s,z_0)\Big(\alpha(s,Z_i(s)) - \alpha(s,z_0)\Big)^2 ds = O_p(b_n^3)$ for each $t \in [0,1]$.

(B5) $na_nc_i(s,z_0) = O_p(1)$ uniformly in s and i.

(B6) $na_nb_n^2 \int_{t-b_n}^{t+b_n} J_1(s)\Big(Z^*(s)'c(s)-z_0\Big)'\Big(Z^*(s)'P(s)Z^*(s)\Big)^{-1}\Big(Z^*(s)'c(s)-z_0\Big) ds \xrightarrow{P} 0$.

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 $A_n(t,z_0)$ and $S_n(t\,|\,z_0)$ 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)*

$$\sqrt{na_n}\Big(A_n(\cdot,z_0) - A(\cdot,z_0)\Big) \xrightarrow{d} U(\cdot,z_0) \quad in \quad D[0,1] \tag{3.2}$$

and

$$\sqrt{na_n}\Big(S_n(\cdot\mid z_0) - S(\cdot\mid z_0)\Big) \xrightarrow{d} S(\cdot\mid z_0)U(\cdot,z_0) \quad in \quad D[0,1], \tag{3.3}$$

where $U(\cdot,z_0)$ is a continuous Gaussian martingale with mean zero and variance function

$$v(t) = \int_0^t g_0(s,z_0)\alpha(s,z_0)ds.$$

Theorem 2 *Let $\alpha_n(t,z_0)$ be the estimator of the hazard function $\alpha(t,z_0)$ 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 $\alpha(t,z_0)$ is twice differentiable with respect to t, that $D = \sup_{t\in[0,1]} |\alpha_t''(t,z_0)| < \infty$, and that $g_0(t,z_0)$ is continuous in $t \in (0,1)$. Then, under conditions (B1)–(B6),*

$$\sqrt{na_nb_n}\Big(\alpha_n(t,z_0) - \alpha(t,z_0)\Big) \xrightarrow{d} \mathcal{N}(0,\sigma_t^2) \quad for\ every\ t \in (0,1),$$

where

$$\sigma_t^2 = g_0(t, z_0)\alpha(t, z_0) \int_{-1}^{1} K^2(u)du. \tag{3.4}$$

4 Asymptotics for Nearest Neighbor 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

- The observations $(N_1, Y_1, Z_1), \ldots, (N_n, Y_n, Z_n)$ are i.i.d.
- The covariates Z_i's are time independent.
- 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, so that the conclusions of Theorems 1 and 2 are satisfied when we use these weights.

Since the weight functions used in (2.4) no longer depend on time, we shall use the notation $W_i(z)$ instead of $W_i(t, z)$. 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 $s \in [0,1]$, $\alpha(s, z)$ is differentiable with respect to z at z_0 and the derivative satisfies

$$\begin{cases} \sup_{s\in[0,1]} ||\frac{\partial \alpha}{\partial z}(s, z_0)|| < \infty \\ |\alpha(s, z) - \alpha(s, z_0) - (\frac{\partial \alpha}{\partial z}(s, z_0))'(z - z_0)| \le K_1 \cdot ||z - z_0||^2 \end{cases}$$

where $\frac{\partial \alpha}{\partial z} \equiv (\frac{\partial \alpha}{\partial z_1}, \ldots, \frac{\partial \alpha}{\partial z_p})'$ and K_1 is a constant that is independent of s.

(R2) For each s, the subdensity function

$$f(z, s) = \frac{\partial^p}{\partial z_1 \cdots \partial z_p} P(Z_1 \le z,\ Y_1(s) = 1) \tag{4.1}$$

of the subdistribution function $P(Z_1 \le z,\ Y_1(s) = 1)$ exists. In addition, there is a constant M that is independent of s such that

$$|f(z_2, s) - f(z_1, s)| \le M||z_2 - z_1|| \tag{4.2}$$

for all (z_1, s), $(z_2, s) \in [0,1]^p \times [0,1]$.

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

Define

$$H(s, z) \equiv P(Y_1(s) = 1 \mid Z_1 = z). \tag{4.3}$$

4.1 Asymptotics for k-NN Estimates

Theorem 3 *Assume the setup described in the beginning of this section. Let $W_i(\boldsymbol{z})$ be the kernel weight function defined by (2.11) where $w(\cdot)$ is a bounded density function satisfying*

(i) for $l = 1, \ldots, p$, w is a symmetric function of u_l when all the other arguments are fixed.

(ii) $w(\boldsymbol{u}) = 0$ *for all* $||\boldsymbol{u}|| > 1$ *(4.4)*

(iii) $w(c\boldsymbol{u}) \geq w(\boldsymbol{u})$ *for any* $0 \leq c \leq 1$ *and* $\boldsymbol{u} \in I\!R^p$. *(4.5)*

Assume that $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$ have a common continuous density function $f(\boldsymbol{z})$ that is positive and continuous at $\boldsymbol{z}_0$. Suppose further that the function $H(s, \boldsymbol{z})$ defined in (4.3) is continuous in $\boldsymbol{z}$ at $\boldsymbol{z}_0$ for each fixed s and satisfies $\inf_{s\in[0,1]} H(s, \boldsymbol{z}_0) > 0$. Let $a_n = k_n/n$, $b_n = (k_n/n)^{\frac{1}{p}}$ and

$$g_\delta(s, \boldsymbol{z}_0) = \left(\frac{2}{H(s, \boldsymbol{z}_0)}\right)^{1+\delta} \int w^{2+\delta}(\boldsymbol{u})d\boldsymbol{u}. \tag{4.6}$$

(1) If $k_n \to \infty$ and $k_n^{p+4}/n^4 \to 0$, then $na_n \to \infty$ and (A1)-(A6) hold with g_δ given by (4.6).

(2) 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 with g_δ given by (4.6).

4.2 Asymptotics for Kernel Estimates

Theorem 4 *Assume the setup described in the beginning of this section. Let $W_i(\boldsymbol{z})$ be the kernel weight function defined by (2.12) where $w(\cdot)$ is a bounded density function satisfying*

(i) $c_1 I_{(||\boldsymbol{u}|| \leq r)} \leq w(\boldsymbol{u}) \leq c_2 I_{(||\boldsymbol{u}|| \leq r)}$ *for some positive constants r, c_1 and c_2* *(4.7)*

(ii) for $l = 1, \ldots, p$, w is a symmetric function of u_l when all the other arguments are fixed.

Suppose that $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$ have a common density function $f(\boldsymbol{z})$ that is positive and continuous at $\boldsymbol{z}_0$. Suppose further that the function $H(s, \boldsymbol{z})$ defined in (4.3) is continuous at $\boldsymbol{z} = \boldsymbol{z}_0$ for each fixed s and satisfies $\inf_{s\in[0,1]} H(s, \boldsymbol{z}_0) > 0$. Let $a_n = h_n^p$, $b_n = h_n$, and

$$g_\delta(s, \boldsymbol{z}) = \frac{1}{\{f(\boldsymbol{z})H(s, \boldsymbol{z})\}^{1+\delta}} \int w^{2+\delta}(\boldsymbol{u})d\boldsymbol{u}. \tag{4.8}$$

(1) If $nh_n^p \to \infty$ and $nh_n^{p+4} \to 0$, then $na_n \to \infty$ and (A1)-(A6) hold with g_δ given by (4.8).

(2) 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 with g_δ given by (4.8).

4.3 Remarks on the Rate of Convergence

We now discuss the rate of convergence of our estimators and we restrict our comments to the estimation of the cumulative hazard rates (Theorem 1 and Part (1) of Theorems 3 and 4); moreover, to keep the notation as simple as possible, we discuss only the case $p = 1$. By "rate of convergence" we shall mean the sequence $\{(na_n)^{-1/2}\}$ appearing in the statement of Theorem 1. Clearly, we wish to take a_n to be as big as possible, subject to satisfying the conditions of the theorems. Our discussion is mostly in terms of the kernel estimates (the situation for k-NN estimates is entirely analogous).

To properly understand our results, it is necessary to briefly review some facts concerning the "ordinary" framework of nonparametric regression, i.e. the framework in which we have independent pairs (Y_i, X_i) all having the same distribution as the generic pair (Y, X), and we wish to estimate the regression function $m(x) = E(Y \mid X = x)$. In that setting, the bandwidth parameter h_n of the kernel density estimator $m_n(x)$ controls the bias and the variance of the estimator: increasing h_n decreases the variance but increases the bias. To choose the bandwidth optimally, in the sense of minimizing the mean squared error, one wishes to balance the variance and the square of the bias, and this is done by choosing h_n to be a certain constant times $n^{-1/5}$. For this choice of h_n, the variance and the square of the bias are constant multiples of each other, i.e. neither is negligible relative to the other, and the root mean squared error of the kernel regression estimator at a point is of the order of $n^{-2/5}$. (For an informal derivation of this see Section 5.1 of Härdle (1991), where some further references are given.) For this reason, the asymptotic distribution of $m_n(x)$ has the form

$$n^{2/5}(m_n(x) - m(x)) \xrightarrow{d} \mathcal{N}(\mu(x), v(x)) \tag{4.9}$$

where $\mu(x) \neq 0$. It is easy to see this heuristically; a formal proof is given in Section 4.2 of Härdle (1990). It is very difficult to use a result such as (4.9) to form confidence intervals for $m(x)$ because the limiting mean $\mu(x)$ is not zero and in general is hard to estimate. To get a limiting distribution which is normal with a mean of 0, one needs to take the bandwidth parameter to be of a smaller order than $n^{-1/5}$; see Schuster (1972). For this reason, a number of authors recommend "undersmoothing", i.e. taking $h_n = o(n^{-1/5})$; see Section 4.2 of Härdle (1990), and Sections 4.4 and 4.5 of Hall (1992).

The assumptions of Theorem 4 of the present paper require that $h_n = o(n^{-1/5})$, giving a rate of convergence just slightly under $n^{-2/5}$, and the remarks above indicate that this is the best rate that one can expect to achieve in weak convergence results of the form (3.2) and (3.3). If one takes the optimal rate for mean squared error, one can expect to get convergence to a Gaussian process with a nonzero mean. (See Stute (1986, p. 641) for a result in this direction, in the context of the ordinary framework of nonparametric regression.) The focus of the present paper is on weak convergence results. We note that it is extremely unlikely that one would be able to calculate asymptotic expressions for the mean squared errors of the estimates $A_n(t, z_0)$ and $S_n(t \mid z_0)$ in our level of generality.

As mentioned in Section 1, Dabrowska (1987) obtained weak convergence results of the form (3.2) and (3.3) for estimates of the type proposed by Beran (1981), for the random censorship model of survival analysis. Her assumptions on the bin width are the same as ours, i.e. her results are valid as long as $h_n = o(n^{-1/5})$.

Returning to the ordinary framework of nonparametric regression, if we use estimates obtained by binning the data, i.e. estimates of the histogram type, the asymptotics are different. The optimal bin width h_n is of the order of $n^{-1/3}$ (smaller than $n^{-1/5}$), and the asymptotic root mean squared error is of the order of $n^{-1/3}$. A precise version of this result in the context of density estimation is given in Diaconis and Freedman (1981). Thus, if we consider the framework of regression in a counting process setting, then by analogy, one does not expect to be able to do as well with histogram-type estimators as with estimators formed by kernel methods. In fact, in their study of histogram versions of Beran's estimators in a counting process setting, McKeague and Utikal (1990) obtain asymptotic results of the form (3.2) and (3.3) with $h_n = o(n^{-1/2})$, for an overall rate of convergence just slightly under $n^{-1/4}$.

5 Illustration on Survival in Diabetics Data

Here 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 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 at the general practitioner and, when relevant, hospital. 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 the potential risk factors. Here we shall focus on the effect of age at diabetes onset on the duration of disease. The date of onset of diabetes is defined to be the first time the physician established the diagnosis.

We first note that 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*, $\delta_i = I(X_i \leq C_i)$ (the indicator function saying that a death is observed if $\delta_i = 1$), 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 $T_i \geq X_i$, i.e. if patient i died before July 1 1973. In addition, the age Z_i at onset of diabetes was recorded for each case i.

Since 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

$$Y_i^{(n)}(t) = I(T_i < t \le \min\{X_i, C_i\}).$$

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 $\lambda_i^{(n)}(t) = Y_i^{(n)}(t)h(t \mid Z_i)$, where

$$h(t \mid z) \equiv \lim_{\Delta t \searrow 0} \frac{1}{\Delta t} P(t < X \le t + \Delta t \mid X > t;\ Z = z)$$

is the conditional hazard function of the survival time X, given that $Z = z$. (See, e.g. Section III.3 of Andersen et al.). Therefore this model falls into the counting process framework described in Section 2. We assume the same probability model for the male group.

Note that we are using date of first diagnosis of diabetes as the time origin since we are interested in studying survival of a patient *after* diagnosis of diabetes. If one is interested in comparing mortalities between patients of the same age, i.e. use date of birth as time origin, then the survival time is defined to be $X' = X + Z$. It is not hard to see that no additional computation is needed for obtaining the survival function of X'. (For fixed z, the conditional survival function $P(X' > t \mid Z = z)$ is obtained by shifting the function $P(X > t \mid Z = z)$ to the right by z units). We also note that, given $Z = z$, a patient has no risk of dying before age z, and hence one should not try to compare $P(X' > t \mid Z = z_1)$ and $P(X' > t \mid Z = z_2)$ for $t \le \max\{z_1, z_2\}$.

The Fyn county diabetes data have been studied by many authors (see e.g. Green et al. (1985) and Andersen et al. (1985)). Most authors focus on comparing mortalities and thus use date of birth as time origin. For the reason described in the end of the previous paragraph, Andersen et al. (1985) analyzed a subset of this data set that included those who had diagnosis established before age 31 years, using models of the proportional hazards type. 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. But 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 Andersen et al. (1985) 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. 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 $S_n(t \mid z)$ (see (2.6)) of the conditional survival function $S(t \mid z) = P(X > t \mid Z = z)$ using the weight function given by (2.11) with $w(u) = \frac{1}{2}I(-1 \le u \le 1)$ and $k = 30\% \times$ *sample size*. Figures 1 and 2 give plots of $S_n(t \mid z)$ as a function of t with $z = 5, 10, 15, \ldots, 80$ for female and male groups respectively. Figure 3 compares 95% simultaneous confidence bands of the conditional survival function $S(t \mid z)$ 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 $S(t \mid z)$ 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 Andersen et al. (1985) 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 $S_n(t \mid z)$ 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, $\frac{S_n(10 \mid 30)}{S_n(10 \mid 45)} = \frac{0.98}{0.83} = 1.18$ compared to $\frac{S_n(30 \mid 30)}{S_n(30 \mid 45)} = \frac{0.64}{0.16} = 4.0$. 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.

Comparison of 95% confidence bands of conditional survival function $S(\cdot \mid z)$ between female and male patients (Figure 3) shows that the survival probabilities of female diabetics are consistently higher than those of male diabetics for the different levels of z. The same conclusion is reached by looking at 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 $\hat{A}(t, z_0)$ is defined by (2.8). Let us write

$$\hat{A}(t, z_0) = \int_0^t \tilde{J}(s) \boldsymbol{c}(s)' d\boldsymbol{N}(s), \tag{6.1}$$

where $\tilde{J}(s) = I(\sum_{i=1}^n c_i(s, \boldsymbol{z}_0) \neq 0)$. 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 $\boldsymbol{h}(s) = (h_1(s, \boldsymbol{z}_0), \ldots, h_n(s, \boldsymbol{z}_0))'$ by

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

and also define

$$R^{(n)}(t, z_0) = \int_0^t J(s) \boldsymbol{h}(s)' d\boldsymbol{N}(s).$$

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

$$A_n(t, \boldsymbol{z}_0) = \hat{A}(t, z_0) + R^{(n)}(t, z_0) + \int_0^t (J(s) - \tilde{J}(s)) \boldsymbol{c}(s)' d\boldsymbol{N}(s). \tag{6.3}$$

Under Assumption (A1), the probability that $\tilde{J}(s)$ is equal to $J(s)$ for all $s \in [0, t]$ tends to 1 as $n \to \infty$, so that for the asymptotics, it is immaterial whether we use $\tilde{J}$ or J. The term $R^{(n)}(t,z_0)$ represents the effect of doing a "local linear fit" instead of doing a "local average".

To obtain (6.3), we recall definitions (2.5), (2.4), the definition of $\boldsymbol{U}(s)$ following (2.3), and write

$$\begin{aligned}
A_n(t,z_0) &= (1,\boldsymbol{z}_0')\boldsymbol{B}_n(t,\boldsymbol{z}_0) \\
&= \int_0^t J(s)(1,\boldsymbol{z}_0')\left(\boldsymbol{U}(s)'\boldsymbol{W}(s,\boldsymbol{z}_0)\boldsymbol{U}(s)\right)^{-1}\boldsymbol{U}(s)'\boldsymbol{W}(s,\boldsymbol{z}_0)d\boldsymbol{N}(s) \\
&= \int_0^t J(s)(1,\boldsymbol{z}_0')\left(\begin{pmatrix} \boldsymbol{1}' \\ \boldsymbol{Z}^*(s)' \end{pmatrix}\boldsymbol{Y}(s)'\boldsymbol{W}(s,\boldsymbol{z}_0)\boldsymbol{Y}(s)(\boldsymbol{1},\boldsymbol{Z}^*(s))\right)^{-1}\begin{pmatrix} \boldsymbol{1}' \\ \boldsymbol{Z}^*(s)' \end{pmatrix} \\
&\qquad \times \boldsymbol{Y}(s)'\boldsymbol{W}(s,\boldsymbol{z}_0)d\boldsymbol{N}(s) \\
&= \int_0^t J(s)(1,\boldsymbol{z}_0')\left(\begin{pmatrix} \boldsymbol{1}' \\ \boldsymbol{Z}^*(s)' \end{pmatrix}C(s)(\boldsymbol{1},\boldsymbol{Z}^*(s))\right)^{-1}\begin{pmatrix} \boldsymbol{1}' \\ \boldsymbol{Z}^*(s)' \end{pmatrix}C(s)d\boldsymbol{N}(s) \qquad (6.4) \\
&= \int_0^t J(s)(1,\boldsymbol{z}_0')\begin{pmatrix} 1 & \boldsymbol{c}(s)'\boldsymbol{Z}^*(s) \\ \boldsymbol{Z}^*(s)'\boldsymbol{c}(s) & \boldsymbol{Z}^*(s)'C(s)\boldsymbol{Z}^*(s) \end{pmatrix}^{-1}\begin{pmatrix} \boldsymbol{c}(s)' \\ \boldsymbol{Z}^*(s)'C(s) \end{pmatrix}d\boldsymbol{N}(s) \\
&= \int_0^t J(s)\boldsymbol{c}(s)'d\boldsymbol{N}(s) + \int_0^t J(s)\boldsymbol{h}(s)'d\boldsymbol{N}(s) \\
&= \hat{A}(t,z_0) + \int_0^t (J(s)-\tilde{J}(s))\boldsymbol{c}(s)'d\boldsymbol{N}(s) + R^{(n)}(t,z_0).
\end{aligned}$$

To see the sixth equality in (6.4) consider first the case $p=1$. In this case (suppressing the argument s), the matrix inverse in the fifth line of (6.4), when it is a genuine inverse, may be written as

$$\begin{aligned}
\begin{pmatrix} 1 & \boldsymbol{c}'\boldsymbol{Z}^* \\ \boldsymbol{Z}^{*\prime}\boldsymbol{c} & \boldsymbol{Z}^{*\prime}C\boldsymbol{Z}^* \end{pmatrix}^{-1} &= \frac{1}{\boldsymbol{Z}^{*\prime}C\boldsymbol{Z}^* - \boldsymbol{Z}^{*\prime}\boldsymbol{c}\boldsymbol{c}'\boldsymbol{Z}^*}\begin{pmatrix} \boldsymbol{Z}^{*\prime}C\boldsymbol{Z}^* & -\boldsymbol{Z}^{*\prime}\boldsymbol{c} \\ -\boldsymbol{c}'\boldsymbol{Z}^* & 1 \end{pmatrix} \\
&= \frac{1}{\boldsymbol{Z}^{*\prime}P\boldsymbol{Z}^*}\begin{pmatrix} \boldsymbol{Z}^{*\prime}C\boldsymbol{Z}^* & -\boldsymbol{Z}^{*\prime}\boldsymbol{c} \\ -\boldsymbol{c}'\boldsymbol{Z}^* & 1 \end{pmatrix}
\end{aligned}$$

Moreover, this inverse is a genuine inverse if and only if $\boldsymbol{Z}^*(s)'P(s)\boldsymbol{Z}^*(s)$ is invertible (since $p=1$ this is simply the condition that $\boldsymbol{Z}^*(s)'P(s)\boldsymbol{Z}^*(s) \neq 0$), so that $J(s) = I(\boldsymbol{Z}^*(s)'P(s)\boldsymbol{Z}^*(s)$ is invertible). We also have $J_1(s) \le J(s)$ from the definitions of $J_1(s)$ and $J(s)$. To see the sixth equality in (6.4) for the case $p \ge 2$, we use the formula

$$\begin{pmatrix} A & B \\ B' & D \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} + A^{-1}BE^{-1}B'A^{-1} & -A^{-1}BE^{-1} \\ -E^{-1}B'A^{-1} & E^{-1} \end{pmatrix}$$

where $E = D - B'A^{-1}B$. See e.g. Problem 2.7 on page 33 of Rao (1973).

To prove the first assertion of Theorem 1, we shall show that

(I) $\sqrt{na_n}\left(\hat{A}(\cdot,z_0) - A(\cdot,z_0)\right) \xrightarrow{d} U$ under (A1)–(A5), where U is defined in the statement of Theorem 1; and

(II) $\sqrt{na_n}\,\sup_{t\in[0,1]}\left|R^{(n)}(t,z_0)\right| \xrightarrow{P} 0$ under (A1)–(A6).

We first prove (I). Note that

$$\sqrt{na_n}\Big(\int_0^t J(s)\boldsymbol{c}(s)'d\boldsymbol{N}(s) - A(t,\boldsymbol{z}_0)\Big) = \sqrt{na_n}\Big(\int_0^t J(s)\sum_{i=1}^n c_i(s,\boldsymbol{z}_0)dN_i(s) - \int_0^t \alpha(s,\boldsymbol{z}_0)ds\Big)$$
$$= X_1^{(n)}(t,\boldsymbol{z}_0) - X_2^{(n)}(t,\boldsymbol{z}_0) + X_3^{(n)}(t,\boldsymbol{z}_0)$$

where

$$\begin{cases} X_1^{(n)}(t,\boldsymbol{z}_0) &=& \sqrt{na_n}\int_0^t J(s)\sum_{i=1}^n c_i(s,\boldsymbol{z}_0)dM_i(s) \\ X_2^{(n)}(t,\boldsymbol{z}_0) &=& \sqrt{na_n}\int_0^t (1-J(s))\alpha(s,\boldsymbol{z}_0)ds \\ X_3^{(n)}(t,\boldsymbol{z}_0) &=& \sqrt{na_n}\int_0^t J(s)\sum_{i=1}^n c_i(s,\boldsymbol{z}_0)\Big(\alpha(s,\boldsymbol{Z}_i(s)) - \alpha(s,\boldsymbol{z}_0)\Big)ds \end{cases}$$

and $M_i(\cdot) = N_i(\cdot) - \int_0^{\cdot} Y_i(s)\alpha(s,\boldsymbol{Z}_i(s))ds$ $(1 \le i \le n)$ are orthogonal locally square integrable martingales. By the version of Rebolledo's martingale central limit theorem stated as Theorem I.2 in Andersen and Gill (1982), we have

$$X_1^{(n)}(\cdot,\boldsymbol{z}_0) = \int_0^{\cdot} J(s)\sum_{i=1}^n \sqrt{na_n}c_i(s,\boldsymbol{z}_0)dM_i(s) \xrightarrow{d} U(\cdot,\boldsymbol{z}_0) \quad \text{in } D[0,1] \tag{6.5}$$

if following two conditions hold.

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

$$\begin{aligned} \langle X_1^{(n)}, X_1^{(n)}\rangle(t) &= \int_0^t \sum_{i=1}^n \Big(J(s)\sqrt{na_n}c_i(s,\boldsymbol{z}_0)\Big)^2 Y_i(s)\alpha(s,\boldsymbol{Z}_i(s))ds \\ &\xrightarrow{P} \int_0^t g_0(s,\boldsymbol{z}_0)\alpha(s,\boldsymbol{z}_0)ds; \end{aligned}$$

(ii) (Lindeberg Condition.) For each $\epsilon > 0$,

$$\int_0^1 \sum_{i=1}^n \Big(J(s)\sqrt{na_n}c_i(s,\boldsymbol{z}_0)\Big)^2 Y_i(s)\alpha(s,\boldsymbol{Z}_i(s))I\Big(J(s)\sqrt{na_n}c_i(s,\boldsymbol{z}_0) > \epsilon\Big)ds \xrightarrow{P} 0.$$

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

$$\sum_{i=1}^n d_i|x_iy_i| = \sum_{i=1}^n \Big(\sqrt{d_i}|x_i|\Big)\Big(\sqrt{d_i}|y_i|\Big) \le \Big(\sum_{i=1}^n d_ix_i^2\Big)^{1/2}\Big(\sum_{i=1}^n d_iy_i^2\Big)^{1/2}$$

pointwise, by the Schwarz inequality (we have omitted the argument s). This implies that

$$\int\sum_{i=1}^n d_i|x_iy_i| \le \int\Big(\sum_{i=1}^n d_ix_i^2\Big)^{1/2}\Big(\sum_{i=1}^n d_iy_i^2\Big)^{1/2} \le \Big(\int\sum_{i=1}^n d_ix_i^2\Big)^{1/2}\Big(\int\sum_{i=1}^n d_iy_i^2\Big)^{1/2}, \tag{6.6}$$

and in particular, if the functions are defined on $[0,1]$ and $\sum_{i=1}^n d_i \equiv 1$, then

$$\int_0^1 \sum_{i=1}^n d_i |x_i| \le \Big(\int_0^1 \sum_{i=1}^n d_i x_i^2 \Big)^{1/2}. \tag{6.7}$$

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

$$\begin{aligned}
&\Big|\langle X_1^{(n)}, X_1^{(n)}\rangle(t) - \int_0^t g_0(s,z_0)\alpha(s,z_0)ds\Big| \\
&\quad \le \; na_n \int_0^1 J(s) \sum_{i=1}^n c_i^2(s,z_0) Y_i(s) \Big|\alpha(s, Z_i(s)) - \alpha(s,z_0)\Big| ds \\
&\qquad + \int_0^1 J(s) \Big| \sum_{i=1}^n na_n c_i^2(s,z_0) Y_i(s)\alpha(s,z_0) - g_0(s,z_0)\alpha(s,z_0)\Big| ds \\
&\qquad + \int_0^t (1 - J(s)) g_0(s,z_0)\alpha(s,z_0) ds \\
&\quad \le \; na_n (\sup_s \max_i c_i(s,z_0)) \int_0^1 \sum_{i=1}^n c_i(s,z_0) \Big|\alpha(s,Z_i(s)) - \alpha(s,z_0)\Big| ds \\
&\qquad + B \int_0^1 \Big| \sum_{i=1}^n na_n c_i^2(s,z_0) - g_0(s,z_0)\Big| ds + \int_0^t (1 - J(s)) g_0(s,z_0)\alpha(s,z_0) ds \\
&\quad \le \; na_n (\sup_s \max_i c_i(s,z_0)) \Big\{ \int_0^1 \sum_{i=1}^n c_i(s,z_0) \Big(\alpha(s,Z_i(s)) - \alpha(s,z_0)\Big)^2 ds \Big\}^{\frac{1}{2}} \\
&\qquad + B \int_0^1 \Big| \sum_{i=1}^n na_n c_i^2(s,z_0) - g_0(s,z_0)\Big| ds + \int_0^t (1 - J_1(s)) g_0(s,z_0)\alpha(s,z_0) ds \\
&\quad \xrightarrow{P} 0
\end{aligned}$$

by (A5), (A4), (A2) and (A1), where the last inequality follows from (6.7). Hence (i) holds.

We now check the Lindeberg condition. For any $\epsilon > 0$, we have

$$\begin{aligned}
&\int_0^1 \sum_{i=1}^n \Big(J(s)\sqrt{na_n} c_i(s,z_0)\Big)^2 Y_i(s) \alpha(s, Z_i(s)) I\Big(J(s)\sqrt{na_n} c_i(s,z_0) > \epsilon\Big) ds \\
&\qquad \le \int_0^1 na_n \sum_{i=1}^n c_i^2(s,z_0)\alpha(s,Z_i(s)) \Big(\frac{J(s)\sqrt{na_n}\, c_i(s,z_0)}{\epsilon} \Big)^\delta ds \\
&\qquad \le \frac{B}{\epsilon^\delta} (na_n)^{-\delta/2} \int_0^1 (na_n)^{1+\delta} \sum_{i=1}^n c_i^{2+\delta}(s,z_0) ds \\
&\qquad \xrightarrow{P} 0
\end{aligned}$$

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

It follows directly from (A1) that

$$\sup_{t \in [0,1]} |X_2^{(n)}(t, z_0)| \le \sqrt{na_n} \int_0^1 (1 - J_1(s))\alpha(s,z_0) ds \xrightarrow{P} 0. \tag{6.8}$$

Moreover, (A3) implies that

$$\sup_{t \in [0,1]} |X_3^{(n)}(t, z_0)| \xrightarrow{P} 0. \tag{6.9}$$

Therefore (I) follows from (6.5), (6.8), and (6.9).

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

$$\sqrt{na_n}R^{(n)}(t, \boldsymbol{z}_0) = \sqrt{na_n}\int_0^t J(s)\boldsymbol{h}(s)'d\boldsymbol{N}(s) = R_1^{(n)}(t, \boldsymbol{z}_0) + R_2^{(n)}(t, \boldsymbol{z}_0) + R_3^{(n)}(t, \boldsymbol{z}_0),$$

where

$$\begin{cases} R_1^{(n)}(t, \boldsymbol{z}_0) &= \sqrt{na_n}\int_0^t J_1(s)\boldsymbol{h}(s)'\boldsymbol{\alpha}_n(s)ds \\ R_2^{(n)}(t, \boldsymbol{z}_0) &= \sqrt{na_n}\int_0^t J_1(s)\boldsymbol{h}(s)'d\boldsymbol{M}(s) \\ R_3^{(n)}(t, \boldsymbol{z}_0) &= \sqrt{na_n}\int_0^t (J(s) - J_1(s))\boldsymbol{h}(s)'d\boldsymbol{N}(s) \end{cases}$$

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

$$\begin{aligned} |R_1^{(n)}(t, \boldsymbol{z}_0)| &= \left|\sqrt{na_n}\int_0^t J_1(s)\boldsymbol{h}(s)'\Big(\boldsymbol{\alpha}_n(s) - 1\alpha(s, \boldsymbol{z}_0)\Big)ds\right| \\ &\le \sqrt{na_n}\int_0^1 \Big|\Big\{J_1(s)\Big(\boldsymbol{z}_0 - \boldsymbol{Z}^*(s)'\boldsymbol{c}(s)\Big)'\Big(\boldsymbol{Z}^*(s)'P(s)\boldsymbol{Z}^*(s)\Big)^{-1}\boldsymbol{Z}^*(s)'P(s)^{\frac{1}{2}}\Big\} \\ &\qquad \times \Big\{P(s)^{\frac{1}{2}}\Big(\boldsymbol{\alpha}_n(s) - 1\alpha(s, \boldsymbol{z}_0)\Big)\Big\}\Big|ds \\ &\le \sqrt{na_n}\Big(\int_0^1 \Big\|J_1(s)P(s)^{\frac{1}{2}}\boldsymbol{Z}^*(s)\Big(\boldsymbol{Z}^*(s)'P(s)\boldsymbol{Z}^*(s)\Big)^{-1}\Big(\boldsymbol{z}_0 - \boldsymbol{Z}^*(s)'\boldsymbol{c}(s)\Big)\Big\|^2 ds\Big)^{\frac{1}{2}} \\ &\qquad \times \Big(\int_0^1 \Big\|P^{\frac{1}{2}}(s)\Big(\boldsymbol{\alpha}_n(s) - 1\alpha(s, \boldsymbol{z}_0)\Big)\Big\|^2 ds\Big)^{\frac{1}{2}} \\ &\le \Big(\sqrt{na_n}\int_0^1 J_1(s)\Big(\boldsymbol{z}_0 - \boldsymbol{Z}^*(s)'\boldsymbol{c}(s)\Big)'\Big(\boldsymbol{Z}^*(s)'P(s)\boldsymbol{Z}^*(s)\Big)^{-1}\Big(\boldsymbol{z}_0 - \boldsymbol{Z}^*(s)'\boldsymbol{c}(s)\Big)ds\Big)^{\frac{1}{2}} \\ &\qquad \times \Big(\sqrt{na_n}\int_0^1 \Big(\boldsymbol{\alpha}_n(s) - 1\alpha(s, \boldsymbol{z}_0)\Big)'C(s)\Big(\boldsymbol{\alpha}_n(s) - 1\alpha(s, \boldsymbol{z}_0)\Big)ds\Big)^{\frac{1}{2}}. \end{aligned} \tag{6.10}$$

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

We now consider $R_2^{(n)}(t, \boldsymbol{z}_0)$. Since for every i, $\sqrt{na_n}J_1(s)h_i(s)$ is a bounded predictable process, we see that $R_2^{(n)}(t, \boldsymbol{z}_0)$ 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 $\epsilon > 0$, we have

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

Now

$$\langle R_2^{(n)}, R_2^{(n)}\rangle(1) = na_n\int_0^1 J_1(s)\sum_{i=1}^n h_i^2(s, \boldsymbol{z}_0)Y_i(s)\alpha(s, Z_i(s))ds$$

$$
\begin{aligned}
&\le Bna_n \int_0^1 J_1(s)\boldsymbol{h}(s)'\boldsymbol{h}(s)ds \\
&= Bna_n \int_0^1 J_1(s)\big(\boldsymbol{z}_0 - \boldsymbol{Z}^*(s)'\boldsymbol{c}(s)\big)' \big(\boldsymbol{Z}^*(s)'P(s)\boldsymbol{Z}^*(s)\big)^{-1}\boldsymbol{Z}^*(s)'P(s) \\
&\qquad \times P(s)\boldsymbol{Z}^*(s)\big(\boldsymbol{Z}^*(s)'P(s)\boldsymbol{Z}^*(s)\big)^{-1}\big(\boldsymbol{z}_0 - \boldsymbol{Z}^*(s)'\boldsymbol{c}(s)\big)ds \qquad (6.11)\\
&\le B\big(na_n \max_i \sup_{s\in[0,1]} c_i(s,\boldsymbol{z}_0)\big) \int_0^1 J_1(s)\big(\boldsymbol{z}_0 - \boldsymbol{Z}^*(s)'\boldsymbol{c}(s)\big)' \\
&\qquad \times \big(\boldsymbol{Z}^*(s)'P(s)\boldsymbol{Z}^*(s)\big)^{-1}\big(\boldsymbol{z}_0 - \boldsymbol{Z}^*(s)'\boldsymbol{c}(s)\big)ds,
\end{aligned}
$$

where in the last step we have used the fact that for any vector $\boldsymbol{x}$

$$
\begin{aligned}
\boldsymbol{x}'P(s)P(s)\boldsymbol{x} &= (\boldsymbol{x}'P(s)^{\frac{1}{2}})P(s)(P(s)^{\frac{1}{2}}\boldsymbol{x}) \\
&\le (\boldsymbol{x}'P(s)^{\frac{1}{2}})C(s)(P(s)^{\frac{1}{2}}\boldsymbol{x}) \\
&\le \big(\max_i \sup_{s\in[0,1]} c_i(s,\boldsymbol{z}_0)\big)(\boldsymbol{x}'P(s)^{\frac{1}{2}})(P(s)^{\frac{1}{2}}\boldsymbol{x}) \\
&= \big(\max_i \sup_{s\in[0,1]} c_i(s,\boldsymbol{z}_0)\big)\boldsymbol{x}'P(s)\boldsymbol{x}.
\end{aligned}
$$

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

Finally, we see that $\sup_{t\in[0,1]}|R_3^{(n)}(t,\boldsymbol{z}_0)| \xrightarrow{P} 0$ from (A1) and the fact that $J_1(s) \le J(s)$. This proves (II), 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 δ-method (see Theorem 3 of Gill (1989)), in which we make the slight modification of replacing $\sqrt{n}$ by $\sqrt{na_n}$.

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

$$
\begin{cases}
\hat{\alpha}(t,\boldsymbol{z}_0) = \dfrac{1}{b_n}\displaystyle\int_0^1 K\Big(\frac{t-s}{b_n}\Big)\hat{A}(ds,z_0) \\
r^{(n)}(t,\boldsymbol{z}_0) = \dfrac{1}{b_n}\displaystyle\int_0^1 K\Big(\frac{t-s}{b_n}\Big)R^{(n)}(ds,z_0)
\end{cases}
\qquad (6.12)
$$

and write

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

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

(I) $\sqrt{na_nb_n}(\hat{\alpha}(t,\boldsymbol{z}_0) - \alpha(t,\boldsymbol{z}_0)) \xrightarrow{d} N(0,\sigma_t^2)$, where σ_t^2 is defined by (3.4); and

(II) $\sqrt{na_nb_n}r^{(n)}(t,\boldsymbol{z}_0) \xrightarrow{P} 0$.

Theorem 2 then follows immediately.

We first prove (I). We have

$$\begin{aligned}\hat{\alpha}(t,z_0) &= \frac{1}{b_n}\int_0^1 J(s)K\Big(\frac{t-s}{b_n}\Big)\sum_{i=1}^n c_i(s,z_0)dM_i(s) - \zeta_n(t,z_0)\\ &\quad + \frac{1}{b_n}\int_0^1 J(s)K\Big(\frac{t-s}{b_n}\Big)\sum_{i=1}^n c_i(s,z_0)\alpha(s,Z_i(s))ds.\end{aligned}$$

For large n we have

$$\begin{aligned}\alpha(t,z_0) &= \Big\{\frac{1}{b_n}\int_0^1 K\Big(\frac{t-s}{b_n}\Big)ds\Big\}\alpha(t,z_0)\\ &= \frac{1}{b_n}\int_0^1 J(s)K\Big(\frac{t-s}{b_n}\Big)\sum_{i=1}^n c_i(s,z_0)\alpha(t,z_0)ds + \frac{1}{b_n}\int_0^1 (1-J(s))K\Big(\frac{t-s}{b_n}\Big)\alpha(t,z_0)ds,\end{aligned}$$

so that

$$\sqrt{na_nb_n}\Big(\hat{\alpha}(t,z_0) - \alpha(t,z_0)\Big) = X^{(n)}(1) + I_{n1} - I_{n2} - \sqrt{na_nb_n}\zeta_n(t,z_0),$$

where

$$\begin{cases} X^{(n)}(\tau) &= (na_n/b_n)^{\frac{1}{2}}\int_0^\tau J(s)K\Big(\frac{t-s}{b_n}\Big)\sum_{i=1}^n c_i(s,z_0)dM_i(s), \qquad \tau\in[0,1]\\ I_{n1} &= (na_n/b_n)^{\frac{1}{2}}\int_0^1 J(s)K\Big(\frac{t-s}{b_n}\Big)\sum_{i=1}^n c_i(s,z_0)\Big(\alpha(s,Z_i(s)) - \alpha(t,z_0)\Big)ds\\ I_{n2} &= (na_n/b_n)^{\frac{1}{2}}\Big\{\int_0^1 (1-J(s))K\Big(\frac{t-s}{b_n}\Big)ds\Big\}\alpha(t,z_0)\end{cases}$$

We shall show that

$$I_{nj}\xrightarrow{P} 0 \quad \text{for } j=1,2; \qquad \text{and} \qquad X^{(n)}(1)\xrightarrow{d} N(0,\sigma_t^2). \tag{6.13}$$

Consider I_{n1}. Let

$$\begin{cases} V_{n1} &= (na_n/b_n)^{\frac{1}{2}}\int_0^1 J(s)K\Big(\frac{t-s}{b_n}\Big)\Big|\sum_{i=1}^n c_i(s,z_0)\Big(\alpha(s,Z_i(s)) - \alpha(s,z_0)\Big)\Big|ds\\ V_{n2} &= (na_n/b_n)^{\frac{1}{2}}\Big|\int_0^1 J(s)K\Big(\frac{t-s}{b_n}\Big)\Big(\alpha(s,z_0) - \alpha(t,z_0)\Big)ds\Big|\end{cases}$$

and note that

$$|I_{n1}| \le V_{n1} + V_{n2}. \tag{6.14}$$

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

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

for every nonnegative function f. Hence, in (6.14), $V_{n1} \xrightarrow{P} 0$ by (B3). Recalling that $D = \sup_{s\in[0,1]} |\alpha_s''(s, z_0)|$, we have for large n

$$\begin{aligned} V_{n2} &\leq (na_n/b_n)^{\frac{1}{2}}\Big|\alpha_t'(t,z_0)\int_0^1 K\Big(\frac{t-s}{b_n}\Big)(s-t)ds\Big| + D(na_n/b_n)^{\frac{1}{2}}\int_0^1 K\Big(\frac{t-s}{b_n}\Big)(s-t)^2ds \\ &\leq (na_n/b_n)^{\frac{1}{2}}\Big|\alpha_t'(t,z_0)b_n^2\int_{-1}^1 uK(u)du\Big| + MD(na_n/b_n)^{\frac{1}{2}}\int_{t-b_n}^{t+b_n}(s-t)^2ds \\ &= 0 + O\Big(\sqrt{nb_n^{p+5}}\Big) \to 0. \end{aligned}$$

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

To that prove $X^{(n)}(1) \xrightarrow{d} N(0, \sigma_t^2)$, we apply Lemma A.4 of the Appendix to $X^{(n)}(\tau) = \int_0^\tau \sum_{i=1}^n H_i(s, z_0)dM_i(s)$, $\tau \in [0,1]$, where

$$H_i(s, z_0) = J(s)(na_n/b_n)^{\frac{1}{2}}K\Big(\frac{t-s}{b_n}\Big)c_i(s, z_0), \quad i = 1, \ldots, n.$$

Thus we need to check the following conditions.

(1°) $\langle X^{(n)}, X^{(n)}\rangle(1) \xrightarrow{P} \sigma_t^2$.

(2°) For each $\epsilon > 0$, $\int_0^1 \sum_{i=1}^n H_i^2(s, z_0)I(H_i(s, z_0) > \epsilon)Y_i(s)\alpha(s, Z_i(s))ds \xrightarrow{P} 0$.

To verify (1°), we write

$$\begin{aligned} \langle X^{(n)}, X^{(n)}\rangle(1) &= \int_0^1 J(s)\frac{na_n}{b_n}K^2\Big(\frac{t-s}{b_n}\Big)\sum_{i=1}^n c_i^2(s, z_0)\alpha(s, Z_i(s))ds \\ &= \int_0^1 J(s)\frac{na_n}{b_n}K^2\Big(\frac{t-s}{b_n}\Big)\sum_{i=1}^n c_i^2(s, z_0)\Big(\alpha(s, Z_i(s)) - \alpha(s, z_0)\Big)ds \\ &\quad + \int_0^1 J(s)\frac{1}{b_n}K^2\Big(\frac{t-s}{b_n}\Big)\Big(na_n\sum_{i=1}^n c_i^2(s, z_0) - g_0(s, z_0)\Big)\alpha(s, z_0)ds \\ &\quad + \int_0^1 J(s)\frac{1}{b_n}K^2\Big(\frac{t-s}{b_n}\Big)g_0(s, z_0)\alpha(s, z_0)ds \\ &\equiv I_1 + I_2 + I_3. \end{aligned}$$

In the above expression,

$$\begin{aligned} |I_1| &\leq \int_0^1 \frac{na_n}{b_n}K^2\Big(\frac{t-s}{b_n}\Big)\sum_{i=1}^n c_i^2(s, z_0)|\alpha(s, Z_i(s)) - \alpha(s, z_0)|ds \\ &\leq M^2\frac{na_n}{b_n}\int_{t-b_n}^{t+b_n}\sum_{i=1}^n c_i^2(s, z_0)|\alpha(s, Z_i(s)) - \alpha(s, z_0)|ds \\ &\leq \frac{M^2}{b_n}\Big(na_n \max_i \sup_s c_i(s, z_0)\Big)\int_{t-b_n}^{t+b_n}\sum_{i=1}^n c_i(s, z_0)|\alpha(s, Z_i(s)) - \alpha(s, z_0)|ds \\ &= O_p(1/b_n)\Big\{\int_{t-b_n}^{t+b_n}\sum_{i=1}^n c_i(s, z_0)\Big(\alpha(s, Z_i(s)) - \alpha(s, z_0)\Big)^2ds\Big\}^{1/2}\Big\{\int_{t-b_n}^{t+b_n}ds\Big\}^{1/2} \\ &= O_p(1/b_n)\Big(O_p(b_n^3)\Big)^{1/2}(2b_n)^{1/2} = O_p(b_n) \end{aligned}$$

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

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

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

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

We now prove (2°). For the $\delta>0$ in (B2), we have

$$\begin{aligned}
&\int_0^1\sum_{i=1}^n H_i^2(s,\boldsymbol{z}_0)I(H_i(s,\boldsymbol{z}_0)>\epsilon)Y_i(s)\alpha(s,\boldsymbol{Z}_i(s))ds\\
&\le \frac{1}{\epsilon^\delta}\int_0^1\sum_{i=1}^n H_i^{2+\delta}(s,\boldsymbol{z}_0)Y_i(s)\alpha(s,\boldsymbol{Z}_i(s))ds\\
&= \frac{1}{\epsilon^\delta}\int_0^1 J(s)\Big((na_n/b_n)^{\frac12}\Big)^{2+\delta}K^{2+\delta}\Big(\frac{t-s}{b_n}\Big)\sum_{i=1}^n c_i^{2+\delta}(s,\boldsymbol{z}_0)Y_i(s)\alpha(s,\boldsymbol{Z}_i(s))ds\\
&\le \frac{BM^{2+\delta}}{\epsilon^\delta}\Big((na_n)^{-\frac{\delta}{2}}b_n^{-1-\delta/2}\Big)(na_n)^{1+\delta}\int_{t-b_n}^{t+b_n}\sum_{i=1}^n c_i^{2+\delta}(s,\boldsymbol{z}_0)ds\\
&= O_p\Big((na_n)^{-\frac{\delta}{2}}b_n^{-1-\delta/2}\Big)O_p(b_n)\\
&= O_p\Big((nb_n^{p+1})^{-\frac{\delta}{2}}\Big)
\end{aligned}$$

where the second equality follows from (B2) and the assumption that $g(s,\boldsymbol{z}_0)$ is continuous in s. The last term converges to 0 in probability since $nb_n^{p+1}\to\infty$. Therefore (6.13) holds and we have proved Part (I).

We now prove (II). For $r^{(n)}(t,\boldsymbol{z}_0)$ defined by (6.12), we have

$$\begin{aligned}
\sqrt{na_nb_n}\,r^{(n)}(t,\boldsymbol{z}_0) &= (na_n/b_n)^{\frac12}\int_0^1 J(s)K\Big(\frac{t-s}{b_n}\Big)\boldsymbol{h}(s)'d\boldsymbol{N}(s)\\
&= r_1^{(n)}+r_2^{(n)}(1)+r_3^{(n)}
\end{aligned}$$

where

$$\begin{cases}
r_1^{(n)} &= (na_n/b_n)^{\frac12}\int_0^1 J_1(s)K\Big(\frac{t-s}{b_n}\Big)\boldsymbol{h}(s)'\boldsymbol{\alpha}_n(s)ds,\\
r_2^{(n)}(\tau) &= (na_n/b_n)^{\frac12}\int_0^\tau J_1(s)K\Big(\frac{t-s}{b_n}\Big)\boldsymbol{h}(s)'d\boldsymbol{M}(s), \quad \tau\in[0,1],\\
r_3^{(n)} &= (na_n/b_n)^{\frac12}\int_0^1 (J(s)-J_1(s))K\Big(\frac{t-s}{b_n}\Big)asdf\boldsymbol{h}(s)'d\boldsymbol{N}(s),
\end{cases}$$

and we recall that $\boldsymbol{\alpha}_n(s)=(\alpha(s,\boldsymbol{Z}_1(s)),\ldots,\alpha(s,\boldsymbol{Z}_n(s)))'$ and $\boldsymbol{h}(s)$ 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 $s \in \mathcal{T}_{\boldsymbol{z}_0}$ we have $\boldsymbol{h}(s)'1 = 0$, we see that

$$
\begin{aligned}
|r_1^{(n)}| &= \left|\sqrt{\tfrac{na_n}{b_n}} \int_0^1 J_1(s) K\Big(\frac{t-s}{b_n}\Big) \boldsymbol{h}(s)'\Big(\boldsymbol{\alpha}_n(s) - 1\alpha(s, \boldsymbol{z}_0)\Big) ds\right| \\
&\leq M\sqrt{\tfrac{na_n}{b_n}} \int_{t-b_n}^{t+b_n} J_1(s) \Big|\boldsymbol{h}(s)'\Big(\boldsymbol{\alpha}_n(s) - 1\alpha(s, \boldsymbol{z}_0)\Big)\Big| ds \\
&\leq M\sqrt{\tfrac{na_n}{b_n}} \Big(\int_{t-b_n}^{t+b_n} J_1(s)\Big(\boldsymbol{z}_0 - \boldsymbol{Z}^*(s)'\boldsymbol{c}(s)\Big)'\Big(\boldsymbol{Z}^*(s)'P(s)\boldsymbol{Z}^*(s)\Big)^{-1}\Big(\boldsymbol{z}_0 - \boldsymbol{Z}^*(s)'\boldsymbol{c}(s)\Big) ds\Big)^{\frac{1}{2}} \\
&\quad \times \Big(\int_{t-b_n}^{t+b_n} \Big(\boldsymbol{\alpha}_n(s) - 1\alpha(s, \boldsymbol{z}_0)\Big)' C(s) \Big(\boldsymbol{\alpha}_n(s) - 1\alpha(s, \boldsymbol{z}_0)\Big) ds\Big)^{\frac{1}{2}} \\
&= O_p\Big(\sqrt{\tfrac{na_n}{b_n}}\Big)\Big(o_p\Big(\tfrac{1}{na_nb_n^2}\Big)\Big)^{1/2}\Big(O_p(b_n^3)\Big)^{\frac{1}{2}} \\
&= o_p(1) \xrightarrow{P} 0
\end{aligned}
$$

where to obtain the second inequality we reason as we did to obtain (6.10), and to obtain the second equality we use (B6) and (B4).

To prove that $r_2^{(n)}(1) \xrightarrow{P} 0$, we apply Lenglart's inequality to $r_2^{(n)}$: For every $\eta > 0$ and $\epsilon > 0$

$$
P\Big(\sup_\tau |r_2^{(n)}(\tau)| > \eta\Big) \leq \frac{\epsilon}{\eta^2} + P\Big(\langle r_2^{(n)}, r_2^{(n)}\rangle(1) > \epsilon\Big). \tag{6.15}
$$

We have

$$
\begin{aligned}
\langle r_2^{(n)}, r_2^{(n)}\rangle(1) &= \frac{na_n}{b_n}\int_0^1 J_1(s) K^2\Big(\frac{t-s}{b_n}\Big) \sum_{i=1}^n h_i^2(s, \boldsymbol{z}_0) Y_i(s) \alpha(s, Z_i(s)) ds \\
&\leq \frac{CM^2}{b_n} na_n \int_{t-b_n}^{t+b_n} J_1(s) \boldsymbol{h}(s)'\boldsymbol{h}(s) ds \\
&\leq \frac{CM^2}{b_n}\Big(na_n \max_i \sup_{s\in[0,1]} c_i(s, \boldsymbol{z}_0)\Big) \int_{t-b_n}^{t+b_n} J_1(s)\Big(\boldsymbol{z}_0 - \boldsymbol{Z}^*(s)'\boldsymbol{c}(s)\Big)' \\
&\quad \times \Big(\boldsymbol{Z}^*(s)'P(s)\boldsymbol{Z}^*(s)\Big)^{-1}\Big(\boldsymbol{z}_0 - \boldsymbol{Z}^*(s)'\boldsymbol{c}(s)\Big) ds \\
&= O_p(1/b_n) O_p(1) o_p(1/na_nb_n^2) \\
&= o_p(1/nb_n^{p+3}) \xrightarrow{P} 0.
\end{aligned} \tag{6.16}
$$

To obtain the second inequality in (6.16) we reason as we did to obtain (6.11). The second equality in (6.16) follows from (B5) and (B6), and the last assertion follows since $nb_n^{p+3} \to \infty$ by assumption. Therefore, $r_2^{(n)}(1) \xrightarrow{P} 0$ by (6.15) and (6.16). Finally, $r_3^{(n)} \xrightarrow{P} 0$ from (B1).

Therefore

$$
\sqrt{na_nb_n}\, r^{(n)} = r_1^{(n)} + r_2^{(n)}(1) + r_3^{(n)} \xrightarrow{P} 0.
$$

To prove Theorem 3, we shall first need to state 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* $(Y,\boldsymbol{X}),(Y,\boldsymbol{X}_1,),\ldots,(Y,\boldsymbol{X}_n,)$ *are i.i.d. random vectors taking values in* $\mathbb{R}\times\mathbb{R}^p$*, and that* $\boldsymbol{X}$ *has a continuous density function* $g(\boldsymbol{x})$*. For each* $\boldsymbol{x}\in\mathbb{R}^p$*, define* $m(\boldsymbol{x}) = E(Y\,|\,\boldsymbol{X}=\boldsymbol{x})$ *and* $m_n(\boldsymbol{x}) = \sum_{i=1}^n W_i(\boldsymbol{x})Y_i$*, where* $W_i(\boldsymbol{x}) = w\left(\frac{\boldsymbol{x}-\boldsymbol{X}_i}{R_n}\right)\Big/\sum_{j=1}^n w\left(\frac{\boldsymbol{x}-\boldsymbol{X}_j}{R_n}\right)$*,* R_n *is the Euclidean distance from* $\boldsymbol{x}$ *to the* k^{th} *closest of* $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n$ *and* $w(\cdot)$ *is a bounded density function on* $\mathbb{R}^p$ *satisfying* $||\boldsymbol{u}||^p w(\boldsymbol{u}) \to 0$ *as* $||\boldsymbol{u}||\to\infty$ *and (4.5). Let* $\boldsymbol{x}_0\in\mathbb{R}^p$ *such that* $g(\boldsymbol{x}_0)>0$*. Assume that* $m(\boldsymbol{x})$ *and* $Var(Y\,|\,\boldsymbol{X}=\boldsymbol{x})$ *exist in a neighborhood of* $\boldsymbol{x}_0$*. Assume further that* $m(\boldsymbol{x})$ *is continuous at* $\boldsymbol{x}_0$ *and* $Var(Y\,|\,\boldsymbol{X}=\boldsymbol{x})$ *is bounded in a neighborhood of* $\boldsymbol{x}_0$*. Then, if* $k_n\to\infty$ *and* $k_n/n\to 0$

$$m_n(\boldsymbol{x}_0) \xrightarrow{P} m(\boldsymbol{x}_0).$$

Proof This is Proposition 1 of Collomb (1980).

Proposition 6.2 *Let* $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n$ *be i.i.d.* $\mathbb{R}^p$*-valued random vectors with bounded density function* $g(\boldsymbol{x})$*. For each* $\boldsymbol{x}\in\mathbb{R}^p$*, define* $\hat{g}_n(\boldsymbol{x}) = \frac{1}{nR_n^p}\sum_{i=1}^n w\left(\frac{\boldsymbol{x}-\boldsymbol{X}_i}{R_n}\right)$ *where* R_n *is defined as in Proposition 6.1 and* $w(\cdot)$ *is a bounded density function on* $\mathbb{R}^p$ *satisfying (4.4), (4.5) and* $w(-\boldsymbol{u}) = w(\boldsymbol{u})$ *for all* $\boldsymbol{u}\in\mathbb{R}^p$*. Then, if* $k_n\to\infty$ *and* $k_n/n\to 0$*, we have*

$$\hat{g}_n(\boldsymbol{x}) \xrightarrow{P} g(\boldsymbol{x}) \quad \text{at every continuity point } \boldsymbol{x} \text{ of } g. \tag{6.17}$$

When $w(\boldsymbol{u}) = \frac{1}{\gamma(p)}I(||\boldsymbol{u}||\le 1)$ *with* $\gamma(p) = \int_{||\boldsymbol{u}||\le 1} d\boldsymbol{u} = 2\pi^{p/2}/(p\Gamma(p/2))$*, (6.17) is the statement*

$$\frac{k_n}{\gamma(p)nR_n^p} \xrightarrow{P} g(\boldsymbol{x}) \quad \text{at every continuity point } \boldsymbol{x} \text{ of } g. \tag{6.18}$$

Proof Let $\boldsymbol{x}\in\mathbb{R}^p$ and $\epsilon>0$. By Theorem 1.1 of Moore and Yackel (1977), there exist $\eta>0$ and a finite set of positive numbers $\alpha_1,\ldots,\alpha_M$ such that $|\hat{g}_n(\boldsymbol{x})-g(\boldsymbol{x})|>\epsilon$ implies that either $|f_n(\boldsymbol{x},\alpha_j)-g(\boldsymbol{x})|>\eta$ or $|g_n(\boldsymbol{x},\alpha_j)-g(\boldsymbol{x})|>\eta$ for at least one j in $\{1,\ldots,M\}$, where

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

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

$$P(|\hat{g}_n(\boldsymbol{x})-g(\boldsymbol{x})|>\epsilon) \le \sum_{j=1}^M P(|f_n(\boldsymbol{x},\alpha_j)-g(\boldsymbol{x})|>\eta) + \sum_{j=1}^M P(|g_n(\boldsymbol{x},\alpha_j)-g(\boldsymbol{x})|>\eta) \to 0,$$

where the convergence statement is a consequence of Theorem 3.1.2 of Rao (1983). Thus, $\hat{g}_n(\boldsymbol{x}) \xrightarrow{P} g(\boldsymbol{x})$.

The following result is needed to prove Theorem 3.

Lemma 6.1 *Under the conditions of Theorem 3, we have*

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

(2) $\left(n(k_n/n)^{\frac{p+1}{p}}\right)\sum_{i=1}^n c_i(s,\boldsymbol{z}_0)(\boldsymbol{Z}_i-\boldsymbol{z}_0)=o_p(1)$ *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 (1) of the lemma since the second part is proved in an identical way. Denote $\alpha_n=\left(n(k_n/n)^{\frac{p-4}{p}}\right)^{\frac{1}{4}}$. Then

$$\begin{aligned}\alpha_n\sum_{i=1}^n c_i(s,\boldsymbol{z}_0)(\boldsymbol{Z}_i-\boldsymbol{z}_0) &= \alpha_n\frac{\sum_{i=1}^n W_i(\boldsymbol{z}_0)Y_i(s)(\boldsymbol{Z}_i-\boldsymbol{z}_0)}{\sum_{j=1}^n W_j(\boldsymbol{z}_0)Y_j(s)}\\ &= \frac{\dfrac{\alpha_n}{nR_n^{p-1}}\displaystyle\sum_{i=1}^n w\Big(\frac{\boldsymbol{Z}_i-\boldsymbol{z}_0}{R_n}\Big)Y_i(s)\Big(\frac{\boldsymbol{Z}_i-\boldsymbol{z}_0}{R_n}\Big)}{\dfrac{1}{nR_n^p}\displaystyle\sum_{j=1}^n w\Big(\frac{\boldsymbol{Z}_j-\boldsymbol{z}_0}{R_n}\Big)\sum_{j=1}^n W_j(\boldsymbol{z}_0)Y_j(s)}.\end{aligned}\qquad(6.19)$$

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

$$\sup_{s\in[0,1]}\Big|\sum_{i=1}^n W_i(\boldsymbol{z}_0)Y_i(s)-H(s,\boldsymbol{z}_0)\Big|\xrightarrow{P}0,\qquad(6.20)$$

where $H(s,\boldsymbol{z}_0)$ is defined by (4.3). For each $i=1,\ldots,n$, 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, for each $s\in[0,1]$

$$\begin{cases}\Big|\displaystyle\sum_{i=1}^n W_i(\boldsymbol{z}_0)Y_{ia}(s)-H_a(s,\boldsymbol{z}_0)\Big|\xrightarrow{P}0\\ \Big|\displaystyle\sum_{i=1}^n W_i(\boldsymbol{z}_0)Y_{ia}(s+)-H_a(s+,\boldsymbol{z}_0)\Big|\xrightarrow{P}0\end{cases}$$

where $H_a(s,\boldsymbol{z}_0)=E(Y_{1a}(s)\,|\,\boldsymbol{Z}_1=\boldsymbol{z}_0)$ and we have used the fact that $E(Y_{1a}(s+)\,|\,\boldsymbol{Z}_1=\boldsymbol{z}_0)=E(\lim_{k\to\infty}Y(s+1/k)\,|\,\boldsymbol{Z}_1=\boldsymbol{z}_0)=\lim_{k\to\infty}E(Y(s+1/k)\,|\,\boldsymbol{Z}_1=\boldsymbol{z}_0)=H_a(s+,\boldsymbol{z}_0)$ by the Bounded Convergence Theorem. Thus, by Lemma A.3 stated in the Appendix,

$$\sup_{s\in[0,1]}\Big|\sum_{i=1}^n W_i(\boldsymbol{z}_0)Y_{ia}(s)-H_a(s,\boldsymbol{z}_0)\Big|\xrightarrow{P}0.$$

Similarly, we have

$$\sup_{s\in[0,1]}\Big|\sum_{i=1}^n W_i(\boldsymbol{z}_0)Y_{ib}(s)-H_b(s,\boldsymbol{z}_0)\Big|\xrightarrow{P}0$$

with $H_b(s,\boldsymbol{z}_0)=E(Y_{1b}(s)\,|\,\boldsymbol{Z}_1=\boldsymbol{z}_0)$. Therefore (6.20) holds since $H(s,\boldsymbol{z}_0)=H_a(s,\boldsymbol{z}_0)-H_b(s,\boldsymbol{z}_0)$.

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

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

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

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

Fix $l \in \{1, \ldots, p\}$. For each $s \in [0,1]$, denote by $\eta_n(s)$ the l^{th} component of the numerator of (6.19). Then since w vanishes outside the unit ball,

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

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

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

where the convergence follows from (6.18) and the assumption that $k_n \to \infty$. Similarly

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

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

$$\eta_n(s) - E(\eta_n(s) \,|\, R_n) \xrightarrow{P} 0 \quad \text{uniformly in } s \in [0,1]. \tag{6.21}$$

Moreover,

$$\begin{aligned}
\Big|E(\eta_n(s) \,|\, R_n)\Big| &= k_n\Big(\frac{\alpha_n}{nR_n^{p-1}}\Big)\Big|E\Big(w\Big(\frac{\tilde{\boldsymbol{Z}}_i - \boldsymbol{z}_0}{R_n}\Big)\tilde{Y}_i(s)\Big(\frac{\tilde{Z}_{il} - z_{0l}}{R_n}\Big) \,\Big|\, R_n\Big)\Big| \\
&= \frac{k_n\alpha_n}{nR_n^{p-1}}\Big|\int w\Big(\frac{\boldsymbol{z} - \boldsymbol{z}_0}{R_n}\Big)\Big(\frac{z_l - z_{0l}}{R_n}\Big)\frac{f(\boldsymbol{z}, s)}{G(R_n)} d\boldsymbol{z}\Big| \\
&= \frac{k_n\alpha_n R_n}{nG(R_n)}\Big|\int w(\boldsymbol{u})u_l f(\boldsymbol{z}_0 + R_n\boldsymbol{u}, s) d\boldsymbol{u}\Big| \\
&= \frac{k_n\alpha_n R_n}{nG(R_n)}\Big|\int w(\boldsymbol{u})u_l\Big(f(\boldsymbol{z}_0 + R_n\boldsymbol{u}, s) - f(\boldsymbol{z}_0, s)\Big) d\boldsymbol{u}\Big|
\end{aligned}$$

$$
\begin{aligned}
&\le M\frac{k_n\alpha_n R_n^2}{nG(R_n)}\int w(\boldsymbol{u})|u_l|\cdot||\boldsymbol{u}||d\boldsymbol{u} \quad \text{(by (4.2))}\\
&= O_p\Big(\frac{k_n\alpha_n R_n^2}{nR_n^p}\Big)\\
&= O_p\Big(\Big(\frac{k_n^{p+4}}{n^4}\Big)^{\frac{1}{4p}}\Big), \quad \text{uniformly in } s\in[0,1]
\end{aligned}
$$

where to obtain the sixth line we have used the fact that as $r\to 0$

$$
G(r)=\gamma(p)f(\boldsymbol{z}_0)r^p+\int_{||z-\boldsymbol{z}_0||\le r}(f(z)-f(\boldsymbol{z}_0))dz=\gamma(p)f(\boldsymbol{z}_0)r^p+o(r^p), \tag{6.22}
$$

and to obtain the last line we have used (6.18). Hence, if $k_n^{p+4}/n^4\to 0$,

$$
\eta_n(s)=\Big(\eta_n(s)-E(\eta_n(s)\,|\,R_n)\Big)+E(\eta_n(s)\,|\,R_n)\xrightarrow{P}0 \quad \text{uniformly in } s\in[0,1].
$$

Therefore

$$
\Big(n(k_n/n)^{\frac{p-4}{p}}\Big)^{\frac{1}{4}}\sum_{i=1}^n c_i(s,\boldsymbol{z}_0)(\boldsymbol{Z}_i-\boldsymbol{z}_0)=o_p(1).
$$

Proof of Theorem 3 Before checking Conditions (A) and (B), we first show that

$$
\frac{1}{R_n^2}\sum_i c_i(s,\boldsymbol{z}_0)(\boldsymbol{Z}_i-\boldsymbol{z}_0)(\boldsymbol{Z}_i-\boldsymbol{z}_0)'\xrightarrow{P}\operatorname{diag}\Big(\int u_1^2w(\boldsymbol{u})d\boldsymbol{u},\ldots,\int u_p^2w(\boldsymbol{u})d\boldsymbol{u}\Big) \tag{6.23}
$$

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

$$
\begin{aligned}
\frac{1}{R_n^2}\sum_{i=1}^n c_i(s,\boldsymbol{z}_0)(Z_{il}-z_{0l})^2 &= \sum_{i=1}^n W_i(\boldsymbol{z}_0)Y_i(s)\Big(\frac{Z_{il}-z_{0l}}{R_n}\Big)^2\Big/\sum_{j=1}^n W_j(\boldsymbol{z}_0)Y_j(s)\\
&= \frac{\frac{1}{nR_n^p}\sum_{i=1}^n w\Big(\frac{\boldsymbol{Z}_i-\boldsymbol{z}_0}{R_n}\Big)\Big(\frac{Z_{il}-z_{0l}}{R_n}\Big)^2Y_i(s)}{\frac{1}{nR_n^p}\sum_{j=1}^n w\Big(\frac{\boldsymbol{Z}_j-\boldsymbol{z}_0}{R_n}\Big)\sum_{j=1}^n W_j(\boldsymbol{z}_0)Y_j(s)}
\end{aligned} \tag{6.24}
$$

It follows from Proposition 6.2 and (6.20) that the denominator in (6.24) converges in probability to $f(\boldsymbol{z}_0)H(s,\boldsymbol{z}_0)=f(\boldsymbol{z}_0,s)$ uniformly in $s\in[0,1]$. We next show that the numerator in (6.24) converges in probability to $f(\boldsymbol{z}_0,s)\int w(\boldsymbol{u})u_l^2d\boldsymbol{u}$. Then it will immediately follow that

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

Denote by $\xi_n(s)$ the numerator in (6.24). Then since w vanishes outside the unit ball,

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

where $(\tilde{Z}_1, \tilde{Y}_1(s)), \ldots, (\tilde{Z}_{k_n}, \tilde{Y}_{k_n}(s))$ are the k_n points among $(Z_1, Y_1(s)), \ldots, (Z_n, Y_n(s))$ such that Z_i lies in the ball centered at z_0 and of radius R_n. Recall that given $R_n = r$, $(\tilde{Z}_1, \tilde{Y}_1(s)), \ldots, (\tilde{Z}_{k_n}, \tilde{Y}_{k_n}(s))$ are conditionally independent and identically distributed with the conditional subdensity function $P(\tilde{Z}_1 \le z,\, \tilde{Y}_1(s) = 1 \,|\, R_n = r)$ given by $\frac{f(z,s)}{G(r)}$ (see the third paragraph in the proof of Lemma 6.1). Hence,

$$\begin{aligned} E(\xi_n(s) \,|\, R_n) &= \frac{k_n}{nR_n^p} E\Big(w\Big(\frac{\tilde{Z}_i - z_0}{R_n}\Big) \tilde{Y}_i(s) \Big(\frac{\tilde{Z}_{il} - z_{0l}}{R_n}\Big)^2 \,\Big|\, R_n \Big) \\ &= \frac{k_n}{nR_n^p} \int w\Big(\frac{z - z_0}{R_n}\Big)\Big(\frac{z_l - z_{0l}}{R_n}\Big)^2 \frac{f(z,s)}{G(R_n)} dz \\ &= \frac{k_n}{nG(R_n)} \int w(u) u_l^2 f(z_0 + R_n u, s) du \\ &\xrightarrow{P} f(z_0, s) \int w(u) u_l^2 du \quad \text{uniformly in } s \in [0,1], \end{aligned}$$

where the convergence statement follows from (6.22), (6.18), and (4.2). Moreover, as in (6.21), we have

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

Therefore

$$\begin{aligned} \xi_n(s) &= (\xi_n(s) - E(\xi_n(s) \,|\, R_n)) + E(\xi_n(s) \,|\, R_n) \\ &\xrightarrow{P} f(z_0, s) \int w(u) u_l^2 du \quad \text{uniformly in } s \in [0,1]. \end{aligned} \tag{6.25}$$

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

$$\frac{1}{R_n^2} \sum_{i=1}^n c_i(s, z_0)(Z_{il} - z_{0l})(Z_{im} - z_{0m}) = \frac{\frac{1}{nR_n^p} \sum_{i=1}^n w\Big(\frac{Z_i - z_0}{R_n}\Big) Y_i(s) \Big(\frac{Z_{il} - z_{0l}}{R_n}\Big)\Big(\frac{Z_{im} - z_{0m}}{R_n}\Big)}{\frac{1}{nR_n^p} \sum_{j=1}^n w\Big(\frac{Z_j - z_0}{R_n}\Big) \sum_{j=1}^n W_j(z_0) Y_j(s)}. \tag{6.26}$$

It follows from Proposition 6.2 and (6.20) that the denominator of (6.26) converges in probability to $f(z_0)H(s, z_0)$ uniformly in $s \in [0,1]$. We need to show that the numerator of (6.26) 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 (6.19) in the proof of Lemma 6.1 (cf. also the proof of (6.25)). Thus the left hand side of (6.26) converges to zero in probability uniformly in $s \in [0,1]$. Therefore (6.23) holds.

Now we prove Part (1) of the theorem. We have

$$\begin{aligned} \frac{1}{R_n^2} Z^{*\prime} P Z^*(s) &= \frac{1}{R_n^2} \sum_{i=1}^n c_i(s, z_0)(Z_i - z_0)(Z_i - z_0)' - \frac{1}{R_n^2}\Big(\sum_{i=1}^n c_i(s, z_0)(Z_i - z_0)\Big) \\ &\quad \times \Big(\sum_{i=1}^n c_i(s, z_0)(Z_i - z_0)\Big)' \\ &= \frac{1}{R_n^2} \sum_{i=1}^n c_i(s, z_0)(Z_i - z_0)(Z_i - z_0)' - \frac{1}{R_n^2}\Big(o_p\Big(\Big(n(k_n/n)^{\frac{p-4}{p}}\Big)^{-\frac{1}{4}}\Big)\Big)^2 \end{aligned} \tag{6.27}$$

$$= \frac{1}{R_n^2}\sum_{i=1}^{n} c_i(s,\boldsymbol{z}_0)(\boldsymbol{Z}_i-\boldsymbol{z}_0)(\boldsymbol{Z}_i-\boldsymbol{z}_0)' - o_p\Big((k_n)^{-\frac{1}{2}}\Big)$$
$$\xrightarrow{P} \operatorname{diag}\Big(\int u_1^2 w(\boldsymbol{u})d\boldsymbol{u},\ldots,\int u_p^2 w(\boldsymbol{u})d\boldsymbol{u}\Big) \quad \text{uniformly in } s\in[0,1],$$

and each integral in the last line of (6.27) is positive. Here, the second equality follows from Part (1) of Lemma 6.1, the third equality from (6.18), and the convergence statement from (6.23). This implies (A1).

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

$$\sup_{s\in[0,1]}\Big|(na_n)^{1+\delta}\sum_{i=1}^{n} c_i^{2+\delta}(s,\boldsymbol{z}_0) - g_\delta(s,\boldsymbol{z}_0)\Big| \xrightarrow{P} 0. \tag{6.28}$$

Let $\tilde{w}(\cdot) = w^{2+\delta}(\cdot)/\int w^{2+\delta}(\boldsymbol{u})d\boldsymbol{u}$ and $\tilde{W}_i(\boldsymbol{z}_0) \equiv \tilde{w}\big(\frac{\boldsymbol{Z}_i-\boldsymbol{z}_0}{h_n}\big)\big/\sum_{j=1}^{n}\tilde{w}\big(\frac{\boldsymbol{Z}_j-\boldsymbol{z}_0}{h_n}\big)$. 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

$$\begin{aligned}
\Big(n\frac{k_n}{n}\Big)^{1+\delta}\sum_{i=1}^{n} W_i^{2+\delta}(\boldsymbol{z}_0)Y_i(s) &= k_n^{1+\delta}\sum_{i=1}^{n} w^{2+\delta}\Big(\frac{\boldsymbol{Z}_i-\boldsymbol{z}_0}{R_n}\Big)Y_i(s)\Big/\Big(\sum_{j=1}^{n} w\Big(\frac{\boldsymbol{Z}_j-\boldsymbol{z}_0}{R_n}\Big)\Big)^{2+\delta} \\
&= \Big(\frac{k_n}{nR_n^p}\Big)^{1+\delta}\Big[\sum_{i=1}^{n}\tilde{W}_i(\boldsymbol{z}_0)Y_i(s)\Big]\left[\frac{\frac{1}{nR_n^p}\sum_{j=1}^{n}\tilde{w}\Big(\frac{\boldsymbol{Z}_j-\boldsymbol{z}_0}{R_n}\Big)}{\Big(\frac{1}{nR_n^p}\sum_{j=1}^{n} w\Big(\frac{\boldsymbol{Z}_j-\boldsymbol{z}_0}{R_n}\Big)\Big)^{2+\delta}}\right]\int w^{2+\delta}(\boldsymbol{u})d\boldsymbol{u} \qquad (6.29)\\
&\xrightarrow{P} \big(\gamma(p)f(\boldsymbol{z}_0)\big)^{1+\delta}H(s,\boldsymbol{z}_0)\Big[\frac{f(\boldsymbol{z}_0)}{f^{2+\delta}(\boldsymbol{z}_0)}\Big]\int w^{2+\delta}(\boldsymbol{u})d\boldsymbol{u} \quad \text{uniformly in } s\in[0,1] \\
&= (\gamma(p))^{1+\delta}H(s,\boldsymbol{z}_0)\int w^{2+\delta}(\boldsymbol{u})d\boldsymbol{u},
\end{aligned}$$

where the convergence of the factor in the first set of brackets in the second line of (6.29) follows from the arguments leading to (6.20) and the convergence of the numerator and denominator in the second set of brackets follows by Proposition 6.2. Thus,

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

and hence (6.28) holds.

By regularity condition (R1),

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

where the last statement follows from Part (1) of Lemma 6.1 and (6.18). Thus

$$\sqrt{k_n}\int_0^1 \Big|\sum_{i=1}^n c_i(s,\boldsymbol{z}_0)\big(\alpha(s,\boldsymbol{Z}_i)-\alpha(s,\boldsymbol{z}_0)\big)\Big|ds = o_p\Big((k_n^{p+4}/n^4)^{\frac{1}{4p}}\Big) + O_p\Big((k_n^{p+4}/n^4)^{\frac{1}{2p}}\Big) \xrightarrow{P} 0$$

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

Since (R1) implies $|\alpha(s,\boldsymbol{z})-\alpha(s,\boldsymbol{z}_0)| \le ||\boldsymbol{z}-\boldsymbol{z}_0||\big(\sup_{s\in[0,1]}\big|\big|\frac{\partial\alpha}{\partial \boldsymbol{z}}(s,\boldsymbol{z}_0)\big|\big|+K_1||\boldsymbol{z}-\boldsymbol{z}_0||\big) = M_1||\boldsymbol{z}-\boldsymbol{z}_0||$ with $M_1 \equiv \sup_{s\in[0,1]}||\frac{\partial\alpha}{\partial \boldsymbol{z}}(s,\boldsymbol{z}_0)|| + K_1||\boldsymbol{z}-\boldsymbol{z}_0|| < \infty$,

$$\begin{aligned}
\sqrt{k_n}\Big|\sum_{i=1}^n c_i(s,\boldsymbol{z}_0)\big(\alpha(s,\boldsymbol{Z}_i)-\alpha(s,\boldsymbol{z}_0)\big)^2\Big| &\le \sqrt{k_n}M_1\sum_{i=1}^n c_i(s,\boldsymbol{z}_0)||\boldsymbol{Z}_i-\boldsymbol{z}_0||^2 \\
&= \sqrt{k_n}O_p(R_n^2) \\
&= O_p\Big((k_n^{p+4}/n^4)^{\frac{1}{2p}}\Big) \quad \text{uniformly in } s\in[0,1]
\end{aligned}$$

where the first equality follows from (6.23) and the second equality from (6.18). 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

$$k_n c_i(s,\boldsymbol{z}_0) = k_n\frac{W_i(\boldsymbol{z}_0)Y_i(s)}{\sum_{j=1}^n W_j(\boldsymbol{z}_0)Y_j(s)} \le \frac{k_n}{nR_n^p}\frac{\sup_{\boldsymbol{u}} w(\boldsymbol{u})}{\frac{1}{nR_n^p}\sum_{j=1}^n w\Big(\frac{\boldsymbol{Z}_j-\boldsymbol{z}_0}{R_n}\Big)\sum_{j=1}^n W_j(\boldsymbol{z}_0)Y_j(s)}. \tag{6.30}$$

Furthermore, by Proposition 6.2 and (6.20), the right-hand side of (6.30) converges in probability to $\gamma(p)\sup_{\boldsymbol{u}} w(\boldsymbol{u})\big/H(s,\boldsymbol{z}_0)$ uniformly in $s\in[0,1]$ and

$$\gamma(p)\sup_{\boldsymbol{u}} w(\boldsymbol{u})\big/H(s,\boldsymbol{z}_0) \le \gamma(p)\sup_{\boldsymbol{u}} w(\boldsymbol{u})\big/\Big(\inf_{s\in[0,1]} H(s,\boldsymbol{z}_0)\Big) < \infty.$$

Therefore

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

Finally, since

$$\begin{aligned}
&\sqrt{k_n}J_1(s)\big(\boldsymbol{Z}^*(s)'\boldsymbol{c}(s)-\boldsymbol{z}_0\big)'\big(\boldsymbol{Z}^*(s)'P(s)\boldsymbol{Z}^*(s)\big)^{-1}\big(\boldsymbol{Z}^*(s)'\boldsymbol{c}(s)-\boldsymbol{z}_0\big) \\
&\quad = J_1(s)\Big(\frac{k_n}{nR_n^p}\Big)^{2/p}\Big(\big(n(k_n/n)^{\frac{p-4}{p}}\big)^{\frac14}\sum_{i=1}^n c(s,\boldsymbol{z}_0)(\boldsymbol{Z}_i-\boldsymbol{z}_0)\Big)' \\
&\qquad \times\Big(\frac{1}{R_n^2}\boldsymbol{Z}^*(s)'P(s)\boldsymbol{Z}^*(s)\Big)^{-1}\Big(\big(n(k_n/n)^{\frac{p-4}{p}}\big)^{\frac14}\sum_{i=1}^n c(s,\boldsymbol{z}_0)(\boldsymbol{Z}_i-\boldsymbol{z}_0)\Big),
\end{aligned}$$

(A6) immediately follows from (6.18), Part (1) of Lemma 6.1, and (6.27).

The proof of Part (2) of the theorem uses Part (2) of Lemma 6.1 and is completely parallel to the proof of Part (1).

To prove Theorem 4, we use some well-known results from the areas of nonparametric regression and density estimation. These are stated in the next two propositions.

Proposition 6.3 *Assume that $(Y,\boldsymbol{X}),(Y_1,\boldsymbol{X}_1),\ldots,(Y_n,\boldsymbol{X}_n)$ are i.i.d. random vectors taking values in $\mathbb{R}\times\mathbb{R}^p$, that $E|Y|<\infty$, and that $\boldsymbol{X}$ has a density function $g(\boldsymbol{x})$. Let $m(\boldsymbol{x}) = E(Y\mid \boldsymbol{X}=\boldsymbol{x})$ and $m_n(\boldsymbol{x})=\sum_{i=1}^n W_i(\boldsymbol{x})Y_i$, where $W_i(\boldsymbol{x}) = w\left(\frac{\boldsymbol{x}-\boldsymbol{X}_i}{h_n}\right)\Big/\sum_{j=1}^n w\left(\frac{\boldsymbol{x}-\boldsymbol{X}_j}{h_n}\right)$ and $w(\cdot)$ is a density function satisfying (4.7). If $h_n\to 0$ and $nh_n^p\to\infty$, then*

$$m_n(\boldsymbol{x}) \xrightarrow{P} m(\boldsymbol{x}) \quad \textit{at every continuity point } \boldsymbol{x} \textit{ of } g \textit{ and } m.$$

Proof The proof is identical to that of Theorem 2.1 of Devroye (1981) except that one replaces the words "for almost all $\boldsymbol{x}(\mu)$" with "for every continuity point $\boldsymbol{x}$ of $g(\cdot)$ and $m(\cdot)$".

Proposition 6.4 *Assume that $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n$ are i.i.d. $\mathbb{R}^p$-valued random vectors with density function $g(\boldsymbol{x})$. Let $g_n(\boldsymbol{x}) = \frac{1}{nh_n^p}\sum_{j=1}^n w\left(\frac{\boldsymbol{x}-\boldsymbol{X}_j}{h_n}\right)$ where $w(\cdot)$ is a bounded density function such that $\lim_{||\boldsymbol{u}||\to\infty}||\boldsymbol{u}||^p w(\boldsymbol{u}) = 0$ and $w(-\boldsymbol{u}) = w(\boldsymbol{u})$ for all $\boldsymbol{u}$. If $h_n\to 0$ and $nh_n^p\to\infty$, then*

$$g_n(\boldsymbol{x}) \xrightarrow{P} g(\boldsymbol{x}) \quad \textit{at every continuity point } \boldsymbol{x} \textit{ of } g.$$

Proof This is a direct consequence of Theorem 3.1.2 of Rao (1983).

The crucial result needed to prove Theorem 4 is the following.

Lemma 6.2 *Under the conditions of Theorem 4, we have*

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

(2) $(nh_n^{p+1})^{1/2}\sum_{i=1}^n c_i(s,\boldsymbol{z}_0)(\boldsymbol{Z}_i-\boldsymbol{z}_0) = o_p(1)$ *uniformly in $s\in[0,1]$, if $nh_n^p\to\infty$ and $nh_n^{p+5}\to 0$.*

Proof By applying Proposition 6.3, Lemma A.3 stated in the Appendix and the arguments leading to (6.20), we see that if $nh_n^p\to\infty$ and $h_n\to 0$, then

$$\sup_{s\in[0,1]}\Big|\sum_{i=1}^n W_i(\boldsymbol{z}_0)Y_i(s) - H(s,\boldsymbol{z}_0)\Big| \xrightarrow{P} 0. \tag{6.31}$$

Write

$$\begin{aligned}\sum_{i=1}^n c_i(s,\boldsymbol{z}_0)(\boldsymbol{Z}_i-\boldsymbol{z}_0) &= \frac{\sum_{i=1}^n W_i(\boldsymbol{z}_0)Y_i(s)(\boldsymbol{Z}_i-\boldsymbol{z}_0)}{\sum_{j=1}^n W_j(\boldsymbol{z}_0)Y_j(s)}\\ &= \frac{\frac{1}{n}\sum_{i=1}^n w\left(\frac{\boldsymbol{Z}_i-\boldsymbol{z}_0}{h_n}\right)Y_i(s)\left(\frac{\boldsymbol{Z}_i-\boldsymbol{z}_0}{h_n^p}\right)}{\frac{1}{nh_n^p}\sum_{j=1}^n w\left(\frac{\boldsymbol{Z}_j-\boldsymbol{z}_0}{h_n}\right)\sum_{j=1}^n W_j(\boldsymbol{z}_0)Y_j(s)}.\end{aligned} \tag{6.32}$$

Then, by (6.31) and Proposition 6.4, the denominator of (6.32) converges in probability to $f(\boldsymbol{z}_0)H(s,\boldsymbol{z}_0)$ uniformly in $s\in[0,1]$. Next we show that the numerator of (6.32) converges in probability to zero uniformly in $s\in[0,1]$.

Recall that $\boldsymbol{Z}_i = (Z_{i1}, \ldots, Z_{ip})'$ and $\boldsymbol{z}_0 = (z_{01}, \ldots, z_{0p})'$. For fixed l $(1 \le l \le p)$, let $\xi_{ni}(s) = w\big(\frac{\boldsymbol{Z}_i - \boldsymbol{z}_0}{h_n}\big) Y_i(s) \big(\frac{Z_{il} - z_{0l}}{h_n^p}\big)$ and $\xi_n(s) = \frac{1}{n}\sum_{i=1}^n \xi_{ni}(s)$. Then the l^{th} component of the numerator of (6.32) is $\xi_n(s)$. Recalling the definition of $f(\boldsymbol{z}, s)$ given in (4.1), note that

$$\begin{aligned}
|E(\xi_{n1}(s))| &= \Big| \int w\Big(\frac{\boldsymbol{z} - \boldsymbol{z}_0}{h_n}\Big)\Big(\frac{z_l - z_{0l}}{h_n^p}\Big) f(\boldsymbol{z}, s) d\boldsymbol{z} \Big| \\
&= \Big| h_n \int w(\boldsymbol{u}) u_l f(\boldsymbol{z}_0 + h_n \boldsymbol{u}, s) d\boldsymbol{u} \Big| \\
&= \Big| h_n \int w(\boldsymbol{u}) u_l (f(\boldsymbol{z}_0 + h_n \boldsymbol{u}, s) - f(\boldsymbol{z}_0, s)) d\boldsymbol{u} \Big| \\
&\le h_n \int w(\boldsymbol{u}) |u_l| \cdot |f(\boldsymbol{z}_0 + h_n \boldsymbol{u}, s) - f(\boldsymbol{z}_0, s)| d\boldsymbol{u} \\
&\le h_n^2 M \int w(\boldsymbol{u}) ||\boldsymbol{u}||^2 d\boldsymbol{u}
\end{aligned}$$

where the last inequality follows from (4.2). Therefore $E(\xi_{n1}(s)) = O(h_n^2)$ uniformly in $s \in [0,1]$. Moreover,

$$\begin{aligned}
E(\xi_{n1}^2(s)) &= \int w^2\Big(\frac{\boldsymbol{z} - \boldsymbol{z}_0}{h_n}\Big)\Big(\frac{z_l - z_{0l}}{h_n^p}\Big)^2 f(\boldsymbol{z}, s) d\boldsymbol{z} \\
&= h_n^{2-p} \int w^2(\boldsymbol{u}) u_l^2 f(\boldsymbol{z}_0 + h_n \boldsymbol{u}, s) d\boldsymbol{u} \\
&= O(h_n^{2-p}) \quad \text{uniformly in } s \in [0,1].
\end{aligned}$$

Thus

$$\begin{aligned}
E(\xi_n^2(s)) &= \mathrm{Var}(\xi_n(s)) + \big(E\xi_n(s)\big)^2 \\
&= \frac{1}{n}\mathrm{Var}(\xi_{n1}(s)) + \big(E\xi_{n1}(s)\big)^2 \\
&= \frac{1}{n}\Big(E\big(\xi_{n1}^2(s)\big) - \big(E\xi_{n1}(s)\big)^2\Big) + \big(E\xi_{n1}(s)\big)^2 \\
&= \frac{1}{n}\Big(O(h_n^{2-p}) - O(h_n^4)\Big) + O(h_n^4) \quad \text{uniformly in } s \in [0,1].
\end{aligned}$$

Therefore $E\Big((nh_n^{p-4})^{\frac{1}{4}}\xi_n(s)\Big)^2 = O\big(1/(nh_n^p)^{\frac{1}{2}}\big) + O\big((nh_n^{p+4})^{\frac{1}{2}}\big)$ and $E\Big((nh_n^{p+1})^{\frac{1}{2}}\xi_n(s)\Big)^2 = O(h_n^3) + O(nh_n^{p+5})$ uniformly in $s \in [0,1]$. The conclusions of both parts of the lemma now follow immediately.

Proof of Theorem 4 Before checking Conditions (A) and (B), we first show that

$$\frac{1}{h_n^2}\sum_{i=1}^n c_i(s, \boldsymbol{z}_0)(\boldsymbol{Z}_i - \boldsymbol{z}_0)(\boldsymbol{Z}_i - \boldsymbol{z}_0)' \xrightarrow{P} \mathrm{diag}\Big(\int u_1^2 w(\boldsymbol{u}) d\boldsymbol{u}, \ldots, \int u_p^2 w(\boldsymbol{u}) d\boldsymbol{u}\Big) \qquad (6.33)$$

uniformly in $s \in [0,1]$ if $h_n \to 0$ and $nh_n^p \to \infty$. For each l $(1 \le l \le p)$

$$\frac{1}{h_n^2}\sum_{i=1}^n c_i(s, \boldsymbol{z}_0)(Z_{il} - z_{0l})^2 = \sum_{i=1}^n W_i(\boldsymbol{z}_0) Y_i(s)\Big(\frac{Z_{il} - z_{0l}}{h_n}\Big)^2 \Big/ \sum_{j=1}^n W_j(\boldsymbol{z}_0) Y_j(s)$$

$$= \sum_{i=1}^n w\Big(\frac{\boldsymbol{Z}_i - \boldsymbol{z}_0}{h_n}\Big)\Big(\frac{Z_{il} - z_{0l}}{h_n}\Big)^2 Y_i(s) \Big/ \sum_{j=1}^n w\Big(\frac{\boldsymbol{Z}_j - \boldsymbol{z}_0}{h_n}\Big) \sum_{j=1}^n W_j(\boldsymbol{z}_0) Y_j(s)$$

$$= \frac{\int u_l^2 w(\boldsymbol{u}) d\boldsymbol{u} \Big\{ \sum_{i=1}^n \tilde{W}_i(\boldsymbol{z}_0) Y_i(s) \frac{1}{nh_n^p} \sum_{i=1}^n \tilde{w}\Big(\frac{\boldsymbol{Z}_i - \boldsymbol{z}_0}{h_n}\Big)\Big\}}{\frac{1}{nh_n^p} \sum_{j=1}^n w\Big(\frac{\boldsymbol{Z}_j - \boldsymbol{z}_0}{h_n}\Big) \sum_{j=1}^n W_j(\boldsymbol{z}_0) Y_j(s)} \tag{6.34}$$

where $\tilde{w}(\boldsymbol{u}) \equiv u_l^2 w(\boldsymbol{u}) / \int u_l^2 w(\boldsymbol{u}) d\boldsymbol{u}$, and $\tilde{W}_i(\boldsymbol{z}_0) \equiv \tilde{w}\big(\frac{\boldsymbol{Z}_i - \boldsymbol{z}_0}{h_n}\big) \big/ \sum_{j=1}^n \tilde{w}\big(\frac{\boldsymbol{Z}_j - \boldsymbol{z}_0}{h_n}\big)$. Note that $\tilde{w}(\cdot)$ is a density function that satisfies the regularity conditions needed to apply Propositions 6.3 and 6.4. By applying Proposition 6.4 and the arguments leading to (6.31) to both numerator and denominator of (6.34), we see that

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

For $l \neq m$ $(1 \leq l,\ m \leq p)$, we have

$$\frac{1}{h_n^2} \sum_{i=1}^n c_i(s, \boldsymbol{z}_0)(Z_{il} - z_{0l})(Z_{im} - z_{0m}) = \frac{\frac{1}{h_n^2} \sum_{i=1}^n W_i(\boldsymbol{z}_0) Y_i(s)(Z_{il} - z_{0l})(Z_{im} - z_{0m})}{\sum_{j=1}^n W_j(\boldsymbol{z}_0) Y_j(s)} \tag{6.35}$$

$$= \frac{\frac{1}{nh_n^p} \sum_{i=1}^n w\Big(\frac{\boldsymbol{Z}_i - \boldsymbol{z}_0}{h_n}\Big) Y_i(s) \Big(\frac{Z_{il} - z_{0l}}{h_n}\Big)\Big(\frac{Z_{im} - z_{0m}}{h_n}\Big)}{\frac{1}{nh_n^p} \sum_{j=1}^n w\Big(\frac{\boldsymbol{Z}_j - \boldsymbol{z}_0}{h_n}\Big) \sum_{j=1}^n W_j(\boldsymbol{z}_0) Y_j(s)}.$$

It follows from Proposition 6.4 and (6.31) that the denominator of (6.35) converges in probability to $f(\boldsymbol{z}_0) H(s, \boldsymbol{z}_0)$. So one only needs to show that the numerator of (6.35) converges in probability to 0. Let

$$\eta_{ni}(s) = \frac{1}{h_n^p} w\Big(\frac{\boldsymbol{Z}_i - \boldsymbol{z}_0}{h_n}\Big) Y_i(s) \Big(\frac{Z_{il} - z_{0l}}{h_n}\Big)\Big(\frac{Z_{im} - z_{0m}}{h_n}\Big), \quad i = 1, \ldots, n.$$

Then $\eta_n(s) \equiv \frac{1}{n} \sum_{i=1}^n \eta_{ni}(s)$ is the numerator in (6.35). Since

$$\begin{aligned}
|E(\eta_{ni}(s))| &= \Big| \frac{1}{h_n^p} \int w\Big(\frac{\boldsymbol{z} - \boldsymbol{z}_0}{h_n}\Big)\Big(\frac{z_l - z_{0l}}{h_n}\Big)\Big(\frac{z_m - z_{0m}}{h_n}\Big) f(\boldsymbol{z}, s) d\boldsymbol{z} \Big| \\
&= \Big| \int w(\boldsymbol{u}) u_l u_m f(\boldsymbol{z}_0 + h_n \boldsymbol{u}, s) d\boldsymbol{u} \Big| \\
&\leq \int w(\boldsymbol{u}) |u_l u_m| \cdot |f(\boldsymbol{z}_0 + h_n \boldsymbol{u}, s) - f(\boldsymbol{z}_0, s)| d\boldsymbol{u} \\
&\leq \int w(\boldsymbol{u}) |u_l u_m| \cdot M ||h_n \boldsymbol{u}|| d\boldsymbol{u} \quad \text{(by (4.2))} \\
&\leq M h_n \int w(\boldsymbol{u}) ||\boldsymbol{u}||^3 d\boldsymbol{u}
\end{aligned}$$

and

$$E(\eta_{ni}^2(s)) = \frac{1}{h_n^{2p}} \int w^2\Big(\frac{\boldsymbol{z} - \boldsymbol{z}_0}{h_n}\Big)\Big(\frac{z_l - z_{0l}}{h_n}\Big)^2 \Big(\frac{z_m - z_{0m}}{h_n}\Big)^2 f(\boldsymbol{z}, s) d\boldsymbol{z}$$

$$
\begin{aligned}
&= \frac{1}{h_n^p}\int w^2(\boldsymbol{u})u_l^2u_m^2 f(\boldsymbol{z}_0+h_n\boldsymbol{u},s)d\boldsymbol{u}\\
&= O_p(1/h_n^p) \quad \text{uniformly in } s\in[0,1],
\end{aligned}
$$

we have

$$
\begin{aligned}
E(\eta_n^2(s)) &= \mathrm{Var}(\eta_n(s)) + \big(E\eta_n(s)\big)^2\\
&= \frac{1}{n}\mathrm{Var}(\eta_{n1}(s)) + \big(E\eta_{n1}(s)\big)^2\\
&= \frac{1}{n}\Big(E(\eta_{n1}^2(s)) - \big(E\eta_{n1}(s)\big)^2\Big) + \big(E\eta_{n1}(s)\big)^2\\
&= \frac{1}{n}\big(O(1/h_n^p) - O(h_n^2)\big) + O(h_n^2)\\
&= O(1/nh_n^p) + O(h_n^2/n) + O(h_n^2)\\
&\longrightarrow 0 \quad \text{uniformly in } s\in[0,1]
\end{aligned}
$$

if $h_n \to 0$ and $nh_n^p \to \infty$. Thus $\eta_n(s) \xrightarrow{P} 0$ uniformly in $s \in [0,1]$. Hence (6.33) holds.

Now we prove Part (1) of the theorem. We have

$$
\begin{aligned}
\frac{1}{h_n^2}(\boldsymbol{Z}^{*'}P\boldsymbol{Z}^*)(s) &= \frac{1}{h_n^2}\sum_{i=1}^n c_i(s,\boldsymbol{z}_0)(\boldsymbol{Z}_i-\boldsymbol{z}_0)(\boldsymbol{Z}_i-\boldsymbol{z}_0)' - \frac{1}{h_n^2}\Big(\sum_{i=1}^n c_i(s,\boldsymbol{z}_0)(\boldsymbol{Z}_i-\boldsymbol{z}_0)\Big)\\
&\quad\times\Big(\sum_{i=1}^n c_i(s,\boldsymbol{z}_0)(\boldsymbol{Z}_i-\boldsymbol{z}_0)\Big)'\\
&= \frac{1}{h_n^2}\sum_{i=1}^n c_i(s,\boldsymbol{z}_0)(\boldsymbol{Z}_i-\boldsymbol{z}_0)(\boldsymbol{Z}_i-\boldsymbol{z}_0)' - \frac{1}{h_n^2}\Big(o_p\big((nh_n^{p-4})^{-\frac{1}{4}}\big)\Big)^2 \qquad (6.36)\\
&= \frac{1}{h_n^2}\sum_{i=1}^n c_i(s,\boldsymbol{z}_0)(\boldsymbol{Z}_i-\boldsymbol{z}_0)(\boldsymbol{Z}_i-\boldsymbol{z}_0)' - o_p\big((nh_n^p)^{-\frac{1}{2}}\big)\\
&\xrightarrow{P} \mathrm{diag}\Big(\int u_1^2w(\boldsymbol{u})d\boldsymbol{u},\ldots,\int u_p^2w(\boldsymbol{u})d\boldsymbol{u}\Big) \quad \text{uniformly in } s\in[0,1],
\end{aligned}
$$

each integral in the last line of (6.36) is positive. In (6.36), the first equality follows from the fact that $P(s)1 = 0$, the second equality follows from Part (1) of Lemma 6.2, and the convergence statement follows from (6.33). This implies (A1).

To verify (A2), we shall prove that for every $\delta \geq 0$

$$
\sup_{s\in[0,1]}\Big|(na_n)^{1+\delta}\sum_{i=1}^n c_i^{2+\delta}(s,\boldsymbol{z}_0) - g_\delta(s,\boldsymbol{z}_0)\Big| \xrightarrow{P} 0. \qquad (6.37)
$$

This will imply (A2) immediately.

Let $\bar{w}(\cdot) = w^{2+\delta}(\cdot)/\int w^{2+\delta}(\boldsymbol{u})d\boldsymbol{u}$ and $\bar{W}_i(\boldsymbol{z}_0) \equiv \bar{w}\big(\frac{\boldsymbol{Z}_i-\boldsymbol{z}_0}{h_n}\big)/\sum_{j=1}^n \bar{w}\big(\frac{\boldsymbol{Z}_j-\boldsymbol{z}_0}{h_n}\big)$. Then, $\bar{w}(\cdot)$ is a density function that satisfies the regularity conditions needed to apply Propositions 6.3 and 6.4. Thus,

$$
(nh_n^p)^{1+\delta}\sum_{i=1}^n W_i^{2+\delta}(\boldsymbol{z}_0)Y_i(s) = (nh_n^p)^{1+\delta}\sum_{i=1}^n w^{2+\delta}\Big(\frac{\boldsymbol{Z}_i-\boldsymbol{z}_0}{h_n}\Big)Y_i(s)\Big/\Big(\sum_{j=1}^n w\Big(\frac{\boldsymbol{Z}_j-\boldsymbol{z}_0}{h_n}\Big)\Big)^{2+\delta}
$$

$$
\begin{aligned}
&= \left[\sum_{i=1}^{n} \bar{W}_i(z_0)Y_i(s)\right] \left[\frac{\frac{1}{nh_n^p}\sum_{j=1}^{n} \bar{w}\left(\frac{Z_j - z_0}{h_n}\right)}{\left(\frac{1}{nh_n^p}\sum_{j=1}^{n} w\left(\frac{Z_j - z_0}{h_n}\right)\right)^{2+\delta}}\right] \int w^{2+\delta}(u)du \\
&\xrightarrow{P} H(s, z_0)\left[\frac{f(z_0)}{f^{2+\delta}(z_0)}\right] \int w^{2+\delta}(u)du \qquad \text{uniformly in } s \in [0,1] \\
&= \frac{H(s, z_0)}{f^{1+\delta}(z_0)} \int w^{2+\delta}(u)du.
\end{aligned} \tag{6.38}
$$

where the convergence of the factor in the first set of bracket follows by the arguments leading to (6.31) and the convergence of the numerator and denominator in the second set of bracket follows by Proposition 6.4. Therefore,

$$
\begin{aligned}
(nh_n^p)^{1+\delta}\sum_{i=1}^{n} c_i^{2+\delta}(s, z_0) &= (nh_n^p)^{1+\delta}\sum_{i=1}^{n} W_i^{2+\delta}(z_0)Y_i(s) \Big/ \Big(\sum_{i=1}^{n} W_i(z_0)Y_i(s)\Big)^{2+\delta} \\
&\xrightarrow{P} \frac{1}{\{f(z_0)H(s, z_0)\}^{1+\delta}} \int w^{2+\delta}(u)du \quad \text{uniformly in } s \in [0,1]
\end{aligned}
$$

from (6.38) and (6.31), so that (6.37) holds.

By regularity condition (R1) and (6.33),

$$
\begin{aligned}
\Big|\sum_{i=1}^{n} c_i(s, z_0)\big(\alpha(s, Z_i) - \alpha(s, z_0)\big)\Big| &\le \Big|\Big(\frac{\partial \alpha}{\partial z}(s, z_0)\Big)' \sum_{i=1}^{n} c_i(s, z_0)(Z_i - z_0)\Big| + O_p(h_n^2) \\
&= o_p\Big((nh_n^{p-4})^{-\frac{1}{4}}\Big) + O_p(h_n^2) \quad \text{uniformly in } s \in [0,1],
\end{aligned}
$$

where the last step follows from (R1) and Part (1) of Lemma 6.2. Thus

$$
\sqrt{nh_n^p}\int_0^1 \Big|\sum_{i=1}^{n} c_i(s, z_0)\big(\alpha(s, Z_i) - \alpha(s, z_0)\big)\Big| ds = o_p\Big((nh_n^{p+4})^{\frac{1}{4}}\Big) + O_p\Big((nh_n^{p+4})^{\frac{1}{2}}\Big) \xrightarrow{P} 0
$$

if $nh_n^{p+4} \to 0$. Therefore (A3) holds.

Since (R1) implies $|\alpha(s, z) - \alpha(s, z_0)| \le ||z - z_0||\Big(\sup_{s\in[0,1]}\Big|\Big|\frac{\partial \alpha}{\partial z}(s, z_0)\Big|\Big| + K_1||z - z_0||\Big)$, we have $\sum_{i=1}^{n} c_i(s, z_0)\big(\alpha(s, Z_i) - \alpha(s, z_0)\big)^2 \le M_1 \sum_{i=1}^{n} c_i(s, z_0)||Z_i - z_0||^2$ for some constant $M_1 > 0$. Hence,

$$
\begin{aligned}
\sqrt{nh_n^p}\Big|\sum_{i=1}^{n} c_i(s, z_0)\big(\alpha(s, Z_i) - \alpha(s, z_0)\big)^2\Big| &\le \sqrt{nh_n^p}M_1 \sum_{i=1}^{n} c_i(s, z_0)||Z_i - z_0||^2 \\
&= \sqrt{nh_n^p}O_p(h_n^2) \\
&= O_p\Big((nh_n^{p+4})^{\frac{1}{2}}\Big) \quad \text{uniformly in } s \in [0,1],
\end{aligned}
$$

where the first equality is from (6.33). This together with the assumption $nh_n^{p+4} \to 0$ implies that (A4) holds.

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

$$\begin{aligned} nh_n^p c_i(s, \boldsymbol{z}_0) &= nh_n^p W_i(\boldsymbol{z}_0) Y_i(s) \Big/ \sum_{j=1}^n W_j(\boldsymbol{z}_0) Y_j(s) \\ &\leq \frac{\sup_{\boldsymbol{u}} w(\boldsymbol{u})}{\frac{1}{nh_n^p} \sum_{j=1}^n w\Big(\frac{\boldsymbol{Z}_j - \boldsymbol{z}_0}{h_n}\Big) \sum_{j=1}^n W_j(\boldsymbol{z}_0) Y_j(s)}. \end{aligned} \tag{6.39}$$

Furthermore, by Proposition 6.4 and (6.31), the right-hand side of (6.39) converges in probability to $\sup_{\boldsymbol{u}} w(\boldsymbol{u})/\{f(\boldsymbol{z}_0)H(s, \boldsymbol{z}_0)\}$ uniformly in $s \in [0,1]$ and

$$\sup_{\boldsymbol{u}} w(\boldsymbol{u}) \Big/ \{f(\boldsymbol{z}_0) H(s, \boldsymbol{z}_0)\} \leq \sup_{\boldsymbol{u}} w(\boldsymbol{u}) \Big/ \Big(f(\boldsymbol{z}_0) \inf_{s \in [0,1]} H(s, \boldsymbol{z}_0)\Big) < \infty.$$

Therefore

$$nh_n^p c_i(s, \boldsymbol{z}_0) = O_p(1) \quad \text{uniformly in } i \text{ and } s \in [0,1]$$

as desired.

Finally, we show that (A6) holds. By Part (1) of Lemma 6.2 and (6.36)

$$\begin{aligned} &\sqrt{nh_n^p} J_1(s) \big(\boldsymbol{Z}^*(s)' \boldsymbol{c}(s) - \boldsymbol{z}_0\big)' \big(\boldsymbol{Z}^*(s)' P(s) \boldsymbol{Z}^*(s)\big)^{-1} \big(\boldsymbol{Z}^*(s)' \boldsymbol{c}(s) - \boldsymbol{z}_0\big) \\ &\quad = J_1(s) \Big((nh_n^{p-4})^{\frac{1}{4}} \sum_{i=1}^n c_i(s, \boldsymbol{z}_0)(\boldsymbol{Z}_i - \boldsymbol{z}_0)\Big)' \Big(\frac{1}{h_n^2} \boldsymbol{Z}^*(s)' P(s) \boldsymbol{Z}^*(s)\Big)^{-1} \\ &\quad\quad \times \Big((nh_n^{p-4})^{\frac{1}{4}} \sum_{i=1}^n c_i(s, \boldsymbol{z}_0)(\boldsymbol{Z}_i - \boldsymbol{z}_0)\Big) \\ &\quad \xrightarrow{P} 0 \quad \text{uniformly in } s \in [0,1]. \end{aligned}$$

This gives (A6).

The proof of Part (2) of the theorem uses Part (2) of Lemma 6.2 and is completely parallel to the proof of Part (1).

Appendix

Lemma A.3 *Assume that $f_1, f_2, \ldots$ is a sequence of left-continuous nondecreasing random functions on $[0,1]$ such that for every $t \in [0,1]$*

$$\begin{cases} f_n(t) \xrightarrow{P} f(t), \\ f_n(t+) - f_n(t-) \xrightarrow{P} f(t+) - f(t-), \end{cases}$$

where f is a left-continuous nondecreasing deterministic function. Then

$$\sup_{t \in [0,1]} |f_n(t) - f(t)| \xrightarrow{P} 0 \quad \text{as } n \to \infty.$$

Proof Lemma A.3 is a standard result. To prove it, we need only show that for any subsequence of $\{n\}$ there exists a further subsequence along which

$$\sup_{t\in[0,1]} |f_n(t) - f(t)| \xrightarrow{a.s.} 0,$$

and this is done by arguing exactly as in the proof of Theorem 5.5.1 of Chung (1974).

Lemma A.4 *For $n = 1, 2, \dots$ let $N^{(n)}$ be a multivariate counting process with n components and intensity process $\lambda^{(n)}$. Let $H^{(n)}$ be a $1 \times n$ vector of locally bounded predictable processes. Define locally square integrable martingales by*

$$W^{(n)}(t) = \int_0^t \sum_{i=1}^n H_i^{(n)}(s)\{dN_i^{(n)}(s) - \lambda_i^{(n)}(s)ds\}.$$

Suppose that

(i) for all $\epsilon > 0$, $\int_0^1 \sum_{i=1}^n H_i^{(n)^2}(s) I(|H_i^{(n)}(s)| > \epsilon)\lambda_i^{(n)}(s)ds \xrightarrow{P} 0,$

(ii) $\int_0^1 \sum_{i=1}^n H_i^{(n)^2}(s)\lambda_i^{(n)}(s)ds \xrightarrow{P} \sigma^2.$

Then

$$W^{(n)}(1) \xrightarrow{d} N(0, \sigma^2) \quad \text{as } n \to \infty.$$

Proof We note that conditions (L_2) and (12) in Corollary 2 of Liptser and Shiryayev (1980) reduce to (i) and (ii) under the setup of Lemma A.4. So Lemma A.4 follows immediately from Remark 1 of Liptser and Shiryayev (1980).

Acknowledgements

We are very 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 Statistics* ***2*** 1–25, Springer-Verlag, 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-Verlag, 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. Statist., Univ. California, Berkeley.

Chambers, J. and Hastie T. (1991). *Statistical Models in S.* Wadsworth and Brooks, Pacific Grove, California.

Chung, K. L. (1974). *A Course in Probability Theory.* Second Edition, 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 Mathematics* ***821*** 159–175, Springer-Verlag, 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.

Devroye, L. (1981). On the almost everywhere convergence of nonparametric regression function estimates. *Ann. Statist.* ***9*** 1310–1319.

Diaconis, P. and Freedman, D. (1981). On the histogram as a density estimator: L_2 theory. *Z. Wahrscheinlichkeitstheorie Verw. Gebiete* ***57*** 453–476.

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.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansions.* Springer-Verlag, New York.

Härdle, W. (1990). *Applied Nonparametric Regression.* Cambridge University Press, Cambridge.

Härdle, W. (1991). *Smoothing Techniques with Implementation in S.* Springer-Verlag, New York.

Hastie, T. and Tibshirani, R. (1990a). *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.

Liptser, R. SH. and Shiryayev, A. N. (1980). A functional central limit theorem for semimartingales. *Theor. Probab. Appl.* ***25*** 667–688.

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.

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

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

Schuster, E. (1972). Joint asymptotic distribution of the estimated regression function at a finite number of distinct points. *Ann. Math. Statist.* ***43*** 84–88.

Stute, W. (1986). Conditional empirical processes. *Ann. Statist.* ***14*** 638-647.

Estimated Conditional Survival Functions S_n(t|z) For Female diabetics
(where z is the Age at Diagnosis of Diabetes)

Figure 1(a)

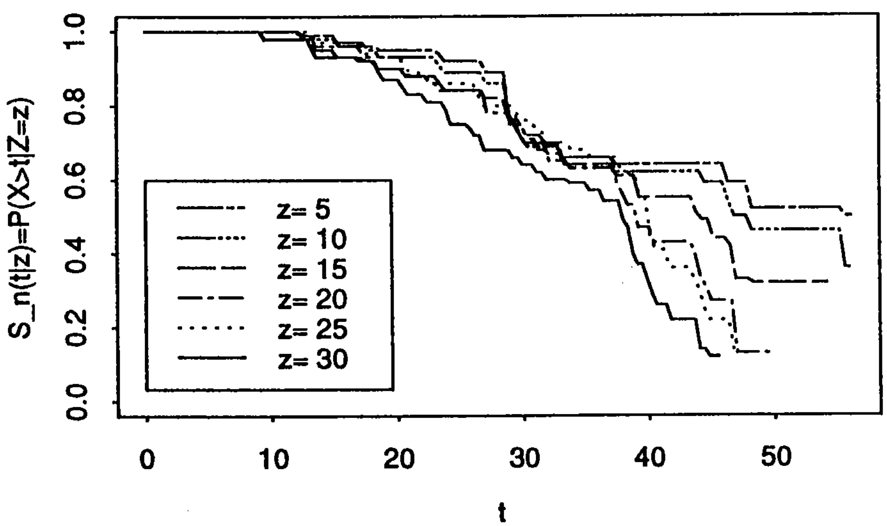


Figure 1(b)

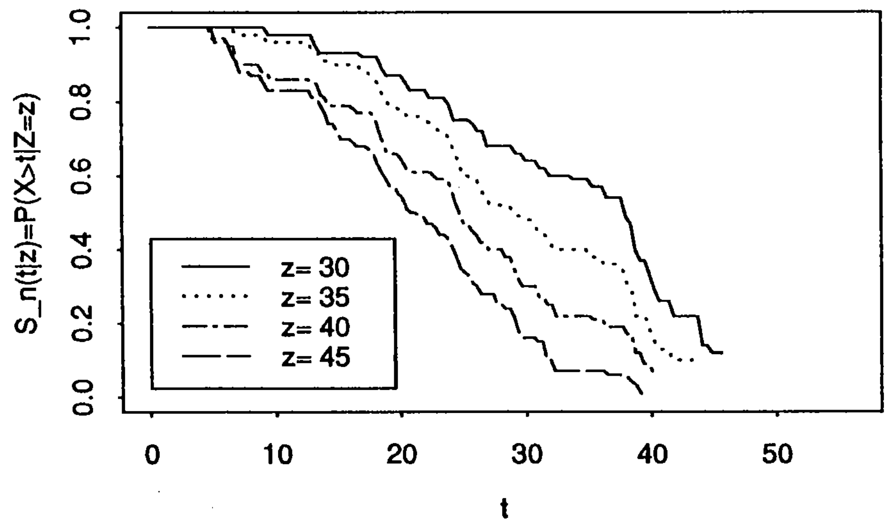


Figure 1(c)

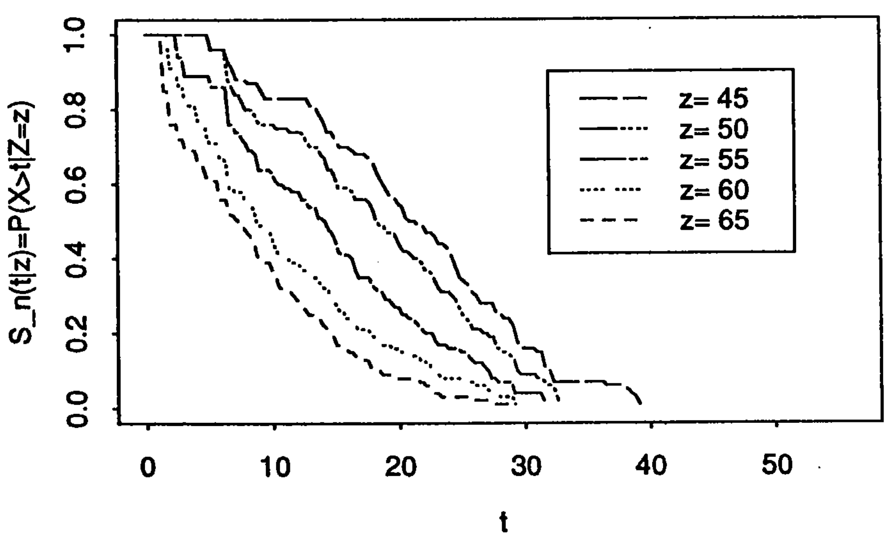
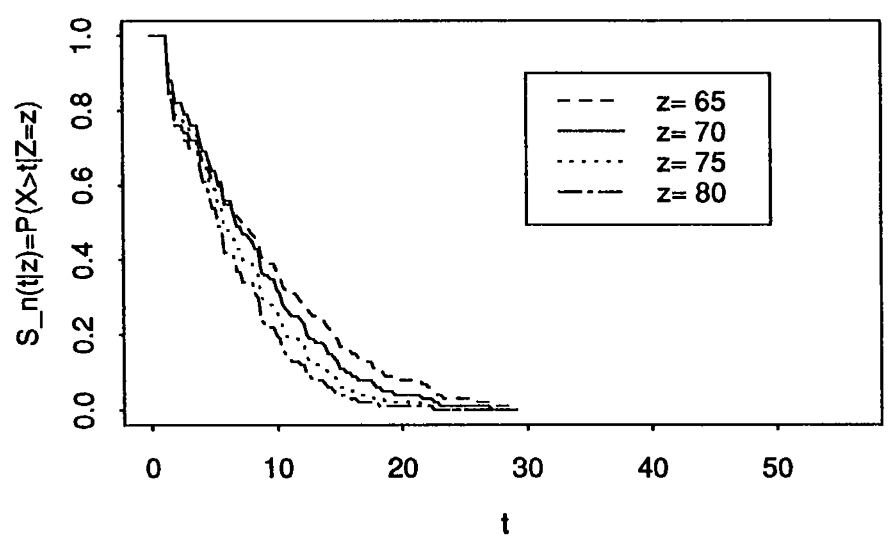


Figure 1(d)



Estimated Conditional Survival Functions S_n(t|z) For Male Diabetics (where z is the Age at Diagnosis of Diabetes)

Figure 2(a)

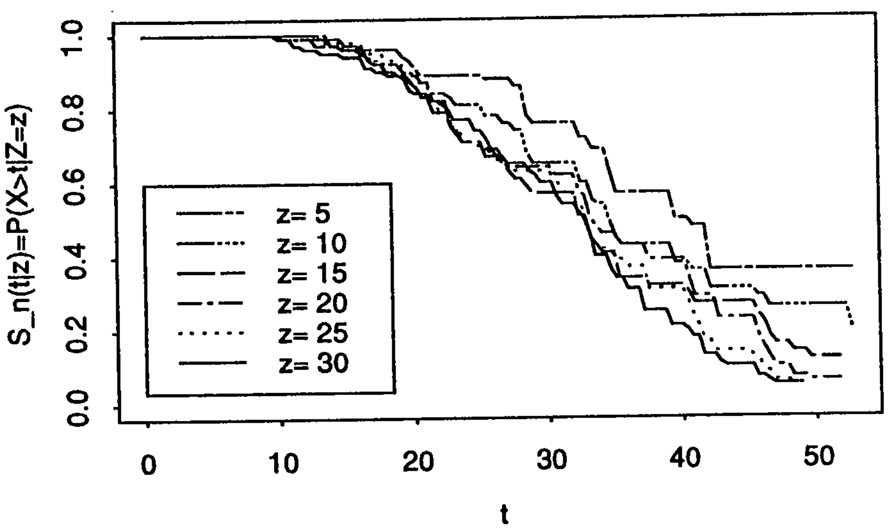


Figure 2(b)

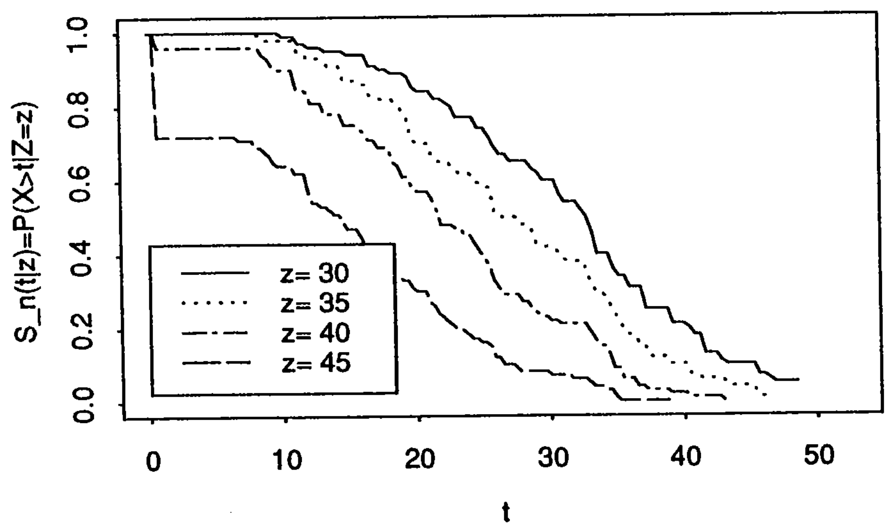


Figure 2(c)

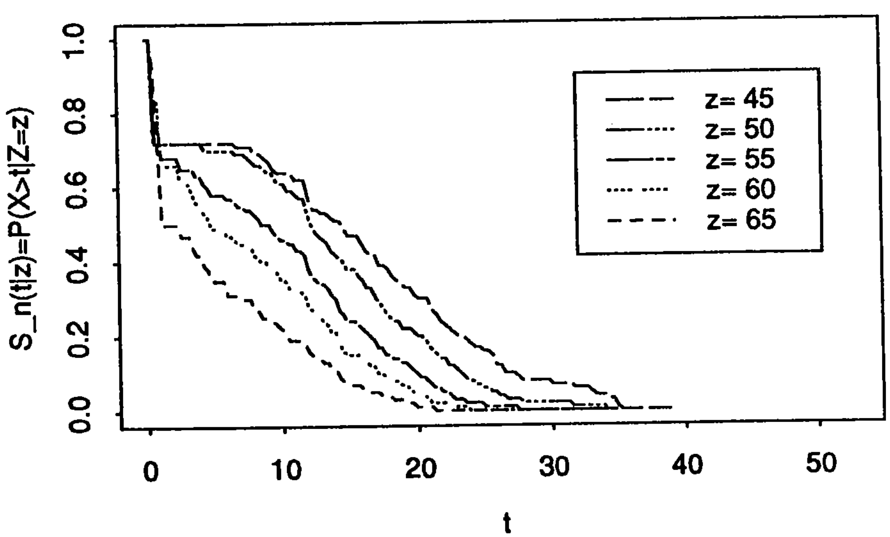


Figure 2(d)

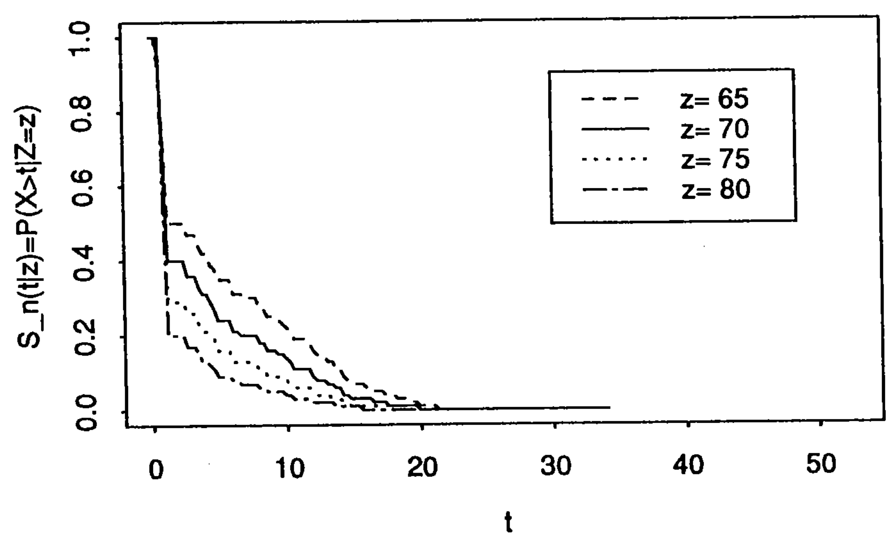


Figure 3: 95% Confidence Bands for Conditional Survival Functions S(t|z)
(Where z Is the Age at Diagnosis of Diabetes)

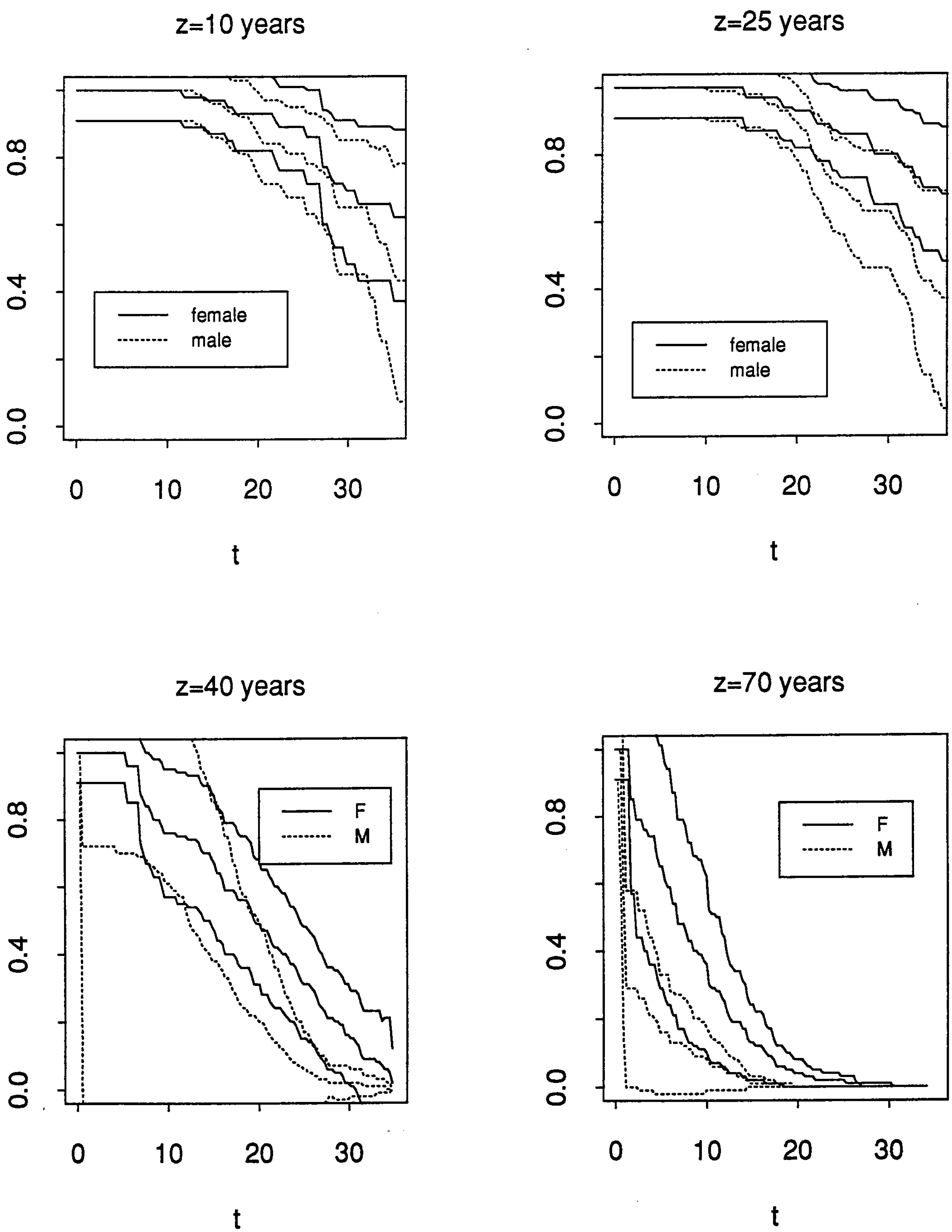


Figure 4: Median Survival Time Plot for Female and Male Diabetics