

SMOOTHED COX REGRESSION¹

BY DOROTA M. DABROWSKA

University of California

Nonparametric regression was shown by Beran and McKeague and Utikal to provide a flexible method for analysis of censored failure times and more general counting processes models in the presence of covariates. We discuss application of kernel smoothing towards estimation in a generalized Cox regression model with baseline intensity dependent on a covariate. Under regularity conditions we show that estimates of the regression parameters are asymptotically normal at rate \sqrt{n} , and we also discuss estimation of the baseline cumulative hazard function and related parameters.

1. Introduction. Nonparametric regression was introduced in survival analysis by Beran (1981) as an alternative to the popular Cox proportional hazard model. In its simplest formulation, the approach aims to recover the conditional survival function

$$(1.1) \quad F(t|z) = P(T > t | Z = z),$$

where T is a possibly censored failure time and Z is a vector of covariates. Under the assumption that the failure time T and censoring time \tilde{T} are conditionally independent given Z , (1.1) is estimated by the sample product integral (Gill and Johansen, 1990)

$$F(t|z) = \prod_{[0, t]} (1 - A(du|z))$$

in which the unknown conditional cumulative hazard function

$$A(t|z) = - \int_{[0, t]} \frac{F(du|z)}{F(u-|z)}$$

is replaced by Beran's conditional Aalen–Nelson estimate:

$$\hat{A}(t|z) = \int_{[0, t]} \frac{\sum_{i=1}^n W_i(z) N_i(ds)}{\sum_{i=1}^n W_i(z) Y_i(s)}.$$

Here (N_i, Y_i) are counting and risk processes,

$$(1.2) \quad N_i(t) = I(T_i \leq t \wedge \tilde{T}_i), \quad Y_i(t) = I(T_i \wedge \tilde{T}_i \geq t),$$

associated with a sample of n individuals under study and $W_i(z)$ are weights dependent only on their covariates Z_1, \dots, Z_n . In particular, kernel estimates

Received May 1996; Revised December 1996.

¹Research supported by NSF Grant DMS 95-04507 and National Cancer Institute Grant 1-R01-95-CA65595-01.

AMS 1991 subject classifications. Primary 62G05, 62M09.

Key words and phrases. Kernel estimation, counting processes, hazard functions estimation.

correspond to the choice

$$W_i(z) = \frac{1}{a} K\left(\frac{z - Z_i}{a}\right) = K_a(z - Z_i),$$

where K is a density, $K_a(\cdot) = a^{-1}K(\cdot/a)$ and $a = a(n)$, $a(n) \rightarrow 0$ is a sequence of bandwidths. The known drawback of Beran's estimate is that it can handle only a small number of covariates. One method to alleviate this problem is to consider models incorporating both parametric and nonparametric constraints on the covariates.

In this paper we discuss parameter estimation in a generalized Cox proportional hazard model. Throughout, (Ω, \mathcal{F}, P) is a complete probability space and $\{\mathcal{F}_t: t \in [0, 1]\}$, $\mathcal{F}_t \subseteq \mathcal{F}$ is a filtration satisfying the "usual conditions," that is, \mathcal{F}_t is a family of right-continuous increasing σ -algebras and \mathcal{F}_0 contains the P -null sets of \mathcal{F} . We assume that $\{N_t: t \in [0, 1]\}$ is an adapted counting process with intensity

$$(1.3) \quad \Lambda(dt) = EN(dt)|\mathcal{F}_{t-} = Y(t)\alpha(t, X_t)\exp[\beta^T Z_t] dt,$$

where $Y(t)$ is a 0–1 predictable process indicating times at which the process N is under observation, and X_t and Z_t are predictable covariates. For the sake of simplicity, X_t is a scalar process whereas Z_t is a p -variate vector. Correspondingly, $\alpha(t, x)$ is an unknown nonnegative deterministic function, and β is a vector of unknown regression coefficients. In the absence of the covariate X_t , (1.3) reduces to the usual Cox proportional hazard model whereas in the absence of the covariate Z_t , it corresponds to the nonlinear counting process regression model in McKeague and Utikal (1990a). In the general case, (1.3) is an extension of the stratified Cox regression model in which, given $X_t = x$, the covariates Z_t follow the proportional hazard assumption, but this assumption may be violated by the X_t covariate.

In the standard survival analysis setting with (N, Y) processes given by (1.2) and time independent (X, Z) covariates, a special case of (1.3) corresponds to the partial linear transformation model in which the underlying failure time T satisfies

$$h(T) = f(X) - \beta^T Z + \varepsilon,$$

where ε is an extreme value error term, and $f(x)$ and $h(t) = \log \int_0^t \alpha_0(u) du$ are unknown functions [Thomas (1983), O'Sullivan (1993)]. This model arises frequently in regression diagnostics. A common approach is to fit Cox regression with the covariate X omitted, and next smooth martingale residuals to recover the shape of the function f [Therneau, Grambsch and Fleming (1990)]. Properties of this procedure are unknown. An alternative approach is to fit the larger model (1.3) and estimate the shape of the function f based on the identity $f'(x) = A'(1|x)/A(1|x)$ where $A'(\cdot|x)$ is the derivative of the baseline cumulative hazard $A(\cdot|x)$ with respect to x . The proportional hazard or parametric regression models can be tested by comparing the doubly cumulative

hazard

$$(1.4) \quad A_d(t, x) = \int_{[0, t] \times [0, x]} \alpha(u, w) du dw,$$

derived under (1.3), with its counterpart appropriate for a specific submodel [McKeague and Utikal (1991)].

Time dependent X_t covariates may arise when the occurrence of one event is recorded on several time scales, or else in analyses of multistate models. For example, let us consider an individual who can be in three states: healthy, sick or dead. Denote these states by 0, 1 and 2, respectively. The clock time t is the age of the individual. He starts in state 0 at time $t=0$ and subsequently makes transitions $0 \rightarrow 1 \rightarrow 2$ or $0 \rightarrow 2$. The disease incidence $\alpha_{01}(t) \exp(\beta_{01} Z_t)$ and the mortality rate for the healthy $\alpha_{02}(t) \exp(\beta_{02} Z_t)$ depend on the age and some covariate Z_t . The mortality rate for the diseased $\alpha_{12}(t, X_t) \exp(\beta_{12} Z_t)$ depends also on X_t , the age of onset of the disease or duration of the disease. This type of a non-Markovian illness process goes back to Fix and Neyman (1951) and was studied in the absence of the Z_t covariate by Andersen, Borgan, Gill and Keiding [(1993), Chapter III] and McKeague and Utikal (1990a).

The model (1.3) was considered by Sasieni (1992) who derived the information bound for estimation of the regression parameters and further proposed to estimate β based on a smoothed profile likelihood. Here we make the assumption that the covariate X_t , $t \in [0, 1]$ has a density $f(t, v)$ with respect to Lebesgue measure on the line. To estimate the regression parameters, we shall use a rectangle in the (time) \times (covariate) space, say $\mathcal{A} = [0, 1] \times [0, 1]$, such that \mathcal{A} is contained in the support of $f(u, v)$. Similarly to the usual Cox regression model, Jacod's (1975) formula for the likelihood of counting processes implies that given a sequence (N_i, Y_i, X_i, Z_i) , $i = 1, \dots, n$ of n iid replica of the (N, Y, X, Z) processes, the log-likelihood of the data is of the form

$$\begin{aligned} L(\beta, A) &= \frac{1}{n} \sum_{i=1}^n \int \beta^T Z_i(u) N_i(du) + \frac{1}{n} \sum_{i=1}^n \int \log[\alpha(u, X_i(u))] N_i(du) \\ &\quad - \frac{1}{n} \sum_{i=1}^n \int Y_i(u) \exp(\beta^T Z_i(u)) \alpha(u, X_i(u)) du. \end{aligned}$$

To alleviate tail problems in smoothing, we shall use observations with covariates $0 \leq X_i(\cdot) \leq 1$ to estimate the regression parameters. Their contribution to the log-likelihood is

$$\begin{aligned} L_{\mathcal{A}}(\beta, A) &= \frac{1}{n} \sum_{i=1}^n \int \beta^T Z_i(u) I(0 \leq X_i(u) \leq 1) N_i(du) \\ &\quad + \frac{1}{n} \sum_{i=1}^n \int \log[\alpha(u, X_i(u))] I(0 \leq X_i(u) \leq 1) N_i(du) \\ &\quad - \frac{1}{n} \sum_{i=1}^n \int Y_i(u) \exp(\beta^T Z_i(u)) I(0 \leq X_i(u) \leq 1) \alpha(u, X_i(u)) du. \end{aligned}$$

Under regularity conditions discussed in Section 2, the right-hand side is up to a term of order $o_P(n^{-1/2})$, equal to

$$\begin{aligned} SML_{\mathcal{A}}(\beta, A) &= \frac{1}{n} \sum_{i=1}^n \int_{\mathcal{A}} \beta^T Z_i(u) \tilde{N}_i(du; x) dx \\ &\quad + \frac{1}{n} \sum_{i=1}^n \int_{\mathcal{A}} \log[\alpha(u, x)] \tilde{N}_i(du; x) dx \\ &\quad - \frac{1}{n} \int_{\mathcal{A}} S^{(0)}(u, x, \beta) \alpha(u, x) du dx, \end{aligned}$$

where

$$S^{(0)}(t, w, \beta) = \sum_{i=1}^n Y_i(t) \exp(\beta^T Z_i(t)) K_a(w - X_i(t))$$

and

$$\tilde{N}_i(t; w) = \int_0^t K_a(w - X_i(u)) N_i(du)$$

are the smoothed risk and counting processes. Analogously to Andersen, Borgan, Gill and Keiding [(1993), Chapter VII], substitution of Beran's estimator

$$\hat{A}^\beta(du; x) = \frac{\sum_{i=1}^n \tilde{N}_i(du; x)}{S^{(0)}(u, x, \beta)}$$

in place of $\alpha(t, x) dt$ yields after some arithmetic $SML_{\mathcal{A}}(\beta, \hat{A}^\beta) = \tilde{C}(\beta) + \text{Rem}$, where Rem is a remainder term independent of β and

$$(1.5) \quad \tilde{C}_{\mathcal{A}}(\beta) = \frac{1}{n} \sum_{i=1}^n \int_{\mathcal{A}} [\beta^T Z_i(u) - \log S^{(0)}(u, x, \beta)] \tilde{N}_i(du; x) dx.$$

As an estimator of the regression parameters we choose the solution $\tilde{\beta}$ to the score equation $\tilde{U}(\beta) = 0$ where $\tilde{U}(\beta) = \nabla \tilde{C}(\beta)$ is the gradient of $\tilde{C}(\beta)$. Under regularity conditions, we show $\tilde{\beta}$ is a consistent estimate of the true parameter value β_0 and $\sqrt{n}[\tilde{\beta} - \beta_0]$ is asymptotically normal with mean zero and covariance matrix that can be estimated consistently by $\tilde{\mathcal{J}}(\tilde{\beta})$ where $\tilde{\mathcal{J}}(\beta) = -\nabla^2 \tilde{C}(\beta)$ is the Hessian of $\tilde{C}(\beta)$. Since evaluation of the smoothed profile likelihood can be quite time consuming in practice, we further give a simpler form of this log-likelihood and verify that under regularity conditions it leads to regression estimates asymptotically equivalent to $\tilde{\beta}$. We also indicate an extension towards higher-dimensional X_t covariates. Finally, we discuss asymptotic normality of Beran's estimate

$$(1.6) \quad \hat{A}(t; w) = \sum_{i=1}^n \int_{[0, t]} \frac{K_a(w - X_i(u))}{S^{(0)}(u, w, \hat{\beta})} N_i(du)$$

of the baseline cumulative hazard function

$$A(t; w) = \int_0^t \alpha(u, w) du$$

and estimates of related parameters such as the baseline hazard function $\alpha(t, w)$ and the doubly cumulative hazard (1.4).

2. Main results. Throughout we denote by \mathcal{B} a compact neighborhood of the true parameter value β_0 . We shall write

$$\xi_{0i}(t, \beta) = Y_i(t) \exp(\beta^T Z_i(t))$$

for short, and correspondingly, we let

$$\xi_{1i}(t, \beta) = Z_i(t) \xi_{0i}(t, \beta), \quad \xi_{2i}(t, \beta) = Z_i^{\otimes 2}(t) \xi_{0i}(t, \beta).$$

Moreover, if $Z_{i\ell}(t)$, $\ell = 1, \dots, p$ is the ℓ th component of the vector $Z_i(t)$, then

$$\xi_{1\ell i}(t, \beta) = Z_{i\ell}(t) \xi_{0i}(t, \beta), \quad \xi_{2\ell i}(t, \beta) = Z_{i\ell}^2(t) \xi_{0i}(t, \beta).$$

We make the assumption that for some small δ , $\delta > 0$, the support of the density $f(u, v)$ of $X(u)$, $u \in [0, 1]$ contains the δ -neighborhood of the rectangle $\mathcal{R} = [0, 1] \times [0, 1]$ of the form

$$\mathcal{R}_\delta = [0, 1] \times [-\delta, 1 + \delta].$$

Further, we let $F_1(u, v)$ and $g_{2\ell}(u, v, \beta)$ be functions on \mathcal{R}_δ and $\mathcal{R}_\delta \times \mathcal{B}$, respectively, such that

$$\begin{aligned} EY_i(u) | X_i(u) &= F_1(u, X_i(u)) \quad \text{a.s.}, \\ E\xi_{1\ell i}^2(u, \beta_0) | X_i(u) &= g_{2\ell}(u, X_i(u)) \quad \text{a.s.} \end{aligned}$$

Under condition I.2 below, these conditional expectations are finite. Finally, for $q = 0, 1, 2$, we set

$$S^{(q)}(u, v, \beta) = \sum_{i=1}^n \xi_{qi}(u, \beta) K_a(v - X_i(u))$$

and we let $s^{(q)}(u, v, \beta)$ be scalar-, vector- and matrix-valued functions on $\mathcal{R}_\delta \times \mathcal{B}$ such that

$$E\xi_{qi}(u, \beta) | X_i(u) = \frac{s^{(q)}(u, X_i(u), \beta)}{f(u, X_i(u))} \quad \text{a.s.}$$

We also let $S^{(q)}(t, x, \beta)$, $q = 1, 2$ be the vector and matrix of first and second derivatives of $S^{(0)}(t, x, \beta)$ with respect to β . To show consistency and asymptotic normality of the regression estimates, we shall use the following assumptions:

- I.1. The density $f(u, v)$ and the function $F_1(u, v)$ are bounded away from 0 on \mathcal{R}_δ .
- I.2. $\max_\ell E \sup\{\xi_{2\ell i}^2(u, \beta) : u \in [0, 1], \beta \in \mathcal{B}\} < \infty$.

- I.3. For $u \in [0, 1]$, the functions $g(u, v) = f(u, v)$ and $\alpha(u, v)$ have a continuous second derivative with respect to v , $v \in [-\delta, 1 + \delta]$. Moreover, the functions $g(u, v)$ and their derivatives with respect to v are bounded uniformly on \mathcal{R}_δ .
- I.4. The functions $g_{2\ell}(u, v, \beta)$ are bounded on $\mathcal{R}_\delta \times \mathcal{B}$. For $q = 0, 1, 2$, the functions $s^{(q)}(u, v, \beta)$ have continuous second derivatives with respect to $v \in [-\delta, 1 + \delta]$ and the derivatives are bounded uniformly on $\mathcal{R}_\delta \times \mathcal{B}$. Moreover, the matrix

$$(2.1) \quad \Sigma(\beta_0) = \int_0^1 \int_0^1 \left[\frac{s^{(2)}(u, w, \beta_0)}{s^{(0)}(u, w, \beta_0)} - e(u, w, \beta_0)^{\otimes 2} \right] \\ \times s^{(0)}(u, w, \beta_0) \alpha(u, w) du dw,$$

where

$$e(u, v, \beta_0) = s^{(1)}(u, w, \beta_0) / s^{(0)}(u, w, \beta_0),$$

is positive semidefinite.

- I.5 The kernel function K is a mean zero left-continuous density of bounded variation on $[-1, 1]$.
- I.6 For some sequence $a = a(n) \rightarrow 0$, $na(n) \rightarrow \infty$,

$$(2.2) \quad \frac{1}{n} S^{(q)}(t, x, \beta) \rightarrow_P s^{(q)}(t, x, \beta)$$

uniformly in $(t, x, \beta) \in \mathcal{R} \times \mathcal{B}$.

Note that if $F(z, d; v, u) = P(Z(u) \leq z, Y(u) = d | X(u) = v)$, $d = 0, 1$ then

$$g_{2\ell}(u, v, \beta) = f(u, v) \int z_\ell^2 \exp(2\beta^T z) F(dz, 1; v, u), \quad z = (z_1, \dots, z_p)^T$$

and conditions I.2–I.4 imply that the functions $s^{(p)}(u, v, \beta)$ are bounded on $\mathcal{R}_\delta \times \mathcal{B}$ and continuous in β for $(u, v) \in \mathcal{R}_\delta$. Condition I.1 implies also that $s^{(0)}(u, v, \beta_0)$ is bounded away from zero.

We consider now in more detail the special case corresponding to observation of an iid sample of random variables $(T \wedge \tilde{T}, I(T \leq \tilde{T}), Z, X)$ where T and \tilde{T} are failure and censoring times, respectively. In this case the smoothed risk process reduces to

$$S^{(0)}(t, x, \beta) = \sum_{i=1}^n I(T_i \wedge \tilde{T}_i \geq t) K_a(x - X_i) \exp(\beta^T Z_i).$$

If the density of the covariate X satisfies condition I.1. and $\log n/na \rightarrow 0$, $na \rightarrow \infty$, then

$$\frac{1}{n} \sum_{i=1}^n I(Z_i \leq z, T_i \wedge \tilde{T}_i \geq t) K_a(x - X_i)$$

is a (strongly) uniformly consistent estimate of $\varphi(t, z, x) = P(Z_i \leq z, T_i \wedge \tilde{T}_i \geq t | X_i = x) f(x)$ where $f(x)$ is the density of X [Dabrowska (1989)]. If

the covariate Z is bounded, then using integration by parts, it is easy to verify that condition I.6 is satisfied. In the general case, under the regularity conditions I.2–I.4, we have $n^{-1}ES^{(q)}(u, v, \beta) = s^{(q)}(u, v, \beta) + O(a^2)$ uniformly on $\mathcal{A} \times \mathcal{B}$. Moreover, condition I.6 is satisfied pointwise in $(u, v, \beta) \in \mathcal{A} \times \mathcal{B}$. By eventually assuming existence of higher moments of the random variables ξ_{2it} , condition I.6 can be verified using Bickel and Wichura’s (1971) bound on the supremum of the process $n^{-1}[S^{(q)} - ES^{(q)}](u, v, \beta)$ in terms of its oscillation modulus. See Nielsen and Linton (1995) for a similar argument.

PROPOSITION 2.1. *Suppose that conditions I.1–I.6 hold.*

(i) *With probability tending to 1, the smoothed profile likelihood*

$$(2.3) \quad \tilde{C}_{\mathcal{A}}(\beta) = \frac{1}{n} \sum_{i=1}^n \int_{\mathcal{A}} [\beta^T Z_i(u) - \log S^{(0)}(u, w, \beta)] \tilde{N}_i(du; w) dw$$

has a unique maximizer $\tilde{\beta}$, and $\tilde{\beta} \rightarrow_P \beta_0$.

(ii) *If in addition $na^4 \rightarrow 0$ then $\sqrt{n}[\tilde{\beta} - \beta_0]$ is asymptotically multivariate normal $\mathcal{N}(0, \Sigma(\beta_0)^{-1})$. Moreover, the covariance matrix $\Sigma(\beta_0)$ can be estimated consistently by $\tilde{\mathcal{I}}(\tilde{\beta})$ where $\tilde{\mathcal{I}}(\beta) = -\nabla^2 \tilde{C}_{\mathcal{A}}(\beta)$ is the Hessian of $\tilde{C}_{\mathcal{A}}(\beta)$.*

The proof in Section 3 will rest on a modification of Andersen and Gill’s (1982) derivation of the asymptotic properties of regression estimates in the standard Cox regression model. Note that evaluation of the profile likelihood, score equation and the matrix $\tilde{\mathcal{I}}(\beta)$ can be time consuming in practice since determining the contribution of $\log S^{(0)}(t, x, \beta)$ and its derivatives requires numerical integration. The following proposition allows reducing this integration to a sum taken over the jump points of the counting processes $N_i(t)$.

PROPOSITION 2.2. *Suppose that conditions I.1–I.6 hold.*

(i) *With probability tending to 1, the profile likelihood*

$$(2.4) \quad C_{\mathcal{A}}(\beta) = \frac{1}{n} \sum_{i=1}^n \int_0^1 [\beta^T Z_i(u) - \log S^{(0)}(u, X_i(u), \beta)] \times I(0 < X_i(u) \leq 1) N_i(du)$$

has a unique maximizer $\hat{\beta}$ and $\hat{\beta} \rightarrow_P \beta_0$.

(ii) *Suppose that $n^{-1}S^{(q)}(u, v, \beta_0)$, $q = 0, 1$, are processes on $[0, 1]^2$ with sample paths of bounded total variation. If $na^4 \rightarrow 0$ and*

$$(2.5) \quad \hat{\psi}(u, w) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_{0i}(u, \beta_0) \left[I(0 < X_i(u) \leq w) - \int_0^w K_a(x - X_i(u)) dx \right] \rightarrow_P 0$$

uniformly in $(t, w) \in [0, 1]^2$, then $\hat{\beta} = \tilde{\beta}_0 + o_P(1/\sqrt{n})$. We also have $I(\hat{\beta}) \rightarrow_P \Sigma(\beta_0)$ where $I(\beta) = -\nabla^2 C_{\mathcal{A}}(\beta)$ is the Hessian of $C_{\mathcal{A}}(\beta)$.

If $na^4 \rightarrow 0$ then under the assumed regularity conditions, (2.5) holds pointwise in $(u, w) \in [0, 1]^2$. If $\widehat{\psi}$ is a tight sequence in $D([0, 1]^2)$, then we also have uniform convergence. In particular in the survival analysis setting with time independent (X, Z) covariates and $Y_i(u)$ processes given by $Y_i(u) = I(T_i \wedge \widetilde{T}_i \geq u)$ where (T_i, \widetilde{T}_i) are the failure and censoring times, the process $\widehat{\psi}$ is tight if the Z covariates are bounded or, for example, if for $p = 2, 4$ the functions $h_p(u, v)$, $(u, v) \in \mathcal{R}_\delta$, $h_p(u, X_i) = E[Y_i(u) \exp(p\beta_0^T Z_i) | X_i]$ a.s., satisfy the Lipschitz condition $|h_p(u, v) - h_p(u', v)| \leq C|u - u'|$ uniformly in $v \in [-\delta, 1 + \delta]$. In the latter case, tightness of $\widehat{\psi}$ is a consequence of Bickel and Wichura's (1971) moment criteria and some elementary algebra.

Both propositions can be extended to include higher-dimensional covariates. We discuss now this point in some more detail by assuming one more time standard censoring model with time independent covariates (X, Z) . In this case, the smoothed counting and risk processes are

$$(2.6) \quad \widetilde{N}_i(t; x) = \frac{1}{a_1 \cdots a_d} K\left(\frac{x_1 - X_{i1}}{a_1}, \dots, \frac{x_d - X_{id}}{a_d}\right) N_i(t)$$

and

$$(2.7) \quad S^{(0)}(t, x, \beta) = \frac{1}{a_1 \cdots a_d} \sum_{i=1}^n Y_i(t) \exp(\beta^T Z_i) K\left(\frac{x_1 - X_{i1}}{a_1}, \dots, \frac{x_d - X_{id}}{a_d}\right).$$

Here $x = (x_1, \dots, x_d)$ is a d -variate covariate and K is a mean zero density of bounded variation on $[-1, 1]^d$. Moreover, $N_i(t) = I(T_i \leq t \wedge \widetilde{T}_i)$ and $Y_i(t) = I(T_i \wedge \widetilde{T}_i \geq t)$ where T_i and \widetilde{T}_i are failure and censoring times conditionally independent given Z_i . If the covariates Z_i are bounded then similarly to the scalar case, a sufficient condition for (strong) uniform consistency of $n^{-1}S^{(q)}(t, x, \beta)$, $q = 0, 1, 2$ is

$$(2.8) \quad \max a_i \rightarrow 0, \quad na_1 \cdots a_d \rightarrow \infty, \quad \frac{\log n}{na_1 \cdots a_d} \rightarrow 0.$$

If $d \leq 3$, then, under the smoothness conditions of assumption I, the estimates $\widetilde{\beta}$ and $\widehat{\beta}$ are asymptotically normal at rate \sqrt{n} , provided $n \max a_i^4 \rightarrow 0$ and (2.8) holds. These conditions are satisfied, for instance, if $a_i \sim n^{-r}$ and $1/4 < r < 1/d$. For higher-dimensional covariates, $d > 3$, the regression estimates $\widetilde{\beta}$ and $\widehat{\beta}$ remain consistent and the Hessian $\widetilde{\mathcal{J}}(\beta)$ of the profile likelihood satisfies $\widetilde{\mathcal{J}}(\beta^*) \rightarrow_p \Sigma(\beta_0)$ for any $\beta^* \rightarrow_p \beta_0$. The regression estimates $\widetilde{\beta}$ are, however, no longer asymptotically normal at rate \sqrt{n} . One possible approach to retain the \sqrt{n} rate of convergence is to consider a one-step MLE estimate

$$\widetilde{\beta} = \beta^* + [n \cdot \widetilde{\mathcal{J}}(\beta^*)]^{-1} \widetilde{U}(\beta^*),$$

where β^* is an initial \sqrt{n} -consistent estimate of β_0 , $\widetilde{\mathcal{J}}(\beta)$ is the Hessian of the profile likelihood $\widetilde{C}(\beta)$ in Proposition 2.1 while $\widetilde{U}(\beta)$ is a score process based

on higher order kernels. Specifically, let K_1 be a function on $[-1, 1]$ such that

$$\begin{aligned} \int K_1(u_1, \dots, u_d) du_1 \dots u_d &= 1, \\ \int u_1^{p_1} \dots u_d^{p_d} K_1(u_1, \dots, u_d) du_1 \dots u_d &= 0 \quad \text{if } \sum p_i \leq 2p - 1, \\ &\neq 0 \quad \text{if } \sum p_i = 2p, \end{aligned}$$

and denote by \tilde{N}_{1i} and $S_1^{(0)}$ the analogues of the processes (2.6) and (2.7) obtained by replacing the kernel K with K_1 . Then, assuming that the differentiability conditions I.2 and I.3 hold for derivatives of order $2p$, a minor modification of arguments in Section 3 can be used to show that the score process

$$\tilde{U}(\beta) = \frac{1}{n} \sum_{i=1}^n \int_{\mathcal{X}} \left[Z_i - \frac{S_1^{(1)}(t, x, \beta)}{S_1^{(0)}(t, x, \beta)} \right] \tilde{N}_{1i}(dt; x) dx$$

satisfies $\sqrt{n}\tilde{U}(\beta_0) \Rightarrow \mathcal{N}(0, \Sigma(\beta_0))$ provided $n \max a_i^{4p} \rightarrow 0$ and (2.8) holds. If $a_i \sim n^{-r}$, then these conditions are satisfied if $1/4p < r < 1/d$. We also have $\tilde{I}(\beta^{**}) \rightarrow_p \Sigma(\beta_0)$ for any $\beta^{**} \rightarrow_p \beta_0$, where $\tilde{I}(\beta) = \nabla \tilde{U}(\beta)$. By Taylor expansion and \sqrt{n} -consistency of the initial estimate β^* , $\sqrt{n}(\tilde{\beta} - \beta_0)$ is then asymptotically $\mathcal{N}(0, \Sigma(\beta_0)^{-1})$. Under suitable smoothness assumptions, an initial \sqrt{n} -consistent estimate of β_0 can be based on averaged derivative estimation.

We return now to the case of univariate X_t covariates and discuss estimation of the doubly cumulative hazard A_d and the hazard function α . First, the baseline cumulative hazard function $A(t; x) = \int_0^t \alpha(u, x) du$ can be recovered using Beran's estimate

$$\hat{A}(t; x) = \sum_{i=1}^n \int_0^t \frac{\tilde{N}_i(du; x)}{S^{(0)}(u, x, \hat{\beta})},$$

where now $\hat{\beta}$ denotes either of the two estimators of the regression parameters β_0 . Further, the baseline doubly cumulative hazard $A_d(t, x) = \int_0^x A(t; w) dw$ is estimated by its sample analogue $\hat{A}_d(t, x)$ obtained by substituting Beran's estimate. Finally, we follow Ramlau-Hansen (1983) and McKeague and Utikal (1990a) and smooth $\hat{A}(t; x)$ in the time direction to get

$$\hat{\alpha}(t, x) = \int_0^{\tau} \tilde{K}_b(t - u) \hat{A}(du; x)$$

as an estimate of the hazard function α . Here \tilde{K} is a kernel function on $[-1, 1]$ and $\tilde{K}_b(\cdot) = b^{-1} \tilde{K}(\cdot/b)$ for some sequence of bandwidths $b = b(n) \rightarrow 0$. In the absence of the covariate Z , Keiding (1990) and Nielsen and Linton (1995) proposed alternative hazard function estimators involving smoothing both counting and risk processes in the (time) \times (covariate) direction. Under regularity

conditions, their estimates are asymptotically equivalent to $\hat{\alpha}$; therefore, we do not consider them in more detail.

PROPOSITION 2.3. *Suppose that the assumptions of Propositions 2.1(ii) or 2.2(ii) hold and set $c(K) = \int K^2(r) dr$, and*

$$\varphi(t, x, \beta_0) = \frac{\alpha(t, x)}{s^{(0)}(t, x, \beta_0)}.$$

(i) *For almost all $x \in [0, 1]$, the process $\hat{V}(t; x) = \sqrt{na}[\hat{A}(t; x) - A_0(t; x)]$ converges weakly in $D[0, 1]$ to a time-transformed Brownian motion $V(t; x)$ with variance function*

$$\text{Var } V(t; x) = c(K) \int_{[0, t]} \varphi(u, x, \beta_0) du$$

and \hat{V} is asymptotically independent of the standardized regression coefficients $\hat{\beta}$.

(ii) *If $n^{-1}S^{(0)}(u, w, \beta_0)$, is a process on $[0, 1]^2$ with sample paths of bounded total variation, then $\hat{V}_d(t, x) = \sqrt{n}[\hat{A}_d - A_d](t, x)$, $(t, x) \in [0, 1]^2$ converges weakly in $D([0, 1]^2)$ to a mean zero Gaussian process $V_d(t, w)$ with covariance function*

$$(2.9) \quad \text{cov}(V_d(t, x), V_d(t', x')) = \int_0^{t \wedge t'} \int_0^{x \wedge x'} \varphi(u, w, \beta_0) du dw \\ + \eta(t, x, \beta_0)^T \Sigma(\beta_0)^{-1} \eta(t', x', \beta_0),$$

where

$$(2.10) \quad \eta(t, x, \beta_0) = \int_0^t \int_0^x e(u, w, \beta_0) A_d(du, dw).$$

The asymptotic covariance of $\sqrt{n}(\hat{\beta} - \beta_0)$ and $\hat{V}_d(t, x)$ is equal to $-\Sigma(\beta_0)^{-1} \eta(t, x)$.

(iii) *Suppose that \tilde{K} is a kernel function satisfying condition I.5 and $\alpha(t, x)$ is continuously twice differentiable in (t, x) , $(t, x) \in \mathcal{R}_\delta$. If $nab \rightarrow \infty$ and $nab^5 \rightarrow 0$, then for almost all $(t, x) \in [\delta, 1 - \delta] \times [0, 1]$, $\sqrt{nab}[\hat{\alpha} - \alpha](t, x)$ is asymptotically mean zero normal with variance $\sigma_\alpha^2(t, x) = c(K)c(\tilde{K})\varphi(t, x, \beta_0)$ and asymptotically independent of $\sqrt{n}[\hat{\beta} - \beta_0]$.*

Note that the asymptotic covariance of $\hat{A}_d(t, x)$ has the same structure as the Aalen–Nelson estimate in the usual Cox regression model; that is, (2.9) represents the asymptotic covariance of the doubly cumulative hazard function estimate appropriate when β_0 is known plus a correction term accounting for estimation of the β_0 . In the remaining two cases, the limiting distributions are not affected by estimation of the regression parameters. Further, the covariances in part (i) and (ii) can be estimated using the substitution method,

whereas Lemma 3.2(iv) implies that

$$(2.11) \quad \hat{\sigma}_\alpha^2(t, x) = c(K)c(\tilde{K}) \sum_{i=1}^n \int_{t-b}^{t+b} \tilde{K}_b(t-u) \frac{n\tilde{N}_i(du; x)}{S^{(0)}(u, x, \hat{\beta})^2}$$

is a consistent estimate of the asymptotic variance of $\hat{\alpha}(t, x)$ provided the sequence of bandwidths satisfies $a^4b^{-1} \rightarrow 0$.

Proposition 2.3 can be used to construct goodness-of-fit tests for specific hypotheses on the form of the baseline hazard function $\alpha(t, x)$. If $\alpha(t, x)$ belongs to some parametric family of hazards, $\mathcal{S} = \{\alpha_\theta(t, x): \theta \in \Theta\}$ then estimation of the (θ, β) parameters can be based for instance on the maximum likelihood method. Under regularity conditions derived by Borgan (1984), such estimates are typically asymptotically normal at rate \sqrt{n} . If $\hat{\alpha}_{\text{MLE}}$ denotes the corresponding maximum likelihood estimate of α , then a conservative test for the hypothesis $\mathcal{H}_0: \alpha \in \mathcal{S}$ can be obtained by checking if $\hat{\alpha}_{\text{MLE}}$ falls into the Bonferroni simultaneous confidence bands of the form $\hat{\alpha}(t, x) \pm z(p)\hat{\sigma}(t, x)[nab]^{-1/2}$, where $\hat{\alpha}(t, x)$ is the estimate of Proposition 2.3, $\hat{\sigma}(t, x)$ is given by (2.11) and $z(p)$ is a percentile of the standard normal distribution.

McKeague and Utikal (1991) proposed chi-squared goodness-of-fit tests obtained by comparing the doubly cumulative hazard function estimate valid under the general model [in our case (1.3)] with its counterpart derived under a specific submodel. Here we consider briefly the test for the hypothesis $\mathcal{H}_0: \alpha(t, x) = \alpha_0(t)$ where α_0 is an unknown function. In the standard survival analysis setting with time independent (X, Z) covariates, \mathcal{H}_0 corresponds to the hypothesis of conditional independence of the failure time T and covariate X given Z , and proportional hazard model assumption on the pair (T, Z) . Under the null hypothesis the doubly cumulative hazard function reduces to $A_d(t, x) = xA_0(t)$ where $A_0(t) = \int_0^t \alpha_0(u) du$. We can estimate it by $\hat{A}_{d0}(t, x) = x\hat{A}_0(t)$,

$$\hat{A}_0(t) = \sum_{i=1}^n \int_0^t \int_0^1 \frac{\tilde{N}_i(du; w)dw}{S_c^{(0)}(u, \hat{\beta}_c)},$$

where

$$S_c^{(0)}(u, \beta) = \int_0^1 S^{(0)}(u, w, \beta) dw$$

is the integrated risk process and $\hat{\beta}_c$ is either the estimate of Propositions 2.1 and 2.2 or else it is the solution of the score equation

$$(2.12) \quad 0 = \tilde{U}_c(\beta) = \frac{1}{n} \sum_{i=1}^n \int_{\mathcal{A}} [Z_i(t) - \nabla \log S_c^{(0)}(u, \beta)] \tilde{N}_i(du; w) dw.$$

Letting $S_c^{(q)}(u, \beta) = \nabla^q S_c^{(0)}(u, \beta)$, $q = 1, 2$, under condition I.6, we have

$$\frac{1}{n} S_c^{(q)}(u, \beta) \rightarrow_P s_c^{(q)}(u, \beta) = \int_0^1 s^{(q)}(u, w, \beta) dw$$

uniformly in $(u, \beta) \in [0, 1] \times \mathcal{B}$. Moreover, by repeating some of the arguments of Section 3, under the null hypothesis, the processes $\widehat{V}_d(t, x) = \sqrt{n}[\widehat{A}_d - A_d](t, x)$ and $\widehat{V}_{d0}(t, x) = \sqrt{n}x[\widehat{A}_0(t) - A_0](t)$ are asymptotically equivalent to mean zero Gaussian processes (V_d, V_{d0}) where V_d is defined in Proposition 2.3 and

$$\begin{aligned} \text{cov}(V_{0d}(t, x), V_{0d}(t', x')) &= \text{cov}(V_{0d}(t, x), V_d(t', x')) \\ &= xx' \int_0^{t \wedge t'} \frac{\alpha_0(u) du}{s_c(u, \beta_0)} + \eta_c(t, x, \beta_0)^T \Sigma_c(\beta_0)^{-1} \eta_c(t', x'). \end{aligned}$$

Here $\eta_c(t, x) = x \int_0^t e_c(u, \beta_0) \alpha_0(u) du$, $e_c(u, \beta_0) = s_c^{(1)}(u, \beta_0) / s_c^{(0)}(u, \beta_0)$. Moreover, $\Sigma_c(\beta_0) = \Sigma(\beta_0)$ if $\widehat{\beta}_c$ is the estimate of Propositions 2.1 and 2.2 and $\Sigma_c(\beta_0)$ is the stochastic limit of $-\nabla \widetilde{U}_c(\beta)$, if $\widehat{\beta}_c$ is the solution of the score equation (2.12). If $\mathcal{D} = \{D_{ij}: i = 1, \dots, p, j = 1, \dots, q\}$, a partition of the parameter set $[0, 1]^2$ into disjoint rectangles of the form $D_{ij} = D_{1i} \times D_{2j}$ where $D_{1i} = (t_{i-1}, t_i]$, $D_{2j} = (x_{j-1}, x_j]$, $0 = t_0 < t_1 < \dots < t_p = 1$, $0 = x_0 < x_1 < \dots < x_q = 1$, then setting $\widehat{R}_{ij} = \sqrt{n}[\widehat{A}_d - \widehat{A}_{d0}](D_{ij})$, under the null hypothesis the vector $\widehat{R} = \{\widehat{R}_{ij}: i \leq p, j \leq q\}$ is approximately mean zero normal with covariance matrix C determined by $\text{As.cov}(\widehat{R}_{ij}, \widehat{R}_{i'j'}) = \text{cov}(V_d(D_{ij}), V_d(D_{i'j'})) - \text{cov}(V_{0d}(D_{ij}), V_{0d}(D_{i'j'}))$. Denoting by \widehat{C} the sample analogue of this matrix, and assuming that its rank is pq , under the null hypothesis the statistic $\widehat{R}^T \widehat{C}^{-1} \widehat{R}$ has approximately a chi-squared distribution with pq degrees of freedom.

Under the additional assumption that $s^{(q)}(u, w, \beta_0) \equiv s_c^{(q)}(u, \beta_0)$, $(u, w) \in [0, 1]^2$, the process $\sqrt{n}[\widehat{A}_d(t, x) - x\widehat{A}_0(t)]$ is under the null hypothesis asymptotically equivalent to a stochastic integral of the form

$$(2.13) \quad \int_0^t \int_0^x \sqrt{h} dB - x \int_0^t \int_0^1 \sqrt{h} dB,$$

where B is a Brownian sheet on $[0, 1]^2$ and $h(u, w) = \alpha_0(u) / s_c(u, \beta_0)$. Under this additional constraint on the expected risk functions $s^{(q)}$, the test for the hypothesis $\mathcal{H}_0: \alpha(t, x) = \alpha_0(t)$ can be based on McKeague, Nikabadze and Sun's (1995) transformation of (2.13) to its innovation Brownian sheet. The resulting test applies in particular to the standard censoring model if, under the null hypothesis, the covariate X is independent of (T, \widetilde{T}, Z) and T and \widetilde{T} are failure and censoring times, conditionally independent of Z .

3. Proofs. We denote by $\|\cdot\|_\infty$ the supremum norm on $\mathcal{B} = [0, 1]^2$. From Section 2 recall that

$$\xi_{0i}(t, \beta) = Y_i(t) \exp(\beta^T Z_i(t))$$

and

$$\xi_{1i}(t, \beta) = Z_i(t) \xi_{i0}(t, \beta), \quad \xi_{2i}(t, \beta) = Z_i^{\otimes 2}(t) \xi_{0i}(t, \beta).$$

We shall also write

$$M_i(t) = N_i(t) - \int_0^t \xi_{0i}(u, \beta_0) \alpha(u, X_i(u)) du$$

for the fundamental martingale associated with the counting process $N_i(t)$, $t \in [0, 1]$ and set

$$\tilde{N}_i(t; x) = \int_0^t K_a(x - X_i(u)) N_i(du),$$

$$\tilde{M}_i(t; x) = \int_0^t K_a(x - X_i(u)) M_i(du).$$

We denote by \bar{K} the cdf associated with the density K and similarly, $\bar{K}_a(\cdot) = \bar{K}(\cdot/a)$ is the cdf corresponding to the density $K_a(\cdot) = a^{-1}K(\cdot/a)$.

In Section 3.1 we give a sequence of lemmas which are next used to modify asymptotic normality results of Andersen and Gill (1982) to the present setting. We shall frequently rely on the fact that the sequence \bar{K}_a , $a \rightarrow 0$ converges weakly to a cdf assigning point mass to $r = 0$. Thus for any bounded continuous functions $\varphi_1(x)$ and $\varphi_2(x_1, x_2)$, we have

$$(3.1) \quad \int \varphi_1(r) \bar{K}_a(dr) \rightarrow \varphi_1(0), \quad \int \varphi_2(r_1, r_2) \bar{K}_a(dr_1) \bar{K}_a(dr_2) \rightarrow \varphi_2(0, 0).$$

More generally, if φ is a function integrable with respect to Lebesgue measure, then

$$(3.2) \quad \int \varphi(u) K_a(x - u) du = \int \varphi(x - r) \bar{K}_a(dr) \rightarrow \varphi(x)$$

for almost all x [Stein (1970)]. Finally, we shall often use boundedness of the kernels K and \tilde{K} implied by the assumption I.5. Thus $aK_a(x - \cdot)$ and $b\tilde{K}_b(t - \cdot)$ are also bounded functions.

3.1. *Some auxiliary results.* In this section we denote by $\hat{h}(u, x)$, $(u, x) \in \mathcal{A}$ a sequence of processes such that for any $x \in [0, 1]$, $\hat{h}(\cdot, x)$ and $\int_0^x \hat{h}(\cdot, w) K(w - X_i(\cdot)) dw$ are \mathcal{F}_t -predictable. In Section 3.2, $\hat{h}(u, x)$ will be either a deterministic function or

$$\hat{h}(u, x) = \frac{S^{(1)}(u, x, \beta_0)}{S^{(0)}(u, x, \beta_0)},$$

$$\hat{h}(u, x) = \log \frac{S^{(0)}(u, x, \beta)}{S^{(0)}(u, x, \beta_0)}$$

or $\hat{h}(u, x) = S^{(0)}(u, x, \beta_0)^{-1}$. In these cases the predictability assumption is satisfied.

The first lemma will be used to verify consistency of the estimate $\tilde{\beta}$ of Proposition 2.1 and asymptotic unbiasedness of the doubly cumulative hazard estimate $\hat{A}_d(t, x)$ and the score process $\tilde{U}(\beta_0) = \nabla \tilde{C}_{\mathcal{A}}(\beta)|_{\beta=\beta_0}$.

LEMMA 3.1. For $p = 0, 1$ let

$$B_{pn}(t, x) = \frac{1}{n} \int_0^t \int_0^x \widehat{h}(u, w) \sum_{i=1}^n \xi_{pi}(u, \beta_0) K_a(w - X_i(u)) \times [\alpha(u, X_i(u)) - \alpha(u, w)] du dw,$$

$$D_{pn}(t, x) = \frac{1}{n} \sum_{i=1}^n \int_0^t \int_0^x \widehat{h}(u, w) Z_i^p(u) \widetilde{N}_i(du; w) dw.$$

(i) Suppose that $\|\widehat{h}\|_\infty = O_P(1)$ and $na \rightarrow \infty$. Then $\|B_{pn}\|_\infty \rightarrow_P 0$. If in addition $na^4 \rightarrow 0$, then $\|\sqrt{n}B_{pn}\|_\infty \rightarrow_P 0$.

(ii) Let h be a bounded function such that $\|\widehat{h} - h\|_\infty \rightarrow_P 0$. Then $\|D_{pn} - D_p\|_\infty \rightarrow_P 0$ where

$$D_p(t, x) = \int_0^t \int_0^x h(u, w) s^{(p)}(u, w, \beta_0) \alpha(u, w) du dw.$$

(iii) If $\|\widehat{h}\|_\infty \rightarrow_P 0$, then for any fixed $x \in [0, 1]$ we have

$$(3.3) \quad \frac{1}{\sqrt{n}} \int_{[0,t] \times [0,x]} \sum_{i=1}^n \widehat{h}(u, w) \widetilde{M}_i(du; w) dw \rightarrow_P 0$$

uniformly in $t \in [0, 1]$.

PROOF. We shall verify part (i) for the process B_{0n} ; each component of the vector B_{1n} can be handled analogously. Define

$$(3.4) \quad \theta_i(u, w, \beta_0) = \xi_{0i}(u, \beta_0) K_a(w - X_i(u)) [\alpha(u, X_i(u)) - \alpha(u, w)].$$

We have $\|B_{0n}^2\|_\infty = O_P(n^{-1}) I_n$, where

$$I_n = \frac{1}{n} \int_0^1 \int_0^1 \left[\sum_{i=1}^n \theta_i(u, w, \beta_0) \right]^2 du dw$$

$$= \frac{1}{n} \sum_{i=1}^n \int_0^1 \int_0^1 \theta_i^2(u, w, \beta_0) du dw + \frac{1}{n} \sum_{i_1 \neq i_2} \int_0^1 \int_0^1 \prod_{j=1}^2 \theta_{i_j}(u, w, \beta_0) du dw$$

$$= I_{1n} + I_{2n}.$$

Therefore it is enough to verify that $EI_{1n} = O(a)$ and $EI_{2n} = O(na^4)$. Using

$$0 \leq \int_0^1 K_a(w - X_i(u)) dw \leq 1$$

and boundedness of $aK_a(\cdot) = K(\cdot/a)$, we have

$$EI_{1n} = E \int_0^1 \int_0^1 [\xi_{0i}(u, \beta_0) K_a(w - X_i(u)) [\alpha(u, X_i(u)) - \alpha(u, w)]]^2 du dw$$

$$= O(a^2) E \int_0^1 \int_0^1 \xi_{0i}^2(u, \beta_0) K_a(w - X_i(u))^2 du dw$$

$$= O(a) E \sup_{u \in [0, 1]} \xi_{0i}^2(u, \beta_0) = O(a).$$

Moreover,

$$EI_{2n} = (n - 1)E \int_0^1 \int_0^1 \prod_{j=1}^2 \left[g_1(u, X_j(u)) K_a(w - X_j(u)) \right. \\ \left. \times [\alpha(u, X_j(u)) - \alpha(u, w)] \right] du dw,$$

where $g_1(u, X_i(u)) = E \xi_{0i}(u, \beta_0) | X_i(u)$ a.s. Under assumption I.4, $g_1(u, w)$ is a bounded function on \mathcal{R}_δ with a bounded second derivative with respect to w . The latter coupled with a two-term Taylor expansion shows that $EI_{2n} = O(n\alpha^4)$.

In part (ii) we consider again the case of $p = 0$ only. We have $\widehat{D}_{0n} - D_0 = \sum_{j=1}^n J_{in}$ where

$$(3.5) \quad J_{1n}(t, x) = \frac{1}{n} \sum_{i=1}^n \int_0^t \int_0^x \widehat{h}(u, w) \widetilde{M}_i(du; w) dw, \\ J_{2n}(t, x) = \frac{1}{n} \sum_{i=1}^n \int_0^t \int_0^x \widehat{h}(u, w) \theta_i(u, w, \beta_0) du dw, \\ J_{3n}(t, x) = \int_0^t \int_0^x \left[\widehat{h}(u, w) S^{(0)}(u, w) - h(u, w) s^{(0)}(u, w) \right] \alpha(u, w) du dw.$$

By part (i) and uniform consistency of $n^{-1}S^{(0)}$, $\|J_{pn}\|_\infty \rightarrow_P 0$ for $p = 2, 3$. Further, by application of Lenglart’s inequality, for any fixed $x \in [0, 1]$, we have

$$(3.6) \quad P\left(\sup_{t \leq 1} |J_{1n}(t, x)| > \eta\right) \leq \frac{\eta^2}{\delta^2} + P(\langle J_{1n} \rangle(1, x) > \delta),$$

where $\langle J_{1n} \rangle(\cdot, x)$ is the predictable variation of the process $J_{1n}(\cdot, x)$, $t \in [0, 1]$. But

$$\langle J_{1n} \rangle(t, x) = \frac{1}{n^2} \sum_{i=1}^n \int_0^t \left[\int_0^x \widehat{h}(u, w) K_a(w - X_i(u)) dw \right]^2 \xi_{0i}(u, \beta_0) \alpha(u, X_i(u)) du \\ \leq \frac{1}{n^2} (\|\widehat{h}\|_\infty)^2 O(1) \sum_{i=1}^n \int_0^1 \left[\int_0^1 K_a(w - X_i(u)) dw \right]^2 \xi_{0i}(u, \beta_0) du.$$

The term in brackets is bounded by one. Therefore the law of large numbers implies that the right-hand side is of order $O_P(1/n)$.

We have shown that for any fixed $x \in [0, 1]$, we have $\widehat{D}_{0n}(t, x) \rightarrow_P D_0(t, x)$ uniformly in $t \in [0, 1]$. If \widehat{h} and h are nonnegative then both \widehat{D}_{0n} and D_0 have nondecreasing paths. By application of the Pólya–Cantelli theorem, we have $\|D_{0n} - D_0\|_\infty \rightarrow_P 0$. In the general case, the Pólya–Cantelli theorem can be applied to \widehat{h}^\pm and h^\pm where $\widehat{h}^+ = \max(\widehat{h}, 0)$, $\widehat{h}^- = \min(-\widehat{h}, 0)$ and h^\pm defined analogously. This completes the proof of part (ii).

Finally, under the assumptions of part (iii), the predictable variation of the sum on the left-hand side of (3.4) is equal to $n \langle J_{1n} \rangle(1, x) = \|\widehat{h}\|_\infty O_P(1) =$

$o_p(1)$ where J_{1n} is given by (3.6). Therefore, Lenglar's inequality (3.6) completes the proof. \square

Using uniform consistency of $n^{-1}S^{(q)}(t, x, \beta)$, $q = 0, 1$ and the strong law of large numbers, it is also easy to verify that if h is a bounded function on \mathcal{R} such that $\|\widehat{h} - h\|_\infty \rightarrow_p 0$ then

$$(3.7) \quad \frac{1}{n} \sum_{i=1}^n \int_0^t \widehat{h}(u, X_i(u)) Z_i^p(u) I(0 < X_i(u) \leq 1) N_i(du) \rightarrow_p D_p(t, x)$$

uniformly in $(t, x) \in [0, 1]^2$ where $D_p(t, x)$ is defined by (3.3). Moreover, if $\|\widehat{h}\|_\infty \rightarrow_p 0$, then for any fixed $x \in [0, 1]$ we have

$$(3.8) \quad \frac{1}{\sqrt{n}} \int_{[0,t]} \sum_{i=1}^n \widehat{h}(u, X_i(u)) I(0 < X_i(u) \leq x) M_i(du) \rightarrow_p 0$$

uniformly in $t \in [0, 1]$.

The next lemma gives two "local" analogues of Lemma 3.1 used in the proof of asymptotic normality of Beran's estimate $\widehat{A}(t; x)$ and hazard function estimate $\widehat{\alpha}(t, x)$.

LEMMA 3.2. For $x \in [0, 1]$ set

$$\widehat{B}_{0n}(t; x) = \frac{1}{n} \int_{[0,t]} \widehat{h}(u, x) \sum_{i=1}^n \theta_i(u, x, \beta_0) du,$$

$$\widehat{D}_{0n}(t; x) = \frac{1}{n} \sum_{i=1}^n \int_0^t \widehat{h}(u, x) \widetilde{N}_i(du; x).$$

where $\theta_i(u, x, \beta_0)$ is given by (3.4). Moreover, for $(t, x) \in [\delta, 1 - \delta] \times [0, 1]$ set

$$\widetilde{B}_{0n}(t; x) = \int_{t-b}^{t+b} \widetilde{K}_b(t-u) \widehat{B}_{0n}(du, x), \quad \widetilde{D}_{0n}(t; x) = \int_{t-b}^{t+b} \widetilde{K}_b(t-u) \widehat{D}_{0n}(du, x).$$

Finally, for any function $h(u, x)$, let $\|h\|_x = \sup_{u \in [0, 1]} |h(u, x)|$.

(i) Suppose that $\|\widehat{h}\|_x = O_p(1)$, and $na \rightarrow \infty$. Then (a) $\|\widehat{B}\|_x \rightarrow_p 0$ and $\sqrt{a} \|\widehat{D}_n\|_x \rightarrow_p 0$; (b) If in addition $na^5 \rightarrow 0$, then $\sqrt{na} \|\widehat{B}_n\|_x \rightarrow_p 0$.

(ii) If $\|\widehat{h} - h\|_x \rightarrow_p 0$ and $\|h\|_x < \infty$, then $\|\widehat{D}_n - D\|_x \rightarrow_p 0$ where

$$D(t; x) = \int_0^t h(u, x) s^{(p)}(u, x, \beta_0) \alpha(u, x) du.$$

(iii) Suppose that $\widehat{h}(u, x)$, $u \in [t - \delta, t + \delta]$ is bounded in probability and $nab \rightarrow \infty$. Then (a) $\sqrt{ab} \widehat{D}_n(t; x) \rightarrow_p O$. Moreover, (b) if $a^4 b^{-1} \rightarrow 0$ then $\widetilde{B}_n(t; x) \rightarrow_p 0$ and (c) if $na^5 \rightarrow 0$, then $\sqrt{nab} \widetilde{B}(t; x) \rightarrow_p 0$.

(iv) If $\widehat{h}(u, x) \rightarrow_p h(u, x)$ uniformly in $u \in [t - \delta, t + \delta]$ and $nab \rightarrow \infty$, $a^4 b^{-1} \rightarrow 0$, then $\widetilde{D}_n(t; x) \rightarrow_p d(t; x)$ where $d(t; x) = h(t, x) s^{(0)}(t, x, \beta_0) \alpha(t, x)$.

In parts (i) and (ii), the assumptions on the \widehat{h} process and the “in probability” limits are meant to hold for almost all $x, x \in [0, 1]$; in parts (iii) and (iv) for almost all $(t, x), (t, x) \in [\delta, 1 - \delta] \times [0, 1]$.

PROOF. The argument is similar to the proof of Lemma 3.1. We indicate merely changes that apply to the terms \widetilde{B}_n and \widetilde{D}_n . For almost all (t, x) , we have $\widetilde{B}_n(t; x) = O_P([nab]^{-1})I_n(t, x)$, where

$$I_n(t; x) = \frac{ab}{n} \int_{t-b}^{t+b} \left[\sum_{i=1}^n \widetilde{K}_b(t-u)\theta_i(u, x, \beta_0) \right]^2 du = I_{1n}(t; x) + I_{2n}(t; x)$$

and

$$I_{1n}(t; x) = \frac{ab}{n} \sum_{i=1}^n \int_{t-b}^{t+b} \widetilde{K}_b^2(t-u)\theta_i^2(u, x, \beta_0) du,$$

$$I_{2n}(t; x) = \frac{ab}{n} \sum_{i_1 \neq i_2} \int_{t-b}^{t+b} \prod_{j=1}^2 \widetilde{K}_b(t-u)\theta_{i_j}(u, x, \beta_0) du.$$

By boundedness of $aK_a(x - \cdot)$ and $b\widetilde{K}_b(t - \cdot)$, we have

$$\begin{aligned} EI_{1n}(t; x) &= O(a^3b)E \int_{t-b}^{t+b} \xi_{0i}^2(u, \beta_0)\widetilde{K}_b^2(t-u)K_a^2(w - X_i(u))^2 du \\ &= O(a)E \sup \xi_{0i}^2(u, \beta_0) \int_{t-b}^{t+b} \widetilde{K}_b(t-u) du = O(a) \end{aligned}$$

and by a two-term Taylor expansion,

$$EI_{2n} = O((n-1)a^5b) \int_{t-b}^{t+b} K_b^2(t-u) du = O(na^5).$$

This implies parts (b) and (c) of (iii). Further, $\widetilde{D}_n(t, x) = \widetilde{D}_{1n}(t; x) + \widetilde{D}_{2n}(t; x)$, (t, x) -a.e. where

$$\widetilde{D}_{1n}(t; x) = \frac{1}{n} \sum_{i=1}^n \int_{t-b}^{t+b} \widehat{h}(u, x)\widetilde{K}_b(t-u)\widetilde{M}_i(du; x),$$

$$\widetilde{D}_{2n}(t; x) = \frac{1}{n} \sum_{i=1}^n \int_{t-b}^{t+b} \widehat{h}(u, x)\widetilde{K}_b(t-u)K_a(x - X_i(u))\xi_{0i}(u, \beta_0)\alpha(u, X_i(u)) du.$$

Using boundedness of α and \widehat{h} , for almost all (t, x) , we have

$$\begin{aligned} \widetilde{D}_{2n}(t; x) &= O_P(1) \sup_{u \in [t-b, t+b]} \frac{S^{(0)}(u, x, \beta_0)}{n} \int_{t-b}^{t+b} \widetilde{K}_b(t-u) du \\ &= O_P(1), \quad (t, x)\text{-a.e.} \end{aligned}$$

Moreover, the predictable variation of the process

$$\widetilde{W}(s; x) = \frac{1}{n} \sum_{i=1}^n \int_{t-b}^s \widehat{h}(u, x)\widetilde{K}_b(t-u)\widetilde{M}_i(du; x), \quad s \in [t-b, t+b],$$

satisfies, for almost all (t, x) ,

$$\begin{aligned} \langle \tilde{W} \rangle(s, x) &= \frac{1}{n^2} \sum_{i=1}^n \int_{t-b}^s \hat{h}^2(u, x) \tilde{K}_b^2(t-u) \xi_{0i}(u, \beta_0) \\ &\quad \times K_a^2(x - X_i(u)) \alpha(u, X_i(u)) du \\ &= O_P((nab)^{-1}) \frac{1}{n} \sum_{i=1}^n \int_{t-b}^{t+b} \tilde{K}_b(t-u) K_a(x - X_i(u)) \xi_{0i}(u, \beta_0) du \\ &= O_P((nab)^{-1}) \sup_{u \in [t-b, t+b]} \frac{S^{(0)}(u, x, \beta_0)}{n} \int_{t-b}^{t+b} \tilde{K}_b(t-u) du \\ &= O_P((nab)^{-1}). \end{aligned}$$

Hence by Lengart's inequality, $\tilde{D}_{1n}(t; x) = o_P(1)$ and $\sqrt{a} \tilde{D}_n(t; x) = o_P(1)$, (t, x) -a.e. Using Stein's result (3.2) and consistency of $n^{-1}S^{(0)}$ and \hat{h} , for almost all (t, x) , we also have $\tilde{D}_n(t; x) - d(t; x) = \tilde{D}_{1n}(t; x) + \tilde{B}_n(t; x) + o(1)$ so that by (iii)(b), the condition $a^4 b^{-1} \rightarrow 0$ implies part (iv). \square

The next lemma will be used to prove weak convergence of the regression estimate of Proposition 2.1 and weak convergence of the finite-dimensional distributions of the doubly cumulative hazard function estimate.

LEMMA 3.3. *Let $h(t, x) = [h_1(t, x), \dots, h_d(t, x)]^T$ be a bounded continuous function on \mathcal{R}_δ . Set*

$$\begin{aligned} \tilde{W}^{(0)}(t, x) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_{[0, t]} \int_{[0, x]} h(u, w) \tilde{M}_i(du; w) dw, \\ \tilde{W}^{(1)}(t, x) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_{[0, t]} \int_{[0, x]} Z_i(u) \tilde{M}_i(du; w) dw. \end{aligned}$$

Then the finite-dimensional distributions of $\tilde{W} = (\tilde{W}^{(0)}, \tilde{W}^{(1)})$ are asymptotically normal with mean zero and covariance

$$\begin{aligned} &\text{cov}(W^{(0)}(t, x), W^{(0)}(t', x')) \\ &= \int_0^{t \wedge t'} \int_0^{x \wedge x'} h^{\otimes 2}(u, w) s^{(0)}(u, w, \beta_0) \alpha(u, w) du dw, \\ &\text{cov}(W^{(0)}(t, x), W^{(1)}(t', x')) \\ (3.9) \quad &= \int_0^{t \wedge t'} \int_0^{x \wedge x'} h(u, w) \otimes s^{(1)}(u, w, \beta_0) \alpha(u, w) du dw, \\ &\text{cov}(W^{(1)}(t, x), W^{(1)}(t', x')) \\ &= \int_0^{t \wedge t'} \int_0^{x \wedge x'} s^{(2)}(u, w, \beta_0) \alpha(u, w) du dw. \end{aligned}$$

PROOF. For $p = 0, 1$ we have

$$(3.10) \quad \tilde{W}^{(p)}(t, x) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^t H_{in}^{(p)}(u, x) M_i(du),$$

where

$$H_{in}^{(0)}(u, x) = \int_0^x h(u, w) K_a(w - X_i(u)) dw,$$

$$H_{in}^{(1)}(u, x) = Z_i(u) \int_0^x K_a(w - X_i(u)) dw.$$

For any fixed $x \in [0, 1]$, $H_{in}^{(p)}(\cdot, x)$ are predictable processes. Moreover, under condition I.2 \tilde{W} is a vector of square integrable martingales. Therefore asymptotic normality of its finite-dimensional distributions will follow from Rebolledo's central limit theorem.

We consider first the process $\tilde{W}^{(1)}$ and verify the Lindeberg condition of Rebolledo's central limit theorem. Let $Z_{i\ell}(u)$ be the ℓ th component of the covariate $Z_i(u)$. Correspondingly, let $H_{i\ell}(u, x)$ be the ℓ th component of $H_i^{(1)}(u, x)$; that is,

$$H_{i\ell}(u, x) = Z_{i\ell}(u) \int_0^x K_a(w - X_i(u)) dw.$$

We have $|H_{i\ell}(u, x)| \leq Z_{i\ell}(u)$ since

$$(3.11) \quad 0 \leq \int_0^x K_a(w - X_i(u)) dw \leq 1.$$

Therefore for any $\varepsilon > 0$,

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \int_0^t H_{i\ell}^2(u, x) I(|H_{i\ell}(u, x)| > \varepsilon) \langle M_i \rangle(du) \\ & \leq \frac{1}{n} \sum_{i=1}^n \int_0^t Z_{i\ell}^2(u) I(|Z_{i\ell}(u)| > \varepsilon) \langle M_i \rangle(du) \\ & = \frac{1}{n} \sum_{i=1}^n \int_0^t \xi_{2\ell i}(u, \beta_0) I(|Z_{i\ell}(u)| > \varepsilon) \alpha(u, X_i(u)) du. \end{aligned}$$

By the boundedness of the hazard function α , the moment condition I.2 and a similar argument as in Andersen and Gill (1982), the right-hand side tends in probability to 0.

Further, let $\widehat{V}(t; x_1, x_2)$ be the predictable covariation matrix of $\tilde{W}^{(1)}(\cdot, x_1)$ and $\tilde{W}^{(1)}(\cdot, x_2)$; that is,

$$\widehat{V}(t; x_1, x_2) = \frac{1}{n} \sum_{i=1}^n \int_0^t \xi_{2i}(u, \beta_0) \alpha(u, X_i(u)) \prod_{j=1}^2 \int_0^{x_j} K_a(w - X_i(u)) dw du.$$

We show that

$$(3.12) \quad \widehat{V}(t; x_1, x_2) \rightarrow_P \int_0^t \int_0^{x_1 \wedge x_2} \alpha(u, v) s^{(2)}(u, v, \beta_0) du dv.$$

If $\widehat{V}_{ll}(t; x, x)$ be the (l, l) entry of the matrix $\widehat{V}(t; x_1, x_2)$ then, by Chebyshev's inequality,

$$\text{Var } \widehat{V}_{ll}(t; x, x) \leq \frac{1}{n} E \int_0^t \xi_{2il}^2(u, \beta_0) \alpha^2(u, X_i(u)) \left[\int_0^x K_a(w - X_i(u)) dw \right]^4 du,$$

where $\xi_{2il}(u, \beta_0) = Y_i(u) Z_{i\ell}^2(u) \exp(\beta_0^T Z_i(u))$. Using boundedness of α , assumption I.2 and (3.11), $\text{Var } \widehat{V}_{ll}(t; x, x) = O(n^{-1})$. Moreover, by Fubini's theorem,

$$\begin{aligned} \widehat{V}(t; x_1, x_2) &= \frac{1}{n} \sum_{i=1}^n \int \int \left[\int_0^t \xi_{2i}(u, \beta_0) \alpha(u, X_i(u)) \right. \\ &\quad \left. \times I(-r_j < X_i(u) \leq x_j - r_j, j = 1, 2) \right] \\ &\quad \times \overline{K}_a(dr_1) \overline{K}_a(dr_2). \end{aligned}$$

Hence

$$\begin{aligned} E \widehat{V}(t; x_1, x_2) &= \frac{1}{n} \sum_{i=1}^n E \int \int \left[\int_0^t I(-r_j < X_i(u) \leq x_j - r_j, j = 1, 2) \right. \\ &\quad \left. \times E[\xi_{2i}(u, \beta_0) | X_i(u)] \alpha(u, X_i(u)) \right] \overline{K}_a(dr_1) \overline{K}_a(dr_2) \\ &= \int \int \left[\int_0^t \int I(-r_j < v \leq x_j - r_j, j = 1, 2) g(u, v) du dv \right] \\ &\quad \times \overline{K}_a(dr_1) \overline{K}_a(dr_2), \end{aligned}$$

where $g(u, v) = \alpha(u, v) s^{(2)}(u, v, \beta_0)$. Under the regularity conditions I.1, I.3 and I.4, the term in brackets is a matrix of bounded continuous functions in (r_1, r_2) . Therefore (3.1) implies

$$E \widehat{V}(t; x_1, x_2) \rightarrow \int_0^t \int_0^{x_1 \wedge x_2} g(u, v) du dv.$$

Since $\text{Var } \widehat{V}_{ll}(t; x, x) \rightarrow 0$, (3.12) holds as well.

Rebolledo's central limit theorem implies now that for any $0 < x_1 < \dots < x_q \leq 1$ the vector $\{\widetilde{W}^{(1)}(\cdot, x_j), j = 1, \dots, q\}$ converges weakly in $D([0, 1])^{p \times q}$ to a mean zero Gaussian process with covariance (3.1). Hence the finite-dimensional distributions of $\widetilde{W}^{(1)}(\cdot, \cdot)$ are asymptotically normal. A similar argument can also be used to verify asymptotic normality of the finite-dimensional distributions of $\widetilde{W}^{(0)}$. The joint asymptotic normality of

the finite dimensional distributions of $\tilde{W} = (\tilde{W}^{(0)}, \tilde{W}^{(1)})$ follows from the Cramér–Wald device. \square

We shall apply now the tightness criteria of Bickel and Wichura (1971) to verify that the process $\tilde{W}^{(0)}$ is C -tight. For this purpose we note that if $\{N_t: t \in [0, 1]\}$ is a counting process whose compensator $\{\Lambda_t: t \in [0, 1]\}$, with respect to a filtration $\{\mathcal{F}_t: t \in [0, 1]\}$, is absolutely continuous with respect to Lebesgue measure, $\Lambda_t = \int_0^t \lambda_u du$, then for any F_t -predictable process $\{H_t: t \in [0, 1]\}$, the martingale

$$R_t = \int_{[0,t]} HdM,$$

$M_t = N_t - \Lambda_t$, satisfies

$$(3.13) \quad ER_t^4 \leq CE \int_0^t H^4(u)[1 + \lambda(u)]\Lambda(du)$$

where C is some universal constant. This can be verified using integration by parts and the Burkholder–Davis–Gundy inequality [Dellacherie and Meyer (1978)]. A slightly different version of this bound is given in McKeague and Utikal (1990b) and McKeague, Nikabadze and Sun (1995).

LEMMA 3.4. *The process $\{\tilde{W}^{(0)}(t, x): (t, x) \in [0, 1]\}$ of Lemma 3.3 is C -tight.*

PROOF. By the tightness conditions of Bickel and Wichura (1971) it is enough to show

$$\limsup_n E[\tilde{W}_\ell^{(0)}((t_1, t_2] \times (x_1, x_2))]^4 = O(1)(t_2 - t_1)^2(x_2 - x_1)^2$$

for any $0 < t_1 < t_2 \leq 1, 0 < x_1 < x_2 \leq 1$. Here $\tilde{W}_\ell^{(0)}$ is the ℓ th component of the process $\tilde{W}^{(0)}$ and $\tilde{W}_\ell^{(0)}(J)$ is its increment on the rectangle J .

From (3.10), $\tilde{W}_\ell^{(0)}(J) = n^{-1/2} \sum_{i=1}^n R_{in}(t_2)$ where

$$R_{in}(t) = \int_{(t_1, t]} H_{in}(u)M_i(du),$$

$$H_{in}(u) = \int_{x_1}^{x_2} h_\ell(u, w)K_a(w - X_i(u)) dw.$$

Using the iid structure of the process $\tilde{W}^{(0)}$, we have

$$E\tilde{W}^{(0)}((t_1, t_2] \times (x_1, x_2))^4 = \frac{n-1}{n} [ER_{in}^2(t_2)]^2 + \frac{1}{n} [ER_{in}^4(t_2)].$$

By (3.13) it is sufficient to verify

$$\limsup_n E \int_{t_1}^{t_2} H_{in}^p(u) \langle M_i \rangle(du) = O(1)(t_2 - t_1)(x_2 - x_1).$$

for $p = 2, 4$ and

$$\limsup E \int_{t_1}^{t_2} H_{in}^4(u) \xi_{0i}(u, \beta_0)^2 \alpha^2(u, X_i(u)) du = O(1)(t_2 - t_1)(x_2 - x_1).$$

We shall show this last identity only. Boundedness of the functions α , h_ℓ and

$$0 \leq \int_{x_1}^{x_2} K_a(w - X_i(u)) dw \leq 1$$

imply that for some constants C_1 and C_2 we have

$$H_{in}^4(u) \leq C_1 \int_{x_1}^{x_2} K_a(w - X_i(u)) dw$$

and

$$\begin{aligned} E \int_{t_1}^{t_2} H_{in}^4(u) \xi_{0i}(u, \beta_0) \alpha^2(u, X_i(u)) du \\ \leq C_2 E \int_{t_1}^{t_2} \int_{x_1}^{x_2} \xi_{0i}^2(u, \beta_0) K_a(w - X_i(u)) dw du \\ = C_2 \int_{t_1}^{t_2} \int_{x_1}^{x_2} g(u, v) K_a(u - v) f(u, v) du dv \\ = C_2 \int \left[\int_{t_1}^{t_2} \int_{x_1-r}^{x_2-r} g(u, v) f(u, v) \right] \bar{K}_a(dr). \end{aligned}$$

Here $g(u, v)$ is a function on \mathcal{R}_δ such that $E \xi_{0i}^2(u, \beta_0) | X_i(u) = g(u, X_i(u))$ a.s. Under conditions I.3. and I.4 $f(u, v)$ and $g(u, v)$ are bounded so that the right-hand side is equal to $C_3(t_2 - t_1)(x_2 - x_1)$ for some constant C_3 . \square

Our final lemma gives “local” asymptotic normality results which we use in the proof of Proposition 2.3.

LEMMA 3.5. *Suppose that for almost all $x \in [0, 1]$, $\hat{h}(t, x) \rightarrow_P h(t, x)$ uniformly in $t \in [0, 1]$ where $h(\cdot, x)$ is the bounded function on $[0, 1]$. If $na \rightarrow \infty$ then for almost all $x \in [0, 1]$,*

$$\widehat{W}_h(t; x) = \sqrt{\frac{a}{n}} \sum_{i=1}^n \int_0^t \hat{h}(u, x) \widetilde{M}_i(du; x)$$

converges weakly in $D([0, 1])$ to a time-transformed Brownian motion $W_h(t; x)$ with variance function

$$\text{Var } W_h(t; x) = c(K) \int_0^t h^2(u, x) s^{(0)}(u, x, \beta_0) \alpha(u, x) du,$$

where $c(K) = \int K^2(r) dr$. Moreover, if $nab \rightarrow \infty$ then, for almost all $(t, x) \in [\delta, 1 - \delta] \times [0, 1]$,

$$\widetilde{W}_h(t; x) = \sqrt{b} \int_{t-b}^{t+b} \widetilde{K}_b(t - u) \widehat{W}_h(du; x)$$

is asymptotically a mean zero normal variable with variance $c(K)c(\tilde{K}) \times h^2(t, x)s^{(0)}(t, x, \beta_0)\alpha(t, x)$.

PROOF. We shall only verify the first part of the lemma; the proof of the second part is analogous. For any $\varepsilon > 0$, let

$$\widehat{W}_{h\varepsilon}(t; x) = \sqrt{\frac{a}{n}} \sum_{i=1}^n \int_0^t \widehat{h}(u, x) I \left[\sqrt{\frac{a}{n}} K_a(x - X_i(u)) |\widehat{h}(u, x)| > \varepsilon \right] \widetilde{M}_i(du; x).$$

The predictable variation of $\widehat{W}_{h\varepsilon}(\cdot; x)$ is

$$\begin{aligned} \langle \widehat{W}_{h\varepsilon} \rangle(t; x) &= \frac{a}{n} \sum_{i=1}^n \int_0^t \widehat{h}^2(u, x) \xi_{0i}(u, \beta_0) K_a^2(x - X_i(u)) \alpha(u, X_i(u)) \\ &\quad \times I[aK_a(x - X_i(u)) |\widehat{h}(u, x)| > \varepsilon \sqrt{na}] du. \end{aligned}$$

Now, $aK_a(x - X_i(u)) \leq C_1$ for some constant C_1 . Therefore, using boundedness of the hazard function α and $n^{-1}S^0$, for almost all x , we have

$$\begin{aligned} \langle \widehat{W}_{h\varepsilon} \rangle(t; x) &= O(1) \frac{1}{n} \int_0^t \widehat{h}^2(u, x) S^{(0)}(u, x, \beta_0) I[C_1 |\widehat{h}(u, x)| > \varepsilon \sqrt{na}] du \\ &= O_P(1) \int_0^t \widehat{h}^2(u, x) I[C_1 |\widehat{h}(u, x)| > \varepsilon \sqrt{na}] du \rightarrow_P 0 \end{aligned}$$

since $\widehat{h}(\cdot, x)$ is bounded in probability, x -a.e.

Further, the predictable variation of the process $\widehat{W}_h(\cdot; x)$ is given by $\langle \widehat{W}_h \rangle(t; x) = \langle \widehat{W}_h \rangle_1(t; x) + \langle \widehat{W}_h \rangle_2(t; x)$, where

$$\begin{aligned} \langle \widehat{W}_h \rangle_1(t; x) &= \frac{a}{n} \int_0^t [\widehat{h}^2(u, x) - h^2(u, x)] \sum_{i=1}^n \xi_{0i}(u, \beta_0) \\ &\quad \times K_a^2(x - X_i(u)) \alpha(u, X_i(u)) du, \\ \langle \widehat{W}_h \rangle_2(t; x) &= \frac{a}{n} \int_0^t h^2(u, x) \sum_{i=1}^n \xi_{0i}(u, \beta_0) K_a^2(x - X_i(u)) \alpha(u, X_i(u)) du. \end{aligned}$$

For almost all x , we have

$$\sup_{t \in [0, 1]} \langle \widehat{W}_h \rangle_1(t; x) \leq \sup_{u \in [0, 1]} |\widehat{h}^2(u, x) - h^2(u, x)| O(1) \sup_{t \in [0, 1]} \frac{1}{n} S^{(0)}(t, x, \beta_0) \rightarrow_P 0,$$

and by Stein's result (3.2),

$$\begin{aligned} E \langle \widehat{W}_h \rangle_2(t; x) &= aE \int_0^t h^2(u, x) E[\xi_{0i}(u, \beta_0) | X_i(u)] \\ &\quad \times \alpha(u, X_i(u)) K_a^2(x - X_i(u)) du \\ &= \int_{-1}^1 \int_0^t h^2(u, x) s^{(0)}(u, x - ar) \alpha(u, x - ar) K^2(r) du dr \\ &\rightarrow \text{Var } W_h(t; x). \end{aligned}$$

By Chebyshev's inequality,

$$\begin{aligned} \text{Var}(\widehat{W}_h)_2(t; x) &\leq \frac{a^2}{n} E \left[\int_0^t h^2(u, x) \xi_{0i}(u, \beta_0) K_a^2(x - X_i(u)) \alpha(u, X_i(u)) du \right]^2 \\ &\leq \frac{a^2}{n} E \int_0^t h^4(u, x) \xi_{0i}^2(u, \beta_0) K_a^4(x - X_i(u)) \alpha^2(u, X_i(u)) du. \end{aligned}$$

Let $g(u, v)$ be a bounded measurable function on \mathcal{A}_δ such that $E \xi_{0i}^2(u, \beta_0) | X_i(u) = g(u, X_i(u))$ a.s. Then

$$E \int_0^t \xi_{0i}^2(u, \beta_0) K_a(x - X_i(u)) du = \int_0^t \int g(u, v) f(u, v) K_a(x - v) du dv = O(1)$$

as $n \rightarrow \infty$. Using boundedness of α , $aK_a(x - \cdot)$, and h , for almost all x , we also have $\text{Var}(\widehat{W}_h)_2(t; x) = O((na)^{-1}) \rightarrow 0$ and hence $\langle \widehat{W}_h \rangle_2(t; x) \rightarrow_P \text{Var} W_h(t; x)$. Rebolledo's central limit theorem completes the proof. \square

3.2. *Proof of Propositions 2.1 and 2.2.* By the monotonicity of the sample paths of the process

$$(3.14) \quad \int_0^t I(0 \leq X_i(u) \leq x) N_i(du)$$

and uniform consistency of $n^{-1}S^{(0)}(u, w, \beta)$, the profile log-likelihood (2.4) satisfies

$$\begin{aligned} C_{\mathcal{A}}(\beta) - C_{\mathcal{A}}(\beta_0) &= \frac{1}{n} \sum_{i=1}^n \int_0^1 \left[(\beta - \beta_0)^T Z_i(u) - \log \frac{s^{(0)}(u, X_i(u), \beta)}{s^{(0)}(u, X_i(u), \beta_0)} \right] \\ &\quad \times I(0 < X_i(u) \leq 1) N_i(du) + o_P(1). \end{aligned}$$

By the weak law of large numbers, we have $C_{\mathcal{A}}(\beta) - C_{\mathcal{A}}(\beta_0) \rightarrow_P C_0(\beta)$, where

$$\begin{aligned} C_0(\beta) &= \int_0^1 \int_0^1 \left[(\beta - \beta_0)^T e(u, w, \beta_0) - \log \frac{s^{(0)}(u, w, \beta)}{s^{(0)}(u, w, \beta_0)} \right] \\ &\quad \times s^{(0)}(u, w, \beta_0) \alpha(u, w) du dw. \end{aligned}$$

For the smoothed profile log-likelihood (2.3), Lemma 3.1(ii) implies

$$\begin{aligned} &[\widetilde{C}_{\mathcal{A}}(\beta) - \widetilde{C}_{\mathcal{A}}(\beta_0)] \\ &= \frac{1}{n} \sum_{i=1}^n \int_0^1 \int_0^1 \left[(\beta - \beta_0)^T Z_i(u) - \log \frac{S^{(0)}(u, w, \beta)}{S^{(0)}(u, w, \beta_0)} \right] \widetilde{N}_i(du; w) dw \\ &\rightarrow_P C_0(\beta). \end{aligned}$$

Both $\widetilde{C}_{\mathcal{A}}(\beta)$ and $C_{\mathcal{A}}(\beta)$ are concave functions and the limit $C_0(\beta_0)$ is a concave function with maximum attained at $\beta = \beta_0$. Therefore Theorem III.1 in

Andersen and Gill (1982) implies uniqueness and consistency of the regression estimates.

We next show that the Hessians $\tilde{\mathcal{J}}(\beta) = -\nabla^2 \tilde{C}_{\mathcal{A}}(\beta)$ and $\mathcal{J}(\beta) = -\nabla^2 C_{\mathcal{A}}(\beta)$ satisfy

$$(3.15) \quad \tilde{\mathcal{J}}(\beta^*) \rightarrow_P \Sigma(\beta_0), \quad \mathcal{J}(\beta^*) \rightarrow_P \Sigma(\beta_0)$$

for any sequence $\beta^* \rightarrow_P \beta_0$. Define

$$V(u, w, \beta) = \frac{S^{(2)}(u, w, \beta)}{S^{(0)}(u, w, \beta)} - E^{\otimes 2}(u, w, \beta),$$

where

$$E(u, w, \beta) = \frac{S^{(1)}(u, w, \beta)}{S^{(0)}(u, w, \beta)}$$

and let

$$v(u, w, \beta) = \frac{s^{(2)}(u, w, \beta)}{s^{(0)}(u, w, \beta)} - \left[\frac{s^{(1)}(u, w, \beta)}{s^{(0)}(u, w, \beta)} \right]^{\otimes 2}.$$

Then

$$\begin{aligned} \tilde{I}(\beta^*) &= \int_0^1 \int_0^1 V(u, w, \beta^*) \frac{1}{n} \sum_{i=1}^n \tilde{N}_i(du; w) dw \\ &= \int_0^1 \int_0^1 [V(u, w, \beta^*) - v(u, w, \beta_0)] \frac{1}{n} \sum_{i=1}^n \tilde{N}_i(du; w) dw \\ &\quad + \int_0^1 \int_0^1 v(u, w, \beta_0) \frac{1}{n} \sum_{i=1}^n \tilde{N}_i(du; w) dw. \end{aligned}$$

Lemma 3.1(ii) applied to the deterministic function $h(u, w) = v(u, w, \beta_0)$ implies that the second-term converges in probability to $\Sigma(\beta_0)$. The first term is bounded by

$$\|V(u, w, \beta^*) - v(u, w, \beta_0)\|_{\infty} \frac{1}{n} \sum_{i=1}^n \int_0^1 \int_0^1 \tilde{N}_i(du; w) dw.$$

By the consistency of $n^{-1}S^{(p)}(u, w, \beta)$, we have $\|V(u, w, \beta^*) - v(u, w, \beta_0)\|_{\infty}$, whereas by Lemma 3.1(ii), $n^{-1}\sum_{i=1}^n \int_0^1 \int_0^1 \tilde{N}_i(du; w) dw = O_P(1)$. This completes the proof of (3.15) for the matrix $\tilde{I}(\beta^*)$. The matrix $I(\beta^*)$ can be handled analogously.

Further, the estimate $\tilde{\beta}$ of Proposition 2.1 is the solution of the score equation $\tilde{U}(\beta) = \nabla \tilde{C}(\beta) = 0$. We have $\sqrt{n}\tilde{U}(\beta_0) = \sqrt{n}[\tilde{\beta} - \beta_0] \cdot \tilde{\mathcal{J}}(\beta^*)$ where β^* lies on the line segment between $\tilde{\beta}$ and β_0 . Using (3.15) and the assumption I.4, it is enough to show that $\sqrt{n}\tilde{U}(\beta_0)$ is asymptotically mean zero normal with

covariance $\Sigma(\beta_0)$. Analogously, to show asymptotic normality of the estimate $\hat{\beta}$ of Proposition 2.2, it is enough to verify that the gradient $U(\beta) = \nabla C(\beta)$ satisfies $\sqrt{n}U(\beta_0) = \sqrt{n}\tilde{U}(\beta_0) + o_P(1)$.

We first deal with the score function $\tilde{U}(\beta)$,

$$\tilde{U}(\beta) = \frac{1}{n} \sum_{i=1}^n \int_0^1 \int_0^1 [Z_i(u) - E(u, w, \beta)] \tilde{N}_i(du; w) dw.$$

For $\beta = \beta_0$, the score process can be written as $\tilde{U}(\beta_0) = \sum_{j=1}^2 \tilde{U}_j(\beta_0)$ where

$$\begin{aligned} \tilde{U}_1(\beta_0) &= \frac{1}{n} \sum_{i=1}^n \int_0^1 \int_0^1 [Z_i(u) - E(u, w, \beta_0)] \tilde{M}_i(du; w) dw, \\ (3.16) \quad \tilde{U}_2(\beta_0) &= \frac{1}{n} \int_{[0,1] \times [0,1]} [Z_i(u) - E(u, w, \beta_0)] \\ &\quad \times K_a(w - X_i(u)) \xi_{0i}(u, \beta_0) \alpha(u, X_i(u)) du dw. \end{aligned}$$

By Lemmas 3.1 (iii) and 3.3,

$$\begin{aligned} \sqrt{n}\tilde{U}_1(\beta_0) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^1 \int_0^1 [Z_i(u) - e(u, w, \beta_0)] \tilde{M}_i(du; w) dw + o_P(1) \\ &\Rightarrow \mathcal{N}(0, \Sigma(\beta_0)). \end{aligned}$$

By adding and subtracting terms, we have

$$\begin{aligned} \tilde{U}_2(\beta_0) &= \frac{1}{n} \sum_{i=1}^n \int_0^1 \int_0^1 \xi_{1i}(u, \beta_0) K_a(w - X_i(u)) [\alpha(u, X_i(u)) - \alpha(u, w)] du dw \\ &\quad - \frac{1}{n} \sum_{i=1}^n \int_0^1 \int_0^1 \xi_{0i}(s, \beta_0) E(u, w, \beta_0) K_a(w - X_i(u)) \\ &\quad \times [\alpha(u, X_i(u)) - \alpha(u, w)] du dw. \end{aligned}$$

Since $E(\cdot, \cdot, \beta_0)$ is bounded in probability, Lemma 3.1(i) implies $n^{1/2}\tilde{U}_3(\beta_0) \rightarrow_P 0$.

Further, the score process $U(\beta)$ is given by

$$U(\beta) = \nabla C(\beta) = \frac{1}{n} \sum_{i=1}^n \int_0^1 [Z_i(u) - E(u, X_i(u), \beta)] I(0 < X_i(u) \leq 1) N_i(du).$$

For $\beta = \beta_0$, we have $U(\beta_0) = \sum_{j=1}^2 U_j(\beta_0)$, where

$$U_1(\beta_0) = \frac{1}{n} \sum_{i=1}^n \int_0^1 [Z_i(u) - E(s, X_i(u), \beta_0)] I(0 < X_i(u) \leq 1) M_i(du)$$

and $U_2(\beta_0) = \tilde{U}_2(\beta_0) + \sum_{j=1}^2 [U_{2j} - \tilde{U}_{2j}](\beta_0)$ with $\tilde{U}_2(\beta_0)$ given by (3.16) and

$$\begin{aligned} \sqrt{n}[U_{21} - \tilde{U}_{21}](\beta_0) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^1 \xi_{1i}(u, \beta_0) \\ &\quad \times \left[I(0 < X_i(u) \leq 1) - \int_0^1 K_a(x - X_i(u)) \right] \\ &\quad \times \alpha(u, X_i(u)) du, \end{aligned}$$

$$\sqrt{n}[U_{22} - \tilde{U}_{22}](\beta_0) = \int_0^1 \int_0^1 E(u, w, \beta_0) \alpha(u, w) \hat{\psi}(u, dw) du,$$

where

$$\hat{\psi}(u, w) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_{0i}(u, \beta_0) \left[I(0 < X_i(u) \leq w) - \int_0^w K_a(x - X_i(u)) dx \right].$$

Using (3.8),

$$\begin{aligned} \sqrt{n}U_1(\beta_0) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^1 [Z_i(u) - e(u, X_i(u), \beta_0)] \\ &\quad \times I(0 < X_i(u) \leq 1) M_i(du) + o_P(1), \end{aligned}$$

so that the standard central limit theorem for sums of iid random variables implies that $\sqrt{n}U_1(\beta_0)$ is asymptotically normal $\mathcal{N}(0, \Sigma(\beta_0))$. To complete the proof, it is enough to show that $\sqrt{n}[U_2(\beta_0) - \tilde{U}_2(\beta_0)]$ converges in probability to 0. First, let $\|\alpha\|_v(u, I)$ and $\|e\|_v(u, I, \beta_0)$ denote the variation of $\alpha(u, \cdot)$ and $e(u, \cdot, \beta_0)$ on an interval $I \subseteq [0, 1]$. Under the assumed regularity conditions, we have $\|\alpha\|_v(u, I) < \infty$ and $\|e\|_v(u, I, \beta_0) < \infty$ uniformly in $u \in [0, 1]$ and $I \subseteq [0, 1]$. Further, by the assumption of Proposition 2.2(ii), the process $E(u, \cdot, \beta_0)$ is of bounded variation uniformly in $u \in [0, 1]$. Since $\|\hat{\psi}\|_\infty \rightarrow 0$, integration by parts or application of Gill's (1989) Hadamard differentiability result for integrals $\int f dg$ implies $\sqrt{n}[U_{22} - \tilde{U}_{22}](\beta_0) \rightarrow_P 0$.

It remains to consider the term $\sqrt{n}[U_{21} - \tilde{U}_{21}](\beta_0)$. Its ℓ th component is a sum of iid random variables $U_{21\ell}(\beta_0) - \tilde{U}_{21\ell}(\beta_0) = n^{-1} \sum_{i=1}^n [U_{21i\ell}(\beta_0) - \tilde{U}_{21i\ell}(\beta_0)]$ and, using a two-term Taylor expansion,

$$\begin{aligned} E[U_{21i\ell} - \tilde{U}_{21i\ell}](\beta_0) &= \int_{-1}^1 \int_0^1 \int_0^1 \left[s_\ell^{(0)}(u, w, \beta_0) \alpha(u, w) \right. \\ &\quad \left. - s_\ell^{(0)}(u, w - ar, \beta_0) \alpha(u, w - ar) \right] du dw \bar{K}(dr) = O(\alpha^2). \end{aligned}$$

Moreover,

$$nE[U_{21\ell}(\beta_0) - \tilde{U}_{21\ell}(\beta_0)]^2 = E[U_{21i\ell}(\beta_0) - \tilde{U}_{21i\ell}(\beta_0)]^2 + (n-1)O(\alpha^4).$$

By the assumption $n\alpha^4 \rightarrow 0$, the second term tends to 0. By Chebyshev's inequality,

$$\begin{aligned} & E[U_{21i\ell}(\beta_0) - \tilde{U}_{21i\ell}(\beta_0)]^2 \\ &= E\left[\int_0^1 Z_{i\ell}(u)\xi_{0i}(u, \beta_0)\alpha(u, X_i(u))\right. \\ &\quad \times \left. \left[I(0 < X_i(u) \leq 1) - \int_0^1 K_a(w - X_i(u)) dw \right] du\right]^2 \\ &\leq E\int_0^1 \xi_{1i\ell}^2(u, \beta_0)\alpha^2(u, X_i(u)) \\ &\quad \times \left[I(0 < X_i(u) \leq 1) - \int_0^1 K_a(w - X_i(u)) dw \right]^2 du \\ &= B_i + \tilde{B}_i - 2D_i, \end{aligned}$$

where

$$\begin{aligned} B_i &= E\int_0^1 \xi_{1i\ell}^2(u, \beta_0)\alpha^2(u, X_i(u))I(0 < X_i(u) \leq 1) du, \\ \tilde{B}_i &= E\int_0^1 \xi_{1i\ell}^2(u, \beta_0)\alpha^2(u, X_i(u))\left[\int_0^1 K_a(w - X_i(u)) dw\right]^2 du, \\ D_i &= E\int_0^1 \xi_{1i\ell}^2(u, \beta_0)\alpha^2(u, X_i(u))I(0 < X_i(u) \leq 1) \\ &\quad \times \left[\int_0^1 K_a(w - X_i(u)) dw\right] du. \end{aligned}$$

Now, by assumption I.4, for some bounded function $g(u, w)$ on \mathcal{R}_δ , we have

$$\begin{aligned} B_i &= \int_0^1 \int_0^1 g(u, w)\alpha^2(u, w)f(u, w) du dw, \\ \tilde{B}_i &= \int_0^1 \int_0^1 g(u, v)\alpha^2(u, v)f(u, v)\left[\int_0^1 K_a(w - v) dw\right]^2 du dv \\ &= \int_{-1}^1 \int_{-1}^1 \left[\int_0^1 \int_{-r_1 \wedge r_2}^{1-r_1 \vee r_2} g(u, v)\alpha^2(u, v)f(u, v) du dv\right] \bar{K}_a(dr_1)\bar{K}_a(dr_2). \end{aligned}$$

The inner integral is a bounded continuous function of (r_1, r_2) so that by (3.1) we have $\tilde{B}_i \rightarrow B_i$. Similarly,

$$\begin{aligned} D_i &= \int_0^1 \int_0^1 g(u, v)\alpha^2(u, v)f(u, v)\left[\int_0^1 K_a(w - v) dw\right] du dv \\ &= \int_{-1}^1 \left[\int_0^1 \int_{-r \wedge 0}^{1-r \vee 0} g(u, v)\alpha^2(u, v)f(u, v) du dv\right] \bar{K}_a(dr) \rightarrow B_i. \end{aligned}$$

This implies $nE[U_{31\ell} - \tilde{U}_{31\ell}]^2(\beta_0) \rightarrow 0$ and $\sqrt{n}(U_{31\ell} - \tilde{U}_{31\ell})(\beta_0) \rightarrow_P 0$. \square

3.3. *Proof of Proposition 2.3.* Similarly to Andersen and Gill (1982), a one-term Taylor expansion implies that $\widehat{A}(t; w) - A(t; w) = \sum_{j=1}^4 I_j(t; w)$ where

$$\begin{aligned}
 I_1(t; w) &= -(\widehat{\beta} - \beta_0) \sum_{i=1}^n \int_0^t \frac{S^{(1)}(u, w, \beta^*) \widetilde{N}_i(du; w)}{S^{(0)}(u, w, \beta_0) S^{(0)}(u, w, \widehat{\beta})}, \\
 I_2(t; w) &= \sum_{i=1}^n \int_0^t \frac{\widetilde{M}_i(du; w)}{S^{(0)}(u, w, \beta_0)}, \\
 I_3(t; w) &= \sum_{i=1}^n \int_0^t \frac{\xi_{i0}(u, \beta_0)}{S^{(0)}(u, w, \beta_0)} [\alpha(u, X_i(u)) - \alpha(u, w)] K_a(w - X_i(u)) du, \\
 I_4(t; w) &= \int_0^t I(S^{(0)}(u, w, \beta_0) = 0) \alpha(u, w) du.
 \end{aligned}$$

Here $\widehat{\beta}^*$ lies on the line segment between $\widehat{\beta}$ and β_0 . For the doubly cumulative hazards, we have $\widehat{A}_d(t, x) - A_d(t, x) = \sum_{j=1}^4 I_{dj}(t, x)$, where

$$I_{dj}(t, x) = \int_0^x I_j(t; w) dw.$$

Finally, setting

$$\widetilde{\alpha}(t, x) = \int_{t-b}^{t+b} \widetilde{K}_b(t-u) A(du; x) du = \int_{t-b}^{t+b} \widetilde{K}_b(t-u) \alpha(u, x) du,$$

a two-term Taylor expansion yields

$$\widetilde{\alpha}(t, x) - \alpha(t, x) = \frac{b^2}{2} \frac{\partial^2}{\partial t^2} \alpha(t, x) \int_0^1 \widetilde{K}(r) r^2 dr + O(b^2).$$

Therefore the condition $nab^5 \rightarrow 0$ implies $\sqrt{nab}[\widetilde{\alpha} - \alpha](t, x) \rightarrow 0$. Moreover, $[\widehat{\alpha} - \widetilde{\alpha}](t, x) = \sum_{j=1}^4 \widetilde{I}_j(t; x)$, where

$$\widetilde{I}_j(t; x) = \int_{t-b}^{t+b} \widetilde{K}_b(t-u) I_j(du; x).$$

Let c_0 be the lower bound of $f(u, v)F_1(u, v)$ on \mathcal{R}_δ , where recall that $F_1(u, X_i(u)) = EY_i(u)|X_i(u)$ a.s. By assumption I.1, we have $c_0 > 0$ and

$$\begin{aligned}
 \sqrt{na} E \sup_{t \in [0, 1]} I_4(t; w) &\leq \sqrt{na} \int_0^1 [1 - P(Y_i(u) = 1, |X_i(u) - w| \leq a)]^n \alpha(u, w) du \\
 &= O(1) \sqrt{na} \exp(-c_0 na) \rightarrow 0.
 \end{aligned}$$

Similarly

$$\begin{aligned}
 \sqrt{nab} E \widetilde{I}_4(t; x) &\leq \sqrt{nab} \int_{t-b}^{t+b} \widetilde{K}_b(t-u) \exp(-nac_0) \alpha(u, x) du \\
 &= O(1) \sqrt{b} (\sqrt{na} \exp(-nac_0)) \rightarrow 0
 \end{aligned}$$

and if $na^4 \rightarrow 0$, then $E\|I_{d4}\| = O(1) \sqrt{n} \exp(-c_0 na) \rightarrow 0$.

Further, by the uniform consistency of $n^{-1}S^{(p)}(u, w, \beta_0)$ and Lemma 3.2, for almost all x , we have $\sqrt{na}|I_3(t; w)| = o_P(1)$ and $\sqrt{na}I_1(t; x) = \sqrt{n}(\hat{\beta} - \beta_0)o_P(1)$ uniformly in $t \in [0, 1]$. Since $\sqrt{n}(\hat{\beta} - \beta_0)$ is bounded in probability, the second term is asymptotically negligible. Similarly, for almost all (t, x) , $\sqrt{ nab}|\tilde{I}_3(t; w)| = o_P(1)$ and $\sqrt{ nab}\tilde{I}_1(t; x) = \sqrt{n}(\hat{\beta} - \beta_0)o_P(1) = o_P(1)$ and parts (i) and (iii) of Proposition 2.3 follow from Lemma 3.5.

Using uniform consistency of $n^{-1}S^{(q)}$, $q = 0, 1$, Lemma 3.1 yields $\sqrt{n}I_{3d}(t, x) = o_P(1)$ and $\sqrt{n}I_{1d}(t, x) = \sqrt{n}[\hat{\beta} - \beta_0]^T \eta(t, x) + o_P(1) = \sqrt{n}\tilde{U}(\beta_0)^T \Sigma(\beta_0)^{-1} \eta(t, x) + o_P(1)$ uniformly in $(t, x) \in [0, 1]^2$. By Lemmas 3.3 and 3.4 and (3.17), $\sqrt{n}\tilde{U}(\beta_0)$ and the process

$$\tilde{W}^{(0)}(t, x) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^t \int_0^x \frac{\tilde{M}_i(du; w) dw}{s^{(0)}(u, w, \beta_0)}$$

are asymptotically independent. Moreover, $\tilde{W}^{(0)}(t, x)$ converges weakly in $D([0, 1]^2)$ to a two-parameter time-transformed Brownian motion with variance function

$$\text{Var } W^{(0)}(t, x) = \int_0^t \int_0^x \frac{\alpha(u, w)}{s^{(0)}(u, w, \beta_0)} du dw.$$

The proof of part (ii) of Proposition 2.3 can be completed using Young's integration by parts formula [Young (1917), Hildebrandt (1963)] and Hadamard differentiability results for weak net integrals $\int f(u, w) dg(u, w)$ in Gill, van der Laan and Wellner (1995). \square

REFERENCES

- ANDERSEN, P. K., BORGAN, O., GILL, R. D. and KEIDING, N. (1993). *Statistical Models Based on Counting Processes*. Springer, New York.
- ANDERSEN, P. K. and GILL, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **10** 1100–1120.
- BERAN, R. (1981). Nonparametric regression with randomly censored survival data. Technical report, Univ. California, Berkeley.
- BICKEL, P. J. and WICHURA, M. J. (1971). Convergence criteria for multiparameter stochastic processes with applications. *Ann. Math. Statist.* **42** 1656–1670.
- BORGAN, Ø. (1984). Maximum likelihood estimation in parametric counting process models, with applications to censored failure time data. *Scand. J. Statist.* **11** 1–16.
- DABROWSKA, D. M. (1989). Uniform consistency of the kernel conditional Kaplan–Meier estimate. *Ann. Statist.* **17** 1157–1167.
- DELLACHERIE, C. and MEYER, P. A. (1978). *Probabilities and Potential*. North-Holland, Amsterdam.
- FIX, E. and NEYMAN, J. (1951). A simple stochastic model for recovery, relapse, death and loss of patients. *Human Biology* **23** 205–241.
- GILL, R. D. (1989). Non- and semiparametric maximum likelihood estimators and the van Mises method, I. *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.
- GILL, R. D., VAN DER LAAN, M. and WELLNER, J. A. (1995). Inefficient estimators of the bivariate survival function in three models. *Ann. Inst. H. Poincaré* **31** 545–597.
- HILDEBRANDT, T. H. (1963). *Introduction to the Theory of Integration*. Academic Press, New York.

- JACOD, J. (1975). Multivariate point processes: predictable projection, Radon–Nikodym derivatives, representation of martingales. *Z. Wahrsch. Verw. Gebiete* **31** 235–254.
- KEIDING, N. (1990). Statistical inference in the Lexis diagram. *Philos. Trans. Roy. Soc. London Ser. A* **332** 487–509.
- MCKEAGUE, I. W., NIKABADZE, A. M. and SUN, Y. (1995). An omnibus test for independence of a survival time from a covariate. *Ann. Statist.* **23** 450–475.
- MCKEAGUE, I. W. and UTIKAL, K. J. (1990a). Inference for a nonlinear counting process regression model. *Ann. Statist.* **18** 1172–1187.
- MCKEAGUE, I. W. and UTIKAL, K. J. (1990b). Identifying nonlinear covariate effects in semi-martingale regression models. *Probab. Theory Related Fields* **87** 1–25.
- 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.
- NIELSEN, J. P. and LINTON, O. B. (1995). Kernel estimation in a nonparametric marker dependent hazard model. *Ann. Statist.* **23** 1735–1749.
- O’SULLIVAN, F. (1993). Nonparametric estimation in the Cox model. *Ann. Statist.* **21** 124–145.
- RAMLAU-HANSEN, H. (1983). Smoothing counting process intensities by means of kernel functions. *Ann. Statist.* **11** 453–466.
- SASIENI, P. (1992). Information bounds for the conditional hazard ratio in a nested family of regression models. *J. Roy. Statist. Soc. Ser. B* **54** 617–635.
- STEIN, E. M. (1970). *Singular Integrals and Differentiability Properties of Functions*. Princeton Univ. Press.
- THERNEAU, T. M., GRAMBSCH, P. M. and FLEMING, T. R. (1990). Martingale-based residuals for survival models. *Biometrika* **77** 147–160.
- THOMAS, D. C. (1983). Nonparametric estimation and tests for fit for dose response relations. *Biometrics* **39** 263–268.
- YOUNG, W. H. (1917). On multiple integration by parts and the second theorem of the mean. *Proc. London Math. Soc. (2)* **16** 273–293.

DEPARTMENT OF BIostatISTICS
UNIVERSITY OF CALIFORNIA
LOS ANGELES, CALIFORNIA 90095-1772
E-MAIL: dorota@mozart.ph.ucla.edu