Statistica Sinica **19** (2009), 981-996

ESTIMATION IN A SEMI-PARAMETRIC TWO-STAGE RENEWAL REGRESSION MODEL

Dorota M. Dabrowska

University of California, Los Angeles

Abstract: We discuss estimation in a two-stage modulated renewal process designed for analysis of paired data on a single individual. We consider a model with proportional hazard intensities. U-process methods are used to show consistency and asymptotic normality of the parameter estimates.

Key words and phrases: Modulated renewal process, proportional hazard model, U-statistics.

1. Introduction

The proportional hazard model remains the most commonly used semi-parametric regression model for analysis of univariate failure time data. The model assumes that the cumulative hazard function $A(t|z)$ of the conditional distribution of a failure time T given a vector of covariates is of the form

$$A(t|z) = A_0(t)e^{\beta^T z}, \tag{1.1}$$

where $A_0(t)$ is an unknown baseline cumulative hazard function and β is a vector of regression coefficients. Several authors have suggested multivariate extensions of this model. In particular, Wei, Lin and Weissfeld (1989) proposed the use of marginal modeling of multivariate failure time data. The approach assumes that the marginal conditional distributions of a failure time vector $T = (T_1, \ldots, T_k)$ satisfy the model (1.1), but the joint conditional distribution is left unspecified. To account for possible dependence among the components of T, Wei, Lin and Weissfeld (1989) and Spiekerman and Lin (1998) developed appropriate modifications of the usual profile likelihood method designed for univariate data. In particular, they showed that that the estimates of the regression coefficients and baseline cumulative hazards are asymptotically normally distributed, and developed robust estimates of their asymptotic covariances. The approach gained much popularity in analyses of multivariate clustered data, in part because in many circumstances only parameters of marginal distributions are of interest. Results of Wei, Lin and Weissfeld (1989) and Spiekerman and Lin (1998) allow

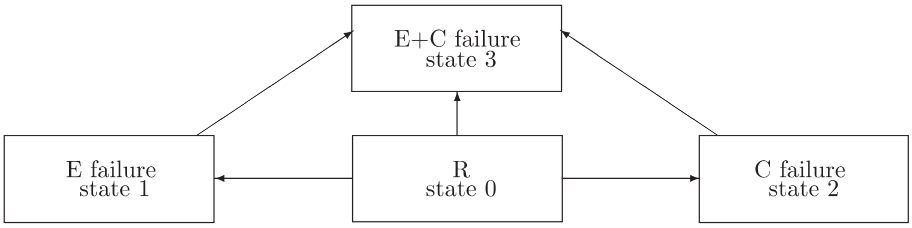

Figure 1. Transition diagram: R = randomization state, E = failure of the organ receiving experimental treatment, C = failure of the organ receiving conventional treatment, E+C = failure of both organs.

for consistent estimation of these parameters without specification of the joint distribution of the vector T.

Models for bivariate failure times can also be defined using marked point processes (Andersen, Borgan, Gill and Keiding (1993) and Hougaard (2000)). Among them, Markov chains and semi-Markov processes provide the simplest example. Both processes can be specified by assuming that transition rates satisfy the proportional hazards model assumption. The transition rates can be further combined to obtain the transition probability matrices of the corresponding processes and, as a by–product, the joint and marginal distribution functions of the failure times. However, the marginal distributions fail to satisfy the proportional hazards model assumption and typically share in common parameters which account for the joint dependence structure of the failure times. These two features of the marginal distributions entail that the estimation approach taken in Wei, Lin and Weissfeld (1989) does not apply in this setting.

In this paper, we consider estimation in a semi-parametric two-stage modulated renewal process (Cox (1975)) which can be used in matched pair experiments on paired organs of a single subject, such as eyes, hands, or kidneys. The model assumes that the two organs are randomized (R) to an experimental (E) and conventional (C) treatment. Failure of both organs (E+C) may be simultaneous or preceded by failure of one of the two organs. The schematic diagram of possible transitions is shown in Figure 1 below.

In Section 2 we define the model to have proportional hazard intensities with general risk functions, and also allow for presence of monotone censoring dependent on the state occupied by the process. In Section 3 we discuss estimation of the Euclidean component of the model and the baseline cumulative hazard function. In analogy to standard Cox regression, for purposes of estimation of the the Euclidean component of the model, we consider solution of the profile likelihood score equation and show that, under mild regularity conditions, the resulting estimates are asymptotically normally distributed and have asymptotic covariance structure similar to standard Cox regression. However, as noted by Oakes

(1981), the usual continuous-time martinagle methods for analysis of stochastic integrals and partial likelihoods fail to apply because time measurements are recorded on a non-chronological time scale. The relevant processes are also not adapted to a common continuous time filtration. Methods for asymptotic analysis of general modulated renewal processes were developed in Oakes and Cui (1994), Dabrowska (1995), and Chang, Hsiung and Wu (1999), among others. Here we use U-process methods to show consistency and asymptotic normality of the estimates under weaker conditions than in these papers.

2. The Model

Let $\mathcal{H}$ be the set of possible one-step transitions corresponding to the multistage model shown in Figure 1. Thus $\mathcal{H} = \mathcal{H}_0 \cup \mathcal{H}_1 \cup \mathcal{H}_2$, where

$$\mathcal{H}_0 = \{(0,i) : i = 1,2,3\} \quad \text{and} \quad \mathcal{H}_i = \{(i,3)\}, \quad i = 1,2. \tag{2.1}$$

We consider a multivariate counting process $\tilde{N} = \{\tilde{N}_h(t) : h \in \mathcal{H}, t \geq 0\}$ given by

$$\begin{aligned} \tilde{N}_h(t) &= 1(\tilde{X}_1 \leq t,\ J = i) && \text{if} \quad h = (0,i),\ i = 1,2,3,\\ &= 1(\tilde{X}_1 < \tilde{X}_2 \leq t,\ J = i) && \text{if} \quad h = (i,3),\ i = 1,2. \end{aligned}$$

Here $\tilde{X}_1$ is the time of the occurrence of the first event and J is its type. The variable $\tilde{X}_2$ represents the time of the occurrence of the second event, and can be observed only if $J \neq 3$. We assume that the cumulative intensity of the process $\tilde{N}$ with respect to the self-exciting filtration is given by $\Lambda = \{\Lambda_h : h \in \mathcal{H}\}$, where

$$\begin{aligned} \Lambda_h(t) &= \int_0^t 1(\tilde{X}_1 \geq u)\alpha_h(u)du && \text{if} \quad h \in \mathcal{H}_0,\\ &= \int_0^t 1(\tilde{X}_2 \geq u > \tilde{X}_1,\ J = i)\alpha_h(u - \tilde{X}_1)du && \text{if} \quad h \in \mathcal{H}_i,\ i = 1,2. \end{aligned}$$

The functions α_h, $h \in \mathcal{H}$ assume the form

$$\alpha_h(u) = q_h(Z_h, u, \theta)e^{r(Z,u,\beta)}\alpha(u), \tag{2.2}$$

where $Z_h = Z$ if $h \in \mathcal{H}_0$ and $Z_h = (X_1, Z)$ if $h \in \mathcal{H}_i, i = 1,2$. In addition,

$$\sum_{h \in \mathcal{H}_0} q_h(Z_h, u, \theta) = 1, \tag{2.3}$$

and $q_h > 0$ if $h \in \mathcal{H}_i$. For $h \in \mathcal{H}_0$, we assume that either $q_h > 0$ for all $h \in \mathcal{H}_0$ or $q_h \equiv 0$ for $h = (0,3)$ and $q_h > 0$ for $h \neq (0,3)$.

The relative risk function $\exp r(Z,u,\beta)$ serves to describe covariate effects which do not depend on the transition type. Simple examples of it may correspond to the choice of $r(Z,u,\beta) = \beta^T Z$ used in standard Cox regression, or relative risk functions considered by Prentice and Self (1983). The function $r(Z,u,\beta)$ may also depend on time (u) and with this choice we can accommodate external time dependent covariates (Andersen et al. (1993)), such as $r(Z,u,\beta) = \beta^T Z f(u)$, where $f(u)$ is a known function or a function dependent on a Euclidean parameter. Other choices include the parametric partial Cox regression model.

The functions $q_h(Z_h,u,\theta)$, $h \in \mathcal{H}$ serve to describe transition-specific effects of covariates. In the case of transitions originating from state 0, we use $Z_h = Z$ and assume that the functions q_h, $h \in \mathcal{H}$, satisfy the constraint (2.3). Multinomial regression models, such as the multinomial logistic or probit models can be used to incorporate this constraint. On the other hand, in the case of transitions originating from states $i = 1,2$, we allow the transition effects to depend also on the length of the sojourn time $\tilde{X}_1$ in state 0.

If both the r and q_h functions depend only on the covariate Z, then (2.2)–(2.3) corresponds to a semi-Markov model. Special cases of it include the semi-Markov extensions of the models of Freund (1961) and Marshall and Olkin (1967). The choice

$$q_{0i}(Z_{0i},u,\theta) = \frac{\exp[\theta_{0i}Z + \theta'_{0i}u]}{1+\sum_{j=1}^{2}\exp[\theta_{0j}Z + \theta'_{0j}u]}, \quad i = 1,2,$$

$q_{03}(Z_{03},u,\theta) = 1 - \Sigma_{j=1}^{2} q_{0j}(Z_{0j},u,\theta)$, and $q_{i3}(Z_{i3},u,\theta) = \exp[\theta_{i3}^T Z + \theta'_{i3}(\tilde{X}_1 + u)]$, $i = 1,2$, provides the example of Wold's process discussed in Cox (1975) and Oakes and Cui (1994). As opposed to semi-Markov models, the transition rates depend in this case both on the sojourn time in each state and calendar time.

We further note that setting

$$A_h(x|z) = \int_0^x q_h(Z_h,u,\theta)e^{r(Z,u,\beta)}\alpha(u)du, \quad h \in \mathcal{H}_0,$$

the survival function of the sojourm time $\tilde{X}_1$, is given by

$$S_0(x|z) = P(\tilde{X}_1 > x|Z = z) = \exp[-\sum_{h\in\mathcal{H}_0} A_h(x|z)] = \exp[-\int_0^x e^{r(Z,u,\beta)}\alpha(u)du],$$

and $P(J = i|\tilde{X}_1 = x, Z = z) = q_{0i}(Z,x,\theta)$ for $i = 1,2,3$. The conditional survival function in state i, $i = 1,2$ is given by

$$S_i(x|x_1,z) = P(\tilde{X}_2 - \tilde{X}_1 > x|Z = z, \tilde{X}_1 = x_1, J = i) = \exp[-A_{i3}(x|x_1,z)],$$

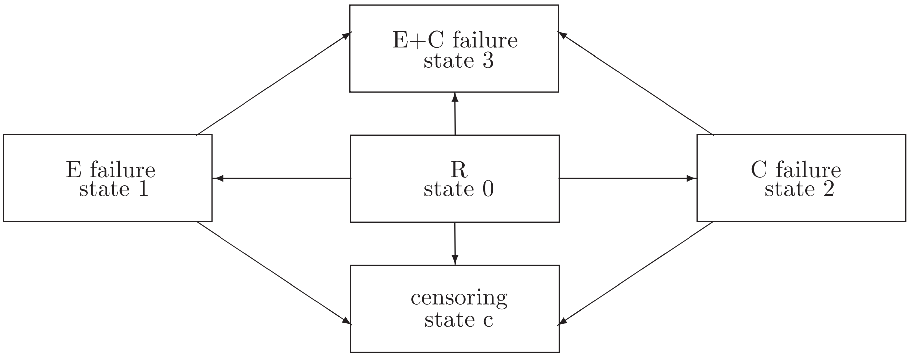

Figure 2. Transition diagram of the censored model. The states R, E, C and E+C are defined as in Figure 1. L = loss-to-follow-up.

where

$$A_{i3}(x|x_1, z) = \int_0^x q_{i3}(Z_{i3}, u, \theta)e^{r(Z,u,\beta)}\alpha(u)du.$$

These two displays can further be combined to obtain the joint and marginal conditional distribution functions of $(\tilde{X}_1, \tilde{X}_2)$ (Dabrowska. and Lee (1996)). However, the marginals do not follow the proportional hazard model.

We assume now that the process is subject to censoring and the censored process can be represented using the transition diagram shown in Figure 2. Thus the observable variables are given by a vector W with entries

$$\begin{aligned} W &= (X_1, J, X_2, \delta_1, \delta_2, Z) && \text{if} \quad J \in \{1,2\} \text{ and } \delta_1 = 1, \\ &= (X_1, J, \delta_1, Z), && \text{if} \quad J = 3 \text{ and } \delta_1 = 1, \\ &= (X_1, \delta_1, Z), && \text{if} \quad \delta_1 = 0, \end{aligned} \tag{2.4}$$

where $X_i = \tilde{X}_i \wedge C_i$ and $\delta_i = 1(X_i = \tilde{X}_i)$, $i = 1, 2$. We further make the following assumptions on the censoring process and the covariate.

Condition 2.1. (i) *The marginal distribution μ of the covariate Z is non-degenerate*; (ii) *conditionally on Z, C_1 and $(\tilde{X}_1, J)$ are independent, and* (iii) *for $i = 1, 2$, $C_2 \geq C_1 \vee \tilde{X}_1$ a.s., and (C_1, C_2) and $\tilde{X}_2$ are conditionally independent given $(Z, \tilde{X}_1)1(\tilde{X}_1 \leq C_1, J_1 = i)$.*

For the sake of convenience, denote by $\mathcal{H}^c$ the collection of possible pairs of one-step transitions in the censored model. Thus $\mathcal{H}^c = \mathcal{H}_0^c \cup \mathcal{H}_1^c \cup \mathcal{H}_2^c$ where

$$\mathcal{H}_0^c = \{(0, j) : j = 1, 2, 3, c\}, \quad \mathcal{H}_1^c = \{(1, 3), (1, c)\}, \quad \mathcal{H}_2^c = \{(2, 3), (2, c)\}\ .$$

Define processes

$$N_h(x) = 1(X_1 \leq x, J = i, \delta_1 = 1) \qquad \text{if} \quad h = (0, i),\ i = 1, 2, 3,$$

$$\begin{aligned}
&= 1(X_1 \le x, \delta_1 = 0) && \text{if} \quad h = (0, c),\\
&= 1(X_2 - X_1 \le x, J = i, \delta_1 = 1, \delta_2 = 1) && \text{if} \quad h = (i, 3),\ i = 1, 2,\\
&= 1(X_2 - X_1 \le x, J = i, \delta_1 = 1, \delta_2 = 0) && \text{if} \quad h = (i, c),
\end{aligned}$$

and set $Y_h(x) = 1(X_1 \ge x)$ for $h \in \mathcal{H}_0^c$, $Y_h(x) = 1(X_2 - X_1 \ge x,\ J_1 = i, \delta_1 = 1)$ for $h \in \mathcal{H}_i^c$, $i = 1, 2$. Finally, let

$$M_h(x) = N_h(x) - \int_0^x Y_h(u) A_h(du|Z_h), h \in \mathcal{H}^c.$$

If $\phi(W) = \{\phi_h(X_1, Z_h),\ \phi_{h'}(X_2, Z_{h'}) : h \in \mathcal{H}_0^c, h' \in \mathcal{H}_j^c, j = 1, 2\}$ is a vector of measurable functions, then the the processes

$$\int_0^x \phi_h(u, Z) M_h(du), \quad x \ge 0,\ h \in \mathcal{H}_0^c,$$

form orthogonal martingales with respect to the filtration $\mathcal{F}_{0x} = \sigma(N_h(u), Z, Y_h(u+) : u \le x, h \in \mathcal{H}_0^c)$, and similarly, the processes

$$\int_0^x \phi_h(u, Z, X_1) M_h(du), \quad x \ge 0,\ h \in \mathcal{H}_i^c,\ i = 1, 2,$$

form orthogonal martingales with respect to the filtration $\mathcal{F}_{ix} = \sigma(N_h(u), Y_h(u+), Z1(\delta_1 = 1, J_1 = i), X_1 1(\delta_1 = 1, J_1 = i) : u \le x, h \in \mathcal{H}_i^c)$. Using direct calculation, it is also easy to verify that the processes are orthogonal. Note, however, that processes originating from the state i, $i = 1, 2$ and processes originating from the state 0 are not adapted to a common filtration.

3. Estimation

We now assume that we have a sample of size n of independent identically distributed (i.i.d.) observations from the censored renewal model, and consider estimation of the parameters (β, θ) and the baseline cumulative hazard function A.

For the sake of convenience, we assume that the functions $r(Z, u, \beta)$ and $q_h(Z_h, u, \theta)$ do not share parameters in common and are differentiable with respect to β and θ, respectively. We also assume that the derivatives $\dot{r}(Z, u, \beta)$ and $\dot{q}_h(Z_h, u, \theta)$ satisfy a certain form of Lipschitz continuity. To avoid cumbersome notation, this Lipschitz continuity assumption is stated in the Appendix. Set $\phi_h(Z_h, u, \xi) = [\phi_{1h}(Z_h, u, \xi), \phi_{2h}(Z_h, u, \xi)]^T$, where

$$\begin{aligned}
\phi_{1h}(Z_h, u, \xi) &= \dot{r}(Z, u, \beta),\\
\phi_{2h}(Z_h, u, \xi) &= \frac{\dot{q}_h}{q_h}(Z_h, u, \theta), \quad h \in \mathcal{H}.
\end{aligned}$$

For $i = 1, \ldots, n$, define processes $S^{(0)}_{hi}(u,\xi) = Y_{hi}(u)e^{r(Z_i,u,\beta)}q_h(Z_{hi},u,\theta)$, $S^{(1)}_{hi}(u,\xi) = \phi_h(Z_{hi},u,\xi)S^{(0)}_{hi}(u,\xi)$ and $S^{(2)}_{hi}(u,\xi) = \phi_h(Z_{hi},u,\xi)^{\otimes 2}S^{(0)}_{hi}(u,\xi)$. Let $S^{(p)}(u,\xi) = n^{-1}\Sigma_i\Sigma_h S^{(p)}_{hi}(u,\xi)$. For $p = 0,1,2$, let $s^{(p)}(u,\xi) = \mathbb{E}\, S^{(p)}(u,\xi)$. We assume that these expectations are finite in a neighborhood of the true parameter ξ_0, and denote by $\Sigma(\xi_0)$ the matrix

$$\Sigma(\xi_0) = \sum_h \int_0^\tau \left[\frac{s^{(2)}}{s^{(0)}}(u,\xi_0) - \left(\frac{s^{(1)}}{s^{(0)}}(u,\xi_0)\right)^{\otimes 2}\right] EN_{hi}(du). \tag{3.1}$$

To estimate the parameter ξ, we use solution to the score equation $\Phi_n(\xi) = o_P(n^{-1/2})$, where

$$\Phi_n(\xi) = \frac{1}{n}\sum_{i=1}^n \sum_h \int_0^\tau [\phi_h(Z_{hi},u,\xi) - \frac{S^{(1)}}{S^{(0)}}(u,\xi)] N_{hi}(du). \tag{3.2}$$

Define also

$$\hat{\Sigma}_n(\xi) = \frac{1}{n}\sum_{i=1}^n \sum_h \int \left[\frac{S^{(2)}}{S^{(0)}}(u,\xi) - \left(\frac{S^{(1)}}{S^{(0)}}(u,\xi)\right)^{\otimes 2}\right] N_{hi}(du). \tag{3.3}$$

Proposition 3.1. *Suppose that the Conditions* 2.1, A.1, A.2 *and* A.3 *hold, and that the matrix* (3.1) *is finite and nonsigular. With probability tending to* 1, *the score equation* $\Phi_n(\xi) = o_P(n^{-1/2})$ *has a root* $\hat{\xi}$ *in the ball* $\mathcal{B}(\xi_0,\epsilon_n)$. *Moreover,* $\sqrt{n}[\hat{\xi} - \xi_0]$ *is asymptotically* $N(0,\Sigma(\xi_0)^{-1})$. *The matrix* $\Sigma(\xi_0)$ *can be estimated consistently by* $\hat{\Sigma}_n(\hat{\xi})$.

We also have a similar asymptotic normality result for the baseline cumulative hazard function A. The estimate is defined by $\hat{A}(x,\hat{\xi})$, where

$$\hat{A}(x,\xi) = \frac{1}{n}\int_0^x \frac{\sum_i \sum_h N_{hi}(du)}{S^{(0)}(u,\xi)} \tag{3.4}$$

is the weighted Nelson-Aalen estimator.

Proposition 3.2. *Suppose that the assumptions of Proposition* 3.1 *and Condition* A.4 *hold. Then the process*

$$\hat{W}(x) = \sqrt{n}[\hat{A}(x,\hat{\xi}) - A(x)] + \sqrt{n}[\hat{\xi} - \xi_0]\int_0^x \frac{S^{(1)}}{S^{(0)}}(u,\hat{\xi})\hat{A}(du,\hat{\xi})$$

converges weakly in $\ell^\infty[0,\tau]$ *to a time transformed Brownian motion with variance function*

$$C(x) = \int_0^x \frac{A(du)}{s^{(0)}(u,\xi_0)}.$$

Moreover, $\hat{W}$ *and* $\sqrt{n}[\hat{\xi} - \xi_0]$ *are asymptotically independent.*

In analogy to the standard Cox regression model, the score function $\Phi_n(\xi)$ is equal to the derivative of the log-profile likelihood obtained by replacing the unknown cumulative hazard function by the weighted Nelson-Aalen estimator (3.4). In he supplement (`http://www.stat.sinica.edu.tw/statistica`), we give a generalization of Propositions 3.1 and 3.2 in Nan, Edmond and Wellner (2004) to verify that $\Sigma(\xi_0)$ forms the information matrix for the parameter ξ_0. Thus, if $\hat{\xi}$ is a solution to the score equation , then the matrix $\hat{\Sigma}_n(\hat{\xi})$ provides an estimate of the Fisher information for the Euclidean parameter of the model. In addition, $\hat{\xi}$ is asymptotically efficient and any other regular estimator of this parameter is asymptotically at least as dispersed as $\hat{\xi}$.

Under regularity conditions assumed in the Appendix, the functions $\phi_h(Z_h, u, \xi)$ are only Lipschitz continuous with respect to ξ. If the functions ϕ_h are differentiable with respect to ξ, then the estimate can be obtained using the conjugate gradient or quasi-Newton algorithm with finite difference or user supplied gradient. Note however, that if $\hat{\xi}_0$ is an arbitrary $\sqrt{n}$-consistent estimator of the parameter ξ, then the one-step estimate $\hat{\xi} = \hat{\xi}_0 + \hat{\Sigma}_n(\hat{\xi}_0)^{-1}\Phi_n(\hat{\xi}_0)$ is asymptotically efficient, so that construction of the estimate $\hat{\xi}$ can be implemented using the Fisher scoring algorithm without differentiation of the score function. Alternatively, the estimate can be obtained using the direct search polytope or simplex algorithms which do not require differentiability of the ϕ_h functions. Standard numerical packages, such as Matlab, IMSL and R provide options for the above mentioned algorithms.

Appendix

Section A.1 collects notation and some results from the theory of U-statistics. They are used to show Propositions 3.1 and 3.2 in Sections A.2 and A.3, respectively.

A.1. Preliminaries

Let $W_1, \ldots, W_n$ be i.i.d. random variables with some distribution P. An (asymmetric) U-statistics with kernel $g(W_1, \ldots, W_m)$ is denoted by

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

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 g satisfies $\mathbb{E}\,|g(W_1, \ldots, W_m)| < \infty$, its Hoeffding projection of degree m is denoted by $\pi_m[g](W_1, \ldots, W_m)$. We have

$$\pi_m[g](W_1, \ldots, W_m) = g(W_1, \ldots, W_m) + \sum_{A\subset\{1,\ldots,m\}} (-1)^{|A|}\mathbb{E}_A g(W_1, \ldots, W_m),$$

where for $\emptyset \neq A = \{i_1, \ldots, i_p\}$, $1 \le p \le m$, $\mathbb{E}_A$ denotes conditional expectation with respect to variables W_j, $j \in A$. Then $\mathbb{U}_{n,m}(\pi_m[g])$ forms a canonical U statistics of degree m.

We now put $\mathbf{W_m} = (W_1, \ldots, W_m)$. The derivation of the asymptotic properties of the estimates uses the following lemma. Its proof can be found in the supplement to this paper at `http://www.stat.sinica.edu.tw/statistica`.

Lemma A.1. *Let $\mathcal{G}$ be a measurable class of P-canonical kernels with envelope $G(\mathbf{W}_m)$, such that $\mathbb{E}\, G^p(\mathbf{W}_m) < \infty$ for some $p \in (1,2)$. Suppose that the class of truncated kernels $\mathcal{G}_n = \{g(\mathbf{W_m})1(G(\mathbf{W}_m) < n^{m/p}) : g \in \mathcal{G}\}$ is Euclidean for the envelope $G(\mathbf{W}_m)1(G(\mathbf{W}_m) < n^{m/p})$. Then*

$$n^{m(p-1)/p} \sup\{|\mathbb{U}_{n,m}(g)| : g \in \mathcal{G}\} \to_P 0. \tag{A.1}$$

Remark A.1. This lemma can be applied also to the U-processes $\mathbb{U}_{n,m}(g_n)$ with kernels g_n varying over a class of functions $\mathcal{G}_n$ dependent on n. In this case, however, we require that the sequence of envelopes $G_n(\mathbf{W}_m)$ be uniformly L^p-integrable.

Remark A.2. If the class $\mathcal{G}$ consists of a single function, then a Marcinkiewicz-Zygmund-type theorem in Teicher (1998) provides a stronger, almost sure convergence result. Unfortunately, we have not been able to extend it to the present setting. On the other hand, if $p \in (0,1)$ and the class $\mathcal{G}$ has an envelope satisfying $\mathbb{E}\, G(\mathbf{W_m})^p < \infty$, then the Marcinkiewicz-Zygmund-type theorem in de la Peña and Giné (1999) implies that (A.1) holds almost surely. If $p = 1$, and the class $\mathcal{G}$ is Euclidean for the envelope $G(\mathbf{W}_m)$ satisfying $\mathbb{E}\, G(\mathbf{W_n}) < \infty$ then the Glivenko-Cantelli Theorem entails that (A.1) holds for U-processes whose kernels have mean zero but are not necessarily P-canonical.

To conclude this section, we denote by

$$\mathbb{V}_{n,m}(g) = \frac{1}{n^m} \sum_{(i_1,\ldots,i_m)} g(W_{i_1}, \ldots, W_{i_m}), \quad g \in \mathcal{G},$$

the V-process corresponding to the kernels g, $g \in \mathcal{G}$. In the following we use V-processes of degree $m \le 4$ and apply Lemma A.1 and Remarks A.1−A.2 to verify that the difference $\sqrt{n}|\mathbb{V}_{n,m}(g) - \mathbb{U}_{n,m}(g)|$ converges to 0 almost surely if g is a fixed function, and in probability if g varies over appropriately chosen classes of functions.

A.2. Proof of Proposition 3.1

The proof of Proposition 3.1 is based on two lemmas. Lemma A.2 shows asymptotic normality of the score process $\Phi_n(\xi_0)$. Lemma A.3 shows consistency and asymptotic normality of the regression estimates.

Condition A.1. Let ϵ_n be a sequence such that $\epsilon_n \sim n^{-\gamma}$, $\gamma \in (0, 1/2)$. Let $B(\xi_0, \epsilon_n) = \{\xi : |\xi - \xi_0| \le \epsilon_n\}$. Then

(i) $\inf\{s^{(0)}(u, \xi) : u \le \tau,\ \xi \in B(\xi_0, \epsilon_n)\} > 0$;

(ii) $\sup\{|S^{(p)}(u, \xi_0) - s^{(p)}(u, \xi_0)| : u \le \tau,\ p = 0, 1, 2\} = o_p(1)$;

(iii) there exist functions $\psi_{1h}(Z_h, u), \psi_{2h}(Z_h, u)$, $h \in \mathcal{H}$ such that for $\xi \in \mathcal{B}(\xi_0, \epsilon_n)$ and $h \in \mathcal{H}$, we have $|\phi_h(Z_h, u, \xi) - \phi_h(Z_h, u, \xi_0)| \le |\xi - \xi_0|\psi_{1h}(Z_h, u)$ and $|[\phi_h b_h](Z_h, u, \xi) - [\phi_h b_h](Z_h, u, \xi_0)| \le |\xi - \xi_0|\psi_{2h}(Z_h, u)$, where $b_h(Z_h, u, \xi) = q_h(Z_h, u, \theta)e^{r(Z,u,\beta)}$;

(iv) for $u \le \tau$, we have $\mathbb{E}\, Y_h(u) g_h(Z_h, u) < \infty$ and

$$\sup\left\{\left|\frac{1}{n}\sum_{i=1}^{n} Y_{hi}(u) g_h(Z_{hi}, u) - \mathbb{E}\, Y_h(u) g_h(Z_h, u)\right| : u \le \tau,\ h \in \mathcal{H}\right\} = o_p(1),$$

where $g_h(Z_h, u) = \psi_{2h}(Z_h, u), [\psi_{h1}\psi_{h2}](Z_h, u), \psi_{h1}(Z_h, u)[b\phi_h](Z_h, u, \xi_0)$ or $g_h(Z_h, u) = \psi_{h2}(Z_h, u)\phi_h(Z_h, u, \xi_0)$.

To simplify notation, we now set $\phi_{hi}(u, \xi) = \phi_h(Z_{hi}, u, \xi)$ for $i = 1, \ldots, n$. Define $\bar{S}_i^{(1)}(u, \xi) = \Sigma_h |\phi_{hi}|(u, \xi) S_{hi}^{(0)}(u, \xi)$ and let $\bar{s}^{(1)}(u, \xi) = \mathbb{E}\, \bar{S}_i^{(1)}(u, \xi)$. In the following we assume that these expectations are finite in a neighborhood of ξ_0. Define

$$\bar{f}_1(W_i, W_j, t) = \sum_h \int_0^t \frac{|\phi_{ih}| S_j^{(0)} + \bar{S}_j^{(1)}}{s^{(0)}}(u, \xi_0) N_{hi}(du),$$

and set $\bar{g}^{(2)}(W_i, W_j) = \bar{f}_1(W_i, W_j, \tau)$,

$$\bar{g}^{(3)}(W_i, W_j, W_k) = \int_0^\tau \frac{S_k^{(0)}}{s^{(0)}}(u, \xi_0)\bar{f}_1(W_i, W_j, du), \tag{A.2}$$

$$\bar{g}^{(4)}(W_i, W_j, W_k, W_l) = \int_0^\tau \left(\frac{S_k^{(0)}}{s^{(0)}}\right)\left(\frac{S_\ell^{(0)}}{s^{(0)}}\right)(u, \xi_0)\bar{f}_1(W_i, W_j, du).$$

We assume the following moment condition.

Condition A.2. $\mathbb{E}\, |\bar{g}^{(2)}(W_i, W_j, \tau)|^{r_2(i,j)} < \infty$, $\mathbb{E}\, |\bar{g}^{(3)}(W_i, W_j, W_k)|^{r_3(i,j,k)} < \infty$ and $\mathbb{E}\, |\bar{g}^{(4)}(W_i, W_j, W_k, W_l)|^{r_4(i,j,k,l)} < \infty$, where $r_m(i_1, \ldots, i_m) = 2d_m(i_1, \ldots, i_m)/(2m-1)$, and $d_m(i_1, \ldots, i_m)$ is the number of distinct indices among $(i_1, \ldots, i_m)$, $m = 2, 3, 4$.

Lemma A.2. *If conditions* A.1 *and* A.2 *are satisfied and the matrix* $\Sigma_0(\xi_0)$ *is non-singular, then* $\sqrt{n}\Phi_n(\xi_0) \implies \mathcal{N}(0, \Sigma(\xi_0))$.

Proof. Define

$$g(W_i, W_j, t) = \sum_h \int_0^t \frac{\phi_{ih} S_j^{(0)} - S_j^{(1)}}{s^{(0)}}(u, \xi_0) N_{hi}(du),$$

and set $g^{(2)}(W_i, W_j) = g(W_i, W_j, \tau)$,

$$g^{(3)}(W_i, W_j, W_k) = -\int_0^\tau \frac{S_k^{(0)} - s^{(0)}}{s^{(0)}}(u, \xi_0) g(W_i, W_j, du),$$

$$g^{(4)}(W_i, W_j, W_k, W_l) = \int_0^\tau \left(\frac{S_k^{(0)} - s^{(0)}}{s^{(0)}}\right)\left(S_\ell^{(0)} - s^{(0)} s^{(0)}\right)(u, \xi_0) \bar{f}_1(W_i, W_j, du).$$

The score function satisfies $\Phi_n(\xi_0) = \sum_{j=2}^4 \Phi_{n,j}(\xi_0)$, where $\Phi_{n,2}(\xi_0) = \mathbb{V}_{n,2}(g^{(2)})$, $\Phi_{n,3}(\xi_0) = \mathbb{V}_{n,3}(g^{(3)})$ and

$$\Phi_{n,4}(\xi_0) = \frac{1}{n}\sum_{i=1}^n \sum_h \int_0^\tau \left[\frac{\phi_{ih} S^{(0)} - S^{(1)}}{s^{(0)}} \frac{(S^{(0)} - s^{(0)})^2}{s^{(0)} S^{(0)}}\right](u, \xi_0) N_{hi}(du).$$

It is easy to verify that for $i \neq j$, we have $\mathbb{E}\, g^{(2)}(W_i, W_j) = 0$, and

$$\mathbb{U}_{n,2}(g^{(2)}) = \frac{1}{n}\sum_i \sum_h \int_0^\tau \frac{\phi_{ih} s^{(0)} - s^{(1)}}{s^{(0)}}(u, \xi_0) M_{ih}(du) + \mathbb{U}_{2,n}(\pi_2[g^{(2)}]).$$

The first term is a sum of i.i.d. variables with mean zero and covariance $\Sigma(\xi_0)$. The second term is a canonical U statistic of degree 2. The moment condition A.2 and Remark A.2 imply that $\mathbb{E}\,|\pi_2[g^{(2)}](W_i, W_j)|^{4/3} < \infty$ so that $\sqrt{n}\mathbb{U}_{n,2}(\pi_2[g^{(2)}]) \to 0$ a.s. By the Central Limit Theorem and Slutzky's Theorem, we have $\sqrt{n}\mathbb{U}_{n,2}(g^{(2)}) \implies \mathcal{N}(0, \Sigma(\xi_0))$. Further, for any sequence (i, j, k, ℓ) of distinct indices, the kernels $g^{(3)}(W_i, W_j, W_k)$ and $g^{(4)}(W_i, W_j, W_k, W_\ell))$ have mean zero. Moreover, their Hoeffding expansion is given by

$$\mathbb{U}_{n,3}(g^{(3)}) = \mathbb{U}_{n,2}(\mathbb{E}_{\{13\}} g^{(3)}) + \mathbb{U}_{n,2}(\mathbb{E}_{\{23\}} g^{(3)}) + \mathbb{U}_{n,3}(\pi_3[g^{(3)}]),$$
$$\mathbb{U}_{n,4}(g^{(4)}) = \mathbb{U}_{n,2}(\mathbb{E}_{\{34\}} g^{(4)}) + \sum_{A=\{134\},\{234\}} \mathbb{U}_{n,3}(\pi_3[\mathbb{E}_A g^{(4)}]) + \mathbb{U}_{n,4}(\pi_4[g^{(4)}])$$

and, in both cases, the Condition A.2 implies that $\sqrt{n}\mathbb{U}_{n,3}(g^{(3)}) \to 0$ a.s. and $\sqrt{n}\mathbb{U}_{n,4}(g^{(4)}) \to 0$ a.s. To complete the proof, we note that Conditions A.2 and Remark A.2 imply that $\sqrt{n}|\mathbb{V}_{n,p}(g^{(p)}) - \mathbb{U}_{n,p}(g^{(p)})| \to 0$ a.s. for $p = 2, 3, 4$. Finally, for any $\epsilon > 0$,

$$\begin{aligned} P(\sqrt{n}|\Phi_{n,4}(\xi_0)| > \epsilon) &\le P\left(\sqrt{n}|\Phi_{n,4}(\xi_0)| > \epsilon,\ \sup_u \frac{s^{(0)}}{S^{(0)}}(u, \xi_0) \le 1 + \epsilon\right) \\ &\quad + P\left(\sqrt{n}|\Phi_{n,4}(\xi_0)| > \epsilon,\ \sup_u \frac{s^{(0)}}{S^{(0)}}(u, \xi_0) > 1 + \epsilon\right) \\ &\le P\left(\sqrt{n}\mathbb{V}_{n,4}(g^{(4)}) > \frac{\epsilon}{(1+\epsilon)}\right) + P\left(\sup_u \left|\frac{S^{(0)}}{s^{(0)}}(u, \xi_0) - 1\right| > \frac{\epsilon}{(1+\epsilon)}\right), \end{aligned}$$

and both terms converge to 0.

We now turn to showing that the equation $\Phi_n(\xi) = o_P(n^{-1/2})$ has a consistent solution. For this purpose, let

$$\bar{f}_{11}(W_i, W_j, t) = \sum_h \int_0^t \frac{[\psi_{1h}(Z_{hi}, u) + \psi_{1h}(Z_{hj}, u)]S_j^{(0)}(u, \xi_0)}{s^{(0)}(u, \xi_0)} N_{hi}(du).$$

The functions $\bar{g}^{(12)}$, $\bar{g}^{(13)}$ and $\bar{g}^{(14)}$ are given by (A.2) with the integrator $\bar{f}_1$ replaced by $\bar{f}_{11}$.

Condition A.3.

$$\begin{aligned} &\mathbb{E}\,[\mathbb{E}\,g^{(12)}(W_i, W_j)|W_i]^{r_1(i)} < \infty, \quad &&\mathbb{E}\,[\mathbb{E}\,g^{(12)}(W_i, W_j)|W_j]^{r_1(j)} < \infty,\\ &\mathbb{E}\,|\bar{g}^{(12)}(W_i, W_j)|^{r_2(i,j)} < \infty, &&\mathbb{E}\,|\bar{g}^{(13)}(W_i, W_j, W_k)|^{r_3(i,j,k)} < \infty,\\ &\mathbb{E}\,|\bar{g}^{(14)}(W_i, W_j, W_k, W_l)|^{r_4(i,j,k,l)} < \infty, \end{aligned}$$

where $r_m(i_1, \ldots, i_m) = 2d_m(i_1, \ldots, i_m)/(2m-1+2\gamma)$, $d_m(i_1, \ldots, i_m)$ is the number of distinct indices among $(i_1, \ldots, i_m)$ and $m = 1, 2, 3, 4$.

Lemma A.3. *Under the assumed regularity conditions, the matrix* $\hat{\Sigma}_n(\xi)$ *defined by* (3.3) *satisfies (i)* $\hat{\Sigma}_n(\xi_0) \to_P \Sigma(\xi_0)$; (ii) $\Phi_n(\xi) - \Phi(\xi_0) = (\xi - \xi_0)^T \hat{\Sigma}_n(\xi_0) + \mathrm{rem}(\xi)$, *where*

$$\sup\left\{\frac{|\mathrm{rem}(\xi) - \mathrm{rem}(\xi_0)|}{n^{-1/2} + |\xi - \xi_0|} : \xi \in \mathcal{B}(\xi_0, \epsilon_n)\right\} = o_P(1);$$

and (iii) $\sup\{|\hat{\Sigma}_n(\xi) - \hat{\Sigma}_n(\xi_0)| : \xi \in B(\xi_0, \epsilon_n)\} \to_P 0$.

By Lemma 5.1 in Dabrowska (2007), these three conditions and asymptotic normality of $\sqrt{n}\Phi_n(\xi_0)$ imply that, with probability tending to 1, the score equation $\Phi_n(\xi) = o_P(n^{-1/2})$ has a root $\hat{\xi}$ in $\mathcal{B}(\xi_0, \epsilon_n)$ and $\sqrt{n}[\hat{\xi} - \xi_0]$ is asymptotically normal with mean zero and covariance $\Sigma(\xi_0)^{-1}$.

Proof. Part (i) and (iii) follows easily from application of Conditions A.1. For $\xi \in \mathcal{B}(\xi_0, \epsilon_n)$, let $\tilde{\phi}_{hi}(u, \xi) = \phi_{hi}(u, \xi) - \phi_{hi}(u, \xi_0)$, and set $\tilde{S}^{(1)}(u, \xi) = n^{-1}\Sigma_i \Sigma_h \tilde{\phi}_{hi}(u, \xi) S_{hi}(u, \xi_0)$. The Condition A.1 and elementary algebra imply that, for $\xi \in \mathcal{B}(\xi_0, \epsilon_n)$, we have $\mathrm{rem}(\xi) - \mathrm{rem}(\xi_0) = \mathrm{rem}_1(\xi) + o_P(|\xi - \xi_0|^2)$, where

$$\mathrm{rem}_1(\xi) = \frac{1}{n}\sum_{i=1}^n \sum_h \int_0^\tau \left[\tilde{\phi}_{hi}(u, \xi) - \frac{\tilde{S}^{(1)}(u, \xi)}{S^{(0)}(u, \xi_0)}\right] N_{ih}(du).$$

We have $\mathrm{rem}_1(\xi_0) = 0$ and part (ii) follows if we show

$$\sup\left\{\sqrt{n}|\mathrm{rem}_1(\xi)| : \xi \in \mathcal{B}(\xi_0, \epsilon_n)\right\} = o_p(1). \tag{A.3}$$

Define

$$f_\xi(W_i, W_j, t) = \sum_h \int_0^t \frac{\tilde{\phi}_{ih}(u,\xi) S_{hj}^{(0)}(u,\xi_0) - \tilde{S}_{hj}^{(1)}(u,\xi)}{s^{(0)}(u,\xi_0)} N_{hi}(du)$$

and set $f_\xi^{(2)}(W_i, W_j) = f_\xi(W_i, W_j, \tau)$,

$$f_\xi^{(3)}(W_i, W_j, W_k) = -\int_0^\tau \frac{S_k^{(0)} - s^{(0)}}{s^{(0)}}(u,\xi_0) f_\xi(W_i, W_j, du),$$

$$f^{(4)}(W_i, W_j, W_k, W_l) = \int_0^\tau \left(\frac{S_k^{(0)} - s^{(0)}}{s^{(0)}}\right)\left(\frac{S_\ell^{(0)} - s^{(0)}}{s^{(0)}}\right)(u,\xi_0) f_{11}(W_i, W_j, du).$$

Here the last term does not depend on ξ. We have $\mathrm{rem}_1(\xi) = \sum_{j=2}^4 \mathrm{rem}_{1j}(\xi)$, where $\mathrm{rem}_{12}(\xi) = \mathbb{V}_{n,2}(f_\xi^{(2)})$, $\mathrm{rem}_{13}(\xi) = \mathbb{V}_{n,3}(f_\xi^{(3)})$ and

$$\mathrm{rem}_{14}(\xi) = \frac{1}{n^2}\sum_{i,j} \int_0^\tau \frac{(S^{(0)} - s^{(0)})^2}{s^{(0)} S^{(0)}}(u,\xi_0) f_\xi(W_i, W_j, du).$$

The U statistic $\mathbb{U}_{n,4}(f^{(4)})$ has a similar Hoeffding decomposition as $\mathbb{U}_{n,4}(g^{(4)})$ in Lemma A.2. The assumed moment conditions and Remark A.2 imply $n^{1/2-\gamma}$ $\mathbb{U}_{n,4}(g^{(4)}) \to 0$ a.s. and $n^{1/2-\gamma}\mathbb{V}_{n,4}(f^{(4)}) \to 0$ a.s. For any $\epsilon > 0$, $P(\sqrt{n}\sup |\mathrm{rem}_{n14}(\xi)| > \epsilon)$ is bounded by

$$P\left(\sqrt{n}\epsilon_n \mathbb{V}_{n,4}(f^{(4)}) > \frac{\epsilon}{(1+\epsilon)}\right) + P\left(\sup_u \left|\left[\frac{S^{(0)}}{s^{(0)}}\right](u,\xi_0) - 1\right| > \frac{\epsilon}{(1+\epsilon)}\right).$$

The right hand side tends to zero because $\epsilon_n \sim n^{-\gamma}$. It follows that the term $\mathrm{rem}_{14}(\xi)$ satisfies (A.2).

Now consider the statistic $\mathbb{U}_{n,2}(f_\xi^{(2)})$. For $\xi \in \mathcal{B}(\xi_0, \epsilon_n)$, we have

$$\begin{aligned}\mathbb{U}_{n,2}(f_\xi^{(2)}) &= \mathbb{U}_{n,1}(\mathbb{E}_{\{1\}} f_\xi^{(2)}) + \mathbb{U}_{n,1}(\mathbb{E}_{\{2\}} f_\xi^{(2)}) + \mathbb{U}_{n,2}(\pi_2[f_\xi^{(2)}]) \\ &= \frac{1}{n}\sum_i \sum_h \int_0^\tau \left[\tilde{\phi}_{ih}(u,\xi) - \frac{\tilde{s}^{(1)}(u,\xi)}{s^{(0)}(u,\xi_0)}\right] M_{ih}(du) + \mathbb{U}_{n,2}\Big(\pi_2\Big[f_\xi^{(2)}\Big]\Big).\end{aligned} \tag{A.4}$$

The class of functions $\mathcal{F}_n = \{f(W_i) = \mathbb{E}\, f_\xi^{(2)}(W_i, W_j | W_i) : \xi \in \mathcal{B}(\xi_0, \epsilon_n)\}$ has envelope $F_n(W_i) = \epsilon_n F(W_i)$, $F(W_i) = \mathbb{E}\, \bar{g}^{(12)}(W_i, W_j)|W_i$. The class of truncated functions $\{f(W_i)1(F(W_i) \le n^\alpha) : f \in \mathcal{F}_n\}$ is Euclidean for the envelope $F_n(W_i) = \epsilon_n[F(W_i)1(F(W_i) \le n^\alpha) + \mathbb{E}\, F(W_i)1(F(W_i) \le n^\alpha)]$, where $\alpha = 1/p$, $p = 2/(1+2\gamma)$. Since $\epsilon_n \sim n^{-\gamma}$ and $\mathbb{E}\, F(W_i)^p < \infty$, it follows that the first term of (A.4) satisfies Condition A.3. A similar argument can be applied to the remaining to terms of (A.4). In particular, in the case of the last term, Lemma A.1 is applied to the class $\mathcal{F}_n = \{f_\xi(W_i, W_j) = \pi_2[f_\xi^{(2)}](W_i, W_j) : \xi \in \mathcal{B}(\xi_0, \epsilon_0)\}$ with envelope $F_n(W_i, W_j) = \epsilon_n F(W_i, W_j)$, $F(W_i, W_j) = \bar{g}^{(12)}(W_i, W_j) + \mathbb{E}_{\{1\}}\bar{g}^{(12)}(W_i, W_j) + \mathbb{E}_{\{2\}}\bar{g}^{(12)}(W_i, W_j) + \mathbb{E}\,\bar{g}^{(12)}(W_i, W_j)$. The term $\mathbb{V}_{n,3}(f_\xi^{(3)})$ can be handled in an analogous fashion.

A.3. Proof of Proposition 3.2

Set $N_{.i} = \Sigma_h N_{hi}$ and define

$$H_1 = 2C(\tau) + 8\mathbb{E}\left[\int_0^\tau \frac{S_i}{s^{(0)}}(u, \xi_0)A(du)\right]^2,$$

$$H_{22}(W_i, W_j) = \int_0^\tau \frac{S_j^{(0)}}{[s^{(0)}]^2}(u, \xi_0)N_{.i}(du),$$

$$H_{23}(W_i, W_j, W_k) = \int_0^\tau \frac{S_j^{(0)}}{s^{(0)}}\frac{S_k^{(0)}}{s^{(0)}}N_{.i}(du).$$

Condition A.4. $\mathbb{E}\,H_{22}(W_i, W_j)^{r_2(i,j)} < \infty$ and $\mathbb{E}\,H_{23}(W_i, W_j, W_k)^{r_3(i,j,k)} < \infty$, where $r_m(i_1, i_2, \ldots, i_m) = 2d_m(i_1, i_2, \ldots, i_m)/(2m-1)$ for $m = 2, 3$ and $d_m(i_1, i_2, \ldots, i_m)$ is the number of distinct indices among $(i_1, \ldots, i_m)$.

We first show that that the process $\{\sqrt{n}I_n(t), t \le \tau\}$,

$$I_n(t) = \frac{1}{n}\int_0^t \frac{\sum_i N_{.i}(du)}{S^{(0)}(u, \xi_0)} - A(t)$$

converges weakly in $\ell^\infty([0, \tau])$ to a time-transformed Brownian motion with variance function $C(t)$ defined in the statement of Proposition 3.2. Set

$$g_t^{(1)}(W_i) = \int_0^t \frac{N_{.i}(du)}{s^{(0)}(u, \xi_0)} - \int_0^t \frac{S_i^{(0)}}{s^{(0)}}(u, \xi_0)\frac{EN_{.i}(du)}{s^{(0)}(u, \xi_0)},$$

$$g_t^{(2)}(W_i, W_j) = -\int_0^t \frac{S_j^{(0)} - s^{(0)}}{[s^{(0)}]^2}(u, \xi_0)[N_{.i} - \mathbb{E}\,N_{.i}](du),$$

$$g^{(3)}(W_i, W_j, W_k) = \int_0^\tau \left(\frac{S_j^{(0)}}{s^{(0)}} - 1\right)\left(\frac{S_k^{(0)}}{s^{(0)}} - 1\right)(u, \xi_0)\frac{N_{.i}(du)}{s^{(0)}(u, \xi_0)}.$$

Then $I_n(t) = \Sigma_{j=1}^3 I_{n,j}(t)$, where $I_{n,1}(t) = n^{-1}\Sigma_i g_t^{(1)}(W_i)$, $I_{n,2}(t) = \mathbb{V}_{n,2}(g_t^{(2)})$ and

$$I_{n,3}(t) = \frac{1}{n}\int_0^t \frac{(S^{(0)} - s^{(0)})^2}{s^{(0)}S^{(0)}}(u, \xi_0)\sum_{i=1}^n N_{.i}(du).$$

The process $\sqrt{n}I_{n,1}(t)$ has uncorrelated increments and $n\text{var}I_{n,1}(t)=C(t)$. Hence its finite dimensional distributions converge weakly to the finite-dimensional distributions of a time-transformed Brownian motion with variance function $C(t)$. Further, the process $I_{n,1}(t)$ can be represented as a difference of two monotone functions (in t) and has envelope

$$H_{n;1}(W_i) = \int_0^\tau \frac{M_{.i}(du)}{s^{(0)}(u,\xi_0)} + 2\int_0^\tau \frac{S_i^{(0)}(u,\xi_0)A(du)}{s^{(0)}(u,\xi_0)}.$$

Using $(a+b)^2 \le 2a^2 + 2b^2$, we have $\mathbb{E}\, H_{n;1}^2(W_i) \le H_1 < \infty$. Therefore equicontinuity of the process $\{\sqrt{n}I_{n,1}(t),\, t \le \tau\}$ follows since $\{I_{n,1}(t),\, t \le \tau\}$ forms an empirical process over a Euclidean class of functions with a square integrable envelope.

Further, set

$$H(W_i,W_j) = \int_0^\tau \frac{S_j^{(0)} + s^{(0)}}{[s^{(0)}]^2}(u,\xi_0)[N_{.i} + \mathbb{E}\, N_{.i}](du)$$

and $I_{n,2}(t) = (n-1)n^{-1}\mathbb{U}_{n,2}(g_t^{(2)}) + n^{-2}\Sigma_{i=1}^n g_t^{(2)}(W_i,W_i)$. The first term is a U-process over the class of functions $\mathcal{G} = \{g_t^{(2)} : t \le \tau\}$ and consists of canonical kernels with envelope $H(W_i,W_j)$ satisfying $\mathbb{E}\, H(W_i,W_j)^p$, $p = 4/3$. The class of truncated kernels $\{g_t^{(2)}1(H < n^{2/p}) : g \in \mathcal{G}\}$ is Euclidean for the envelope $H1(H < n^{2/p})$ since each function $g_t^{(2)}1(H < n^{2/p})$ can be written as a linear combination of four monotone functions bounded by $H1(H < n^{2/p})$. Hence by Lemma A.1, we have $\sqrt{n}\sup\{|\mathbb{U}_{n,2}(g_t^{(2)})| : t \le \tau\} = o_P(1)$. Application of Remark A.2 shows also that $\sqrt{n}\sup\{|\mathbb{V}_{n,2}(g_t^{(2)})| : t \le \tau\} = o_P(1)$.

Remark A.2, Condition A.4 and similar algebra as in Lemma A.2 show that $\sqrt{n}\mathbb{V}_{n,3}(g^{(3)}) \to 0$ a.s. For any $\epsilon > 0$, $P(\sup_t \sqrt{n}I_{n3}(t) > \epsilon)$ is bounded by $P(\sqrt{n}\mathbb{V}_{n,3}(g^{(3)}) > \epsilon/(1+\epsilon)) + P(\sup_u |[S^{(0)}/s^{(0)}](u,\xi_0) - 1| > \epsilon/(1+\epsilon)) \to 0$. Taylor expansion also yields $A_n(t,\hat{\xi}) - A(t,\xi_0) = [A_n - A](t,\xi_0) - (\hat{\xi}-\xi_0)\dot{A}_n(t,\xi^*)$, where ξ^* is on the line segment between $\hat{\xi}$ and ξ_0 and

$$\dot{A}_n(t,\xi) = -n^{-1}\int_0^t \left[\frac{S^{(1)}}{(S^{(0)})^2}\right](u,\xi)\sum_{i=1}^n N_{.i}(du).$$

The condition A.1 implies that $\sup\{|S^{(p)} - s^{(p)}(u,\xi)| : \xi \in \mathcal{B}(\xi_0,\epsilon_n),\ u \le \tau,\ p = 0,1\} = o_P(1)$ and, using this, it is easy to see that $\sup\{|\dot{A}_n(t,\xi_0) - \dot{A}(t,\xi_0)| : t \le \tau\} = o_P(1)$ and $\sup\{|\dot{A}_n(t,\xi) - \dot{A}_n(t,\xi_0)| : t \le \tau,\ \xi \in B(\xi_0,\epsilon_n)\} = o_P(1)$.

Acknowledgement

I thank reviewers and editors for their comments. Research supported by the NIAID grant 5R01AI67943.

References

Andersen, P. K., Borgan, 0., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer Verlag.

Chang, I.-S., Hsiung, C. A. and Wu, S. M. (1999). Estimation in a proportional hazard model for semi-Markov counting process. *Statist. Sinica* **9** 1257-1266.

Cox, D. R. (1975). The statistical analysis of dependencies in point processes. In *Symposium on Point Processes* (Edited by P. A. W. Lewis). Wiley, New York.

Dabrowska, D. M. (1995). Estimation of transition probabilities and bootstrap in a semi-parametric Markov renewal model. *J. Nonparametr. Stat.* **7**, 75-103.

Dabrowska, D. M. and Lee, W. (1996). Nonparametric estimation of transition probabilities in a two-stage duration model. *J. Nonparametr. Stat.*, **7**, 75-103.

Dabrowska, D. M. (2007). Information bounds and efficient estimation in a class of censored transformation models. *Acta Appl. Math.* **96**,177-201.

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

Freund, J. E. (1961). A bivariate extension of the exponential distribution. *J. Amer. Statist. Assoc.* **56**, 971-977.

Hougaard, P. (2000). *Analysis of Multivariate Survival Data.* Springer.

Marshall, A. W. and Olkin, I. (1967). A multivariate exponential distribution. *J. Amer. Statist. Assoc.* **62**, 30-44.

Nan, B., Edmond, M. and Wellner, J. A. (2004). Information bounds for Cox regression with missing data. *Appl. Statist.* **32**, 723-753.

Oakes, D. (1981). Survival analysis: aspects of partial likelihood (with discussion). *Internat. Statist. Rev.* **49**, 235-264.

Oakes, D. and Cui, L. (1994). On semiparametric inference for modulated renewal processes. *Biometrika* **91**, 83-91.

Prentice, R. L. and Self, S. G. (1983). Asymptotic distribution theory for Cox-type regression models with general relative risk form. *Ann. Statist.* **11**, 804-813.

Spiekerman, C. F. and Lin, D. Y. (1998). Marginal regression models for multivariate failure time data. *J. Amer. Statist. Assoc.* **93**, 1164-1175.

Teicher, H. (1998). On the Marcinkiewicz-Zygmund strong law for U-statistics. *J. Theoret. Probab.* **11**, 279-288.

Wei, L. J., Lin, D. Y. and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J. Amer. Statist. Assoc.* **84**, 1068-1073.

Department of Biostatistcs, University of California. Los Angeles, CA 90095-1772, USA.

E-mail: dorota@ucla.edu

(Received January 2007; accepted May 2008)