ESTIMATING MULTIPLICATIVE AND ADDITIVE HAZARD FUNCTIONS BY KERNEL METHODS*

Oliver Linton[1]
London School of Economics and Political Science

Jens Perch Nielsen[2]
Codan, Copenhagen

Sara van de Geer
University of Leiden

Contents:

Discussion Paper
No.EM/01/411
February 2001

The Suntory Centre
Suntory and Toyota International Centres
for Economics and Related Disciplines
London School of Economics and Political Science
Houghton Street
London, WC2A 2AE
Tel.: 020-7504 7686

* We thank three referees and the editor for helpful comments. [1] Thanks to the NSF for financial support. [2] Thanks to Bergiafonden for financial support.

Abstract

We propose new procedures for estimating the univariate quantities of interest in both additive and multiplicative nonparametric marker dependent hazard models. We work with a full counting process framework that allows for left truncation and right censoring. Our procedures are based on kernels and on the idea of marginal integration. We provide a central limit theorem for our estimator.

Keywords: Additive model; censoring; kernel; proportional hazards; survival analysis.

JEL No.: C14

1 Introduction

Suppose that the conditional hazard function

$$\lambda(t|Z_i) = \lim_{\epsilon \downarrow 0} \frac{1}{\epsilon} P\left(T_i \leq t + \epsilon | T_i > t; (Z_i(s), s \leq t)\right)$$

for the survival time T_i of an individual i with the covariate or marker process $Z_i = (Z_i(t))$ has the form

$$\lambda(t|Z_i) = \alpha(t, Z_i(t)), \tag{1}$$

where α is an unknown function of time t and the value of the covariate process of the individual at time t only. Inference for this general class of models was initiated by Beran (1981), and extended by Ramlau-Hansen (1983), Dabrowska (1987), McKeague and Utikal (1990), Hjort (1994), and Nielsen and Linton (1995), who established asymptotic normality and uniform convergence of their estimators of $\alpha(t, z)$ in the case where one observes a sample of mutually independent individuals and their covariate processes, subject perhaps to some (non-informative) censoring and truncation. Unfortunately, the rate of convergence of estimators of $\alpha(t, z)$ increases rapidly with the number of covariates, Stone (1980). Furthermore, it is hard to visualize the model in high dimensions.

This motivates the study of separable structures, and in particular additive and multiplicative models. These models can be used in their own right or as an aid to further model specification. They allow for the visual display of the components of high dimensional models and for a clean interpretation of effects. Also, the optimal rate of convergence in additive and other separable regression models has been shown to be better than in the unrestricted case, see Stone (1985,1986). In this paper, we consider additive and multiplicative sub-models of (1). Multiplicative separability of the baseline hazard from the covariate effect has played a central role in survival analysis as is evident from the enormous literature inspired by Cox (1972); see Andersen, Borgan, Gill, and Keiding (1992, Chapter 7) for a discussion of semiparametric and nonparametric hazard models, and see Sasieni (1992), O'Sullivan (1993) , Lin and Yang (1995), Dabrowska (1997), Nielsen, Linton, and Bickel (1998), and Huang (1999) for some recent extensions. Additive models are perhaps less common, but have been studied in Aalen (1980) and McKeague and Utikal (1991). Our focus here is on fully nonparametric models where we do not specify either the time or the covariate effects.

We propose a class of kernel-based marginal integration estimators for the components in additive and multiplicative models. This methodology has been developed in Linton and Nielsen (1995) [see also Auestad and Tjøstheim (1991), Tjøstheim and Auestad (1994) and Newey (1994)] for regression. We extend this literature to counting process models, which allow for a wide range of censoring and truncation. The estimation idea involves integrating out a high dimensional estimator that does not impose the separable structure; in our setting this is provided by the Nielsen and Linton (1995) estimator. The averaging (or integration) reduces variance and hence permits faster convergence rates. We establish the pointwise and uniform consistency properties of our marginal integration estimators, and give their limiting distributions. Mammen, Linton, and Nielsen (1999) show that it is possible to improve the variance of marginal integration estimators in regression, and it is possible that this result can be extended to the current situation although the nonlinear form of the appropriate estimating function makes a rigorous proof of the central limit theorem for this procedure somewhat harder. Alternative estimation methods include the log-spline methods developed in Kooperberg, Stone, and Truong (1995) and Huang, Kooperberg, Stone, and Truong (2000).

One major theoretical problem we encounter in deriving the asymptotic properties of our procedures is the so-called predictability issue. We use the solution to this problem provided by Nielsen, Linton and Bickel (1998) and improved in Nielsen (1999). We also provide a new result on uniform convergence of kernel hazard estimators in the counting process framework. This result is fundamental to the proofs of limiting properties of many nonparametric and semiparametric procedures, including our own. The result contained herein greatly improves and extends the result contained in Nielsen and Linton (1995) and gives essentially the optimal rate. Our proof makes use of the recently derived exponential inequality for martingales obtained in van de Geer (1995). The methods developed in this paper have already been applied in Felipe, Guillen, and Nielsen (2000), a study of longevity in different European countries over time.

For any vectors $x = (x_1, \ldots, x_k)$ and $a = (a_1, \ldots, a_k)$ of common length k, we let $x^a = x_1^{a_1} \cdots x_k^{a_k}$ and $|a| = \sum_{j=1}^k a_j$. Finally, for any function g: $\mathbb{R}^k \to \mathbb{R}$, let

$$D^a g(x) = \frac{\partial^{|a|}}{\partial x_1^{a_1} \cdots \partial x_k^{a_k}} g(x).$$

Note that all integrals in the sequel are pathwise Riemann-Stieltjes because we have piecewise continuous integrators and continuous integrands.

2 The marker dependent hazard model

2.1 The Counting Process Formulation

We work with a sampling framework laid down in Aalen (1978) that is based on counting processes. This framework is very general and can be shown to accommodate a wide variety of censoring mechanisms; in the next section we describe one such data generation mechanism. Let $\mathbf{N}^{(n)}(t) = (N_1(t), \ldots, N_n(t))$ be a n-dimensional counting process with respect to an increasing, right-continuous, complete filtration $\mathcal{F}_t^{(n)},\ t \in [0, T]$, i.e., $\mathbf{N}^{(n)}$ is adapted to the filtration and has components N_i, which are right-continuous step-functions, zero at time zero, with jumps of size one such that no two components jump simultaneously. Here, $N_i(t)$ records the number of observed failures for the i'th individual during the time interval $[0, t]$, and is defined over the whole period [taken to be $[0, T]$, where T is finite]. Suppose that N_i has intensity

$$\lambda_i(t) = \lim_{\epsilon \downarrow 0} \frac{1}{\epsilon} P\left(N_i(t+\epsilon) - N_i(t) = 1 | \mathcal{F}_t^{(n)}\right) = \alpha(t, Z_i(t)) Y_i(t), \tag{2}$$

where Y_i is a predictable process taking values in $\{0, 1\}$, indicating (by the value 1) when the i^{th} individual is observed to be at risk, while Z_i is a d-dimensional predictable covariate process with support in some compact set $\mathcal{Z} \subseteq \mathbb{R}^d$. The function $\alpha(t, z)$ represents the failure rate for an individual at risk at time t with covariate $Z_i(t) = z$.

We assume that the stochastic processes $(N_1, Z_1, Y_1), \ldots, (N_n, Z_n, Y_n)$ are independent and identically distributed (i.i.d.) for the n individuals. In the sequel we therefore drop the n superscript for convenience. This simplifying assumption has been adopted in a number of leading papers in this field, for example Andersen and Gill (1982, section 4), and McKeague and Utikal (1990, section 4). Let $\mathcal{F}_{t,i} = \sigma\{N_i(u), Z_i(u), Y_i(u);\quad u \le t\}$ and $\mathcal{F}_t = \bigvee_{i=1}^n \mathcal{F}_{t,i}$. With these definitions, λ_i is predictable with respect to $\mathcal{F}_{t,i}$ and hence $\mathcal{F}_t$, and the processes $M_i(t) = N_i(t) - \Lambda_i(t)$, $i = 1, \ldots, n$, with compensators $\Lambda_i(t) = \int_0^t \lambda_i(u) du$, are square integrable local martingales with respect to $\mathcal{F}_{t,i}$ on the time interval $[0, T]$. Hence, $\Lambda_i(t)$ is the compensator of $N_i(t)$ with respect to both the filtration $\mathcal{F}_{t,i}$ and the filtration $\mathcal{F}_t$. This model formulation is adopted because it allows us to use the convenient solution to the so-called predictability problem given in Nielsen (1999), see the discussion in the appendix. In fact, rather than observing the whole covariate process Z_i, it is sufficient to observe Z_i at times when the individual is at risk, i.e., when $Y_i(s) = 1$. We shall not provide a complete model of the evolution of the covariate process, but see Jewell and Nielsen (1993) for some discussions on the

2.2 The Observable Data

The above counting process framework is general enough to include many models for survival data including complicated left truncation and right censoring patterns given an appropriate choice of (N, Z, Y), see Andersen, Borgan, Gill, and Keiding (1992, Chapter 3) for some discussions. In this section, we just outline the leading case where the data are right-censored survival times.

Specifically, let T be the survival time and let $\widetilde{T} = \min\{T, C\}$, where C is the censoring time. Suppose that T and C are conditionally independent given the left-continuous covariate process Z, and suppose that the conditional hazard of T at time t given $\{Z(s),\ s \leq t\}$ is $\alpha(t, Z(t))$. For each of n independent copies (T_i, C_i, Z_i), $i = 1, \ldots, n$ of (T, C, Z), we observe $\widetilde{T}_i, \delta_i = 1(T_i \leq C_i)$ and $Z_i(t)$ for $t \leq T_i$. Define also $Y_i(t) = 1(\widetilde{T}_i \leq t)$, the indicator that the individual is observed to be at risk at time t, and $N_i(t) = 1(\widetilde{T}_i \leq t, \delta_i = 1)$. Then, $\mathbf{N}(t) = (N_1(t), \ldots, N_n(t))$ is a multivariate counting process, and N_i has intensity (2).

We can allow quite general covariate types including time invariant covariates, variables common to all individuals, and both discrete and continuous variables [although our regularity conditions given below rule out discrete variables, this is just for notational convenience. If the covariates are discrete then better rates of convergence result, because one doesn't need to smooth at all. However, discrete variables are handled by different methods and combining them with continuous variables makes for notationally complex proofs without significantly affecting the arguments].

It might be objected that full knowledge of the covariate process is possible only if it changes deterministically between observed time points. This is the case for some but not all covariates. Trivially it applies to time invariant covariates. A prominent example of a time-varying covariate to which this applies would be when Z is the time since a certain event, such as onset of disability, which arises in many actuarial applications. In other cases, one must interpolate or extrapolate between the points in order to compute the required integrals in the estimation routines. The error in doing this depends on the frequency of observation and on the variability of the covariate process between observation points. See Nielsen (1999) for some discussion.

2.3 Separable Models and Estimands

For notational convenience we combine time and the covariates into one vector, i.e., we write $x = (t, z)$ and $X_i(t) = (t, Z_i(t))$, and label the components of x as $0, 1, \ldots, d$, with $x_0 = t$. Let $x_{-j} = (x_0, \ldots, x_{j-1}, x_{j+1}, \ldots, x_d)$ be the $d \times 1$ vector of x excluding x_j and likewise for $X_{-ji}(s)$.

The main object of interest is the hazard function $\alpha(\cdot)$ and functionals computed from it. We consider the case that α is restricted to be separable either additively or multiplicatively. The multiplicative model is that

$$\alpha(x) = c_M \prod_{j=0}^{d} h_j(x_j) \tag{3}$$

for some constant c_M and functions h_j, $j = 0, 1, \ldots, d$. The additive model is

$$\alpha(x) = c_A + \sum_{j=0}^{d} g_j(x_j) \tag{4}$$

for some constant c_A and functions g_j, $j = 0, 1, \ldots, d$. The functions $h_j(.)$ and $g_j(.)$ and constants c_A and c_M are not separately identified, and we need to make a further restriction in both cases to obtain uniqueness.

Let Q be a given absolutely continuous c.d.f. and define the marginals $Q_j(x_j) = Q(\infty, \ldots, \infty, x_j, \infty, \ldots, \infty)$ and $Q_{-j}(x_{-j}) = Q(x_0, \ldots, x_{j-1}, \infty, x_{j+1}, \ldots, x_d)$, $j = 0, 1, \ldots, d$. We identify the models (3) and (4) through these probability measures. Specifically, we suppose that in the additive case $\int g_j(x_j)dQ_j(x_j) = 0$, while in the multiplicative case $h_j(x_j) \int \prod_{k \neq j} h_k(x_k)dQ_{-j}(x_{-j}) = 1$ for each $j = 0, \ldots, d$. These restrictions ensure that the model components $(c_A, g_0, \ldots, g_d)$ and $(c_M, h_0, \ldots, h_d)$ respectively are well-defined. Now consider the following contrasts:

$$\alpha_{Q_{-j}}(x_j) = \int \alpha(x)dQ_{-j}(x_{-j}), \tag{5}$$

$j = 0, \ldots, d$. In the additive model, $\alpha_{Q_{-j}}(x_j) = g_j(x_j) + c_A$, while in the multiplicative model, $\alpha_{Q_{-j}}(x_j) = h_j(x_j)c_M$ for some constants $c_A(Q)$ and $c_M(Q)$ defined below. It follows that $\alpha_{Q_{-j}}(\cdot)$ is, up to a constant factor, the univariate component of interest in both additive and multiplicative structures. Now define the constants $c_{*j} = \int \alpha(x)dQ_{-j}(x_{-j})dQ_j(x_j)$ and $c_\dagger = \int \alpha(x)dQ_0(x_0)\cdots dQ_d(x_d)$. In the additive case (4), $c_A = c_\dagger = \sum_{j=0}^{d} c_{*j}/(d+1)$ and $\alpha(x) = \sum_{j=0}^{d} \alpha_{Q_{-j}}(x_j) - d\sum_{j=0}^{d} c_{*j}/(d+1)$. In the multiplicative case (3), $c_M = \left(\prod_{j=0}^{d} c_{*j}/c_\dagger\right)^{1/d}$ and $\alpha(x) = \prod_{j=0}^{d} \alpha_{Q_{-j}}(x_j)c_\dagger / \prod_{j=0}^{d} c_{*j}$. The quantities

$$\alpha_A(x) = \sum_{j=0}^{d} \alpha_{Q_{-j}}(x_j) - d \cdot c_A \text{ and } \alpha_M(x) = \prod_{j=0}^{d} \alpha_{Q_{-j}}(x_j)/c_M, \tag{6}$$

respectively, are both equal to $\alpha(x)$ in the corresponding submodel. In the general model (2), the constants c_A and c_M and functions $\alpha_A(x)$, $\alpha_M(x)$ are not of much interest, but $\alpha_{Q_{-j}}(.)$ can be interpreted as an average of the higher dimensional surface with respect to Q_{-j}; one can also interpret $\alpha_{Q_{-j}}(.)$ as a projection, albeit with respect to a product measure, see Nielsen and Linton (1998).

The Cox model is a special case of (3) with $h_j(z_j) = \exp(\beta_j z_j)$ for some parameters β_j. Separable nonparametric models have been investigated previously in regression by Hastie and Tibshirani (1990) and in hazard estimation by Andersen, Borgan, Gill, and Keiding (1992).

3 Estimation

We first define a class of estimators of the unrestricted conditional hazard function $\alpha(x)$. Defining the bandwidth parameter b and product kernel function $K_b(u_0, \ldots, u_d) = \prod_{j=0}^d k_b(u_j)$, where $k(\cdot)$ is a one-dimensional kernel with $k_b(u_j) = b^{-1}k(u_j/b)$, we let

$$\widehat{\alpha}(x) = \frac{\frac{1}{n}\sum_{i=1}^n \int_0^T K_b(x - X_i(s))dN_i(s)}{\frac{1}{n}\sum_{i=1}^n \int_0^T K_b(x - X_i(s))Y_i(s)ds} \equiv \frac{\widehat{o}(x)}{\widehat{e}(x)} \tag{7}$$

be our estimator of $\alpha(x)$, a ratio of local occurrence $\widehat{o}(x)$ to local exposure $\widehat{e}(x)$. The estimator $\widehat{\alpha}(x)$ was introduced in Nielsen and Linton (1995) who gave some statistical properties of (7) for general d. When the bandwidth sequence is chosen of order $n^{-1/(2r+d+1)}$, the random variable $\widehat{\alpha}(x) - \alpha(x)$ is asymptotically normal with rate of convergence $n^{-r/(2r+d+1)}$, where r is an index of smoothness of $\alpha(x)$. This is the optimal rate for regression without the separability restrictions, see Stone (1980). We shall be using $\widehat{\alpha}(x)$ as an input into our procedures and will refer to it as the 'pilot' estimator.

We now define a method of estimating the components in (3) and (4) based on the principle of marginal integration. We estimate the contrast by replacing the unknown quantities in (5) by estimators, thus,

$$\widehat{\alpha}_{Q_{-j}}(x_j) = \int \widehat{\alpha}(x)d\widehat{Q}_{-j}(x_{-j}), \tag{8}$$

where $\widehat{\alpha}(x)$ is the unrestricted estimator (7). Here, $\widehat{Q}$ is a probability measure that converges in probability to the distribution Q, while $\widehat{Q}_j$ and $\widehat{Q}_{-j}$, $j = 0, \ldots, d$, are the corresponding marginals. We assume that $\widehat{Q}$ and its marginals are continuous except at a finite number of points, which implies that the integral in (8) is well-defined because $\widehat{\alpha}(.)$ is continuous when K is. Finally, we take $\widehat{c}_{*j} = \int \widehat{\alpha}_{Q_{-j}}(x_j)d\widehat{Q}_j(x_j)$ and $\widehat{c}_{\dagger} = \int \widehat{\alpha}(x)d\widehat{Q}_0(x_0)\cdots d\widehat{Q}_d(x_d)$, and then let

$$\widehat{\alpha}_A(x) = \sum_{j=0}^d \widehat{\alpha}_{Q_{-j}}(x_j) - \frac{d}{d+1}\sum_{j=0}^d \widehat{c}_{*j} \quad \text{and} \quad \widehat{\alpha}_M(x) = \frac{\prod_{j=0}^d \widehat{\alpha}_{Q_{-j}}(x_j)\widehat{c}_{\dagger}}{\prod_{j=0}^d \widehat{c}_{*j}}. \tag{9}$$

The quantities $\widehat{\alpha}_A(x)$ and $\widehat{\alpha}_M(x)$ estimate $\alpha_A(x)$ and $\alpha_M(x)$, respectively, which are both equal to $\alpha(x)$ in the corresponding submodel. For added flexibility, we suggest using a different bandwidth sequence in the estimators $\widehat{c}_{*j}, \widehat{c}_{\dagger}$, this is because we can expect to estimate the constants at rate root-n because the target quantities are integrals over the entire covariate vector.

The distribution $\widehat{Q}$ can essentially be arbitrary, although its support should be contained within the support of the covariates. The most obvious choices of Q seem to be Lebesgue measure on some compact set I or an empirical measure similarly restricted. In this case, one would, for example, compute $\sum_{i=1}^{n} \widehat{\alpha}(t, Z_i(t))Y_i(t)/\sum_{i=1}^{n} Y_i(t)$, or some trimmed version thereof [note that this is asymptotically equivalent to $\int \alpha(t,z)f_t(z)dz$]. There has been some investigation of the choice of weighting in regression, see for example Linton and Nielsen (1995) and Fan, Mammen, and Härdle (1997). Finally, the marginal integration procedures we have proposed are based on high dimensional smoothers, and can suffer some small sample problems if the dimensions are high. Sperlich, Linton, and Härdle (1999) compared the performance of integration and backfitting estimators on simulated data in regression and concluded that both methods suffered from some finite sample deterioration: the backfitting estimators performed better in estimating the total response in highly correlated designs and in boundary regions, while the marginal integration method performed better when estimating the individual effects and in less highly correlated designs.

4 Asymptotic Properties

We derive the asymptotic distribution of the marginal integration estimators $\widehat{\alpha}_{Q_{-j}}$ at interior points under the general sampling scheme (2), i.e., we do not assume either of the separable structures holds. However, when either the additive or multiplicative submodels (3) or (4) are true, our results are about the corresponding univariate components. We are assuming an i.i.d. set-up throughout. We could weaken this along the lines of McKeague and Utikal (1990, condition A), but at the cost of quite complicated notations. We shall assume that the support of $Z_i(s)$ does not depend on s, and is rectangular. This is just to avoid a more complicated notation. We also assume that the estimation region is a strict rectangular subset of the covariate support, and so ignore boundary effects.

For functions $g : \mathbb{R}^d \mapsto \mathbb{R}$, define the Sobolev norm of order s,

$$\|g\|_{d,s,\mathcal{I}}^2 = \sum_{a:|a|\leq s} \int_{\mathcal{I}} \{D^a g(z)\}^2 dz,$$

where $\mathcal{I} \subseteq \mathbb{R}^d$ is a compact set, and let $\mathcal{G}_{d,s}(\mathcal{I})$ be class of all functions with domain $\mathcal{I}$ with Sobolev norm of order s bounded by some constant C. An important step in our argument is to replace $\widehat{Q}$ by Q; we shall use empirical process arguments to show this. Define the stochastic process $\nu_n(\cdot)$ by

$$\nu_n(g) = \sqrt{n}\left\{\int_{I_{-j}} g(z)d\widehat{Q}_{-j}(z) - \int_{I_{-j}} g(z)dQ_{-j}(z)\right\}$$

for any $g \in \mathcal{G}_{d,s}(I_{-j})$, where the set I_{-j} is specified below. We make the following assumptions:

(A1) The covariate process is supported on the compact set $\mathcal{X} = [0,T] \times \mathcal{Z}$, where $\mathcal{Z} = \mathcal{Z}_1 \times \cdots \times \mathcal{Z}_d$. For each $t \in [0,T]$, define the conditional [given $Y_i(s) = 1$] distribution function of the observed covariate process $F_t(z) = \Pr(Z_i(t) \le z | Y_i(t) = 1)$, and let $f_t(z)$ the corresponding density with respect to Lebesgue measure. Let $x = (t,z) \in I$ and define the exposure $e(x) = f_t(z)y(t)$, where $y(t) = E[Y_i(t)]$. The functions $t \mapsto y(t)$ and $t \mapsto f_t(z)$ are continuous on $[0,T]$ for all $z \in \mathcal{Z}$.

(A2). The probability measure Q is absolutely continuous with respect to Lebesgue measure and has density function q. It has support on the compact interval $I = I_0 \times \cdots \times I_d$, strictly contained in $\mathcal{Z}$. Define also the marginals Q_j, Q_{-j} and their continuous densities q_j and q_{-j} whose supports are I_j, I_{-j} respectively. Furthermore, $0 < \inf_{x_j \in I_j} q_j(x_j), \inf_{x_{-j} \in I_{-j}} q_{-j}(x_{-j})$.

(A3) The functions $\alpha(\cdot)$ and $e(\cdot)$ are r-times continuously differentiable on I and satisfy $\inf_{x \in I} e(x) > 0$ and $\inf_{x \in I} \alpha(x) > 0$. The integer r satisfies $(2r+1)/3 > (d+1)$.

(A4) The kernel k has support $[-1,1]$, is symmetric about zero, and is of order r, that is, $\int_{-1}^{1} k(u)u^j du = 0$, $j = 1, \ldots, r-1$ and $\int_{-1}^{1} k(u)u^r du \in (0,\infty)$, where $r \ge 2$ is an even integer. The kernel is also $r-1$ times continuously differentiable on $[-1,1]$ with Lipschitz remainder, i.e., there exists a finite constant k_{lip} such that $\left|k^{(r-1)}(u) - k^{(r-1)}(u')\right| \le k_{lip}|u-u'|$for all u, u'. Finally, $k^{(j)}(\pm 1) = 0$ for $j = 0, \ldots, r-1$. Define the kernel moments: $\mu_r(k) = \int_{-1}^{1} u^r k(u) du$ and $\|k\|_2^2 = \int_{-1}^{1} k(u)^2 du$.

(A5) The probability measure $\widehat{Q}$ has support on I and satisfies $\sup_{x \in I} |\widehat{Q}(x) - Q(x)| = O_p(n^{-1/2})$. Furthermore, for some s with $r \ge s > d/2$, the empirical process $\{\nu_n(\cdot) : n \ge 1\}$ is stochastically equicontinuous on $\mathcal{G}_{d,s}(I_{-j})$ at $g_0(\cdot) = \alpha(x_j, \cdot)$, i.e., for all $\epsilon, \eta > 0$ there exists $\delta > 0$ such that

$$\limsup_{n\to\infty} \mathbf{P}^* \left[\sup_{g \in \mathcal{G}_{d,s}(I_{-j}), \|g - g_0\|_{d,s,I_{-j}} \le \delta} |\nu_n(g) - \nu_n(g_0)| > \eta \right] < \epsilon, \tag{10}$$

where $\mathbf{P}^*$ denotes outer probability.

The smoothness and boundedness conditions in A1-A4 are fairly standard in nonparametric estimation. Our assumptions are strictly stronger than those of McKeague and Utikal (1990), and indeed imply the conditions of their Proposition 1. In particular, we assume smoothness of e with respect to all its arguments rather than just continuity. This is necessary for the arguments we use to establish the limiting distribution of our estimator, although for consistency only continuity in t would suffice, and perhaps for other estimators like local linear etc. we might require less smoothness. In any case, this condition is likely to hold for a large class of covariate processes. Certainly, time invariant covariates can be expected to satisfy this condition. When Z is the time since a certain event, such as onset of disability, we can model the stochastic process $Z_i(t)$ as $Z_i(t) = t - Z_{0i}$ for some random variable Z_{0i} that represents the age at which disability occurred. This is essentially as in McKeague and Utikal (1990, Example 5, p 1180 especially), and under smoothness conditions on their α_{jk} we obtain the smoothness of (in our notation) the corresponding exposure $e(x)$. The restriction on (r, d) is used to ensure that in the expansion of $\widehat{\alpha} - \alpha$ the second order terms are small in comparison with the leading terms; these terms are of order $n^{-1}b^{-(d+1)}\log n + b^{2r}$, so we must have $r > d$. We require slightly stronger restrictions in order to deal with the passage from $\widehat{Q}$ to Q. The stochastic equicontinuity condition in A5 is satisfied under conditions on the entropy of the class of functions, see van de Geer (2000).

We have

$$(\widehat{\alpha} - \alpha)(x) = (\widehat{\alpha} - \alpha^*)(x) + (\alpha^* - \alpha)(x) = \frac{V_n(x) + B_n(x)}{\widehat{e}(x)},$$

where the compensator $\alpha^*(x)$ is

$$\alpha^*(x) = \frac{\sum_{i=1}^n \int_0^T K_b(x - X_i(s))\alpha(X_i(s))Y_i(s)ds}{\sum_{i=1}^n \int_0^T K_b\,(x - X_i(s))Y_i(s)\,ds},$$

while:

$$V_n(x) = \tfrac{1}{n}\textstyle\sum_{i=1}^n \int_0^T K_b(x - X_i(s))\,dM_i(s)$$

$$B_n(x) = \tfrac{1}{n}\textstyle\sum_{i=1}^n \int_0^T K_b(x - X_i(s))\,[\alpha(X_i(s)) - \alpha(x)]\,Y_i(s)ds.$$

Both $B_n(x)$ and $\widehat{e}(x)$ are sums of independent variables and satisfy the conditions of the law of large numbers. We can apply a martingale limit theorem to $V_n(x)$ because the integrand is predictable. This decomposition was used in Nielsen and Linton (1995) to establish the limiting behavior of $(\widehat{\alpha} - \alpha)(x)$: they established pointwise asymptotic normality and also showed that

$$\sup_{x \in I} |\widehat{\alpha}(x) - \alpha(x)| = O_P(\frac{1}{n^{1/2} b^{(d+3)/2}}) + O_P(b^r). \tag{11}$$

We improve on their rate in Lemma 3 in the appendix - we achieve the rate $(\log n) n^{-1/2} b^{-(d+1)/2} + b^r$.

We now turn to the behavior of $\widehat{\alpha}_{Q_{-j}}$. It suffices to work with the stochastic integrator $\widehat{Q}_{-j}$ replaced by the deterministic Q_{-j}, since under our conditions

$$\int \widehat{\alpha}(x) d\widehat{Q}_{-j}(x_{-j}) - \int \widehat{\alpha}(x) dQ_{-j}(x_{-j}) = \int \alpha(x) d\widehat{Q}_{-j}(x_{-j}) - \int \alpha(x) dQ_{-j}(x_{-j}) + o_p(n^{-1/2}) \tag{12}$$

$$= O_P(n^{-1/2}). \tag{13}$$

The proof of (12) is given in the appendix, while (13) follows directly from our assumptions about α and $\widehat{Q}$. By crude bounding we have

$$\begin{aligned} \sup_{x_j \in I_j} \left| \widehat{\alpha}_{Q_{-j}}(x_j) - \alpha_{Q_{-j}}(x_j) \right| &= \sup_{x_j \in I_j} \left| \int_{I_{-j}} \widehat{\alpha}(x) dQ_{-j}(x_{-j}) - \int_{I_{-j}} \alpha(x) dQ_{-j}(x_{-j}) \right| + O_P(n^{-1/2}) \\ &\leq \sup_{x \in I} |\widehat{\alpha}(x) - \alpha(x)| \times \sup_{x_{-j} \in I_{-j}} |q_{-j}(x_{-j})| \times \mathrm{Vol}(I) + O_P(n^{-1/2}), \end{aligned}$$

i.e., $\widehat{\alpha}_{Q_{-j}}(x_j)$ has no worse rate of uniform convergence than the multidimensional estimator $\widehat{\alpha}(x)$. We are interested in obtaining the faster one-dimensional rate of convergence for $\widehat{\alpha}_{Q_{-j}}(x_j)$, which requires a more careful analysis. Defining the marginal compensators $\alpha^*_{Q_{-j}}(x_j) = \int_{I_{-j}} \alpha^*(x) dQ_{-j}(x_{-j})$, we have

$$(\widehat{\alpha}_{Q_{-j}} - \alpha_{Q_{-j}})(x_j) = (\widehat{\alpha}_{Q_{-j}} - \alpha^*_{Q_{-j}})(x_j) + (\alpha^*_{Q_{-j}} - \alpha_{Q_{-j}})(x_j) = V_{Q_{-j}}(x_j) + B_{Q_{-j}}(x_j), \tag{14}$$

where

$$V_{Q_{-j}}(x_j) = \frac{1}{n} \sum_{i=1}^{n} \int_0^T H_i^{(n)}(x_j, s) dM_i(s) \quad ; \quad B_{Q_{-j}}(x_j) = \int_{I_{-j}} \frac{B_n(x)}{\widehat{e}(x)} dQ_{-j}(x_{-j}),$$

with

$$H_i^{(n)}(x_j,s)=\int_{I_{-j}}\frac{K_b(x-X_i(s))}{\widehat{e}(x)}dQ_{-j}(x_{-j}).$$

The bias expression $B_{Q_{-j}}(x_j)$ is really just the integrated bias of the pilot estimator, and it is of the same magnitude in probability.

We now outline the asymptotics of $V_{Q_{-j}}(x_j)$. Let us start with the following naive heuristics. First, it is verified that $\sup_{x_{-j}\in I_{-j}}|\widehat{e}(x)-e(x)|=o_P(1)$, then this fact is used to obtain the approximation

$$H_i^{(n)}(x_j,s)=\int_{I_{-j}}\frac{K_b(x-X_i(s))}{\widehat{e}(x)}dQ_{-j}(x_{-j})=\widetilde{H}_i^{(n)}(x_j,s)\left\{1+o_P(1)\right\},$$

where

$$\widetilde{H}_i^{(n)}(x_j,s)=\int_{I_{-j}}\frac{K_b(x-X_i(s))}{e(x)}dQ_{-j}(x_{-j}).$$

It therefore seems reasonable to approximate $V_{Q_{-j}}(x_j)$ by $\widetilde{V}_{Q_{-j}}(x_j)=n^{-1}\sum_{i=1}^n\int_0^T\widetilde{H}_i^{(n)}(x_j,s)dM_i(s)$, which can be analyzed by standard martingale theory, since $\widetilde{H}_i^{(n)}$ is predictable with respect to the chosen filtration. The last approximation is a bit tricky, because M_i is a signed measure. The correct proof requires an argument that takes into account the nature of signed measures (or martingales). The solution to the predictability issue presented in Nielsen, Linton and Bickel (1998), and updated in Nielsen (1999), provides us with such an argument. The asymptotic variance of $\widetilde{V}_{Q_{-j}}(x_j)$ is obtained from a change of variable argument and dominated convergence.

We next state our main result, which follows from the above arguments and some more detailed calculations presented in the appendix. Our main theorem gives the pointwise distribution of the marginal integration estimator $\widehat{\alpha}_{Q_{-j}}(x_j)$ and the corresponding additive and multiplicative reconstructions $\widehat{\alpha}_A(x),\widehat{\alpha}_M(x)$. As discussed earlier, we do not maintain either separability hypothesis in this theorem, and so the result is about the functionals of the underlying function $\alpha(x)$ defined in (6). We first need some notation. For any function $f(x)$ of a vector argument x, let $f_j^{(r)}(x)=\partial^r f(x)/\partial x_j^r$ and define for any integer r and $j=0,\ldots,d$:

$$\beta_j^{(r)}(x)=[(\alpha\cdot e)_j^{(r)}(x)-\alpha(x)\cdot e_j^{(r)}(x)]/e(x) \tag{15}$$

and $\beta^{(r)}(x)=\sum_{j=0}^d\beta_j^{(r)}(x)$.

Theorem 1. *Suppose that assumptions A1-A4 hold and that* $n^{1/(2r+1)}b\to\gamma$ *for some* $0<\gamma<\infty$. *Then,*

$$n^{r/(2r+1)}(\widehat{\alpha}_{Q_{-j}} - \alpha_{Q_{-j}})(x_j) \Longrightarrow N\left[m_j(x_j), \upsilon_j(x_j)\right], \tag{16}$$

where

$$m_j(x_j) = \frac{\mu_r(k)}{r!}\gamma^r \int_{I_{-j}} \sum_{l=0}^{d} \beta_l^{(r)}(x) dQ_{-j}(x_{-j}) \quad ; \quad \upsilon_j(x_j) = \gamma^{-1} \|k\|_2^2 \int_{I_{-j}} \frac{\alpha(x) q_{-j}^2(x_{-j})}{e(x)} dx_{-j}. \tag{17}$$

Suppose also that $\widehat{c}_{*j}, \widehat{c}_{\dagger}$ *are root-n consistent, then*

$$n^{r/(2r+1)}(\widehat{\alpha}_A - \alpha_A)(x) \quad \Longrightarrow \quad N\left[m(x), v(x)\right] \tag{18}$$

$$n^{r/(2r+1)}(\widehat{\alpha}_M - \alpha_M)(x) \quad \Longrightarrow \quad N\left[\overline{m}(x), \overline{v}(x)\right], \tag{19}$$

where $m(x) = \sum_{j=0}^{d} m_j(x_j)$ *and* $v(x) = \sum_{j=0}^{d} \upsilon_j(x_j)$*, while* $\overline{m}(x) = \sum_{j=0}^{d} m_j(x_j) s_j(x_{-j})$ *and* $\overline{v}(x) = \sum_{j=0}^{d} \upsilon_j(x_j) s_j^2(x_{-j})$*, where* $s_j(x_{-j}) = \prod_{k \neq j} \alpha_{Q_{-k}}(x_k) / c_M^d$ *, Furthermore,* $\widehat{\upsilon}_j(x_j) \to_p \upsilon_j(x_j)$*,* $\widehat{v}(x) \to_p v(x)$*, and* $\widehat{\overline{v}}(x) \to_p \overline{v}(x)$*, where* $\widehat{\upsilon}_j(x_j) = n^{-1} \sum_{i=1}^{n} \int_0^T \{H_i^{(n)}(x_j, s)\}^2 dN_i(s)$*, while* $\widehat{v}(x) = \sum_{j=0}^{d} \widehat{\upsilon}_j(x_j)$ *and* $\widehat{\overline{v}}(x) = \sum_{j=0}^{d} \widehat{\upsilon}_j(x_j) \widehat{s}_j^2(x_{-j})$ *with* $\widehat{s}_j(x_{-j}) = \prod_{k \neq j} \widehat{\alpha}_{Q_{-k}}(x_k) / \widehat{c}_M^d$.

The bandwidth rate $b \sim n^{-1/(2r+1)}$ gives an optimal [pointwise mean squared error] rate of convergence for $\widehat{\alpha}_{Q_{-j}}(x_j)$, $\widehat{\alpha}_A(x)$, and $\widehat{\alpha}_M(x)$ [i.e., this is the same rate as the optimal rate of convergence in one-dimensional kernel regression estimation, see Stone (1980)].

An earlier version of this paper was the starting point for a number of actuarial applications, see for example Nielsen and Voldsgaard (1996). The working paper version of this paper, Linton, Nielsen and van de Geer (1999), contains some numerical results and comparison with a backfitting extension of our procedure.

5 Appendix

We first state and prove three preliminary results that are needed in the proof of our theorem. Lemma 1 and Lemma 2 are used to establish the uniform convergence of the pilot estimator $\widehat{\alpha}$ and the exposure $\widehat{e}$ in Lemma 3. The uniform convergence of $\widehat{e}$ is used throughout the proof of Theorem 1 to establish that the remainder terms in the expansion are of smaller order. The uniform convergence of $\widehat{\alpha}$ and its derivatives are needed to ensure that (12) is true.

We use the following convenient notation: for two random variables X_n, Y_n, we say that $X_n \simeq Y_n$ whenever $X_n = Y_n(1 + o_p(1))$ as $n \to \infty$.

Preliminary Results

We first establish the following exponential inequality, which is a version of Bernstein's inequality for sums of independent martingales. This is used in establishing the uniform convergence of $\widehat{\alpha}$, which is the second result of this section.

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability triple, and let $\{\mathcal{F}_t\}_{t \geq 0}$ be a filtration satisfying the *usual conditions*. Consider n independent martingales $M_1, \ldots, M_n$. Let $V_{2,i}$ be the predictable variation of M_i, and let $V_{m,i}$ be the m^{th} order variation process of M_i, $i = 1, \ldots, n$, $m = 3, 4, \ldots$.

LEMMA 1. *Fix $0 < T \leq \infty$ and suppose that for some $\mathcal{F}_T$-measurable random variable $R_n^2(T)$ and some constant $0 < K < \infty$, one has*

$$\sum_{i=1}^{n} V_{m,i}(T) \leq \frac{m!}{2} K^{m-2} R_n^2(T). \tag{20}$$

Then, for all $a > 0$, $b > 0$,

$$\Pr\left(\sum_{i=1}^{n} M_i(T) \geq c \text{ and } R_n^2(T) \leq d^2 \right) \leq \exp[-\frac{c^2}{2(cK + d^2)}]. \tag{21}$$

PROOF. Define for $0 < \lambda < 1/K$, $i = 1, \ldots, n$,

$$Z_i(t) = \lambda M_i(t) - S_i(t), \ t \geq 0,$$

where S_i is the compensator of

$$W_i = \frac{1}{2}\lambda^2 < M_i^c, M_i^c > + \sum_{s \leq \cdot} \left(\exp[\lambda|\Delta M_i(s)|] - 1 - \lambda|\Delta M_i(s)| \right).$$

Then $\exp Z_i$ is a supermartingale, $i = 1, \ldots, n$ [see the proof of Lemma 2.2 in van de Geer (1995)]. So $E \exp Z_i(T) \leq 1, \ i = 1, \ldots, n$. But then also $E \exp[\sum_{i=1}^{n} Z_i(T)] \leq 1$. One easily verifies that

$$\sum_{i=1}^{n} S_i(T) \leq \frac{\lambda^2 R_n^2(T)}{2(1 - \lambda K)}.$$

So on the set

$$A = \left\{ \sum_{i=1}^{n} M_i(T) \geq c \text{ and } R_n^2(T) \leq d^2 \right\},$$

one has

$$\exp[\sum_{i=1}^{n} Z_i(T)] \geq \exp[\lambda c - \frac{\lambda^2 d^2}{2(1-\lambda K)}].$$

Therefore,

$$\Pr(A) \leq \exp[-\lambda c + \frac{\lambda^2 d^2}{2(1-\lambda K)}].$$

The result follows by choosing

$$\lambda = \frac{c}{d^2 + Kc}.$$

■

This result is formulated for fixed T, and K may depend on T and n. If the conditions of Lemma 1 hold for all T, n, then it can be extended to stopping times [see section 8.2 in van de Geer (2000) for related results].

In the next lemma, we assume as in the main text that T is fixed and finite, and write $\int = \int_0^T$. We also assume that the $\Lambda_i^n(t)$ exist, and are bounded by a (nonrandom) constant $\overline{\Lambda}$ for all $1 \leq i \leq n$ and $0 \leq t \leq T$.

LEMMA 2. *Let Θ be a bounded subset of $\mathbb{R}^{d+1}$, and for each $\theta \in \Theta$, consider independent predictable functions $g_{1,\theta}, \ldots, g_{n,\theta}$. Suppose that for some constants L_n, K_n, and $\rho_n \geq 1$, we have*

$$|g_{i,\theta}(t) - g_{i,\tilde{\theta}}(t)| \leq L_n |\theta - \tilde{\theta}|, \text{ for all } \theta, \tilde{\theta} \in \Theta, \text{ and all } i \geq 1 \text{ and } t \geq 0, \tag{22}$$

$$|g_{i,\theta}(t)| \leq K_n, \text{ for all } \theta \in \Theta, \text{ and all } i \geq 1 \text{ and } t \geq 0,$$

$$\frac{1}{n} \sum_{i=1}^{n} \int |g_{i,\theta}(t)|^2 dt \leq \rho_n^2, \text{ for all } \theta \in \Theta, \text{ and all } n > 1,$$

$$L_n \leq n^{\nu}, \text{ for all } n > 1, \text{ and some } \nu < \infty, \tag{23}$$

and

$$K_n \leq \sqrt{\frac{n}{\log n}} \rho_n, \text{ for all } n > 1. \tag{24}$$

Then for some constant c_0, we have for all $C \geq c_0$, and $n > 1$

$$\Pr\left(\sup_{\theta \in \Theta} \frac{1}{\sqrt{n}} \left| \sum_{i=1}^{n} \int g_{i,\theta} d(N_i^{(n)} - \Lambda_i^{(n)}) \right| \geq C \rho_n \sqrt{\log n} \right) \leq c_0 \exp[-\frac{C \log n}{c_0}].$$

PROOF. From Lemma 1, we know that for each $\theta \in \Theta$, $a > 0$ and $R > 0$

$$\Pr\left(\frac{1}{\sqrt{n}} \left| \sum_{i=1}^{n} \int g_{i,\theta} d(N_i^{(n)} - \Lambda_i^{(n)}) \right| \geq a \text{ and } \frac{1}{n} \sum_{i=1}^{n} \int g_{i,\theta}^2 d\Lambda_i^{(n)} \leq R^2 \right) \tag{25}$$

$$\leq 2\exp[-\frac{a^2}{2(aK_n n^{-\frac{1}{2}} + R^2)}].$$

Let $\epsilon > 0$ to be chosen later, and let $\{\theta_1, \ldots, \theta_N\} \subset \Theta$ be such that for each $\theta \in \Theta$, there is a $j(\theta) \in 1, \ldots, N$, such that $|\theta - \theta_{j(\theta)}| \leq \epsilon$. Then, by the Lipschitz condition (22), one has

$$\frac{1}{\sqrt{n}}|\sum_{i=1}^{n} \int (g_{i,\theta} - g_{i,\theta_{j(\theta)}}) d(N_i^{(n)} - \Lambda_i^{(n)})| \leq \sqrt{n} L_n \epsilon (1 + \bar{\Lambda}),$$

where $\bar{\Lambda}$ is an upper bound for $\Lambda_i^{(n)}(t)$, $1 \leq i \leq n$, $n \geq 1$, $t \geq 0$.

Now, in (25), take $a = C\rho_n \sqrt{\log n}/2$, and $R_n^2 = \rho_n^2 \bar{\lambda}$, with $\bar{\lambda}$ an upper bound for $\lambda_i^{(n)}(t)$, $1 \leq i \leq n$, $n \geq 1$, $t \geq 0$. Moreover, take $\epsilon = a/(\sqrt{n} L_n (1 + \bar{\Lambda}))$. With these values, we find

$$\Pr\left(\sup_{\theta \in \Theta} \frac{1}{\sqrt{n}}|\sum_{i=1}^{n} \int g_{i,\theta} d(N_i^{(n)} - \Lambda_i^{(n)})| \geq C\rho_n \sqrt{\log n}\right)$$

$$= \Pr\left(\sup_{\theta \in \Theta} \frac{1}{\sqrt{n}}|\sum_{i=1}^{n} \int g_{i,\theta} d(N_i^{(n)} - \Lambda_i^{(n)})| \geq 2a\right)$$

$$\leq \Pr\left(\max_{j=1,\ldots,N} \frac{1}{\sqrt{n}}|\sum_{i=1}^{n} \int g_{i,\theta_j} d(N_i^{(n)} - \Lambda_i^{(n)})| \geq a\right)$$

$$\leq 2\exp[\log N - \frac{a^2}{2(aK_n n^{-\frac{1}{2}} + \rho_n^2 \bar{\lambda}}].$$

Because Θ is a bounded, finite-dimensional set, we know that for some constant c_1,

$$\log N \leq c_1 \log(\frac{1}{\epsilon}).$$

By our choice $\epsilon = C\rho_n \sqrt{\log n}/(2\sqrt{n} L(1 + \bar{\Lambda}))$, and using condition (23), we see that for $C \geq 1$ (say) and some constant c_2,

$$\log N \leq c_2 \log n.$$

Invoking moreover condition (24), we arrive at

$$\Pr\left(\sup_{\theta \in \Theta} \frac{1}{\sqrt{n}}|\sum_{i=1}^{n} \int g_{i,\theta} d(N_i^{(n)} - \Lambda_i^{(n)})| \geq C\rho_n \sqrt{\log n}\right)$$

$$\leq 2\exp[c_2 \log n - \frac{C^2 \rho_n^2 \log n}{8(C\rho_n \sqrt{\log n} K_n n^{-\frac{1}{2}}/2 + \rho_n^2 \bar{\lambda})}]$$

$$\leq 2\exp[c_2\log n - \frac{C^2\log n}{8(C/2+\bar{\lambda})}]$$

$$\leq 2\exp[c_2\log n - \frac{C\log n}{8}] \leq 2\exp[-\frac{C\log n}{16}],$$

where in the last two steps, we take $C \geq 2\bar{\lambda}$, and $C \geq 16c_2$.

■

Note that by the continuity of (22) and the boundedness of Θ, the statement of Lemma 2 does not give rise to measurability problems. Note moreover that (22)-(24) imply that K_n, ρ_n, and L_n cannot be chosen in an arbitrary manner. Most important here is that the sup-norm should not grow too fast as compared to the L_2 norm.

LEMMA 3. *Suppose that the assumptions stated in Theorem 1 hold. Then, for any $a = (a_0, \ldots, a_d)$ with $|a| \leq r-1$, we have*

(a) $\sup_{x\in I}|D^a\widehat{e}(x) - D^a e(x)| = O_P\left(b^{r-|a|}\right) + O_P\left\{\left(\frac{\log n}{nb^{d+1+2|a|}}\right)^{1/2}\right\}$

(b) $\sup_{x\in I}|D^a\widehat{\alpha}(x) - D^a\alpha(x)| = O_P\left(b^{r-|a|}\right) + O_P\left\{\left(\frac{\log n}{nb^{d+1+2|a|}}\right)^{1/2}\right\}.$

PROOF. We write $D^a\widehat{e}(x) - D^a e(x) = D^a\widehat{e}(x) - ED^a\widehat{e}(x) + ED^a\widehat{e}(x) - eD^a(x)$, a decomposition into a 'stochastic' part $D^a\widehat{e}(x) - ED^a\widehat{e}(x)$ and a 'bias' part $ED^a\widehat{e}(x) - D^a e(x)$. Nielsen and Linton (1995,) showed, for the case $a = 0$, that $ED^a\widehat{e}(x) - D^a e(x) = O\left(b^r\right)$ for any interior point x. By identical distribution we have

$$ED^a\widehat{e}(x) = E\int_0^T D^a K_b(x - X_i(s))Y_i(s)ds = \int D^a K_b(x-x')e(x')dx',$$

where the last integral is over $\mathcal{X}$. Integrating by parts [using the fact that $k^{(j)}(\pm 1) = 0$] and changing variables $x' \longmapsto u = (x'-x)/b$, we have

$$ED^a\widehat{e}(x) = \int K_b(x-x')D^a e(x')dx' = \int_{[-1,1]^{d+1}} K(u)D^a e(x+bu)du,$$

where the last equality follows for large enough n because x is an interior point and so eventually the range of u contains the support of K. For notational simplicity we now suppose that $a = 0$. By Taylor expansion

$$e(x+bu) - e(x) = \sum_{j=1}^{r}\sum_{\{c:|c|=j\}}\frac{b^c}{c!}u^c D^c e(x) + \sum_{\{c:|c|=r\}}\frac{b^c}{c!}u^c\left\{D^c e(x^*(u)) - D^c e(x)\right\},$$

where $x^*(u) = (x_0^*(u), \ldots, x_d^*(u))$ are intermediate values satisfying $|x_j^* - x_j| < b|u_j|$, for $j = 0, \ldots, d$. Therefore, we have

$$
\begin{aligned}
E\widehat{e}(x) - e(x) &= \sum_{j=1}^{r} \sum_{\{c:|c|=j\}} \frac{b^c}{c!} D^c e(x) \int_{[-1,1]^{d+1}} u^c K(u) du \\
&\quad + \sum_{\{c:|c|=r\}} \frac{b^c}{c!} \int_{[-1,1]^{d+1}} \{D^c e(x^*(u)) - D^c e(x)\} K(u) u^c du,
\end{aligned}
$$

where

$$
\sum_{j=1}^{r} \sum_{\{c:|c|=j\}} \frac{b^c}{c!} D^c e(x) \int_{[-1,1]^{d+1}} u^c K(u) du = \frac{b^r}{r!} \sum_{j=1}^{r} \frac{\partial^r e(x)}{\partial x_j^r} \int_{-1}^{1} s^r k(s) ds
$$

by assumption A4, while

$$
\begin{aligned}
&\left| \sum_{\{c:|c|=r\}} \frac{b^c}{c!} \int_{[-1,1]^{d+1}} u^c \{D^c e(x^*(u)) - D^c e(x)\} K(u) du \right| \\
\leq & \sum_{\{c:|c|=r\}} \frac{b^c}{c!} \int_{[-1,1]^{d+1}} u^c |D^c e(x^*(u)) - D^c e(x)| K(u) du \\
= & \; o(b^r),
\end{aligned}
$$

because of the continuity of $D^c e(x)$ for $|c| = r$ [we can apply dominated convergence because $|D^c e(x^*(u)) - D^c e(x)|$ is bounded for all x, u]. In conclusion

$$
\left| E\widehat{e}(x) - e(x) - \frac{b^r}{r!} \sum_{j=1}^{r} \frac{\partial^r e(x)}{\partial x_j^r} \int_{-1}^{1} s^r k(s) ds \right| = o(b^r),
$$

and this result holds uniformly over $x \in I$ because $D^c e(.)$ is uniformly continuous over I. The bias term for $D^a \widehat{e}$ is of order $b^{r-|a|}$ because the partial integration claims some of the kernel moments.

We now turn to the stochastic part of $\widehat{e}(x)$. We claim that

$$
\begin{aligned}
\sup_{x \in I} |\widehat{e}(x) - E\widehat{e}(x)| &= \sup_{x \in I} \left| \frac{1}{n} \sum_{i=1}^{n} \int_0^T [K_b(x - X_i(s)) Y_i(s) - E(K_b(x - X_i(s)) Y_i(s))] ds \right| \\
&= O_P\left\{ \left(\frac{\log n}{nb^{d+1}} \right)^{1/2} \right\}. \qquad (26)
\end{aligned}
$$

The pointwise result [without the logarithmic factor] is given in Nielsen and Linton (1995). The uniformity [at the cost of the logarithmic factor] follows by standard arguments, the key component of which is the application of an exponential inequality like that obtained in Lemma 2 above. We write $\widehat{e}(x) - E\widehat{e}(x) = \sum_{i=1}^{n} \zeta_{n,i}^{c}(x)$, where $\zeta_{n,i}^{c}(x) = \zeta_{n,i}(x) - E\zeta_{n,i}(x)$ with $\zeta_{n,i}(x) = n^{-1}\int_0^T K_b(x - X_i(s))Y_i(s)ds$. Note that $\zeta_{n,i}^{c}(x)$ are independent and mean zero random variables with $m_n = \sup_{x,i}|\zeta_{n,i}^{c}(x)| = c_1 n^{-1} b^{-(d+1)}$ for some constant c_1; thus m_n is uniformly bounded because $nb^{d+1} \to \infty$ by assumption. Following Nielsen and Linton (1995), we have $\sigma_{ni}^2 = \mathrm{var}[\zeta_{n,i}^{c}(x)] \leq c_2 n^{-1} b^{-(d+1)}$ for some constant c_2. Let $\{B(x_1, \epsilon_L), \ldots, B(x_L, \epsilon_L)\}$ be a cover of I, where $B(x_\ell, \epsilon)$ is the ball of radius ϵ centered at x_ℓ. Hence, $\epsilon_L = c_3/L$ for some constant c_3. We have for some constant c_4

$$\begin{aligned}
\sup_{x\in I}\left|\sum_{i=1}^{n}\zeta_{n,i}^{c}(x)\right| &\leq \max_{1\leq \ell\leq L}\left|\sum_{i=1}^{n}\zeta_{n,i}^{c}(x_\ell)\right| + \max_{1\leq \ell\leq L}\sup_{x\in B(x_\ell,\epsilon)}\sum_{i=1}^{n}\left|\zeta_{n,i}^{c}(x_\ell) - \zeta_{n,i}^{c}(x)\right| \\
&\leq \max_{1\leq \ell\leq L}\left|\sum_{i=1}^{n}\zeta_{n,i}^{c}(x_\ell)\right| + \frac{c_4\epsilon_L}{b^{2d+2}}
\end{aligned}$$

using the differentiability of k. Provided

$$\epsilon_L\sqrt{\frac{n}{b^{3d+3}\log n}} \to 0, \tag{27}$$

we have by the Bonferroni and Bernstein inequalities

$$\begin{aligned}
\Pr\left(\sqrt{\frac{nb^{d+1}}{\log n}}\max_{1\leq \ell\leq L}\left|\sum_{i=1}^{n}\zeta_{n,i}^{c}(x_\ell)\right| > \lambda\right) &\leq \sum_{\ell=1}^{L}\Pr\left(\left|\sum_{i=1}^{n}\zeta_{n,i}^{c}(x_\ell)\right| > \lambda\sqrt{\frac{\log n}{nb^{d+1}}}\right) + o(1) \\
&\leq \sum_{\ell=1}^{L}\exp\left(-\frac{\lambda^2\frac{\log n}{nb^{d+1}}}{2c_2\frac{1}{nb^{d+1}} + \frac{c_1}{nb^{d+1}}\lambda\sqrt{\frac{\log n}{nb^{d+1}}}}\right) \\
&= \sum_{\ell=1}^{L}\exp\left(-(\log n^{\lambda^2/2c_2})\right).
\end{aligned}$$

By taking λ large enough the latter probability goes to zero fast enough to kill $L(n) = n^{\kappa}$ with $\kappa = 1 + \eta + (3d+3)/(2r+1)$ for some $\eta > 0$, and this choice of L satisfies (27). The result for general a follows the same pattern; differentiation to order a changes K to K^a and adds an additional bandwidth factor of order $b^{-2|a|}$.

To establish (b) we first write $\widehat{\alpha}(x) = \widehat{o}(x)/\widehat{e}(x)$ and $\alpha(x) = o(x)/e(x)$, where $o(x) = \alpha(x)e(x)$. Then by the chain rule we have

$$D^a\widehat{\alpha}(x) - D^a\alpha(x) = \sum_{|c|+|d|=|a|}\kappa_{c,d}\{D^c\widehat{o}(x)D^d\widehat{e}^{-1}(x) - D^c o(x)D^d e^{-1}(x)\}$$

$$= \sum_{|c|+|d|=|a|} \kappa_{c,d}\{D^c\widehat{o}(x) - D^c o(x)\} D^d e^{-1}(x) +$$
$$\sum_{|c|+|d|=|a|} \kappa_{c,d}\{D^d\widehat{e}^{-1}(x) - D^d e^{-1}(x)\} D^c o(x) +$$
$$\sum_{|c|+|d|=|a|} \kappa_{c,d}\{D^d\widehat{e}^{-1}(x) - D^d e^{-1}(x)\}\{D^c\widehat{o}(x) - D^c o(x)\}$$

for vectors c, d and positive finite constants $\kappa_{c,d}$. By further application of the chain rule and, Lemma 3(a), and the assumption that $e(x)$ is bounded away from zero on I, we see that the second term is of the same magnitude as $D^c\widehat{e}(x) - D^c e(x)$. Therefore,

$$\sup_{x\in I} |D^a\widehat{\alpha}(x) - D^a\alpha(x)| \leq \kappa \sum_{|c|\leq|a|} \sup_{x\in I} |D^c\widehat{o}(x) - D^c o(x)| + O_P\left(b^{r-|a|}\right) + O_P\left\{\left(\frac{\log n}{nb^{d+1+2|a|}}\right)^{1/2}\right\}$$

for some positive finite constant κ, and it suffices to establish the result for the numerator statistic $D^c\widehat{o}(x) - D^c o(x)$ only. Again, we shall just work out the details for the case $a = 0$. The bias calculation $E\widehat{o}(x) - o(x)$ is as for $E\widehat{e}(x) - e(x)$ discussed above. Therefore, it suffices to show that $\sup_{x\in I} |\widehat{o}(x) - E\widehat{o}(x)| = \sup_{x\in I} |V_n(x)|$ is the stated magnitude, where $V_n(x) = n^{-1}\sum_{i=1}^n \int_0^T K_b(x - X_i(s))d(N_i(s) - \Lambda_i(s))$. We now apply Lemma 2 with $g_{i,\theta}(t) = K_b(x - X_i(t))$, $\theta = x$, and $\Theta = I$. Conditions (22)-(24) hold with probability tending to one for some constant γ and: $K_n = \gamma \cdot b^{-(d+1)}$, $L_n = \gamma \cdot b^{-2(d+1)}$, and $\rho_n^2 = \gamma \cdot b^{-(d+1)}$ by the boundedness and differentiability of the kernel. It now follows that for some constant c_0 we have for all $C \geq c_0$ and $n > 1$

$$\Pr\left[\sup_{x\in I} \sqrt{\frac{nb^{d+1}}{\log n}} |V_n(x)| \geq C\right] \leq c_0 \exp(-C\log n/c_0)$$

as required. ■

In the proof of Theorem 1 we have to deal with random variables of the form

$$\overline{M}_t = \sum_{i=1}^n \int_0^t h_i^{(n)}(u)dM_i(u),$$

where M_i is a martingale, but $h_i^{(n)}$ is not a predictable process according to the usual definition. We must replace $h_i^{(n)}$ by $\widetilde{h}_i^{(n)}$, where the $\widetilde{h}_i^{(n)}{}'s$ are predictable processes, and then to apply standard martingale theory to $\sum_{i=1}^n \int_0^t \widetilde{h}_i^{(n)}(u)dM_i(u)$. We use the solution provided by Nielsen, Linton and Bickel (1998). We need the following definition

Definition A.1. *The sequence of processes* $\{h^{(n)}_{i_1,\ldots,i_k}\}$ *is of the leave-k-out type if* $h^{(n)}_{i_1,\ldots,i_k}$ *is predictable with respect to the filtration given by*

$$\mathcal{F}_t^{(i_1,\ldots,i_k;n)} = \bigvee_{j\notin\{i_1,\ldots,i_k\}} \mathcal{F}^{(n)}_{j,1} \bigvee_{l=1}^{k} \mathcal{F}^{(n)}_{i_l,t}.$$

We use below the facts that: $h_j^{(n)}(t)$ is predictable with respect to $\mathcal{F}_t^{(j;n)}$, that $h_j^{(n)}(t) - h_{i,j}^{(n)}(t)$ is predictable with respect to $\mathcal{F}_t^{(i,j;n)}$, and that $M_j = N_j - \Lambda_j$ is a martingale with respect to both filtrations. These are consequences of the i.i.d. set up we adopted.

Lemma 4. *Suppose that the processes* $\{h_i^{(n)}(u)\}$ *and* $\{h_{i,j}^{(n)}(u)\}$, $i, j = 1, \ldots, n$ *are cadlag and of the leave-one-out and leave-two-out types respectively and that* $h_{i,j}^{(n)} = h_i^{(n)}$, *and that the process* $h_{i,j}^{(n)}$ *is independent of the* σ-field $\mathcal{F}^{(n)}_{j,T}$. *Then*

$$E(\overline{M}_t^2) \leq 3n \sum_{i=1}^{n} E \int_0^t \{h_i^{(n)}(u) - h_{i,j}^{(n)}(u)\}^2 d\Lambda_i(u) + \sum_{i=1}^{n} E \int_0^t h_i^{(n)}(u)^2 d\Lambda_i(u).$$

Proof. See Nielsen, Linton, and Bickel (1998) and Nielsen (1999). ■

Proof of Theorem 1. We first prove (12), i.e., $\nu_n(\widehat{\alpha}(x_j, .)) - \nu_n(\alpha(x_j, .)) \xrightarrow{p} 0$. We have for all $\eta, \delta > 0$

$$\begin{aligned}
&\Pr\left[|\nu_n(\widehat{\alpha}(x_j, .)) - \nu_n(\alpha(x_j, .))| > \eta\right] \\
\leq\ & \Pr\left[|\nu_n(\widehat{\alpha}(x_j, .)) - \nu_n(\alpha(x_j, .))| > \eta, \widehat{\alpha}(x_j, .) \in \mathcal{G}_{d,s}(I_{-j}), \|\widehat{\alpha}(x_j, .) - \alpha(x_j, .)\|_{d,s,I_{-j}} \leq \delta\right] \\
& + \Pr\left[\widehat{\alpha}(x_j, .) \notin \mathcal{G}_{d,s}(I_{-j})\right] + \Pr\left[\|\widehat{\alpha}(x_j, .) - \alpha(x_j, .)\|_{d,s,I_{-j}} > \delta\right] \\
\leq\ & \mathbf{P}^* \left[\sup_{g\in\mathcal{G}_{d,s}(I_{-j}), \|g-g_0\|_{d,s,I_{-j}}\leq\delta} |\nu_n(g) - \nu_n(g_0)| > \eta\right] \\
& + \Pr\left[\widehat{\alpha}(x_j, .) \notin \mathcal{G}_{d,s}(I_{-j})\right] + \Pr\left[\|\widehat{\alpha}(x_j, .) - \alpha(x_j, .)\|_{d,s,I_{-j}} > \delta\right].
\end{aligned}$$

As a consequence of A5 it suffices to show that: (i) $\Pr(\widehat{\alpha}(x_j, .) \in \mathcal{G}_{d,s}(I_{-j})) \to 1$ and (ii) $\|\widehat{\alpha}(x_j, .) - \alpha(x_j, .)\|_{d,s,I_{-j}} \xrightarrow{p} 0$. The second condition follows by the uniform convergence of $D^a\widehat{\alpha}$

on I for $|a| \leq s$ established in Lemma 3(b). The first condition follows provided the function class $\mathcal{G}_{d,s}(I_{-j})$ is big enough. Note that for any a, x, $|D^a\widehat{\alpha}(x)| \leq |D^a\alpha(x)| + |D^a\widehat{\alpha}(x) - D^a\alpha(x)|$ by the triangle inequality. By assumption $|D^a\alpha(x)|$ is bounded, while $|D^a\widehat{\alpha}(x) - D^a\alpha(x)|$ is uniformly $o_p(1)$ for any $s \leq r$ by Lemma 3(b). Since $r > d/2$, the result follows.

The proof of (16) is divided into the proofs of the following two results:

$$n^{r/(2r+1)}V_{Q_{-j}}(x_j) \implies N(0,\ v_j^I(x_j)) \tag{28}$$

$$n^{r/(2r+1)}B_{Q_{-j}}(x_j) \longrightarrow {}_p m_j^I(x_j), \tag{29}$$

which are given below.

Proof of (28). Define:

$$\widehat{h}_i^{(n)}(x_j, s) = \int_{I_{-j}} \frac{W_{ni}(x,s)}{\widehat{e}(x)} dQ_{-j}(x_{-j}) \quad ; \quad \ddot{h}_i^{(n)}(x_j, s) = \int_{I_{-j}} \frac{W_{ni}(x,s)}{\widehat{e}_{-i}(x)} dQ_{-j}(x_{-j}) \quad ;$$

$$\widetilde{h}_i^{(n)}(x_j, s) = \int_{I_{-j}} \frac{W_{ni}(x,s)}{e(x)} dQ_{-j}(x_{-j}),$$

where $\widehat{e}_{-i}(x) = n^{-1}\sum_{j\neq i}\int_0^T K_b(x - X_j(s))Y_j(s)ds$ is the leave one out exposure estimator, while

$$W_{ni}(x, s) = \left(\frac{b}{n}\right)^{1/2} K_b(x - X_i(s)).$$

Then write

$$(nb)^{1/2}\widetilde{V}_{Q_{-j}}(x_j) = \sum_{i=1}^n \int_0^T \widetilde{h}_i^{(n)}(x_j, s)dM_i(s).$$

The proof of (28) is given in a series of lemmas below. We approximate $V_{Q_{-j}}(x_j)$ by $\widetilde{V}_{Q_{-j}}(x_j)$ and then apply a Martingale Central Limit theorem to this quantity. Lemma 5 gives the CLT for $\widetilde{V}_{Q_{-j}}(x_j)$, while Lemmas 6 and 7 show that the remainder terms are of smaller order.

LEMMA 5.

$$(nb)^{1/2}\widetilde{V}_{Q_{-j}}(x_j) \Longrightarrow N(0, v^I(x_j)). \tag{30}$$

PROOF. Since the $\widetilde{h}_i^{(n)}$ processes are predictable, asymptotic normality (30) follows by an application of Rebolledo's central limit theorem for martingales [see Proposition 4.2.1 of Ramlau-Hansen

(1983)]. Specifically, we must show that for all $\epsilon > 0$:

$$\sum_{i=1}^{n} \int_0^T \{\widetilde{h}_i^{(n)}(x_j, s)\}^2 1\left(|\widetilde{h}_i^{(n)}(x_j, s)| > \epsilon\right) d\langle M_i\rangle(s) \rightarrow {}_p 0 \tag{31}$$

$$\sum_{i=1}^{n} \int_0^T \{\widetilde{h}_i^{(n)}(x_j, s)\}^2 d\langle M_i\rangle(s) \rightarrow {}_p v_j^I(x_j), \tag{32}$$

where $\langle M\rangle$ is the quadratic variation of a process M, in our case $\langle M_i\rangle(s) = \Lambda_i(s) = \alpha(s, Z_i(s))Y_i(s)$. We shall approximate $\widetilde{h}_i^{(n)}(x_j, s)$ by

$$\overline{h}_i^{(n)}(x_j, s) = \frac{1}{\sqrt{nb}} k\left(\frac{x_j - X_{ji}(s)}{b}\right) \frac{q_{-j}(X_{-ji}(s))}{e(x_j, X_{-ji}(s))}.$$

We have

$$\begin{aligned}\sup_{0\leq s\leq T} \left|\{\widetilde{h}_i^{(n)}(x_j, s)\}^2 - \{\overline{h}_i^{(n)}(x_j, s)\}^2\right| &\leq 2 \sup_{0\leq s\leq T} |\overline{h}_i^{(n)}(x_j, s)| \sup_{0\leq s\leq T} \left|\widetilde{h}_i^{(n)}(x_j, s) - \overline{h}_i^{(n)}(x_j, s)\right| \\ &\quad + \sup_{0\leq s\leq T} \left|\widetilde{h}_i^{(n)}(x_j, s) - \overline{h}_i^{(n)}(x_j, s)\right|^2,\end{aligned}$$

where for some constant $\overline{k}$

$$\sup_{0\leq s\leq T} |\overline{h}_i^{(n)}(x_j, s)| \leq \frac{\overline{k}}{\sqrt{nb}}.$$

To analyze the term $\widetilde{h}_i^{(n)}(x_j, s) - \overline{h}_i^{(n)}(x_j, s)$, we make the change of variables $x_{-j} \longmapsto (x_{-j} - X_{-ji}(s))/b = u_{-j}$, which transforms the integration region I_{-j} to some set that eventually (for n large enough) contains $[-1, 1]^d$, because x_{-j} is an interior point of $\mathcal{X}_{-j}$. Therefore, for n large enough we have

$$\begin{aligned}&\left|\widetilde{h}_i^{(n)}(x_j, s) - \overline{h}_i^{(n)}(x_j, s)\right| \\ &= \frac{1}{\sqrt{nb}} \left| k\left(\frac{x_j - X_{ji}(s)}{b}\right) \int_{[-1,1]^d} \left\{\frac{q_{-j}(X_{-ji}(s) + bu_{-j})}{e(x_j, X_{-ji}(s) + bu_{-j})} - \frac{q_{-j}(X_{-ji}(s))}{e(x_j, X_{-ji}(s))}\right\} \prod_{\ell\neq j} k_b(u_\ell) du_\ell \right| \\ &\leq \frac{1}{\sqrt{nb}} \left| k\left(\frac{x_j - X_{ji}(s)}{b}\right)\right| \times o_p(1),\end{aligned}$$

by the dominated convergence theorem. This bound is uniform in s for $0 \leq s \leq T$. This implies that

$$\sum_{i=1}^{n} \int_0^T \left[\{\widetilde{h}_i^{(n)}(x_j, s)\}^2 - \{\overline{h}_i^{(n)}(x_j, s)\}^2\right] d\langle M_i\rangle(s) \rightarrow_p 0.$$

Also, we have

$$\sum_{i=1}^{n}\int_0^T \{\overline{h}_i^{(n)}(x_j,s)\}^2 d\langle M_i\rangle(s) = \frac{1}{nb}\sum_{i=1}^{n}\int_0^T k^2\left(\frac{x_j - X_{ji}(s)}{b}\right)\frac{q_{-j}^2(X_{-ji}(s))}{e^2(x_j, X_{-ji}(s))}\alpha(s, Z_i(s))Y_i(s)ds$$

$$\to {}_pE\left[\int_0^T \frac{1}{b}k^2\left(\frac{x_j - X_{ji}(s)}{b}\right)\frac{q_{-j}^2(X_{-ji}(s))}{e^2(x_j, X_{-ji}(s))}\alpha(s, Z_i(s))Y_i(s)ds\right]$$

by the law of large numbers for independent random variables. The above expectation is approximately equal to $\upsilon_j^I(x_j)$, by an application of Fubini's theorem, a change of variables and dominated convergence. Specifically,

$$\begin{aligned}
&E\left[\int_0^T \frac{1}{b}k^2\left(\frac{x_j - X_{ji}(s)}{b}\right)\frac{q_{-j}^2(X_{-ji}(s))}{e^2(x_j, X_{-ji}(s))}\alpha(s, Z_i(s))Y_i(s)ds\right] \\
&= \int \frac{1}{b}k^2\left(\frac{x_j - x_j'}{b}\right)\frac{q_{-j}^2(x_{-j}')}{e^2(x_j, x_{-j}')}\alpha(x')e(x')dx' \\
&= \int k^2(u)\frac{q_{-j}^2(x_{-j}')}{e^2(x_j, x_{-j}')}\alpha(x_j + bu, x_{-j}')e(x_j + bu, x_{-j}')dudx_{-j}' \\
&\simeq \int k^2(u)du\int \frac{q_{-j}^2(x_{-j}')}{e^2(x_j, x_{-j}')}\alpha(x_j, x_{-j}')e(x_j, x_{-j}')dx_{-j}'.
\end{aligned}$$

We now turn to the proof of (31). It suffices to show that for all $\epsilon > 0$

$$\text{Leb}\left\{s : \sum_{i=1}^{n}|\widetilde{h}_i^{(n)}(x_j,s)| > \epsilon\right\} = o_p(1).$$

This is true because $\sup_{s\in[0,T]}|\widetilde{h}_i^{(n)}(x_j,s)| \le \overline{k}/\sqrt{nb}$ for some constant $\overline{k} < \infty$.

■

To complete the proof of (28), we now must show that

$$(nb)^{1/2}\{\widetilde{V}_{Q_{-j}}(x_j) - V_{Q_{-j}}(x_j)\} \longrightarrow_p 0. \tag{33}$$

By the triangle inequality,

$$\begin{aligned}
(nb)^{1/2}\left|\widetilde{V}_{Q_{-j}}(x_j) - V_{Q_{-j}}(x_j)\right| &\le \left|\sum_{i=1}^{n}\int_0^T \widehat{h}_i^{(n)}(x_j,s)dM_i(s) - \sum_{i=1}^{n}\int_0^T \ddot{h}_i^{(n)}(x_j,s)dM_i(s)\right| \\
&+ \left|\sum_{i=1}^{n}\int_0^T \ddot{h}_i^{(n)}(x_j,s)dM_i(s) - \sum_{i=1}^{n}\int_0^T \widetilde{h}_i^{(n)}(x_j,s)dM_i(s)\right|.
\end{aligned}$$

Therefore, it suffices to show that each of these terms goes to zero in probability. This is shown in Lemmas 6 and 7 below.

LEMMA 6.
$$\sum_{i=1}^{n}\int_0^T \widehat{h}_i^{(n)}(x_j,s)dM_i(s)-\sum_{i=1}^{n}\int_0^T \overset{..}{h}_i^{(n)}(x_j,s)dM_i(s)\longrightarrow_p 0. \tag{34}$$

PROOF. By the Cauchy-Schwarz inequality

$$\begin{aligned}\left|\widehat{h}_i^{(n)}(x_j,s)-\overset{..}{h}_i^{(n)}(x_j,s)\right| &= \left|\int_{I_{-j}} W_{ni}(x,s)\frac{\widehat{e}_i(x)-\widehat{e}(x)}{\widehat{e}(x)\widehat{e}_{-i}(x)}dQ_{-j}(x_{-j})\right| \\ &\leq \frac{\left[\int_{I_{-j}} W_{ni}^2(x,s)dQ_{-j}(x_{-j})\cdot\int_{I_{-j}}\left\{\widehat{e}_{-i}(x)-\widehat{e}(x)\right\}^2 dQ_{-j}(x_{-j})\right]^{1/2}}{\inf_{x\in I}|\widehat{e}(x)\widehat{e}_{-i}(x)|},\end{aligned}$$

where $\widehat{e}_{-i}(x)-\widehat{e}(x)=n^{-1}\int_0^T K_b\{x-X_i(t)\}Y_i(t)dt$. Therefore, if for some $\epsilon>0$:

$$\inf_{x\in I}|\widehat{e}(x)\widehat{e}_{-i}(x)| \geq \epsilon+o_p(1) \tag{35}$$

$$\sup_{0\leq s\leq T}\left|\int_{I_{-j}} W_{ni}^2(x,s)dQ_{-j}(x_{-j})\right| = O_P(\frac{1}{nb^{d+1}}) \tag{36}$$

$$\int_{I_{-j}}\left\{\widehat{e}_{-i}(x)-\widehat{e}(x)\right\}^2 dQ_{-j}(x_{-j}) = O_P(\frac{1}{n^2b^{d+1}}), \tag{37}$$

then

$$\begin{aligned}\left|\sum_{i=1}^{n}\int_0^T \widehat{h}_i^{(n)}(x_j,s)dM_i(s)-\sum_{i=1}^{n}\int_0^T \overset{..}{h}_i^{(n)}(x_j,s)dM_i(s)\right| &\leq n\cdot O_P(\frac{1}{nb^{(d+1)/2}})\cdot O_P(\frac{1}{n^{1/2}b^{(d+1)/2}}) \\ &= o_P(1),\end{aligned} \tag{38}$$

because $nb^{2(d+1)}\to\infty$. It remains to prove (35)-(37) for some $\epsilon>0$.

The proof of (35) follows directly from the uniform convergence of the exposure estimator and the triangle inequality. Specifically, by the triangle inequality and Lemma 4,

$$\inf_{x\in I}|\widehat{e}(x)\widehat{e}_{-i}(x)|\geq\inf_{x\in I}e^2(x)-\sup_{x\in I}\left|\widehat{e}(x)\widehat{e}_{-i}(x)-e^2(x)\right|\geq\inf_{x\in I}e^2(x)-o_p(1),$$

which establishes (35) because $e(x)$ is bounded strictly away from zero by assumption A3. Furthermore, by a change of variables and the boundedness of q_{-j} and k, we have

$$\begin{aligned}\int_{I_{-j}} W_{ni}^2(x,s)dQ_{-j}(x_{-j}) &= \frac{b}{n}\int_{I_{-j}} \left[K_b(x - X_i(s))\right]^2 dQ_{-j}(x_{-j}) \\ &\leq \frac{1}{nb^{d+1}}k^2\left(\frac{x_j - X_{ji}(s)}{b}\right)\int_{[-1,1]^d}\prod_{\ell\neq j}\left\{k^2(u_\ell)du_\ell\right\}\cdot \sup_{x_{-j}\in I_{-j}}|q_{-j}(x_{-j})| \\ &= O_p(\frac{1}{nb^{d+1}}),\end{aligned}$$

which establishes (36). The proof of (37) is as follows. By the boundedness of k and e, and Fubini's theorem

$$\begin{aligned}E\left[|\widehat{e}_{-i}(x) - \widehat{e}(x)|\right] &= \frac{1}{n}E\left|\int_0^T [K_b\{x - X_i(t)\}Y_i(t)dt\right| \leq \frac{1}{n}\int_{\mathcal{X}} |K_b(x-x')|e(x')dx' \\ &\leq \frac{1}{n}\int_{\mathcal{X}}|K_b(x-x')|dx'\cdot\sup_{x'\in\mathcal{X}} e(x') \leq \frac{1}{n}\int_{[-1,1]^{d+1}}|K(u)|\,du\cdot\sup_{x'\in\mathcal{X}} e(x') \\ &= O(1/n)\end{aligned}$$

for n large. Furthermore,

$$\begin{aligned}\sup_{x\in I}|\widehat{e}_{-i}(x) - \widehat{e}(x)| &= \frac{1}{n}\sup_{x\in I}\left|\int_0^T [K_b\{x - X_i(t)\}Y_i(t)dt\right| \\ &\leq T\sup_{u\in[-1,1]^{d+1}}|K(u)|\,\frac{1}{nb^{d+1}}.\end{aligned}$$

Therefore, since

$$\int_{I_{-j}}\left\{\widehat{e}_{-i}(x) - \widehat{e}(x)\right\}^2 dQ_{-j}(x_{-j}) \leq \sup_{x\in I}|\widehat{e}_{-i}(x) - \widehat{e}(x)|\int_{I_{-j}}|\widehat{e}_{-i}(x) - \widehat{e}(x)|\,dQ_{-j}(x_{-j}),$$

the result (37) is true. This concludes the proof of (34). ■

Lemma 7.

$$\sum_{i=1}^n \int_0^T \ddot{h}_i^{(n)}(x_j,s)dM_i(s) - \sum_{i=1}^n \int_0^T \widetilde{h}_i^{(n)}(x_j,s)dM_i(s) \longrightarrow_p 0. \tag{39}$$

PROOF. Writing

$$\ddot{h}_i^{(n)}(x_j,s) - \widetilde{h}_i^{(n)}(x_j,s) = \int_{I_{-j}} W_{ni}(x,s)\frac{e(x)-\widehat{e}_{-i}(x)}{e(x)\widehat{e}_{-i}(x)}dQ_{-j}(x_{-j}),$$

we see that the left hand side of (39) is of the general form $\overline{M}_t = \sum_{i=1}^n \int_0^t \ddot{h}_i^{(n)}(u)dM_i(u)$, where the M_i process is a martingale, but $\ddot{h}_i^{(n)}(u)$ is not a predictable process according to the usual definition. Furthermore, the random denominator $\widehat{e}_{-i}(x)$ can take negative values. We write

$$\begin{aligned}\overline{M}_t = \overline{M}_{t1} + \overline{M}_{t2} + \overline{M}_{t3} &= \sum_{i=1}^n \int_0^T \left\{\int_{I_{-j}} W_{ni}(x,s)\frac{e(x)-E(\widehat{e}_{-i}(x))}{e^2(x)}dQ_{-j}(x_{-j})\right\}dM_i(s) \quad (40)\\ &\sum_{i=1}^n \int_0^T \left\{\int_{I_{-j}} W_{ni}(x,s)\frac{E(\widehat{e}_{-i}(x))-\widehat{e}_{-i}(x)}{e^2(x)}dQ_{-j}(x_{-j})\right\}dM_i(s)\\ &+\sum_{i=1}^n \int_0^T \left\{\int_{I_{-j}} W_{ni}(x,s)\frac{\{e(x)-\widehat{e}_{-i}(x)\}^2}{e^2(x)\widehat{e}_{-i}(x)}dQ_{-j}(x_{-j})\right\}dM_i(s).\end{aligned}$$

We first examine $\overline{M}_{t1}$. We have $\{E[\widehat{e}_{-i}(x)] - e(x)\}/e^2(x) = b^r\gamma_n(x)$ for some bounded continuous function γ_n. Then, by the change of variable used above we have

$$\begin{aligned}\int_{I_{-j}} W_{ni}(x,s)\frac{E_i\left[\widehat{e}_{-i}(x)\right]-e(x)}{e^2(x)}dQ_{-j}(x_{-j}) &= b^r\left(\frac{b}{n}\right)^{1/2}\int_{I_{-j}}\gamma_n(x)K_b(x-X_i(s))dQ_{-j}(x_{-j})\\ &= \frac{b^r}{\sqrt{nb}}k\left(\frac{x_j-X_{ji}(s)}{b}\right)\gamma_n^*(x_j,X_{-ji}(s)).\end{aligned}$$

For large n, $\gamma_n^*(x_j,X_{-ji}(s)) = \int_{[-1,1]^d}\gamma_n(x_j,X_{-ji}(s)+bu_{-j})q_{-j}(X_{-ji}(s)+bu_{-j})du_{-j}$ is also bounded. Therefore,

$$\begin{aligned}&\sum_{i=1}^n \int_0^T \left\{\int_{I_{-j}} W_{ni}(x,s)\frac{E_i\left[\widehat{e}_{-i}(x)\right]-e(x)}{e^2(x)}dQ_{-j}(x_{-j})\right\}dM_i(s)\\ \simeq\ &\frac{b^r}{\sqrt{nb}}\sum_{i=1}^n\int_0^T\gamma_n^*(x_j,X_{-ji}(s))k\left(\frac{x_j-X_{ji}(s)}{b}\right)q_{-j}(X_{-ji}(s))dM_i(s)\\ =\ &O_p(b^r),\end{aligned}$$

which follows by the same arguments used in the proof of Theorem 1 of Nielsen and Linton (1995) because this term is like the normalized stochastic part of a one-dimensional kernel smoother multiplied by b^r. Therefore, $\overline{M}_{t1} = o_p(1)$.

We now deal with the stochastic term $\overline{M}_{t2}$. Let

$$h_i^{(n)}(u) = \int_{I_{-j}} W_{ni}(x,u)\frac{\widehat{e}_{-i}(x) - E\left[\widehat{e}_{-i}(x)\right]}{e^2(x)}dQ_{-j}(x_{-j}) = \sum_{\substack{l=1 \\ l\neq i}}^{n}\{a_{nil}(u) - E_i a_{nil}(u)\}, \qquad (41)$$

where E_i denotes conditional expectation given $X_i(u)$, while

$$a_{nil}(u) = \left(n^3 b^{-1}\right)^{-1/2}\int_{I_{-j}}\int_0^T \frac{K_b(x - X_i(u))K_b(x - X_l(s))}{e^2(x)}Y_l(s)dsdQ_{-j}(x_{-j}).$$

Let also $h_{i,j}^{(n)}(u) = \sum_{\substack{l=1 \\ l\neq i,j}}^{n}\{a_{nil}(u) - E_i a_{nil}(u)\}$. From Lemma 4 we have

$$E\left(\overline{M}_{t1}^{*2}\right) \leq 3n\sum_{i=1}^{n} E\int_0^T \{h_i^{(n)}(u) - h_{i,j}^{(n)}(u)\}^2 d\Lambda_i(u) + \sum_{i=1}^{n} E\int_0^T \{h_i^{(n)}(u)\}^2 d\Lambda_i(u). \qquad (42)$$

We now investigate each term in the bound (42). By the law of iterated expectations we have

$$E\left[\{h_i^{(n)}(u)\}^2\right] = nE\left[E_i(a_{nil}^2(u)) - E_i^2(a_{nil}(u))\right] \leq nE\left[E_i(a_{nil}^2(u))\right] \qquad (43)$$

since $a_{nil}(u)$ and $a_{nik}(u)$ are conditionally independent given $X_i(u)$. Therefore, we must calculate $E_i a_{nil}^2(u)$. Let $(k*k)_b(t) = (k*k)(t/b)/b$, where $(k*k)(s) = \int k\left(t+s\right)k(t)dt$. The convolution kernel $k*k$ is symmetric about zero and differentiable to the same degree as k. Then,

$$a_{nil} = \frac{b^{1/2}}{n^{3/2}}k_b(x_j - X_{ji}(u))\int_0^T k_b(x_j - X_{jl}(s))Y_l(s)\mathcal{I}_{nil}(x_j, s)ds$$

where

$$\begin{aligned}
\mathcal{I}_{nil}(x_j, s) &= \int_{I_{-j}}\prod_{j'\neq j} k_b(x_j - X_{j',i}(u))k_b(x_j - X_{j',l}(s))\frac{q_{-j}(x_{-j})}{e(x)}dx_{-j} \\
&= b^{-d}\int_{[-1,1]^d}\prod_{j'\neq j} k(v_{j'})k\left(v_{j'} + \frac{X_{j',l}(s) - X_{j',i}(u)}{b}\right)\frac{q_{-j}(X_{-j,l}(s) + bv_{-j})}{e(x_j, X_{-j,l}(s) + bv_{-j})}dv_{-j} \\
&\simeq \frac{q_{-j}(X_{-j,l}(s))}{e(x_j, X_{-j,l}(s))}\prod_{j'\neq j}(k*k)_b(X_{j',l}(s) - X_{j',i}(u))
\end{aligned}$$

by changing variables $x_{-j} \longmapsto (x_{-j} - X_{-j,l}(s))/b = v_{-j}$ and using dominated convergence. For large n, we have for some constants c, c'

$$
\begin{aligned}
E_i[\{a_{nil}(u)\}^2] &\leq \frac{c}{n^3 b} k^2\left(\frac{x_j - X_{ji}(u)}{b}\right) E_i\left[\int_0^T k_b(x_j - X_{jl}(s))Y_l(s)\prod_{j'\neq j}(k*k)_b(X_{j',l}(s) - X_{j',i}(u))ds\right]^2 \\
&\leq \frac{c'}{n^3 b} k^2\left(\frac{x_j - X_{ji}(u)}{b}\right) E_i\int_0^T \left|k_b(x_j - X_{jl}(s))Y_l(s)\prod_{j'\neq j}(k*k)_b(X_{j',l}(s) - X_{j',i}(u))\right|^2 ds,
\end{aligned}
$$

by the Cauchy-Schwarz inequality. After changing variables again $x' \mapsto ((x_j, X_{-j,i}(u)) - (x'_j, x'_{-j}))/b = v$, we obtain the bound

$$
\begin{aligned}
E_i[\{a_{nil}(u)\}^2] &\leq \frac{c'}{n^2 b^{d+1}}\frac{1}{nb} k^2\left(\frac{x_j - X_{ji}(u)}{b}\right)\int_{[-1,1]^{d+1}} k^2(v_j)e(x_j, X_{-ji}(u) + bv_{-j})\prod_{j'\neq j}(k*k)^2(v_{j'})dv. \\
&\leq \frac{c'}{n^2 b^{d+1}}\frac{1}{nb} k^2\left(\frac{x_j - X_{ji}(u)}{b}\right).
\end{aligned}
$$

Therefore,

$$
E[\{h_i^{(n)}(u)\}^2] \leq \frac{c'}{nb^{d+1}}\frac{1}{nb} Ek^2\left(\frac{x_j - X_{ji}(u)}{b}\right)
$$

and so

$$
\begin{aligned}
\sum_{i=1}^n E\int_0^T \{h_i^{(n)}(u)\}^2 d\Lambda_i(u) &\leq \frac{c'}{nb^{d+1}}\frac{1}{nb}\sum_{i=1}^n \int_0^T Ek^2\left(\frac{x_j - X_{ji}(u)}{b}\right) d\Lambda_i(u) \\
&= O(n^{-1}b^{-(d+1)}) = o_p(1),
\end{aligned}
$$

because $nb^{(d+1)} \to \infty$. Furthermore, $h_i^{(n)}(u) - h_{i,j}^{(n)}(u) = a_{nij}(u) - E_i a_{nij}(u)$, so that similar arguments show that

$$
E\int_0^T \{h_i^{(n)}(u) - h_{i,j}^{(n)}(u)\}^2 d\Lambda_i(u) \leq O(n^{-3}b^{-(d+1)}).
$$

In conclusion, we have established that $E[\overline{M}_{t2}^2] = o(1)$, as required.

The term $\overline{M}_{t3}$ in (40) is handled by direct methods using the uniform convergence of $\widehat{e}_{-i}(x)$. Thus

$$
\left|\int_{I_{-j}} W_{ni}(x,s)\frac{\{e(x) - \widehat{e}_{-i}(x)\}^2}{e^2(x)\widehat{e}_{-i}(x)} dQ_{-j}(x_{-j})\right|
$$

$$\leq \frac{\left\{\int_{I_{-j}} W_{ni}^2(x,s)dQ_{-j}(x_{-j})\right\}^{1/2} \sup_{x\in I}|e(x)-\widehat{e}_{-i}(x)|^2}{\inf_{x\in I}|e^2(x)\widehat{e}_{-i}(x)|}$$

$$= O_P(\frac{1}{n^{1/2}b^{(d+1)/2}})\cdot\left\{O_P(\frac{1}{nb^{(d+1)}})+O_P(b^{2r})\right\}.$$

by the Cauchy-Schwarz inequality and Lemma 3(a). Therefore,

$$\sum_{i=1}^{n}\int_0^T\left\{\int_{I_{-j}} W_{ni}(x,s)\frac{\{e(x)-\widehat{e}_{-i}(x)\}^2}{e^2(x)\widehat{e}_{-i}(x)}dQ_{-j}(x_{-j})\right\}dM_i(s)$$

$$= O_P(\frac{1}{n^{1/2}b^{3(d+1)/2}})+O_P(n^{1/2}b^{2r-(d+1)/2}).$$

This concludes the proof of (39). ■

Proof of (29). We have

$$\int_{I_{-j}}\frac{B_n(x)}{\widehat{e}(x)}dQ_{-j}(x_{-j}) = \int_{I_{-j}}\frac{B_n(x)}{e(x)}dQ_{-j}(x_{-j}) + \int_{I_{-j}} B_n(x)\frac{\widehat{e}(x)-e(x)}{\widehat{e}(x)e(x)}dQ_{-j}(x_{-j}),$$

where

$$\left|\int_{I_{-j}} B_n(x)\frac{\widehat{e}(x)-e(x)}{\widehat{e}(x)e(x)}dQ_{-j}(x_{-j})\right| \leq \frac{\sup_{x_{-j}\in I_{-j}}|B_n(x)|\cdot\sup_{x_{-j}\in I_{-j}}|\widehat{e}(x)-e(x)|}{\inf_{x_{-j}\in I_{-j}}|\widehat{e}(x)e(x)|}$$

$$= O_P(b^r)O_P(b^r) = o_P(b^r)$$

by the uniform convergence result of Lemma 3(a). After a change of variables $[x' \longmapsto (x'-x)/b]$ and using the structure of $B_n(x)$, we have

$$E\left[\int_{I_{-j}}\frac{B_n(x)}{e(x)}dQ_{-j}(x_{-j})\right] = \int_{I_{-j}}\frac{E\int_0^T K_b(x-X_i(s))\left[\alpha(X_i(s))-\alpha(x)\right]Y_i(s)ds}{e(x)}dQ_{-j}(x_{-j})$$

$$= \int_{I_{-j}}\int_{\mathcal{X}}\frac{K_b(x-x')\left[\alpha(x')-\alpha(x)\right]e(x')dx'}{e(x)}dQ_{-j}(x_{-j})$$

$$= \int_{I_{-j}}\int_{[-1,1]^{d+1}} K(u)\left\{\alpha(x-bu)-\alpha(x)\right\}\frac{e(x-bu)}{e(x)}dQ_{-j}(x_{-j})du,$$

where $u = (u_0, \ldots, u_d)$. By Taylor expansion

$$\begin{aligned}(\alpha \cdot e)(x+bu) &= \sum_{j=0}^{r} \sum_{\{a:|a|=j\}} \frac{b^a}{a!} u^a D^a(\alpha \cdot e)(x) + \sum_{\{a:|a|=r\}} \frac{b^a}{a!} u^a \left\{D^a(\alpha \cdot e)(x^*) - D^a(\alpha \cdot e)(x)\right\}, \\ e(x+bu) &= \sum_{j=0}^{r} \sum_{\{a:|a|=j\}} \frac{b^a}{a!} u^a D^a e(x) + \sum_{\{a:|a|=r\}} \frac{b^a}{a!} u^a \left\{D^a e(x^*) - D^a e(x)\right\},\end{aligned}$$

where $x_j^*(u)$ are intermediate values satisfying $|x_j^*(u) - x_j| < b|u_j|$, for $j = 0, \ldots, d$. Using assumption A4 and the fact that x is an interior point of $\mathcal{X}$, we have

$$E\left[\int_{I_{-j}} \frac{B_n(x)}{e(x)} dQ_{-j}(x_{-j})\right] = \frac{\mu_r(k)}{r!} b^r \sum_{j=0}^{d} \int_{I_{-j}} \beta_j^{(r)}(x) dQ_{-j}(x_{-j})\{1 + o(1)\},$$

by continuity and dominated convergence. Finally, we verify that the variance of the random variable $\int (B_n(x)/e(x)) dQ_{-j}(x_{-j})$ is of smaller order. We have

$$\int_{I_{-j}} \frac{B_n(x)}{e(x)} dQ_{-j}(x_{-j}) = \frac{1}{n} \sum_{i=1}^{n} X_{ni},$$

where X_{ni} are independent random variables of the form $\int_0^T g_n(X_i(s)) Y_i(s) ds$ with

$$g_n(X_i(s)) = \int_{I_{-j}} \frac{K_b(x - X_i(s)\left\{\alpha(X_i(s)) - \alpha(x)\right\}}{e(x)} dQ_{-j}(x_{-j}).$$

By a change of variables $x_\ell \longmapsto u_\ell = (X_{\ell i}(s) - x_\ell)/b$, $\ell \neq j$, we have for large n

$$\begin{aligned} &|g_n(X_i(s))| \\ = &\left| k_b(x_j - X_{ji}(s)) \int_{[-1,1]^d} \frac{\prod_{\ell \neq j} k(u_\ell)\left\{\alpha(X_i(s)) - \alpha(x_j, X_{-ji}(s) + bu_{-j})\right\} q(X_{-ji}(s) + bu_{-j}))}{e(x_j, X_{-ji}(s) + bu_{-j})} du_{-j} \right| \\ \leq &\ |k_b(x_j - X_{ji}(s))| \times \int_{[-1,1]^d} \prod_{\ell \neq j} |k(u_\ell)| \left|\alpha(X_i(s)) - \alpha(x_j, X_{-ji}(s) + bu_{-j})\right| du_{-j} \times \sup_{x \in I} \left|\frac{q(x_{-j})}{e(x)}\right|.\end{aligned}$$

Now we can apply expression (10) from Nielsen and Linton (1995, p 1728) to get

$$\text{var}\left[\int_{I_{-j}} \frac{B_n(x)}{e(x)} dQ_{-j}(x_{-j})\right] \leq \frac{T}{n}\left\{\int_0^T g_n^2(z,s) e(z,s) y(s) ds\right\} = O(\frac{1}{nb})$$

by the same dominated convergence arguments used above. This concludes the proof of (16). ■

Proof of (18) and (19). Recall that $\widehat{c}_{*j}$ and $\widehat{c}_{\dagger}$ satisfy $\widehat{c}_{\dagger} - c_{\dagger} = O_p(n^{-1/2})$ and $\widehat{c}_{*j} - c_{*j} = O_p(n^{-1/2})$. Therefore, by Taylor expansion:

$$\widehat{\alpha}_A(x) - \alpha_A(x) = \sum_{j=0}^{d} \left\{ \widehat{\alpha}_{Q_{-j}}(x_j) - \alpha_{Q_{-j}}(x_j) \right\} + O_P(n^{-1/2})$$

$$\begin{aligned}\widehat{\alpha}_M(x) - \alpha_M(x) &= \frac{1}{c^d} \sum_{j=0}^{d} \left\{ \widehat{\alpha}_{Q_{-j}}(x_j) - \alpha_{Q_{-j}}(x_j) \right\} \prod_{k \neq j} \alpha_{Q_{-k}}(x_k) + O_P(n^{-1/2}) \\ &\quad + O_P\left(\sum_{j=0}^{d} \left| \widehat{\alpha}_{Q_{-j}}(x_j) - \alpha_{Q_{-j}}(x_j) \right|^2 \right)\end{aligned}$$

for some constant c. We next substitute in the expansions for $\widehat{\alpha}_{Q_{-j}}(x_j) - \alpha_{Q_{-j}}(x_j)$, which were obtained above. To show that $\widehat{\alpha}_{Q_{-j}}(x_j) - \alpha_{Q_{-j}}(x_j)$ and $\widehat{\alpha}_{Q_{-k}}(x_k) - \alpha_{Q_{-k}}(x_k)$ are uncorrelated it suffices to show that the leading stochastic terms are so. We have

$$\operatorname{cov}\Big(\sum_{i=1}^{n} \int_0^T \widetilde{h}_i^{(n)}(x_j, s) dM_i(s), \sum_{i=1}^{n} \int_0^T \widetilde{h}_i^{(n)}(x_k, s) dM_i(s)\Big) \tag{44}$$

$$\begin{aligned}&= b \int_{\mathcal{X}} \Bigg[\int_{I_{-j}} \frac{k_b(x_j - w_j) k_b(x'_k - w_k) \prod_{m \neq j,k} k_b(x'_m - w_m)}{e(x_j, x'_{-j})} dQ_{-j}(x'_{-j}) \\ &\qquad \int_{I_{-k}} \frac{k_b(x'_j - w_j) k_b(x_k - w_k) \prod_{m \neq j,k} k_b(x'_m - w_m)}{e(x_k, x'_{-k})} dQ_{-k}(x'_{-k}) \Bigg] e(w) dw d\left\langle M_i(s) \right\rangle \\ &\simeq b \int_{\mathcal{X}} k_b(x_j - w_j) k_b(x_k - w_k) \frac{q_{-j}(w_{-j})}{e(x_j, w_{-j})} \frac{q_{-k}(w_{-k})}{e(x_k, w_{-k})} e(w) \alpha(w) dw ds \\ &= o(1),\end{aligned}$$

which establishes the result. The first equality follows by the independence of the processes, while the second equality replaces the integrals over I_{-j} and I_{-k} by their limits using the changes of variable $x_{-j} \mapsto (x_{-j} - X_{-j})/b$ and $x_{-k} \mapsto (x_{-k} - X_{-k})/b$ and dominated convergence. Note that this shows that the covariance between the normalized component estimators is $O(b)$ - and hence the covariance between the unnormalized estimators would be $O(1/n)$.

Finally, the standard errors are consistent by the same reasoning as Nielsen and Linton (1995, pp1741-1742).

■

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. and Borgan, O., Gill, R.D., and N. Keiding (1992). Statistical models based on counting processes. Springer-Verlag, New-York.

Auestad, B. and Tjøstheim, D. (1991). Functional identification in nonlinear time series, in Proceedings of the NATO Advanced Study Institute on Nonparametric Functional Estimation and Related Topics, Spetzes, August 1990, ed. G.H. Roussas, pp. 493-507.

Beran, R. (1981). Nonparametric regression with randomly censored survival data. Technical Report, Dept. of Statistics, University of California, Berkeley.

Bickel, P.J., (1975). One-step Huber estimates in the linear model. J. Amer. Statist. Assoc. 70, 428-434.

Breiman, L. and J.H. Friedman (1985). Estimating optimal transformations for multiple regression and correlation (with discussion). J. Amer. Statist. Assoc. 80, 580-619.

Cox, D.R. (1972). Regression models and life tables. J. Roy. Statist. Soc. Ser. B. 34 187-220.

Dabrowska, D.M . (1987). Nonparametric regression with censored survival time data. Scand. J. Statist. 14 1811-1977.

Dabrowska, D.M . (1997). Smoothed Cox Regression. Ann. Statist. 25, 1510-1540.

Fahrmeir, L., and A. Klinger (1998). A nonparametric multiplictive hazard model for event history analysis. Biometrika 85, 581-592.

Felipe, A., M. Guillen, and J.P. Nielsen (2000). Longevity studies based on kernel hazard estimation. Forthcoming in Insurance, Mathematics, and Economics.

Fusaro, R., Nielsen, J.P., and Scheike, T. (1993). Marker dependent hazard estimation. An application to Aids. Statistics in Medicine 12 843-865.

Hall, P., and I. Johnstone (1992). Empirical functional and efficient smoothing parameter selection. J. Roy. Statist. Soc Ser. B. 54, 475-530.

Hastie, T.J. and R.J. Tibshirani (1990). Generalized Additive Models. Chapman and Hall.

Hjort, N.L. (1994). Dynamic likelihood hazard rate estimation. Biometrika. To appear.

Huang, J., (1999). Efficient estimation of the partly linear additive Cox model. Ann. Statist. 27, 1536-1563.

Huang, J.Z., C. Kooperberg, C.J. Stone, and Y.K. Truong (2000). Functional ANOVA modeling for proportional hazards regression. Ann. Statist. 28, .

Jewll, N.P., and J.P. Nielsen (1993). Consistent prediction rules based on markers. Biometrika 80, 153-164.

Kooperberg, C., C.J. Stone, and Y.K. Truong (1995). L2 Rate of Convergence for Hazard Regression. Scan J. Statist (22), 143-158.

Lin, D.Y., and Z. Ying (1995). Semiparametric Analysis of General Additive-Multiplicative Hazard Models for Counting Processes. Ann. Statist. 23, 1712-1734.

Linton, O.B.(1997). Efficient estimation of additive nonparametric regression models. Biometrika 84, 469-474.

Linton, O.B., Nielsen, J.P., and S. van de Geer (1999). Estimating Multiplicative and Additive Marker Dependent Hazard Functions by Backfitting with the Assistance of Marginal Integration. Technical Report, LSE.

Linton, O.B. and Nielsen, J.P. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. Biometrika 82, 93-101.

Mammen, E., Linton, O. B., and Nielsen, J. P. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. Ann. Statist. 27, 1443-1490.

Masry, E., (1996): Multivariate regression estimation: Local polynomial fitting for time series. Stochastic Processes and their Applications 65, 81-101.

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-offit Tests for Additive Hazards and Proportional Hazards Models. Scan J. Statist. 18, 177-195.

Newey, W.K. (1994). Kernel estimation of partial means. Econometric Theory. 10, 233-253.

Nielsen, E. (1994). Marginal mortality estimation. Master's Thesis, Laboratory of Actuarial Mathematics, University of Copenhagen.

Nielsen, J.P. (1990). Kernel estimation of densities and hazards: A counting process approach. PhD thesis. University of California at Berkeley.

Nielsen, J.P. (1999). Super-efficient hazard estimation based on high-quality marker information. Biometrika 86, 227-232.

Nielsen, J.P. and O.B. Linton. (1995). Kernel estimation in a nonparametric marker dependent hazard model. Ann. Statist. 23, 1735-1748.

Nielsen, J.P. and O.B. Linton. (1998). An optimisation interpretation of integration and backfitting estimators for separable nonparametric models. J. Roy. Statist. Soc., Ser. B. 60, 217-222.

Nielsen, J.P., O.B. Linton and Bickel, P.J. (1998). On a semiparametric survival model with flexible covariate effect. Ann. Statist. 26, 215-241.

Nielsen, J.P. and P. Voldsgaard. (1996). Structured nonparametric marker dependent hazard estimation based on marginal integration. An application to health dependent mortality. Proceedings of 27'th Astin Conference in Copenhagen. 2, 634-641.

Opsomer, J. D. and D. Ruppert (1997). Fitting a bivariate additive model by local polynomial regression. Ann. Statist. 25, 186 - 211.

O'Sullivan, F. (1993). Nonparametric estimation in the Cox Model. Ann., Statist. 23 1735-1749.

Ramlau-Hansen, H., (1983). Smoothing counting process intensities by means of kernel functions. Ann. Statist. 11, 453-466.

Sasieni, P. (1992). Non-orthogonal projections and their application to calculating the information in a partly linear Cox model. Scand. J. Statist. 19, 215-234.

Sperlich, S., W. Härdle and O.Linton (1999). A Simulation comparison between the Backfitting and Integration methods of estimating Separable Nonparametric Models. TEST 8, 419-458.

Stone, C.J. (1980). Optimal rates of convergence for nonparametric estimators. Ann. Statist. 8, 1348-1360.

Stone, C.J. (1985). Additive regression and other nonparametric models. Ann. Statist., 13, 685-705.

Stone, C.J. (1986). The dimensionality reduction principle for Generalized additive models. Ann. Statist. 14, 592-606.

Tjøstheim, D., and B. Auestad (1994). Nonparametric identification of nonlinear time series: projections. J. Am. Stat. Assoc. 89, 1398-1409.

van de Geer, S., (1995). Exponential inequalities for martingales, with application to maximum likelihood estimation for counting processes. Ann. Statist. 23, 1779-1801.

van de Geer, S., (2000). Empirical Processes in M-Estimation. Cambridge Series in Statistical and Probabilistic Mathematics.

Wand, M.P. & Jones, M.C.(1996). Kernel Smoothing. Chapman and Hall, London.