Statistica Sinica **16**(2006), 93-119

ESTIMATION IN A SEMIPARAMETRIC MODULATED RENEWAL PROCESS

Dorota M. Dabrowska and Wai Tung Ho

University of California, Los Angeles and SPSS Inc.

Abstract: We consider parameter estimation in a regression model corresponding to an i.i.d. sequence of censored observations of a finite state modulated renewal process. The model assumes a similar form as in Cox regression except that the baseline intensities are functions of the backwards recurrence time of the process and a time dependent covariate. As a result of this it falls outside the class of multiplicative intensity counting process models. We use kernel estimation to construct estimates of the regression coefficients and baseline cumulative hazards. We give conditions for consistency and asymptotic normality of estimates. Data from a bone marrow transplant study are used to illustrate the results.

Key words and phrases: Modulated renewal process, kernel estimation, U processes.

1. Introduction

In medical and engineering applications it is common to consider a Markov renewal process to model the lengths of time spent in consecutive stages of a disease or lifetime of a piece of equipment. Denoting by $\mathbb{J} = \{1, \ldots, k\}$ the set of possible states, the process is described by a sequence of random variables $(T, J) = (T_m, J_m)_{m \geq 0}$, such that $T_0 < T_1 < T_2 < \cdots$ are consecutive times of entrances into states $J_0, \ldots, J_m \in \mathbb{J}$. Under assumption of the Markov renewal process, the sequence $J = \{J_m : m \geq 0\}$ of states visited forms a Markov chain and given J, the sojourn times T_1, $T_2 - T_1, \ldots$ are independent with distributions depending on the adjoining states. Associated with the sequence (T, J) is a counting process $\{\widetilde{N}_{ij}(t) : t \geq 0, i, j \in \mathbb{J}\}$ whose components register each direct $i \rightarrow j$ transition,

$$\widetilde{N}_{ij}(t) = \sum_{m \geq 0} 1(T_{m+1} \leq t, J_{m+1} = i, J_m = j) \ .$$

Its compensator $\{\Lambda_{ij}(t) : t \geq 0, i, j \in \mathbb{J}\}$ relative to the self-exciting filtration is given by

$$\Lambda_{ij}(t) = \int_0^t 1(J(s-) = i) dA_{ij}(L(s)) \ ,$$

where $J(t), t \geq 0$, is the state occupied at time t, $L(t) = t - T_{\widetilde{N}(t-)}$, $\widetilde{N}(t) = \sum_{ij \in E} \widetilde{N}_{ij}(t)$ is the backwards recurrence time, and $[A_{ij}(x)]_{i,j \in \mathbb{J}}$ is a matrix of unknown deterministic functions representing cumulative hazards of one-step transitions. Nonparametric estimation of this matrix and the associated semi-Markov kernel of the process was considered by Lagakos, Sommer and Zelen (1978), Gill (1980), Voelkel and Crowley (1984) and Phelan (1990), among others.

In this paper we consider estimation in a modulated renewal process, assuming that components of the counting process $\{\widetilde{N}_{ij} : (i,j) \in \mathbb{J}\}$ have intensities of the form

$$\Lambda_{ij}(t) = \int_0^t 1(J(s-) = i) e^{\beta^T Z_{ij}(s)} \alpha_{ij}(L(s), X(s)) ds \ , \tag{1.1}$$

where $X(s)$ is a time dependent covariate, $Z = \{Z_{ij}(t) : t \geq 0, i, j \in \mathbb{J}\}$ is a vector of external transition specific covariates, and $[\alpha_{ij}]$ is a matrix of two-parameter baseline hazards. A model of this kind may arise for instance in medical applications where survival status of a patient is characterized by an illness process with baseline intensities dependent on the length of time spent in each stage of a disease and a covariate $X(s)$, possibly changing with time. In the absence of this covariate, the model reduces to the modulated renewal process proposed by Cox (1973) with cumulative intensities

$$\Lambda_{ij}(t) = \int_0^t 1(J(s-) = i) e^{\beta^T Z_{ij}(s)} \alpha_{i,j}(L(s)) ds \ . \tag{1.2}$$

Both models have several interesting features. The first one is that the event times can be viewed as recorded on two simultaneously evolving time scales. In the case of (1.2), the covariates depend on the calendar time t, whereas the matrix α of baseline hazards depends on the duration scale. In the case of the (1.1), the latter matrix depends both on the duration and calendar time scale. Further, if α corresponds to a matrix of functions depending only on a Euclidean parameter θ, then estimation of the pair (β, θ) based on an i.i.d. sample of modulated renewal processes can be carried out using a counting process framework for analysis of maximum likelihood or M estimates. However, if the matrix α is completely unspecified, then its nonparametric maximum likelihood estimate falls outside the class of statistics taking the form of stochastic integrals with respect to counting processes (Gill (1980)). Similarly, in the case of (1.2), estimation of the regression coefficient β can be in principle based on the solution to the score equation

$$\Phi_n(\beta) = \sum_{\ell=1}^n \sum_{i,j \in \mathbb{J}} \int [Z_{ij\ell}(t) - \frac{S^{(1)}}{S^{(0)}}(t, \beta)] \widetilde{N}_{ij\ell}(dt) = 0 \ , \tag{1.3}$$

where $S^{(p)}(t, \beta) = \sum_{\ell=1}^n 1(J_\ell(t-) = i) Z_{ij\ell}^p(t) e^{\beta Z_{ij\ell}(t)}, p = 0, 1$. However, as a result of the dependence of the compensators on the backwards recurrence time, the score function in (1.3), evaluated at the true parameter value β_0, fails to satisfy the identity $\mathbb{E}\Phi_n(\beta_0) = o_P(1)$, and consequently the estimate of the regression coefficient obtained by solving the equation $\Phi_n(\beta) = 0$ cannot be consistent. Several authors considered also the special case of the one-jump process (1.1) and showed that estimation of regression coefficients requires smoothing (Sasieni (1992), Dabrowska (1997), Nielsen, Linton and Bickel (1998) and Pons and Vissier (2000)).

To circumvent difficulties arising in the analyses of renewal processes, Gill (1980) and Oakes and Cui (1994) proposed the use of a random time-change approach which replaces the calendar time scale t by the duration scale. Here we consider an extension of this approach to analyse a simple case of (1.1), assuming that the covariate $X(s)$ is constant between the jumps of the process $\widetilde{N}(t) = \sum_{ij} \widetilde{N}_{ij}(t)$, and $\{Z_{ij}(t) : i, j \in \mathbb{J}, t \geq 0\}$ is a vector of external covariates. In Section 3 we discuss kernel estimation in single-type models. In Section 4 we give examples multi-type models with a "small" state space to which the results can also be applied. We use data from a bone marrow transplant study to illustrate the results.

2. The Model

Throughout the paper we assume that $(\Omega, \mathcal{F}, P)$ is a complete probability space and $(T_m, V_m)_{m\geq 0}$ is a marked point process defined on it, with marks taking on values in a measurable space $(E, \mathrm{E}\,)$ and enlarged by the empty mark Δ. Thus $T_0 < T_1 < \cdots T_m < \cdots$ is a sequence of random time points registering occurrence of some events in time, and such that T_m are almost surely distinct and $T_m \uparrow \infty$ P-a.s. At time T_m we observe a variable V_m such that $V_m \in E$ if $T_m < \infty$, and $V_m = \Delta$ if $T_m = \infty$.

For any $B \in \mathrm{E}\,$, let $\widetilde{N}(t, B) = \sum_{m\geq 0} 1(T_{m+1} \leq t, V_{m+1} \in B)$ be the process counting observations falling into the set $[0, t] \times B$. The internal history of the process, $\{\mathcal{F}^N_t\}_{t\geq 0}$, represents information collected on N until time t, and is given by

$$\mathcal{F}^N_t = \sigma(1(T_m \leq s, V_m \in B) : m \geq 0, s \leq t, B \in \mathrm{E}\,)\ .$$

Then $\{\mathcal{F}^N_t\}_{t\geq 0}$ forms an increasing family of right-continuous σ-fields. Let $\mathcal{F}_t = \mathcal{F}_0 \vee \mathcal{F}^N_t$ be the self-exciting filtration associated with the process $\widetilde{N}$, obtained by adjoining to the internal history of the process, the P-null sets. The compensator of the process $\widetilde{N}(t, B)$, with respect to $\mathcal{F}_t$ is given by

$$\Lambda(t, B) = \Lambda(T_m, B) + \int_{(T_m,t]} \frac{P_m(d(s, v))}{P_m([s, \infty); E \cup \Delta)} \quad \text{for} \quad t \in (T_m, T_{m+1}]\ ,$$

where $P_m(d(s, v))$ is a version of a regular conditional distribution of (T_{m+1}, V_{m+1}) given $\mathcal{F}_{T_m}$ (Jacod (1975)).

In this paper we assume that the marks V_m have the form $V_m=(J_m, X_m, \widetilde{Z}_m)$, where $J_m \in \mathbb{J}$ is the state visited at time T_m and $(\widetilde{Z}_m, X_m)$ are covariates taking on value in $E_1 = R^d \times [0,\tau], \tau < \infty$. The pair $(\widetilde{Z}_m, X_m)$ may represent some measurements taken upon entrance into the state J_m. For any Borel set B of E_1, let $\mu_{m+1}(B,t,j) = Pr((\widetilde{Z}_{m+1}, X_{m+1}) \in B | T_{m+1} = t,\ J_{m+1} = j,\ (T_l, J_l, \widetilde{Z}_l, X_l)_{l=0}^m)$ and suppose that

$$\begin{aligned}&Pr(T_{m+1} - T_m \le s, J_{m+1} = j | (T_\ell, J_\ell, \widetilde{Z}_\ell, X_\ell)_{\ell=0}^m)\\ &= 1(J_m = i) \int_{[0,s]} \exp[-\sum_\ell \int_0^u e^{\beta^T Z_{i\ell m}(v)} \alpha_{i\ell}(v, X_m) dv] e^{\beta^T Z_{ijm}(u)} \alpha_{ij}(u, X_m) du\ ,\end{aligned}$$

where $Z_{ijm}(u) = f_m(u, T_l, J_l, \widetilde{Z}_l, X_l : l = 0, \ldots, m)$ is a fixed deterministic function f_m, left continuous in u. The process $\widetilde{N}_{ij}(t,B) = \sum_{m\ge 0} 1(T_{m+1} \le t,$ $J_{m+1} = j,\ J_m = i,\ (\widetilde{Z}_{m+1}, X_{m+1}) \in B)$ has compensator given by

$$\begin{aligned}\Lambda_{ij}(t,B) &= \Lambda_{ij}(T_m, B)\\ &\quad + \int_{(T_m,t]} \mu_{m+1}(B,u,j) 1(J_m = i) e^{\beta^T Z_{ijm}(u-T_m)} \alpha_{ij}(u - T_m, X_m) du\ .\end{aligned}$$

In particular, setting $B = E_1$ and using $\mu_{m+1}(E_1, T_{m+1}, j) 1(T_{m+1} < \infty) = 1$,

$$\Lambda_{ij}(t) = \Lambda_{ij}(t,E) = \Lambda_{ij}(T_m) + \int_{(T_m,t]} 1(J_m = i) e^{\beta^T Z_{ijm}(u-T_m)} \alpha_{ij}(u - T_m, X_m) du$$

is the compensator of the counting process $\widetilde{N}_{ij}(t) = \widetilde{N}_{ij}(t,E) = \sum_{m\ge 0} 1(T_{m+1} \le t, J_{m+1} = j, J_m = i)$, registering transitions among the adjacent states of the model.

In the following we assume the random censorship model of Gill (1980). Thus the times at which the process is observed is determined a process $C(s) = \sum_{m\ge 1} 1(C_{m-1} < t \le C_m)$, where $0 = C_0 \le C_1 \le \cdots \le C_m \cdots$ is an increasing sequence such that $C_m \in [T_m, T_{m+1}]$ are stopping times with respect to the history $\{\mathcal{F}_t\}_{t\ge 0}$, and $(C_m)_{m\ge 0}$ is conditionally independent of $\{(T_m, J_m, \widetilde{Z}_m, X_m)\}_{m\ge 0}$ given $(J_0, \widetilde{Z}_0, X_0)$. If $T_m = C_m$, then no information is available on either the sojourn time $T_{m+1} - T_m$, the states (J_m, J_{m+1}) or the covariates $(\widetilde{Z}_m, X_m)$, $(\widetilde{Z}_{m+1}, X_{m+1})$. If $C_m = T_{m+1}$, then the sojourn time $T_{m+1} - T_m$, the adjoining states (J_m, J_{m+1}) and the covariates $(\widetilde{Z}_m, X_m), (\widetilde{Z}_{m+1}, X_{m+1})$ are observable. Finally, if $T_m < C_m < T_{m+1}$, then the state J_m and the covariates $(\widetilde{Z}_m, X_m)$ are visible while the sojourn time $T_{m+1} - T_m$ is only known to exceed $C_m - T_m$. We also assume that the censoring process is monotone in the sense that $T_m \le C_m < T_{m+1} \Rightarrow C_{m'} = T_{m'}$ for all $m' \ge m$. This condition stipulates that the process terminates once censoring takes place. To construct estimates of the unknown parameters, we use a time transformation which replaces the chronological (or calendar) time scale by the duration scale (Gill (1980) and Oakes and Cui (1994)). For $m \geq 0$, let

$$\begin{aligned} N_{ijm}(v) &= 1(T_{m+1} - T_m \leq v, J_m = i, J_{m+1} = j, T_m = C_{m+1})\ , \\ Y_{im}(v) &= 1(T_{m+1} - T_m \geq v, C_m - T_m \geq v, J_m = i)\ , \\ M_{ijm}(v) &= N_{ijm}(v) - \int_0^v Y_{im}(u) e^{\beta^T Z_{ijm}(u)} \alpha(u, X_m) du\ . \end{aligned}$$

Lemma 2.1. *Suppose that $\{\varphi_m(v), m \geq 0, v \geq 0\}$ is a sequence of left-continuous random functions such that the process $\varphi \circ L(t) = \sum_{m \geq 0} \varphi_m(t - T_m) 1(T_m < t \leq T_{m+1})$ is predictable with respect to the filtration $\{\mathcal{F}_t\}_{t \geq 0}$ and $\mathbb{E} \int_0^\infty [\varphi \circ L]^2(s) \Lambda_{ij}(ds) < \infty$. Then*

$$\begin{aligned} \mathbb{E} \sum_m \int_0^\infty \varphi_m(u) N_{ijm}(du) &= \mathbb{E} \sum_m \int_0^\infty Y_m(u) \varphi_m(u) e^{\beta_0^T Z_{mij}(u)} \alpha_{ij}(u, X_m) du, \\ \mathbb{E}[\sum_m \int_0^\infty \varphi_m(u) M_{ijm}(du)]^2 &= \mathbb{E} \sum_m \int_0^\infty Y_m(u) \varphi_m^2(u) e^{\beta_0^T Z_{mij}(u)} \alpha_{ij}(u, X_m) du. \end{aligned}$$

In addition, if $\{\varphi_{1m} : m \geq 0\}$ and $\{\varphi_{2m} : m \geq 0\}$ are two such sequences, then

$$\mathbb{E}[\sum_m \int \varphi_{1m}(u) M_{ijm}(du)][\sum_m \int \varphi_{2m}(u) M_{klm}(du)] = 0$$

for pairs $(i, j) \neq (k, \ell)$.

Much in the same way as in Gill (1980), this lemma follows from the Dominated Convergence Theorem, martingale properties of the processes $\widetilde{M}_{ij}$, and

$$\begin{aligned} \int_0^\infty [\varphi \circ L](s) C(s) \widetilde{N}_{ij}(s) &= \sum_{m \geq 0} \int_0^\infty \varphi_m(u) N_m(du) \\ \int_0^\infty [\varphi \circ L]^k(s) C(s) \Lambda_{ij}(ds) &= \sum_{m \geq 0} \int_0^\infty \varphi_m(u)^k Y_m(u) e^{\beta_0^T Z_{ijm}(u)} \alpha_{ij}(u, X_m) du\ . \end{aligned}$$

The identies hold almost surely for $k = 1, 2$. We omit the details.

3. Estimation in Single-Type Event Processes

In this section we assume that all events are of a single type. To estimate the baseline cumulative hazard function, we use conditional Aalen-Nelson estimator (Beran (1981))

$$\widehat{A}(v; x, \beta) = \frac{1}{na} \int_0^v \frac{N_i(du, x)}{S_{-i}^{(0)}(u, \beta, x)}\ ,$$

where $S^{(0)}_{-i}(u,\beta,x) = 1/[(n-1)a]\sum_{j\neq i} S^{(0)}_j(u,\beta,x)$ and for each $i = 1,\ldots,n$,

$$N_i(u,x) = \sum_m K_n(x,X_{im})N_{im}(u),$$
$$S^{(0)}_i(u,\beta,x) = \sum_m Y_{im}(u)e^{\beta^T Z_{im}(u)}K_n(x,X_{im})\ .$$

Here $K_n(x,w)$ is the boundary kernel of Müller and Wang (1994), given by

$$K_n(x,w) = \begin{cases} 1(x-a\le w\le x+a)K_{11}(\frac{x-w}{a}) & \text{if } a<x<\tau-a \\ 1(w\le (1+q)a)K_{1q}(q-\frac{w}{a}) & \text{if } 0<x\le a \\ 1(\tau-a\le w\le (1+a)p)K_{p1}(\frac{\tau-w}{a}-p) & \text{if } \tau-a\le x\le\tau, \end{cases}$$

where

$$\begin{aligned} &K_{11}(r) = 2C(\mu)(\frac{1}{2})^{2\mu+2}(1+r)^\mu(1-r)^\mu && \text{(central region)}\\ &K_{pq}(r) && \text{(left boundary region)}\\ &= C(\mu)(\frac{1}{p+q})^{2\mu+2}(p+r)^\mu(q-r)^{\mu-1}[2r((p-q)\mu-q)+\mu(p-q)^2+2q^2] \\ &K_{pq}(r) && \text{(right boundary region)}\\ &= C(\mu)(\frac{1}{p+q})^{2\mu+2}(p+r)^{\mu-1}(q-r)^\mu[2r((p-q)\mu+p)+\mu(p-q)^2+2p^2]\ , \end{aligned}$$

$p,q\in(0,1)$, and $C(\mu) = 2(2\mu+1)\binom{2\mu-1}{\mu}$. The kernels K_{pq} are Jacobi polynomials, and for $(p,q) = (1,1),(1,q)$ and $(p,1)$, we have

$$\int_{-p}^{q} K_{pq}(u)du = 1,\ \int_{-p}^{q} uK_{pq}(u)du = 0,\ \int_{-p}^{q} u^2K_{pq}(u)du < \infty\ .$$

Table 3.1 gives the form of these kernels for polynomials of degree 2, 4 and 6.

Table 3.1. Polynomial kernels of degree 2, 4 and 6.

$\mu=1$	interior	$(3/4)(1-x^2)$
	left	$6(p+x)(p+q)^{-4}[p^2-2pq+3q^2+2x(p-q)]$
	right	$6(q-x)(p+q)^{-4}[3p^2-2pq+q^2+2x(2p-q)]$
$\mu=2$	interior	$(15/16)(1-x^2)^2$
	left	$60(q-x)(p+x)^2(p+q)^{-6}[p^2-2pq+2q^2+(2p-3q)x]$
	right	$60(q-x)^2(p+x)(p+q)^{-6}[2p^2-2pq+q^2+(3p-2q)x]$
$\mu=3$	interior	$(35/32)(1-x^2)^3$
	left	$140(q-x)^2(p+x)^3[3p^2-6pq+5q^2+2(3p-4q)x]$
	right	$140(q-x)^3(p+x)^2[5p^2-6pq+3q^2+2(4p-3q)x]$

In the following we assume that $(u, x) \in \mathcal{R} = [0, \tau_0] \times [0, \tau], \tau_0 < \infty, \tau < \infty$. To control the bias of the risk process and the Aalen-Nelson estimator, we need the following regularity conditions.

Condition A

(i) The variables X_{im} have densities $f_m(x)$ with respect to Lebesgue measure on $[0, \tau]$.

(ii) There exists a bounded open neighbourhood $\mathcal{B}$ of the true parameter value β_0 such that $\mathbb{E}\sum_m [Z_{im}(u)]^{\otimes k} Y_{im}(u) \exp[\beta^T Z_{im}(u)] < \infty$, for $k = 0, 1, 2$.

(iii) For $k = 0, 1, 2$, and $\beta \in \mathcal{B}$, the functions

$$s^{(k)}(u, \beta, w) = \sum_m \mathbb{E}([Z_{im}(u)]^{\otimes p} Y_{im}(u) \exp[\beta^T Z_{im}(u)] | X_m = w) f_m(w)$$

are uniformly bounded and twice differentiable with respect to β. In addition $\nabla s^{(0)}(u, \beta, w) = s^{(1)}(u, \beta, w), \nabla^2 s^{(0)}(u, \beta, w) = s^{(2)}(u, \beta, w)$, and the functions $s^{(k)}(u, \beta, w), k = 0, 1, 2$ are uniformly Lipshitz continuous in β.

(iv) The function $\alpha(u, w), (u, w) \in \mathcal{R}$ is bounded.

(v.1) The functions $s(u, \beta, w) = s^{(k)}(u, \beta, w), k = 0, 1, 2$, and $\alpha(u, w)$ satisfy $\sup\{|\alpha(u, w_1) - \alpha(u, w_2)| : (u, w_j) \in \mathcal{R},\ |w_1 - w_2| \leq a,\ j = 1, 2\} = O(a)$ and $\sup\{|s(u, \beta, w_1) - s(u, \beta, w_2)| : (u, w_j) \in \mathcal{R},\ |w_1 - w_2| \leq a,\ \beta \in \mathcal{B}, j = 1, 2\} = O(a)$.

(v.2) $s(u, \beta, w)$ and $\alpha(u, w)$ are twice differentiable with respect to w with a uniformly bounded second derivatives $s''(u, \beta, w)$, $\alpha''(u, w)$ such that $\sup\{|\alpha''(u, w_1) - \alpha''(u, w_2)| : (u, w_j) \in \mathcal{R},\ |w_1 - w_2| \leq a,\ j = 1, 2\} = O(1)$ and $\sup\{|s''(u, \beta, w_1) - s''(u, \beta, w_2)| : (u, w_j) \in \mathcal{R}, |w_1 - w_2| \leq a,\ \beta \in \mathcal{B}, j = 1, 2\} = O(1)$.

We refer to this condition as A.1 or A.2, depending on whether the assumption (v.1) or (v.2) is in force. For $k = 1, 2$, let $S_{-i}^{(k)}(u, \beta, x) = \nabla^k S_{-i}^{(0)}(u, \beta, x)$ be the vector and matrix of first and second derivatives of the risk process $S_{-i}^{(0)}$ with respect to β. Set $\overline{s}^{(k)}(u, \beta, x) = a^{-1}\mathbb{E}S_i^{(k)}(u, \beta, x)$, $\overline{n}(u, x) = \mathbb{E}N_i(u, x)$ and

$$\overline{A}(v; x, \beta_0) = \int_0^v \frac{\overline{n}(du; x)}{\overline{s}^{(0)}(u, \beta_0, x)} \,.$$

Proposition 3.2. *Under assumptions A we have* $\overline{s}^{(k)}(u, \beta, x) - s^{(k)}(u, \beta, w) = O(a^r)$ *for* $k = 0, 1, 2$, *uniformly in* $(u, x) \in \mathcal{R}$ *and* $\beta \in \mathcal{B}$, *and* $\overline{A}(v; x, \beta_0) - A_0(v; x) = O(a^r)$ *uniformly in* $(v, x) \in \mathcal{R}$. *Here* $r = 1$ *under condition* A.1 *and* $r = 2$ *under condition* A.2.

Proof. Dropping the superscript k, in the central region we have

$$\frac{1}{a}\mathbb{E}S_i(u,\beta,x)=a^{-1}\int_{x-a}^{x+a}K_{11}(\frac{x-w}{a})s(u,\beta,w)dw=\int_{-1}^{1}K_{11}(r)s(u,\beta,x-ra)dr.$$

In the left and right boundary regions, the expectation $a^{-1}\mathbb{E}S_i(u,\beta,x)$ is

$$a^{-1}\int_{x-qa}^{x+a}K_{1q}(\frac{x-w}{a})s(u,\beta,w)dw=\int_{-1}^{q}K_{1q}(r)s(u,\beta,x-ra)dr,$$
$$a^{-1}\int_{x-a}^{x-pa}K_{p1}(\frac{x-w}{a})s(u,\beta,w)dw=\int_{-p}^{1}K_{p1}(r)s(u,\beta,x-ra)dr.$$

In the left boundary region, $q=x/a$ and in the right-boundary region, $p=(\tau-x)/a$. Under condition (v.1), we have $|a^{-1}\mathbb{E}S_i(u,\beta,x)-s(u,\beta,x)|=O(a)$, uniformly in $(u,x)\in\mathcal{R}$ and $\beta\in\mathcal{B}$. Under condition (v.2), we have

$$a^{-1}\mathbb{E}S_i(u,\beta,x)-s(u,\beta,x)=\frac{a^2}{2}s''(u,\beta,x)\int_{-p}^{q}r^2K_{pq}(r)dr+O(a^2)\ .$$

Similarly $\overline{n}(u,x)=\int_0^u\int_{-p}^q s^{(0)}(v,\beta_0,x-ra)\alpha(v,w-ra)K_{pq}(r)drdv$. Therefore, if one of the two functions (s or α) is Lipschitz of order 1, then $\overline{n}(u,x)-\int_0^u s(v,\beta_0,x)\alpha(v,x)dv=O(a)$, whereas if both functions are twice differentiable in x, then the bias is

$$\frac{a^2}{2}\int_0^u\left\{\frac{\partial^2}{\partial x^2}[s^{(0)}(v,\beta_0,x))\alpha(v,x)]\right\}dv\int_{-p}^{q}r^2K_{pq}(r)dr+O(a^2).$$

We also have $\overline{A}(v;x,\beta_0)-A(v;x)=\int_0^v\gamma(u,x)A(du;x)$, where $\gamma(u,x)=[\overline{n}(du,x)/\overline{s}^{(0)}(u,\beta_0,x)\alpha(u,w)]-1$. Thus the bias is of order $O(a^r), r=1,2$.

We turn now to estimation of the regression coefficients. The first method corresponds to an M-estimator obtained by solving the score equation $\widetilde{\Phi}_n(\beta)=0$, where

$$\widetilde{\Phi}_n(\beta)=\frac{1}{n}\sum_{i=1}^{n}\sum_{m}\int_0^{\tau_0}[Z_{im}(u)S_{-i}^{(0)}(u,\beta,X_{im})-S_{-i}^{(1)}(u,\beta,X_{im})]N_{im}(du)\ .$$

The analysis of this score equation requires only smoothness conditions A.1 and second moment bounds on the risk processes. For the sake of convenience, these moment bounds are given in the appendix. Let

$$V(u,\beta,x)=[\frac{s^{(2)}}{s^{(0)}}-(\frac{s^{(1)}}{s^{(0)}})^{\otimes 2}](u,\beta,x).$$

Proposition 3.3. *Suppose that the conditions* A.1 *and* D.2 (i)−(ii) *hold. Let* $\Sigma_1(\beta_0) = \int_{\mathcal{R}}(V[s^{(0)}]^2)(u,\beta_0,x)\alpha(u,x)dudx$ *and* $\Sigma_2(\beta_0) = \int_{\mathcal{R}}(V[s^{(0)}]^3)(u,\beta_0,x)$ $\alpha(u,x)dudx$. *Suppose that* $\Sigma_1(\beta_0)$ *is a non-singular matrix, that* $na^2 \downarrow 0$ *and* $na \uparrow \infty$. *With probability tending to* 1, *the score equation* $\widetilde{\Phi}(\beta) = 0$ *has a unique root* $\widetilde{\beta}$ *and* $\sqrt{n}(\widetilde{\beta} - \beta_0)$ *converges in distribution to a mean zero normal variable with covariance* $\Sigma_1^{-1}(\beta_0)\Sigma_2(\beta_0))[\Sigma_1^{-1}(\beta_0)]^T$.

The proof is given in Appendix D. The next Proposition deals with asymptotic normality of the Aalen-Nelson estimator. We need the following consistency assumption on the risk function.

Condition B. Suppose that $\inf\{s^{(0)}(u,\beta,w) : u \leq \tau_0, \beta \in \mathcal{B}, w \in [0 \vee x - a_n, x + a_n \wedge \tau]\} > 0$. Moreover, that under assumption A.$r, r = 1, 2$, we have

$$\max_i \mathbb{E} \sup_{\beta\in\mathcal{B}, u\leq\tau_0} |\frac{S_{-i}^{(0)} - \overline{s}^{(0)}}{s^{(0)}}|(u,\beta,x) \to 0$$

for a bandwidth sequence $a = a_n \downarrow 0$ such that $na \uparrow \infty$ and $na^{2r+1} \downarrow 0$.

Proposition 3.4. *Suppose that conditions A.r*$(r = 1, 2)$, *B and D.*1 *are satisfied. For any root-n consistent estimate* $\widehat{\beta}$ *of the parameter* β_0, *the process* $[\sqrt{na}[A(v;x,\widehat{\beta}) - A(v;x)], v \leq \tau_0]$ *converges weakly in* $\ell^\infty([0,\tau_0])$ *to a mean zero Gaussian process* $G(v,x)$ *with covariance*

$$\mathrm{Cov}\,[G(v,x), G(v',x)] = d_{p(x),q(x)}(K) \int_{[0,v\wedge v']} \frac{A(du,x)}{s^{(0)}(u,\beta_0,x)} \;.$$

Here $r = 1$ *under condition A.*1 *and* $r = 2$ *under assumptions of condition A.*2. *Moreover,* $d_{pq}(K) = \int_{-p}^{q} K_{pq}^2(w)dw$ *and* $p(x) = q(x) = 1$ *if* $a < x < \tau - a$, $p = 1, q(x) = a^{-1}x$ *if* $0 < x < a$ *and* $p(x) = a^{-1}(\tau - x), q(x) = 1$ *if* $\tau - a < x < \tau$.

Finally, we consider a partial score likelihood estimate of the regression coefficient. It is obtained by solving the the score equation $\Phi_n(\beta) = 0$, where

$$\Phi_n(\beta) = \frac{1}{n}\sum_m \sum_{i=1}^n \int_0^{\tau_0} [Z_{im}(u) - \frac{S_{-i}^{(1)}}{S_{-i}^{(0)}}(u,\beta,X_{im})]N_{im}(du) \;.$$

Note that this score function is similar to that arising in the standard Cox regression, except that we use leave-one-out risk processes. The choice of risk processes $S^{(k)} = \sum_{j=1}^n S_j^{(k)}, k = 1, 2$, is also possible. In both cases the resulting score functions form an approximate V process of degree 4 and the difference between them converges in probability to 0, but only under stronger moment conditions than those considered in Appendix D.

To analyze the score function $\Phi_n(\beta)$, we require condition A.2, moment conditions, and the following uniform consistency assumption.

Condition C. Suppose that $\inf\{s^{(0)}(u,\beta,x) : (u,x) \in \mathcal{R}, \beta \in \mathcal{B}\} > 0$. Moreover, that

$$\max_i \mathbb{E} \sup_{(u,x)\in\mathcal{R},\beta\in\mathcal{B}} \left|\frac{S_{-i}^{(0)} - \overline{s}^{(0)}}{s^{(0)}}\right|(u,\beta,x)| \to 0$$

for a bandwidth sequence $a_n \downarrow 0, na_n^2 \uparrow \infty, na_n^4 \downarrow 0$.

Proposition 3.5. *Suppose that conditions A.2, C, D.2 are satisfied and the matrix* $\Sigma(\beta_0) = \int_{\mathcal{R}} (Vs^{(0)})(u,\beta_0,x)\alpha(u,x)dudx$ *is non-singular. With probability tending to* 1*, the score equation* $\Phi_n(\beta) = 0$ *has a unique root* $\widehat{\beta}$*, and* $\sqrt{n}(\widehat{\beta} - \beta_0)$ *converges in distribution to a mean zero normal variable with covariance* $\Sigma^{-1}(\beta_0)$*.*

The proofs of these propositions are given in Appendices B$-$D. Similar to the approach of Pons and Visser (2000), we use U-process theory. Whereas in their setting asymptotic normality results for the estimate $\widehat{\beta}$ were obtained based on analysis of U-statistics of degree 2, in our case the term R_{1n} of their Proposition 3 satisfies only $R_{1n} = \sqrt{n}O_p(1)\sup_{\beta,(u,x)} |S^{(0)} - \overline{s}^{(0)}|(u,\beta,x)$. (Here $S^{(0)} = (na)^{-1}\sum S_i^{(0)}$.) In the case of one jump processes with bounded time independent covariates, say, results of Einmahl and Mason (2000) imply that the supremum is of order $O(\sqrt{\log a^{-1}/na})$ a.s., so that the term R_{1n} diverges to infinity. In the following we therefore use expansions of higher order.

Except for moment bounds, the proofs of these propositions do not use any special properties of the Z process, and we do not require uniform consistency of the derivatives $S_{-i}^{(k)}$, $k = 1, 2$. On the other hand, assumptions B and C require a more detailed specification of the covariate Z in order to apply inequalities from empirical process theory. The following proposition gives one set of conditions under which these assumptions hold. We consider the assumption C only. Let $\mathcal{R}_{1n} = \{(u,x) \in \mathcal{R} : a \leq x \leq \tau - a\}$, $\mathcal{R}_{2n} = \{(u,x) \in \mathcal{R} : 0 < x \leq a\}$ and $\mathcal{R}_{3n} = \{(u,x) \in \mathcal{R} : \tau - a < x \leq \tau\}$. Let $\mathcal{H}_{pn} = \{h(u,\beta,x) : (u,x) \in \mathcal{R}_{pn}, \beta \in \mathcal{B}\}$, $p = 1, 2, 3$, where $h(u,\beta,x) = s^{-1}(u,\beta,x)\sum_m Y_m(u)e^{\beta^T Z_m(u)}K_n(x,X_m)$. Note that for large n

$$\max_i \mathbb{E} \sup_{\substack{(u,x)\in\mathcal{R}_{pn}\\ \beta\in\mathcal{B}}} \left|\frac{S_{-i}^{(0)} - \overline{s}^{(0)}}{s^{(0)}}\right|(u,\beta,x)$$

is of the same order as $\mu_{pn} = \mathbb{E}\sup\{|h - \mathbb{E}h|(u,\beta,x) : (u,x) \in \mathcal{R}_{pn}, \beta \in \mathcal{B}\}$.

Proposition 3.6. *Suppose that for some* $r > 2$ *the bandwidth sequence satisfies* $a_n \downarrow 0$*,* $na_n \uparrow \infty$ $b_n = \log a_n^{-1}/(na_n) \downarrow 0$*,* $a_n^{-1}b_n^{r/2-1} = O(1)$ *and there exists* *a random variable H_{1n}, such that* (1) $\mathbb{E}H_{1n}^r = O(1)$; (2) $\|h(u,\beta,x)\|_{L_2(P)} \leq \sqrt{a}_n\|H_{1n}\|_{L_2(P)}$ *and* (3) $N_{[]}(\varepsilon\|H_{1n}\|_{L_2(P)}, \mathcal{H}_{1n}, \|\cdot\|_{L_2(P)}) \leq [A\varepsilon^{-1}]^V$ *for some finite constants A and V not depending on n and $\varepsilon \in (0,1)$. Then $\mu_{1n} = O(\sqrt{b}_n)$. If in addition there exist random variables H_{pn}, $p = 2,3$, such that* (4) $\mathbb{E}H_{pn}^2 = O(a)$ *and* (5) $N_{[]}(\varepsilon\|H_{pn}\|_{L_2(P)}, \mathcal{H}_{pn}, \|\cdot\|_{L_2(P)}) \leq [A_p\varepsilon^{-1}]^{V_p}$ *for some finite constants A_p and V_p not depending on n and $\varepsilon \in (0,1)$, then in the boundary regions we have $\mu_{pn} = O((na_n)^{-1/2})$, $p = 2,3$.*

Here $\|\cdot\|_{L_2(P)}$ is the $L_2(P)$ norm, and $N_{[]}(\eta, \mathcal{H}_{pn}, \|\cdot\|_{L_2(P)})$ is the minimal number of brackets of $L_2(P)$-size η covering the class $\mathcal{H}_{pn}$.

Proof. By Theorem 2.14.2 in van der Vaart and Wellner (1996, p.240), in the central region we have

$$\mu_{1n} \leq \frac{1}{a_n\sqrt{n}} J_{[]}(\sqrt{a}_n, \mathcal{H}_{1n}, \|\cdot\|_{L_2(P)}) + a_n^{-1}\mathbb{E}H_{1n}1(H_{1n} \geq \sqrt{n}c(\sqrt{a}_n)), \tag{3.1}$$

where $J_{[]}(\delta, \mathcal{H}, \|\cdot\|_{L_2(P)}) = \int_0^\delta [1+\log N_{[]}(\varepsilon\|H\|_{L_2(P)}, \mathcal{H}, \|\cdot\|_{L_2(P)})]^{1/2} d\varepsilon$ and $c(\delta) = \delta\|H\|_{L_2(P)}/[1+\log N_{[]}(\delta\|H\|_{L_2(P)}, \mathcal{H}, \|\cdot\|_{L_2(P)})]^{-1/2}$. For $\delta = \sqrt{a_n}$ the first term of (3.1) is of order $O(\sqrt{b}_n)$. Since $c(\sqrt{a_n}) = O(\sqrt{a_n/\log a_n^{-1}})$, the second term is bounded by $a_n^{-1}(\sqrt{n}c(\sqrt{a}_n))^{1-r}\mathbb{E}H_{1n}^r = O(\sqrt{b}_n)O(a_n^{-1}b_n^{r/2-1}) = O(\sqrt{b}_n)$. The same theorem in van der Vaart and Wellner (1996) implies that in the boundary regions we have $\mu_{pn} = n^{-1/2}a_n^{-1}O(J_{[]}(1, \mathcal{H}_{np}, \|\cdot\|_{L_2(P)})\|H\|_{L_2(P)} = O((na_n)^{-1/2})$, $p = 2,3$.

Using a somewhat tedious argument, it is not difficult to show that conditions of this proposition are satisfied in the case of covariates not dependent on u. Under added envelope conditions, the proposition is also satisfied by Lipshitz continuous covariates, covariates that form functions of bounded variation, etc.

4. Multi-Type Event Processes

The results of the previous section extend to the multistate setting provided the state space of the process is "small". An example is provided by an illness-death process in which a person in "healthy" state (0) can either progress to a "death" state (2), or can first develop a reversible disease (state 1) and subsequently die. In the absence of censoring, the cumulative transitions rates are given by

$$\Lambda_{ij}(t) = \Lambda_{ij}(T_m) + 1(J_m = i)\int_{(T_m,t]} e^{\beta^T Z_{ijm}(s-T_m)}\alpha_{ij}(s-T_m, X_m)ds$$

for $t \in (T_m, T_{m+1}]$. Similarly to multi-type processes in Andersen et al. (1993), estimation of regression coefficients can be based on the score function

$$\Phi_n(\beta) = \frac{1}{n} \sum_{i=1}^{n} \sum_{h} \sum_{m} \int [Z_{ihm}(u) - \frac{S^{(1)}_{-ih}}{S^{(0)}_{-ih}}(u, \beta, X_{im})] N_{ihm}(du) \ ,$$

where the sum extends over pairs $h = (0,1), (0,2), (1,2), (2,1)$ of possible one-step transitions,

$$S^{(0)}_{-ih}(u, \beta, x) = \frac{1}{a_h} \sum_{j \neq i} Y_{jhm}(u) e^{\beta^T Z_{jhm}(u)} K_n(x, X_{im}) \ ,$$

and $S^{(1)}_{ih}$ is the derivative of this process with respect to β. Note that the bandwidth sequence $a_h = a_{nh}$ is taken here to depend on the transition type h. The orthogonality relations of Lemma 2.1 imply that the score function is asymptotically normal with covariance matrix $\sum_h \Sigma_h(\beta)$, where matrices Σ_h assume a similar form as in Proposition 3.5. The M-estimator of Proposition 3.3 provides an alternative estimate.

Another example of a multi-type process is provided by progressive multi-state models. In this case a subject may move among a finite number of transient states, but each such state can be visited at most once. As an example of such a model we consider data on 3020 bone marrow transplant (BMT) recipients for acute myelogeneous leukemia (AML) and acute lymphoblastic leukemia (ALL). The data were collected by the International Bone Marrow Transplant Registry (IBMTR) during the period 1991-2000. Only first transplants in remission are considered and all patients received transplant from an HLA-identical sibling. Transplant recipients first received high doses of chemotherapy and radiation to destroy malignant cells in bone marrow and elsewhere. To rescue them from the toxicity of this therapy, they subsequently received bone marrow cells from a suitably matched donor.

In the following we donote by TX the transplant state. It can be followed by a number of complications, among them graft–versus–host disease (GVHD), relapse and death in remission. Two forms of GVHD are usually distinguished. Acute GVHD (AGVHD) occurs in the first 2–3 months following transplant, whereas chronic GVHD (CGVHD) occurs later in time. We use time independent covariates corresponding to X= square root of patient's age at transplant, and binary covariates represting donor–recipient sex–match, (Z), disease type and GVHD prophylaxis treatment. The square-root transformation of age serves to reduce skewness of the data. Removal of T–cells from the donor's bone marrow and posttransplant administration of immune supressive drugs are the major GVHD prophylactic treatments.

We are interested in the dependence of the intensities of one-step transitions on age. In Figures 4.1-4.3 we show plots of the baseline cumulative hazards $A_{ij}(v|x)$ as functions of x . Note that for fixed x, $A_{ij}(v|x)$ is an increasing function of v, but for fixed v this function may assume a variety of forms. Figure 4.1 shows that cumulative hazards of transitions TX $\rightarrow$ AGVHD, TX $\rightarrow$ CGVHD and AGVHD $\rightarrow$ CGVHD are increasing functions of age, and this monotonicity pattern is most pronounced in the case of transitions into the CGVHD state. The cumulative hazards of transitions TX $\rightarrow$ death and CGVHD $\rightarrow$ death are both U-shaped functions, suggesting higher incidence of death among older and very young patients. Finally, the graphs of cumulative hazards of transitions into the relapse state are decreasing functions of age, though nearly constant in age in the upper tail. Note that in the case of transitions originating from the TX state, all 3020 subjects enter into the risk process. However, transitions originating from the GVHD states use only those subjects who progress to the AGVHD and/or CGVHD state. In particular, a total of 560 patients progressed into the CGVHD state. Subsequently 100 developed relapse and 170 died in remission. Thus transitions from the CGVHD state are heavily censored. The relatively small number of relapses accounts for the noisy graphs of the cumulative hazards of the CGVHD $\rightarrow$ relapse state.

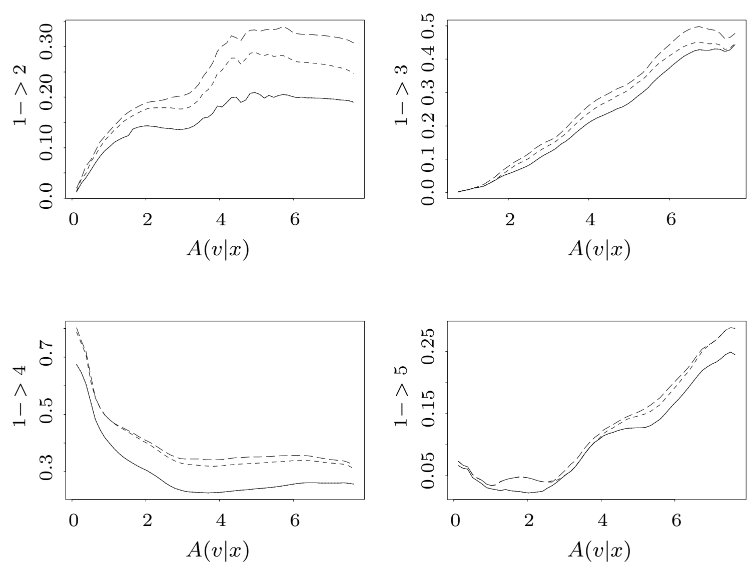

Figure 4.1 Baseline cumulative hazards of transitions originating from the transplant state versus age. The labels of states are 1 − transplant (TX), 2 − AGVHD, 3 − CGVHD, 4 − relapse and 5 − death.

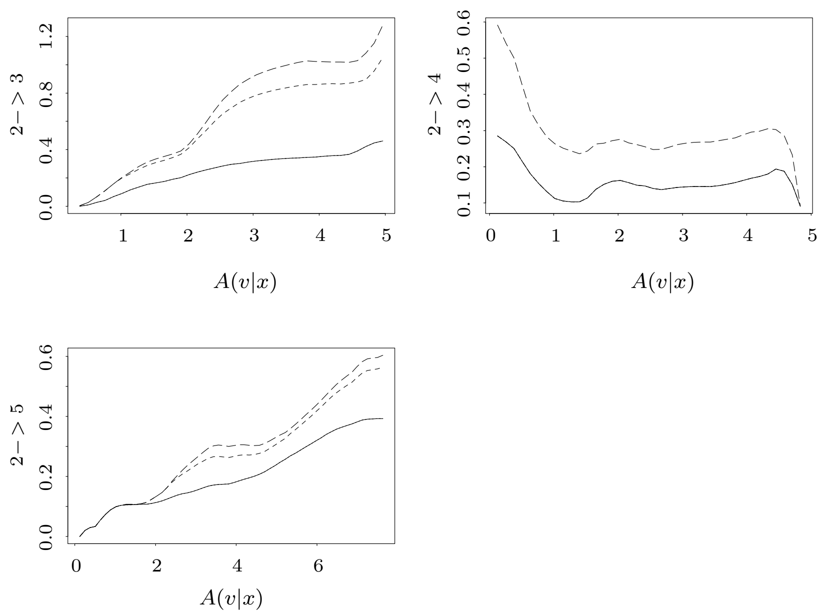

Figure 4.2 Baseline cumulative hazards of transitions originating from the AGVHD state versus age. The labels of states are 2 − AGVHD, 3 − CGVHD, 4 − relapse and 5 − death.

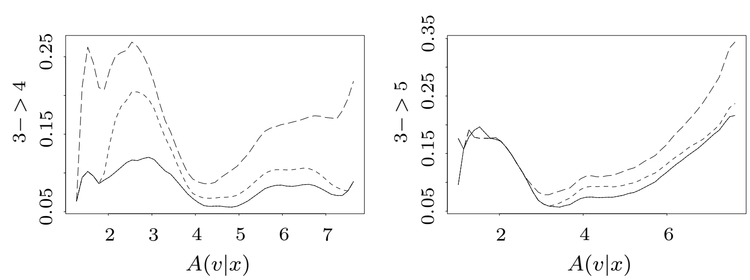

Figure 4.3 Baseline cumulative hazards of transitions originating from the CGVHD state versus age. The labels of states are 3 − CGVHD, 4 − relapse and 5 − death.

The regression coefficients for the model are reported in Table 4.1. As in any multistate analysis based on the proportional hazard model, the regression coefficients do not have a clear meaning. For example, male recipients receiving transplant from a female donor are at higher risk for progression from the transplant state into the AGVHD and CGVHD state, but are also at lower risk for direct (one-step) transition from the transplant into the relapse state. The overall effect of this covariate on the occurrence of death in remission or relapse cannot be, however, directly assessed based on regression coefficients because patients who develop AGVHD are at higher risk for death in remission, and also female-to-male transplant increases the risk of CGVHD to relapse transition. Likewise, the direction of the regression coefficients corresponding to each of the GVHD prophylactic treatments varies from one transition to another. Examples of parameters which can be used to summarize effects of covariates on the occurrence of endpoint events were discussed in Klein, Keiding and Copelan (1993), Arjas and Eerola (1993) and Dabrowska, Sun and Horowitz (1994). Their extension to the present setting is beyond the scope of this paper.

Table 4.1. Regression estimates and standard errors of direct transitions.

	TX → AGVHD	TX → CGVHD	AGVHD → CGVHD
sex–match	0.08 (0.05)	0.12 (0.05)	
CSA	0.46 (0.08)	0.18 (0.12)	
Trem	-0.58 (0.13)	-0.42 (0.15)	-0.28 (0.24)
MTX	0.38 (0.20)	-0.40. (0.31)	
disease	0.12 (0.07)		-0.13 (0.13)

	TX → relapse	AGVHD → relapse	CGVHD → relapse
sex–match	-0.11 (0.10)		0.21 (0.10)
CSA		-0.52 (0.31)	
Trem	0.20 (0.12)	-0.75 (0.59)	0.40 (0.31)
MTX	0.32 (0.23)	0.67 (0.56)	
disease	0.14 (0.10)		

	TX → death	AGVHD → death	CGVHD → death
Trem	0.23 (0.15)	0.57 (0.20)	0.48 (0.25)
CSA	-0.25 (0.18)		
MTX	-1.06 (0.58)		-0.82 (0.71)
disease		0.21 (0.13)	
prior AGVHD			0.75 (0.16)

The covariates are binary 0-1 variables: Sex–match = 1 if the donor is a female and the recipient is a male; Disease = 1 if the disease type is ALL; Prior AGVHD = 1 if AGVHD occurs prior to CGVHD. The GVHD prophylactic treatments are labeled as cyclosporin (CSA =1), T cell removal (Trem = 1) and methotraxate (MTX=1).

Appenidx A. Preliminaries

Let $W_1, \ldots, W_n$ be i.i.d. random variables with some distribution P. An (asymmetric) U statistics of degree $m, m \geq 1$ is denoted by

$$\mathbb{U}_{n,m}(h) = \frac{(n-m)!}{n!} \sum_{(i_1,\ldots,i_m)\in I_n^m} h(W_{i_1}, \ldots, W_{i_m}),$$

where I_n^m is the collection of vectors $(i_1, \ldots, i_m)$ with distinct coordinates, each in $\{1, \ldots, n\}$. Assuming that the kernel h satisfies $\mathbb{E}|h(W_1, \ldots, W_m)| < \infty$, the Hoeffding projection of degree m of the kernel h is denoted by $\pi_m h(W_1, \ldots, W_m)$. We have $\pi_m h(W_1, \ldots, W_m) = \sum_{A \subset \{1,\ldots,m\}} (-1)^{m-|A|} \mathbb{E}_A h(W_1, \ldots, W_m)$, where for $\emptyset \neq A = \{i_1, \ldots, i_p\}, 1 \leq p \leq m$, $\mathbb{E}_A$ denotes conditional expectation with respect to variables $\{W_j, j \in A\}$ and $\mathbb{E}_\emptyset h(W_1, \ldots, W_m) = \mathbb{E}h(W_1, \ldots, W_m)$. Then $\mathbb{U}_{n,m}(\pi_m h)$ forms a canonical U statistics of degree m. For canonical U-processes indexed by classes of kernels changing with n, Lemma 3.5.2, Remarks 3.5.4 and inequality (5.4.3) in de la Peña and Gine (1999) provide the following.

Lemma A.7. *Let* $\{\mathbb{U}_{n,m}(h) : h \in \mathcal{H}_n\}$ *be a canonical U–process over a measurable class class* $\mathcal{H}_n$ *of (asymmetric) kernels of degree* m*. If* $\mathcal{H}_n$ *forms a Euclidean class of functions for a square integrable envelope* H_n*, then* $\mathbb{E}n^{m/2} \|\mathbb{U}_{n,m}(h)\|_{\mathcal{H}_n} = O(\mathbb{E}[H_n(W_1, \ldots, W_m)^2]^{1/2})$*.*

A measurable class of functions $\mathcal{H}$ defined on some measure space $(\Omega, \mathcal{A})$ is Euclidean for envelope H is $h \leq H$ for all $h \in \mathcal{H}$, and there exist constants A and V such that $N(\varepsilon \|H\|_{L_2(P)}, \mathcal{H}, \|\cdot\|_{L_2(P)}) \leq (A/\varepsilon)^V$ for all $\varepsilon \in (0, 1)$ and all probability measures P such that $\|H\|_{L_2(P)} < \infty$ (Nolan and Pollard (1987)). Here $\|\cdot\|_{L_2(P)}$ is the $L_2(P)$ norm and $N(\eta, \mathcal{H}, \|\cdot\|_{L_2(P)})$ is the minimal number of $L_2(P)$–bals of radius η covering the class $\mathcal{H}$. In the case of classes $\mathcal{H}_n$ changing with n, the Euclidean constants A and V are taken to be independent of n.

In the following we use U processes of degree $m \leq 1, 2, 3, 4$. Finally, in our case for each subject i, the sequence W_i represents the total number of events observed in the interval $[0, \tau_0]$, their times of the occurrence, types and covariates observed at each jump time. The Euclidean property of the classes of functions appearing in the remainder of the text can be easily verified based on results of Nolan and Pollard (1987), Pakes and Pollard (1989) and Giné and Guillou (1999).

Appendix B. Regularity Conditions and Two Lemmas

We give some additional regularity conditions.

Condition D.0 (i) For sequences $(m) = (m_1, m_2)$, $m_1 \neq m_2$, of nonnegative integers, the variables $X_{(m)} = (X_{m_1}, X_{m_2})$ have joint density $f_{(m)}$ with respect to Lebesgue measure on $[0, \tau]^2$.

(ii) For sequences $[m] = (m_1, m_2, m_3)$ of distinct nonnegative integers, the variables $X_{[m]} = (X_{m_1}, X_{m_2}, X_{m_3})$ have joint densities $f_{[m]}$ with respect to Lebesgue measure on $[0, \tau]^3$.

For any vector, we denote by $|\cdot|$ the ℓ_1 norm. Without loss of generality we assume that the neighbourhood $\mathcal{B}$ surrounding the true parameter β_0 corresponds to a ball $\mathcal{B} = \{\beta : |\beta - \beta_0| \leq c_B\}$.

For nonnegative integers p and m define $\overline{\theta}_m^{(p)}(u) = |Z_m(u)|^p Y_m(u) e^{[|\beta_0|+c_B]|Z_m(u)|}$ and $\theta_m^{(p)}(u,\beta) = |Z_m(u)|^p Y_m(u) e^{\beta^T Z_m(u)}$. For $u \in [0,\tau_0]$, $\underline{u} = (u_1,u_2) \in [0,\tau_0]^2$, $\overline{u} = (u_1,u_2,u_3) \in [0,\tau_0]^3$, and $w \in [0,\tau]$, $\underline{w} = (w_1,w_2) \in [0,\tau]^2$, $\overline{w} = (w_1,w_2,w_3) \in [0,\tau]^3$, let

$$\begin{aligned}
\sigma_{p_1,p_2}(\underline{u},w) &= \sum_m \mathbb{E}[\prod_{j=1}^{2} \overline{\theta}_m^{(p_j)}(u_j)|X_m = w] f_m(w) \ , \\
\rho_{p_1,p_2}(\underline{u},\underline{w}) &= \sum_{(m)} \mathbb{E}[\prod_{j=1}^{2} \overline{\theta}_m^{(p_j)}(u_j)|X_{(m)} = \underline{w}] f_{(m)}(\underline{w}) \ , \\
\kappa_{1;p}(\overline{u},w) &= \sum_m \mathbb{E}[\theta_m^{(p)}(u_1,\beta_0) \prod_{j=2}^{3} \theta_m^{(0)}(u_j,\beta_0)|X_m = w] f_m(w) \ , \\
\kappa_{2;p}(\overline{u},\underline{w}) &= \sum_{(m)} \mathbb{E}[\theta_{m_1}^{(p)}(u_3,\beta_0) \prod_{j=1}^{2} \theta_{m_j}^{(0)}(u_j,\beta_0)|X_{(m)} = \underline{w}] f_{(m)}(\underline{w}) \ , \\
\kappa_{3;p}(\overline{u},\overline{w}) &= \sum_{[m]} \mathbb{E}[\theta_{m_1}^{(p)}(u_1,\beta_0) \prod_{j=2}^{3} \theta_{m_j}^{(0)}(u_j,\beta_0)|X_{[m]} = \overline{w}] f_{[m]}(\overline{w}) \ , \\
s_{0;2}(u,\underline{w}) &= \sum_{(m)} \mathbb{E}[\theta_{m_1}^{(0)}(u,\beta_0)|X_{(m)} = \underline{w}] f_{(m)}(\underline{w}) \ , \\
s_{0;3}(u,\overline{w}) &= \sum_{[m]} \mathbb{E}[\theta_{m_1}^{(0)}(u,\beta_0)|X_{[m]} = \overline{w}] f_{[m]}(\overline{w}) \ .
\end{aligned}$$

Under conditions D.1 and D.2 these expectations exist, at least in local neighbourhoods of a point $x \in [0,\tau]$. Such local neighbourhoods correspond to sets $\mathcal{R}(x) = \{(u,w) \in \mathcal{R} : |w-x| \le a\}$.

Conidtion D.1 (i) The condition D.0 (i) is satisfied and, for integers p_1, p_2 such that $p_j \ge 0, p_1+p_2 \le 4$, we have

$$\begin{aligned}
&\sup\{\sigma_{p_1,p_2}(\underline{u},w) : (u_1,w) \in \mathcal{R}(x), (u_2,w) \in \mathcal{R}(x)\} = O(1) \ , \\
&\sup\{|\rho_{p_1,p_2}(\underline{u},\underline{w}))| : (u_j,w_j) \in \mathcal{R}(x), j=1,2\} = O(1) \ .
\end{aligned}$$

(ii) The condition D.0 (ii) is sastisfied, and

$$\begin{aligned}
&\sup\{\kappa_{1;0}(\overline{u},w) : (u_j,w) \in \mathcal{R}(x), j=1,2,3\} = O(1) \ , \\
&\sup\{\kappa_{2;0}(\overline{u},\underline{w}) : (u_1,w_1) \in \mathcal{R}(x), (u_2,w_2) \in \mathcal{R}(x), (u_3,w_1) \in \mathcal{R}(x)\} = O(1) \ , \\
&\sup\{\kappa_{3;0}(\overline{u},\overline{w}) : (u_j,w_j) \in \mathcal{R}(x), j=1,2,3\} = O(1) \ , \\
&\sup\{s_{0;2}(u,\underline{w}) : (u,w_j) \in \mathcal{R}(x), j=1,2\} = O(1) \ , \\
&\sup\{s_{0;3}(u,\overline{w}) : (u,w_j) \in \mathcal{R}(x), j=1,2,3\} = O(1) \ .
\end{aligned}$$

Condition D.2 (i) The condition D.0 (i) is satisfied and, for integers p_1, p_2 such that $p_j \geq 0, p_1 + p_2 \leq 4$, we have

$$\begin{aligned}
&\sup\{\sigma_{p_1,p_2}(\underline{u}, w) : (u_1, w) \in \mathcal{R}, (u_2, w) \in \mathcal{R}\} = O(1) \ , \\
&\sup\{|\rho_{p_1,p_2}(\underline{u}, \underline{w}))| : (u_j, w_j) \in \mathcal{R}, |w_2 - w_1| \leq a, j = 1, 2\} = O(1) \ .
\end{aligned}$$

(ii) The condition D.0 (ii) is satisfied and, for $p = 0, 1$, we have

$$\begin{aligned}
&\sup\{\kappa_{1;p}(\overline{u}, w) : (u_j, w) \in \mathcal{R}, j = 1, 2, 3\} = O(1) \ , \\
&\sup\{\kappa_{2;p}(\overline{u}, \underline{w}) : (u_1, w_1) \in \mathcal{R}, (u_2, w_2) \in \mathcal{R}, (u_3, w_1) \in \mathcal{R}, |w_2 - w_1| \leq a\} = O(1) \ , \\
&\sup\{\kappa_{3;p}(\overline{u}, \overline{w}) : (u_j, w_j) \in \mathcal{R}, |w_2 - w_1| \leq a, |w_3 - w_2|, j = 1, 2, 3\} = O(1) \ .
\end{aligned}$$

We now give two lemmas which collect bounds on certain random variables arising in the analysis of the Aalen-Nelson estimate. Both lemmas can be verfied using elementary algebra, Hölder's inequality and conditions A and D.

Lemma B.8. *Suppose that* $\inf\{s^{(0)}(u, \beta, x) : \beta \in \mathcal{B}, u \leq \tau_0\} > 0$. *For* $k = 0, 1, 2$, *let* $\overline{f}_{kni}(u, \beta, x) = [s^{(0)}(u, \beta, x)]^{-1} \sum_m \theta^{(k)}(u, \beta)|K_n(x, X_{im})|$ *and* $f^*_{kni}(u, \beta, x) = [s^{(0)}(u, \beta, x)]^{-1} \sum_m \theta^{(k)}(u, \beta)\alpha(u, X_{im})|K_n(x, X_{im})|$. *If conditions* A *and* D.1 (i) *hold, then* $a^{-1}\mathbb{E} \prod_{p=1}^2 \overline{f}_{k_p ni}(u_p, \beta, x) = O(1)$ *and* $a^{-1}\mathbb{E} \prod_{p=1}^2 f^*_{k_p ni}(u_p, \beta, x) = O(1)$, *uniformly in* $u_1, u_2 \leq \tau_0$ *and* $\beta \in \mathcal{B}$. *If in addition the condition* D.1 (ii) *holds, then* $a^{-1}\mathbb{E} \prod_{p=1}^3 f^*_{k_p ni}(u_p, \beta_0, x) = O(1)$ *uniformly in* $u_1, u_2, u_3 \leq \tau_0$. *If* $\inf\{s^{(0)}(u, \beta, x) : \beta \in \mathcal{B}, (x, u) \in \mathcal{R}\} > 0$ *and conditions* D.2 *hold, then these bounds are also uniform in* $x, x \in [0, \tau]$.

Lemma B.9. *Supose that* $\inf\{s^{(0)}(u, \beta_0, w) : u \leq \tau, \beta \in \mathcal{B}, w \in [x - a_n \vee 0, x + a_n \wedge \tau]\} > 0$. *Set*

$$\begin{aligned}
\overline{\overline{f}}_{ni}(u, x) &= [\overline{s}^{(0)}(u, x)]^{-2}[\overline{S}_i^{(2)}(u, x) + \overline{S}_i^{(1)}(u, x)\overline{s}^{(1)}(u, x)] \ , \\
\overline{S}_i^{(p)}(u, x) &= \sum_m \overline{\theta}_{im}^{(p)}(u)|K_n(x, X_{im})| \ , \\
\overline{s}^{(0)}(u, x) &= \sum_m \mathbb{E}Y_{im}(u) \exp([-|\beta_0| - c_B]|Z_{im}(u))|X_{im} = x)f_m(x) \ , \\
\overline{s}^{(1)}(u, x) &= \sum_m \mathbb{E}(\overline{\theta}_m^{(p)}|X_{im} = x)f_m(x) \ , \\
\overline{\overline{g}}_{nj}(v, x) &= \sum_m \int_0^v |K_n(x, X_{jm})|[\overline{s}^{(0)}(u, x)]^{-1} N_{jm}(du) \ ,
\end{aligned}$$

$$\begin{aligned}
\overline{g}_{in}(u,\beta,x) &= \sum_m \int_0^u |K_n(x,X_{im})|[s^{(0)}(u,\beta,x)]^{-1}N_{im}(du)\ ,\\
H_{0n}(W_i) &= a^{\frac{-1}{2}}[\overline{g}_{ni}(\tau_0,\beta_0,x)+\int_0^{\tau_0} f^*_{0ni}(u,\beta_0,x)du]\ ,\\
H_{1n}(W_i) &= a^{\frac{-1}{2}}\sum_m \int_0^{\tau_0} \frac{|K_n(x,X_{im})|}{s^{(0)}(u,\beta_0,x)} Y_{im}(u)e^{\beta_0^T Z_{im}(u)}\\
&\quad\times\quad |\alpha(u,X_{im})-\alpha(u,x)|du\ ,\\
H_{2n}(W_i,W_j) &= \frac{1}{a\sqrt{na}}\int_0^{\tau_0} \overline{f}_{0ni}(u,\beta_0,x)\overline{g}_{nj}(u,\beta_0,x)\ ,\\
H_{3n}(W_i) &= a^{\frac{-1}{2}}\int_0^{\tau_0} \frac{|\overline{s}^{(0)}-s^{(0)}|}{s^{(0)}}(u,\beta_0,x)\overline{g}_{ni}(du,\beta_0,x)\ ,\\
H_{4n}(W_i) &= a^{\frac{-1}{2}}\int_0^{\tau_0} \overline{f}_{ni}(u,\beta_0,x)|\alpha(u,x)du-\frac{\mathbb{E}N(du,x)}{s^{(0)}(u,\beta_0,x)}|\ ,\\
H_{5n}(W_i,W_j) &= \frac{1}{na^2}\int_0^{\tau_0} \overline{f}_{1ni}(u,\beta_0,x)\overline{g}_{nj}(du,\beta_0,x)\ ,\\
&\quad+\frac{c_B}{na^2}\int_0^{\tau_0} \overline{\overline{f}}_{ni}(u,x)\overline{\overline{g}}_{nj}(du,x)\ .
\end{aligned}$$

If conditions A.$r(r=1,2)$ *and* D.1 *hold, then* $\mathbb{E}H^2_{0n}(W_1)=O(1)$, $\mathbb{E}H^3_{0n}(W_1)=O(a^{-1/2})$, $\mathbb{E}H^2_{1n}(W_1)=O(a^2)$, $\mathbb{E}H_{3n}(W_1)^2=O(a^{2r})$ *and* $\mathbb{E}H_{4n}(W_1)^2=O(a^{2r})$. *We also have* $\mathbb{E}H^2_{2n}(W_1,W_2)=O((na)^{-1})$, $\mathbb{E}H^2_{5n}(W_1,W_2)=O((na)^{-2})$ *and* $n\mathbb{E}[\mathbb{E}_{\{1\}}H_{5n}(W_1,W_2)]^2=O((na)^{-1})=n\mathbb{E}[\mathbb{E}_{\{2\}}H_{5n}(W_1,W_2)]^2$.

Appendix C. Proof of Proposition 3.4

Set

$$b(v,x)=\int_0^v \frac{\alpha(u,x)}{s^{(0)}(u,\beta_0,x)}[\gamma(u,x)-\overline{s}^{(0)}(u,\beta_0,x)]du\ ,$$

where $\overline{s}^{(0)}(u,\beta_0,x)=\mathbb{E}S^{(0)}_{-i}(u,\beta_0,x)$, $\gamma(u,x)=\overline{n}(du,x)/\alpha(u,x)$, and $\overline{n}(v,x)=\Sigma_m\mathbb{E}N_{im}(v)K_n(x,X_{im})$. Then $\sqrt{na}[\widehat{A}(v;x,\beta_0)-A_0(v,x)-b(v,x)]=\widehat{Z}_n(v,x)+R_n(v,x)$, where

$$\widehat{Z}_n(v,x)=\sqrt{\frac{n}{a}}\sum_{i=1}^n\sum_m\int_0^v \frac{K_n(x,X_{im})}{s^{(0)}(u,\beta_0,x)}M_{im}(du)+R_n(v,x)\ ,$$

and $R_n(v,x)$ is a remainder term given below. Under conditions A.$r, r=1,2$, we have $\sqrt{na}b(v,x)=O(\sqrt{na}a^r)=o(1)$. Therefore it is enough to show that the process $\widehat{Z}_n(v,x)$ converges in $\ell^\infty([0,\tau_0])$ to a time transformed Brownian motion and the remainder term R_n is asymptotically negligible.

We have $\widehat{Z}_n(v,x) = \sqrt{n}[P_n - P]h_{n,v}$ where

$$h_{n,v}(W_i) = a^{\frac{-1}{2}} \sum_m \int_0^v \frac{K_n(x, X_{im})}{s^{(0)}(u,\beta_0,x)} M_{im}(du) \ .$$

The class $\mathcal{H}_n = \{h_{n,v} : v \leq \tau_0\}$ consists of functions that can be represented as a linear combination of at most four monotone functions with respect to v and has envelope $4H_{0n}(W_i)$. By Lemma B.9 we have (i) $\mathbb{E}H_{0n}^2(W_1) = O(1)$, and (ii) $\mathbb{E}H_{0n}(W_1)1(H_{0n}(W_1) > \eta\sqrt{n}) \leq \mathbb{E}H_{0n}^3(W_1)(\eta\sqrt{na})^{-1} \to 0$ for any $\eta > 0$. Also (iii) for any $0 < v_1 < v_2 \leq \tau_0$, the difference $|h_{nv_1} - h_{nv_2}|(W_i)$ is bounded by

$$a^{\frac{-1}{2}} \sum_m \int_{v_1}^{v_2} \frac{|K_n(x, X_{im})|}{s^{(0)}(u,\beta_0,x)} M_{im}(du) + \frac{2}{\sqrt{a}} \int_{v_1}^{v_2} f_{0ni}^*(u,\beta_0,x)du \ .$$

Using $(x+y)^2 \leq 2(x^2+y^2)$ and Lemma 2.1, $\mathbb{E}|h_{nv_1} - h_{nv_2}|^2(W_1)$ is bounded by

$$\frac{2}{a}\int_{v_1}^{v_2}\int_0^\tau \frac{[s^{(0)}\alpha](u,\beta_0,w)}{s^{(0)}(u,\beta_0,x)^2} K_n^2(x,w)dwdu + \frac{8}{a}\int_{v_1}^{v_2}\int_{v_1}^{v_2} \mathbb{E}\prod_{p=1}^2 f_{0ni}^*(u_p,\beta_0,x)du_1du_2 \ ,$$

and is of order $O(|v_2 - v_1| + |v_2 - v_1|^2)$. Lemmas 2.1 and 3.2, imply (iv)

$$\mathrm{Var}\,[\widehat{Z}_n(v_1,x)] = d_{p(x),q(x)}(K) \int_0^{v_1} \frac{\alpha(u,x)}{s^{(0)}(u,x,\beta_0)} du + O(a) \ ,$$

and $cov[\widehat{Z}_n(v_1,x), \widehat{Z}_n(v_2,x) - \widehat{Z}_n(v_1,x)] = O(a)$. Finally, (v) the class of functions $\{h_{nv} : v \leq \tau_0\}$ has polynomial bracketing number. Properties (i)$-$(v) and Theorem 2.11.23 in van der Vaart and Wellner (1996) imply that $\{\widehat{Z}_n(v,x) : v \in \tau_0\}$ converges weakly $\ell^\infty([0,\tau_0])$ to a tight Gaussian process.

The remainder term $R_n(v,x)$ is given by $R_n(v,x) = \sum_{j=1}^5 R_{jn}(v,x)$, where

$$R_{1n}(v,x) = \frac{1}{\sqrt{na}} \sum_{i=1}^n \int_0^v \widetilde{f}_{ni}(u,x)du - \sqrt{na}b(v,x) \ ,$$

$$R_{2n}(v,x) = -\frac{\sqrt{na}}{n(n-1)a^2} \sum_{i \neq j} \int_0^v [f_{ni} - \mathbb{E}f_{ni}](u,x)[g_{nj} - \mathbb{E}g_{nj}](du,x)] \ ,$$

$$R_{3n}(v,x) = -\frac{1}{\sqrt{na}} \sum_{i=1}^n \int_0^v [\frac{\overline{s}^{(0)} - s^{(0)}}{s^{(0)}}](u,\beta_0,x)[g_{ni} - \mathbb{E}g_{ni}](du,x) \ ,$$

$$R_{4n}(v,x) = \frac{1}{\sqrt{na}} \sum_{i=1}^n \int_0^v ([f_{ni} - \mathbb{E}f_{ni}]s^{(0)})(u,\beta_0,x)\alpha(u,x)du - \mathbb{E}N(du,x)] \ ,$$

$$R_{5n}(v,x) = \frac{1}{\sqrt{na}} \int_0^v [\overline{s}^{(0)} - s^{(0)}](u,\beta_0,x)\left[\alpha(u,x)du - \frac{\mathbb{E}N(du,x)}{s^{(0)}(u,\beta_0,x)}\right] \ ,$$

$$R_{6n}(v,x) = \sqrt{na} \sum_{i=1}^n \int_0^v \frac{[S_{-i}^{(0)}(u,\beta_0,x) - s^{(0)}(u,\beta_0,x)]^2}{S_{-i}^{(0)}(u,\beta_0,x)s^2(u,\beta_0,x)} \overline{N}_i(du,x) \ ,$$

where

$$f_{ni}(u,x) = [s^{(0)}(u,\beta_0,x)]^{-1}\sum_m Y_{im}(u)e^{\beta_0^T Z_{im}(u)}K_n(x,X_{im})\ ,$$
$$\widetilde{f}_{ni}(u,x) = [s^{(0)}(u,\beta_0,x)]^{-1}\sum_m [\alpha(u,X_{im})-\alpha(u,x)]Y_{im}(u)e^{\beta_0^T Z_{im}(u)}K_n(x,X_{im}),$$
$$g_{ni}(u,x) = \sum_m \int_0^u K_n(x,X_{im})[s^{(0)}(v,\beta_0,x)]^{-1}N_{im}(dv)\ .$$

The term R_{1n} has mean zero. By decomposing the integrands and the integrators into their positive and negative parts, we have $(na)^{-1/2}R_{1n}(v,x) + b(v,x) = \mathbb{P}_n h_{1nv}$, where $h_{1nv}(W_i)$ is a sum of four monotone functions, bounded by $H_{1n}(W_i)$. Thus $R_{1n}(v,x)$ is a normalized empirical process over a Euclidean class of functions for envelope $4H_{1n}(W_i)$. By Lemmas B.9 and A.7, we have $\mathbb{E}H_{1n}(W_1)^2 = O(a^2)$ and $\mathbb{E}\sup_v |R_{1n}(v,x)| = O(a)$. Similarly, using envelopes H_{3n} and H_{4n}, we can show that $\mathbb{E}\sup_v |R_{3n}(v,x)| = O(a^r) = \mathbb{E}\sup_v |R_{4n}(v,x)|$ and $R_{5n}(v,x) = O(\sqrt{na}a^{2r})$ a.s.uniformly in $v \le \tau_0$. The term term R_{2n} is easily seen to form a canonical U-process of degree 2 over a Euclidean class of functions with envelope $H'_{2n}(W_i,W_j) = H_{2n}(W_i,W_j) + \mathbb{E}_{\{1\}}H_{2n}(W_i,W_j) + \mathbb{E}_{\{2\}}H_{2n}(W_i,W_j)+\mathbb{E}H_{2n}(W_i,W_j)$. Lemmas B.9 and A.7 imply $\mathbb{E}\sup_v |R_{2n}(v,x)| = O((na)^{-1/2}))$, since $\mathbb{E}H_{2n}^2(W_i,W_j) = O((na)^{-1})$ and $\mathbb{E}[H'_{2n}]^2(W_1,W_2)$ is of the same order.

Next define

$$\begin{aligned} R_{7n} &= \frac{1}{\sqrt{na}}\sum_{i=1}^n \int_0^{\tau_0} (\frac{S_{-i}^{(0)} - s^{(0)}}{s^{(0)}})^2(u,\beta_0,x)\overline{g}_{ni}(du,\beta_0,x) \\ &\le 2\sqrt{na}O(a^{2r})\frac{1}{na}\sum_{i=1}^n \overline{g}_{ni}(\tau_0,\beta_0,x) + O(1)(R_{7n;1}+R_{7n;2})\ , \end{aligned}$$

where $R_{7n;1} = \sqrt{na}a^{-3}\mathbb{U}_{n,3}(h)$, $R_{7n;2} = \sqrt{na}(na^3)^{-1}\mathbb{U}_{n,2}(\overline{h})$,

$$h(W_i,W_j,W_k) = \int_0^{\tau_0} [(f_{0nj}-\mathbb{E}f_{0nj})(f_{0nk}-\mathbb{E}f_{0nk})](u,\beta_0,x)\overline{g}_{ni}(du,\beta_0,x)$$

and $\overline{h}(W_i,W_j) = h(W_i,W_j,W_j)$. The first term is of order $O_p(\sqrt{na}a^{2r})$. We have $\mathbb{E}H(W_1,W_2,W_3) = 0$ and, using Lemmas B.9 and A.7, $\mathbb{E}(na)^{1/2}a^{-3}|\mathbb{U}_{n,3}(\pi_3 h)| = O((na)^{-1})$ and $\mathbb{E}(na)^{1/2}a^{-2}|\mathbb{U}_{n,2}(\pi_2[\mathbb{E}_{\{23\}}h])| = O((na)^{-1/2})$. The remaining projections are 0. In the case of the term $R_{7n;2}$, we have $\mathbb{E}R_{7n;2} = O((na)^{-1/2})$ and the expected $\mathbb{E}|R_{7n;2}|$ is of the same order.

We consider now term R_{6n}. For $\varepsilon \in (0,1)$, define

$$\Omega_n(\varepsilon) = \left\{\frac{1}{1+\varepsilon} \le \min_i \inf_{\substack{u\le\tau_0\\ \beta\in\mathcal{B}}} \frac{s^{(0)}}{S_{-i}^{(0)}}(u,\beta,x) \le \max_i \sup_{\substack{u\le\tau_0\\ \beta\in\mathcal{B}}} \frac{s^{(0)}}{S_{-i}^{(0)}}(u,\beta,x) \le \frac{1}{1-\varepsilon}\right\}.$$

We have $P(\Omega_n(\varepsilon)) \leq \min_i P(\sup_{u\leq\tau} |S^{(0)}_{-i}/s^{(0)} - 1|(u,\beta,x)| \leq \varepsilon) \to 1$, by condition B and Markov's inequality. On the event $\Omega_n(\varepsilon)$, we also have $\sup_{v\leq\tau_0} |R_{6n}(v,x)| \leq (1-\varepsilon)^{-1} R_{7n}$. Therefore $P(\sup_{v\leq\tau_0} |R_{6n}(v,x)| > \eta) \leq P(\Omega^c_n(\varepsilon)) + P(\sup_{v\leq\tau_0} |R_{6n}(v,x)| > \eta, \Omega_n(\varepsilon)) \leq P(\Omega^c_n(\varepsilon)) + P(R_{7n} > (1-\varepsilon)\eta) \to 0$ for any $\eta > 0$.

Finally, suppose that $\widehat{\beta}$ is a $\sqrt{n}$ consistent estimate of the parameter β. Then

$$\begin{aligned}&\sqrt{na}[\widehat{A}(v,x,\widehat{\beta}) - \widehat{A}(v,x,\beta_0)] \\ &= \sqrt{n}[\widehat{\beta} - \beta_0]\sqrt{a}\int_0^v \frac{S^{(1)}_{-i}}{[S^{(0)}_{-i}]^2}(u,\beta^*,x)N_i(du,x)\ ,\end{aligned} \tag{C.1}$$

where β^* is between β_0 and $\widehat{\beta}$. Let $I_n(\beta) = a^{-2}U_{n2}(h_\beta)$, where $h_\beta(W_i,W_j) = \int_0^{\tau_0} \overline{f}_{1ni}(u,\beta,x)\overline{g}_{nj}(du,\beta,x)$. It is easy to see that $\mathbb{E}I_n(\beta) = O(1)$. By Lipschitz continuity of the function h_β with respect to β, $I_n(\beta)$ is a U-process of degree 2 over a Eulidean class of functions for envelope $H_{5n}(W_i,W_j)$. By Lemmas B.9 and A.7, $\mathbb{E}\sup_{\beta\in\mathcal{B}} |a^{-2}\mathbb{U}_{n,2}(\pi_2 h_\beta)| = O([\mathbb{E}H_{5n}(W_1,W_2)^2]^{1/2}) = O((na)^{-1})$, $\mathbb{E}\sup_{\beta\in\mathcal{B}} |a^{-2}U_{1n}(\pi_1\mathbb{E}_{\{1\}}h_\beta)| = O((na)^{-1/2}))$, $\mathbb{E}\sup_{\beta\in\mathcal{B}} |a^{-2}U_{1n}(\pi_1\mathbb{E}_{\{2\}}h_\beta)| = O((na)^{-1/2}))$. Therefore $\sup_{\beta\in\mathcal{B}} |I_n(\beta)| = O_p(1)$. Further, if $\widehat{\beta}$ is a $\sqrt{n}$ consistent estimate of β_0, then $\sqrt{n}[\widehat{\beta} - \beta_0] = O_p(1)$. To show that the right–hand side of (6.1) is of order $O_p(\sqrt{a})$, it is enough to note that for any $\varepsilon \in (0,1)$, the supremum $\sup\{|\int_0^v S^{(1)}_{-i}[S^{(0)}_{-i}]^{-2}(u,\beta,x)N_i(du,x)| : v \leq \tau_0, \beta \in \mathcal{B}\}$ is bounded by $(1-\varepsilon)^{-2}\sup_{\beta\in\mathcal{B}} I_n(\beta)$ on the event $\Omega(\varepsilon)$.

Appendix D. Proof of Propositions 3.3 and 3.5

Define

$$\begin{aligned}\widetilde{\Phi}_{0n}(\beta_0) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n\sum_m \int [Z_{im}(u)s^{(0)}_{-i}(u,\beta_0,X_{im}) - s^{(1)}_{-i}(u,\beta_0,X_{im})]M_{im}(du)\ ,\\ \widetilde{\Sigma}_n(\beta) &= \frac{1}{n}\sum_{i=1}^n\sum_m \int [S^{(2)}(u,\beta,X_{im}) - Z_{im}(u)\otimes S^{(1)}(u,\beta,X_{im})]N_{im}(du)\ ,\\ \Phi_{0n}(\beta_0) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n\sum_m \int [Z_{im}(u) - \frac{s^{(1)}}{s^{(0)}}(u,\beta_0,X_{im})]M_{im}(du)\ ,\\ \Sigma_{0n}(\beta) &= \frac{1}{n}\sum_{i=1}^n\sum_m \int [\frac{s^{(2)}}{s^{(0)}} - \left(\frac{s^{(1)}}{s^{(0)}}\right)^{\otimes 2}](u,\beta,X_{im})N_{im}(du)\ ,\end{aligned}$$

$$\Phi_{1n}(\beta_0) = \frac{1}{\sqrt{n}} \sum_m \int_0^{\tau_0} [Z_{im}(u) \frac{S_{-i}^{(0)}}{s^{(0)}}(u, \beta_0, X_{im}) - \frac{S_{-i}^{(1)}}{s^{(0)}}(u, \beta_0, X_{im})] N_{im}(du) \ ,$$

$$\Phi_{2n}(\beta_0) = -\frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_m \int_0^{\tau_0} [Z_{im}(u) - \frac{s^{(1)}}{s^{(0)}}(u, \beta_0, X_{im})]$$

$$\times \quad (\frac{S_{-i}^{(0)} - s^{(0)}}{s^{(0)}})(u, \beta_0, X_{im}) N_{im}(du) \ ,$$

$$\Phi_{3n}(\beta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_m \int_0^{\tau_0} (\frac{S_{-i}^{(1)} - s^{(1)}}{s^{(0)}})(\frac{S_{-i}^{(0))} - s^{(0)}}{s^{(0)}})(u, \beta_0, X_{im}) N_{im}(du) \ ,$$

$$\Phi_{4n}(\beta_0) = -\frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_m \int_0^{\tau_0} \frac{S_{-i}^{(1)}}{s} [(\frac{S_{-i}^{(0)}}{s^{(0)}} - 1)^2 \frac{s^{(0)}}{S_{-i}^{(0)}}](u, \beta_0, X_{im}) N_{im}(du) \ ,$$

$$\Sigma_{1n}(\beta) = \frac{1}{n} \sum_{i=1}^n \sum_m \int [\frac{S_{-i}^{(2)} - s^{(2)}}{s^{(0)}} - \psi_{-i} - \psi_{-i}^T](u, \beta, X_{im}) N_{im}(du) \ ,$$

$$\Sigma_{2n}(\beta) = -\frac{1}{n} \sum_{i=1}^n \sum_m \int [\frac{S_{-i}^{(1)} - s^{(1)}}{s^{(0)}}]^{\otimes 2}(u, \beta, X_{im}) N_{im}(du) \ ,$$

$$\Sigma_{3n}(\beta) = \frac{1}{n} \sum_{i=1}^n \sum_m \int [\widehat{V}_{-i} - \widetilde{V}_{-i}](u, \beta, X_{im}) N_{im}(du) \ ,$$

where $\widehat{\psi}_{-i} = (S_{-i}^{(1)} - s^{-1}) \otimes s^{(1)}/[s^{(0)}]^2$, $\widehat{V}_{-i} = S_{-i}^{(2)}/S_{-i}^{(0)} - (S_{-i}^{(1)}/S_{-i}^{(0)})^{\otimes 2}$ and $\widetilde{V}_{-i} = S_{-i}^{(2)}/s^{(0)} - (S_{-i}^{(1)}/s^{(0)})^{\otimes 2}$.

Under assumptions of Proposition 2.3, $\widetilde{\Sigma}_n(\beta)$ is the negative derivative of the score function $\widetilde{\Phi}_n(\beta)$. Similarly, under assumptions of Proposition 2.5, we have $\Phi_n(\beta_0) = \Sigma_{j=1}^4 \Phi_{jn}(\beta_0)$ and $\Sigma_n(\beta) = \sum_{j=1}^3 \Sigma_{nj}(\beta)$ is the negative derivative of the score function $\Phi_n(\beta)$. The proof of both propositions amounts to application of the following lemma and results of Bickel et al. (1993, p.517).

Lemma D.10.

(i) *Under assumptions of Proposition* 2.3 *we have* $\widetilde{\Phi}_{0n}(\beta_0) \Rightarrow \mathcal{N}(0, \Sigma_2(\beta_0))$, $\widetilde{\Sigma}_n(\beta_0) \to_P \Sigma_1(\beta_0)$, $\widetilde{\Phi}_n(\beta_0) - \widetilde{\Phi}_{0n}(\beta_0) \to_P 0$, *and* $\sup\{|\widetilde{\Sigma}_n(\beta) - \widetilde{\Sigma}(\beta_0)| : |\beta - \beta_0| \leq \varepsilon_n\} \to_P 0$.

(ii) *Under assumptions of Proposition* 2.5 *we have* $\Phi_{0n}(\beta_0) \Rightarrow \mathcal{N}(0, \Sigma(\beta_0))$, $\Sigma_{0n}(\beta_0) \to_P \Sigma(\beta_0)$, $\Phi_{1n}(\beta_0) - \Phi_{0n}(\beta_0) \to_P 0$, $\Phi_{kn}(\beta_0) \to_P 0$ *for* $k = 2, 3, 4$, $\Sigma_{kn}(\beta_0) \to_P 0$ *for* $k = 1, 2, 3$, *and* $\sup\{|\Sigma_{kn}(\beta) - \Sigma_{kn}(\beta_0)| : |\beta - \beta_0| \leq \varepsilon_n\} \to_P 0$ *for* $k = 0, 1, 2, 3$.

Proof. First note that under the assumed regularity conditions, asymptotic normality of the terms $\widetilde{\Phi}_{0n}(\beta_0)$ and $\Phi_{0n}(\beta_0)$ follows from the CLT.

We show that $\widetilde{\Phi}_n(\beta_0) - \widetilde{\Phi}_{0n}(\beta_0) \to_P 0$ and $\Phi_{1n}(\beta_0) - \Phi_{0n}(\beta_0) \to_P 0$. For any bounded function $\varphi(u,x)$, let $G^{\varphi}_{ij} = G^{\varphi}(W_i, W_j)$ be given by

$$G^{\varphi}_{ij} = \sum_m \int_0^{\tau_0} \varphi(u, X_{im})[Z_{im}(u) S^{(0)}_j(u,\beta,X_{im}) - S^{(1)}_j(u,\beta,X_{im})] N_{im}(du).$$

Under assumptions of Proposition 2.3, we have $\widetilde{\Phi}_n(\beta_0) = \widetilde{\Phi}_{0n}(\beta_0) + O_P(\sqrt{n}a) + \mathbb{U}_{n,2}(\pi_2 G^{\varphi}))$ for $\varphi \equiv 1$. Similarly, under assumptions of Proposition 2.5, we have $\Phi_{1n}(\beta_0) = \Phi_{0n}(\beta_0) + O_P(\sqrt{n}a^2) + \mathbb{U}_{n,2}(\pi_2 G^{\varphi})$ for $\varphi(u,x) = [s^{(0)}(u,\beta_0,x)]^{-1}$. Thus it is enough to show that in both cases $\mathbb{E}\mathbb{U}_{n,2}(\pi_2 G^{\varphi}) = O((na)^{-1/2})$. Choose $\varphi = [s^{(0)}]^{-1}$ for instance, and define

$$\begin{aligned}
&\overline{G}_n(W_i, W_j)) = a^{-1} \sum_{p=0}^{1} \sum_m \int |Z_{im}(u)|^p \overline{f}_{1-pjn}(u,\beta_0,X_{im}) M_{im}(du) \\
&+ a^{-1} \sum_{p=0}^{1} \int \sum_m |Z_{im}|^p(u) Y_{im}(u) e^{\beta_0 Z_{im}(u)} \overline{f}_{1-p,jn}(u,\beta_0,X_{im}) \alpha(u,X_{im}) du \\
&a^{-1} \sum_m \int_0^{\tau_0} [|Z_{im}(u)| \overline{f}_{0nj}(u,\beta_0,X_{im}) + \overline{f}_{1,nj}(u,\beta_0,X_{im})] N_{im}(du)\ .
\end{aligned}$$

We have $\mathbb{E}\mathbb{U}_{n,2}(\pi_2 G^{\varphi})) = O(n^{-1/2}(\mathbb{E}\overline{G}^2_n(W_1,W_2))^{1/2}) = O((na)^{-1/2})$ because, by Lemma 2.1, the expectation $\mathbb{E}\overline{G}^2_n(W_1,W_2)$ is bounded by

$$\begin{aligned}
&\frac{4}{a^2} \sum_{p=0}^{1} \int_0^{\tau_0} \int_0^{\tau} \sigma_{p,p}(u,u,x) \mathbb{E}[\overline{f}_{1-p,jn}(u,\beta_0,x)]^2 \alpha(u,x) du dx \\
&+ \int_0^{\tau_0} \int_0^{\tau_0} \int_0^{\tau} \sigma_{p,p}(u_1,u_2,x) \mathbb{E}[\prod_{l=1}^{2} \overline{f}_{1-p,jn}(u_l,\beta_0,x)] \prod_{l=1}^{2} \alpha(u_l,x) du_1 du_2 dx \\
&+ \int_0^{\tau_0} \int_0^{\tau_0} \int_0^{\tau} \int_0^{\tau} \rho_{p;p}(\underline{u},\underline{x}) \mathbb{E}[\prod_{l=1}^{2} \overline{f}_{1-p,jn}(u_l,\beta_0,x_l)] \prod_{l=1}^{2} \alpha(u_l,x_l) du_l dx_l.
\end{aligned}$$

Here in the last line $\underline{u} = (u_1,u_2)$ and $\underline{x} = (x_1,x_2)$. By Lemma B.8, the bound is of order $O(a^{-1})$. It follows now that $\Phi_{1n}(\beta_0) - \Phi_{0n}(\beta_0) \to_P 0$.

The same argument applied to the function $\varphi(u,x) \equiv 1$ shows that $\widetilde{\Phi}_n(\beta_0) - \widetilde{\Phi}_{0n} \to_P 0$. Changing the risk processes $S^{(k)}_j$ $-S^{(k+1)}_j$, $k = 0,1$, in the definition of $G^{\varphi}(W_i, W_j)$, we also obtain

$$\begin{aligned}
&\widetilde{\Sigma}_n(\beta_0) - \Sigma_1(\beta_0) \\
&= n^{-1} \sum_m \int_0^{\tau_0} Z_{im}(u) s^{(1)}(u,\beta_0,X_{im}) - s^{(2)}(u,\beta_0,X_{im}) M_{im}(du) + o_p(1)
\end{aligned}$$

The Strong Law of Large Numbers implies that $\widetilde{\Sigma}(\beta_0) \to_P \Sigma_1(\beta_0)$. Components of the matrix $\widetilde{\Sigma}(\beta)$ are Lipschitz continuous in β, and it is easy to verify that $|\widetilde{\Sigma}_n(\beta) - \widetilde{\Sigma}_n(\beta')| \leq |\beta - \beta'| \mathbb{U}_{n,2}(G_{2n})$ where G_{2n} is a kernel degree 2 satisfying $\mathbb{E}|\mathbb{U}_{n,2}(G_{2n})| = O(1)$. This completes the proof of the first part of the proposition.

Further, the terms $\Phi_{2n}(\beta_0)$ and $\Sigma_{1n}(\beta_0)$ are U-statistics of degree 2. Using similar algebra as in the case of the difference $\Phi_{1n} - \Phi_{0n}$, we can show that they converge to 0 in probability.

Next define

$$H_{1n} = \frac{1}{n} \sum_{i=1}^{n} \sum_{m} \int_0^{\tau_0} \varphi(u, X_{im}) \prod_{k=1}^{2} \left(\frac{S_{-i}^{(p)} - s^{(p)}}{s^{(0)}} \frac{S_{-i}^{(q)} - s^{(q)}}{s^{(0)}}\right)(u, \beta_0, X_{im}) N_{im}(du),$$

where $\varphi(u,x)$ is a bounded function and $p, q = 0$ or 1. We have $\sqrt{n} H_{1n} = O_p(\sqrt{n} a^2) + O(1)[\sqrt{n} a^{-2} \mathbb{U}_{n,3}(H) + \sqrt{n}(na^2)^{-1} \mathbb{U}_{n,2}(\overline{H})]$, where

$$H(W_i, W_j, W_k) = \sum_{m} \int_0^{\tau_0} \varphi(u, X_{im}) [f_{pjn} - \mathbb{E} f_{pjn}][f_{qkn} - \mathbb{E} f_{qkn}](u, \beta_0, X_{im}) N_{im}(du)$$

and $\overline{H}(W_i, W_j) = H(W_i, W_j, W_j)$. We have $\mathbb{E} H(W_1, W_2, W_3) = 0$. Lemmas A.7, B.8 and B.9 imply that $\mathbb{E}\sqrt{n} a^{-2} |\mathbb{U}_{n,3}(\pi_3 H)| = O((na)^{-1})$ and $O((na)^{-1/2}) = \mathbb{E}\sqrt{n}(na^2)^{-1} |\mathbb{U}_{n,2}(\pi_2 \mathbb{E}_{\{23\}} H)|$, while the remaining projections are 0. Further, $\sqrt{n}(na^2)^{-1} \mathbb{E}\mathbb{U}_{n,2}(\overline{H}) = O((na^2)^{-1/2}) = \sqrt{n}(na^2)^{-1} \mathbb{U}_{n,2}(|\overline{H}|)$, so that the condition $na^2 \uparrow \infty$ implies asymptotic negligibility of the third term of $\sqrt{n} H_{1n}$.

The choice of $\varphi \equiv 1$, $p = 1, q = 0$ implies that if $na^4 \downarrow 0$ and $na^2 \uparrow \infty$ then $\Phi_{3n}(\beta_0) \to_P 0$. The choice of $\varphi \equiv 1$ and $p = q = 1$ implies $\Sigma_{2n}(\beta_0) \to_P 0$.

To handle the term $\Phi_{4n}(\beta_0)$ define

$$H_{2n} = \frac{1}{n} \sum_{i=1}^{n} \sum_{m} \int_0^{\tau_0} \left(\frac{\overline{f}_{1n}}{s^{(0)}} \left(\frac{S_{-i}^{(0)} - s^{(0)}}{s^{(0)}}\right)^2\right)(u, \beta_0, X_{im}) N_{im}(du) \,.$$

Using $(x+y)^2 \leq 2x^2 + 2y^2$, we have $\sqrt{n}|H_{2n}| \leq 2 O_p(\sqrt{n} a^4) + 2\sqrt{n} H_{2n;1} + 2\sqrt{n} H_{2n;2}$, where $H_{2n;1}$ corresponds to the sum H_{1n} applied with function $\varphi = \mathbb{E}\overline{f}_{1ni}/s^{(0)}$, and $H_{2n;2}$ is a V statistics of degree 4: $H_{2n;2} = O(1)[a^{-3}\mathbb{U}_{n,4}(h) + (a^3 n)^{-1} \mathbb{U}_{n,3}(\overline{h}) + 2(na^3)^{-1} \mathbb{U}_{n,3}(h') + (n^2 a^3)^{-1} \mathbb{U}_{n,2}(h'')]$, where

$$h(W_i, W_j, W_k, W_l) = \sum_{m} \int_0^{\tau_0} [\overline{f}_{1jn} - \mathbb{E}\overline{f}_{1jn}] \prod_{p=k,\ell} [f_{0pn} - \mathbb{E} f_{0pn}](u, \beta_0, X_{im}) N_{im}(du)$$

and $\overline{h}(W_i, W_j, W_k) = h(W_i, W_j, W_k, W_k)$, $h'(W_i, W_j, W_k) = h(W_i, W_j, W_j, W_k)$, $h''(W_i, W_j) = h(W_i, W_j, W_j, W_j)$. We have $\mathbb{E}|\sqrt{n}(n^2 a^3)^{-1} \mathbb{U}_{n,2}(h'')| \leq \sqrt{n}(na^3)^{-1}$ $\mathbb{E}\mathbb{U}_{n,2}|h''|$, which is bounded by

$$\frac{\sqrt{n}}{n^2 a^3} \int_{\mathcal{R}} \mathbb{E}|[f_{1jn} - \mathbb{E}\overline{f}_{1jn}][f_{0jn} - \mathbb{E} f_{0jn}]^2|(u, \beta_0, x) s^{(0)}(u, \beta_0, x) \alpha(u, x) du dx.$$

Under conditions D.2 (ii), this bound is of order $O(n^{-3/2}a^{-2})$ and tends to 0 if $na^2 \uparrow \infty$. A similar argument shows also that the second and third term of $\sqrt{n}H_{2n;2}$ have expectation tending to 0 when $na^2 \uparrow \infty$ and $na^4 \downarrow 0$. The first term has expectation 0. By Lemmas A.7 and B.9, we have $\mathbb{E}\sqrt{n}a^{-3}|U_{4n}\pi_4 h| = O((na)^{-3/2})$, $\mathbb{E}\sqrt{n}a^{-3}|\mathbb{U}_{n,3}(\pi_3\mathbb{E}_{\{234\}}h)| = O((na)^{-1})$, while the remaining projections are 0.

Further, for $\varepsilon \in (0,1)$, define

$$\Omega_n(\varepsilon) = \Big\{\frac{1}{1+\varepsilon} \leq \min_i \inf_{\substack{(u,x)\in\mathcal{R}\\ \beta\in\mathcal{B}}} \frac{s^{(0)}}{S^{(0)}_{-i}}(u,\beta,x) \leq \max_i \sup_{\substack{(u,x)\in\mathcal{R}\\ \beta\in\mathcal{B}}} \frac{s^{(0)}}{S^{(0)}_{-i}}(u,\beta,x) \leq \frac{1}{1-\varepsilon}\Big\}.$$

As in the proof of Proposition 3.4, the condition C implies $P(\Omega_n(\varepsilon)) \to 1$. Also on the event $\Omega_n(\varepsilon)$, the term $\Phi_{4n}(\beta_0)$ satisfies $|\Phi_{4n}(\beta_0| \leq (1-\varepsilon)^{-1}\sqrt{n}H_{2n}$. For any $\eta > 0$, we have $P(|\Phi_{4n}(\beta_0)| > \eta) \leq P(\sqrt{n}H_{2n} > \eta, \Omega(\varepsilon)) + P(\Omega_n^c(\varepsilon)) \leq P(\sqrt{n}H_{2n} > (1-\varepsilon)\eta) + P(\Omega_n^c(\varepsilon)) \to 0$.

Application of the condition C shows also that $\Sigma_{3n}(\beta_0) \to_P 0$. Finally, it is easy to verify that the matrices $\Sigma_{nk}, k = 0,1,2,3$, satisfy $|\Sigma_{nk}(\beta) - \Sigma_{nk}(\beta_0)| \leq |\beta - \beta'|O_P(1)$, which completes the proof of the lemma.

Acknowledgement

We thank an anonymous reviewer and Editor Jane Ling Wang for comments. Research was supported by grants from the National Science Foundation and National Cancer Institute. The data presented here were obtained from the Statistical Center of the International Blood and Marrow Transplant Registry. The analysis has not been reviewed or approved by the Advisory Committee of the IBMTR.

References

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

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **10**, 1100-1120.

Arjas, E. and Eerola, M. (1993). On predictive causality in longitudinal studies. *J. Statist. Plann. Inference* **34**, 361-384.

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

Bickel, P. J., Klaassen, C., Ritov, Y. and Wellner, J. A. (1993). *Efficient and Adaptive Estimation in Transformation Models.* Johns Hopkins University Press.

Cox, D. R. (1973). The statistical analysis of dependencies in point processes. In *Symposium on Point Processes* (Edited by Lewis, P. A. W.). Wiley, New York.

Dabrowska, D. M., Sun, G. W. and Horowitz, M. M. (1994). Cox regression in a Markov renewal model: an application to the analysis of bone marrow transplant data. *J. Amer. Statist. Assoc.* **89**, 867-877.

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

de la Peña, V. and Giné, E. (1999). *Decoupling: From Dependence to Independence.* Springer Verlag.

Einmahl, U. and Mason, D. (2000). An empirical process approach to uniform consistency of kernel-type function estimators. *J. Theoret. Probab.* **13**, 1-37.

Gill, R. D. (1980). Nonparametric estimation based on censored observations of a Markov renewal process. *Z. Wahrscheinlichkeitstheorie verv. Gebiete* **53**, 97-116.

Giné, E. and Guillou, A. (1999) Laws of iterated logarithm for censored data. *Ann. Probab.* **27**, 2042-2067.

Jacod, J. (1975). Multivariate point processes: predictable projection, Radon-Nikodym derivatives, representation of martingales. *Z. Wahrscheinlichkeitstheorie verv. Gebiete* **31**, 235-254.

Klein, J. P., Keiding, N. and Copelan, E. A. (1993). Plotting summary predictions in multistate survival models: probabilities of relapse and death in remission for bone marrow transplant patients. *Statist. Medicine* **12**, 2315-2332.

Lagakos, S. W., Sommer, C. J. and Zelen, M. (1978). Semi-Markov models for partially censored data. *Biometrika* **65**, 311-317.

Müller, H. G. and Wang, J. L. (1994). Hazard rate estimation under random censoring with varying kernels and bandwidths. *Biometrics* **50**, 61-76.

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

Nolan, D. and Pollard, D. (1987). U-processes: rates of convergence. *Ann. Statist.* **15**, 780-799.

Oakes, D. and Cui, L. (1994). On semi-parametric inference for modulated renewal processes. *Biometrika* **81**, 83-90.

Pakes, A. and Pollard, D. (1989). Simulation and the asymptotics of the optimization estimators. *Econometrica* **57**, 1027-1057.

Phelan, M. J. (1990). Bayes estimation from a Markov renewal process. *Ann. Statist.* **18**, 603-616.

Pons, 0. and Visser, M. (2000). A non-stationary Cox model. *Scand. J. Statist.* **27**, 619-641.

Sasieni, P. (1992) Information bounds for conditional hazard ratio in a nested family of regression models. *J. Roy. Statist. Soc. Ser. B* **54**, 617-635.

Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence of Empirical Processes with Applications.* Springer Verlag.

Voelkel, J. G. and Crowley, J. J. (1984). Nonparametric inference for a class of semi-Markov processes with censored observations. *Ann. Statist.* **12**, 142-160.

Department of Biostatistics, University of California, Los Angeles, CA 90095-1772, U.S.A.

E-mail: dorota@ucla.edu

SPSS Inc., 233 South Wacker Drive, 11th Floor, Chicago, IL 60606, U.S.A.

E-mail: who@spss.com

(Received March 2004; accepted October 2004)