$$h_{n\theta}^{(r)}(j,k) = \frac{\sum_i \sum_m 1(R_{im} = k) 1((J_{im}, J_{im+1}) = j)}{\sum_i \sum_m 1(R_{im} \geq k, J_{im} = j_1) \alpha_j (g_{n\theta}^{(r-1)}(\cdot, k-1, \theta, Z_{im})}$$

for $j = (j_1, j_2) \in \mathscr{J}_0$. The indicators $1(T_k(\omega_0) \leq x)$ are jointly measurable with respect to $\mathscr{S}_n \otimes \mathscr{B}(\mathscr{T})$ and by Lemma 5.1, so are the weights $h_{n\theta}$ and $g_{n\theta}$. Therefore the graph of $H_{n\ell}$ is $\mathscr{S}_n \otimes \mathscr{B}(\mathscr{T})$ is measurable and

$$\{(\omega_0, x, \theta) : \Gamma_{n\theta}(\omega_0, x) \in B\} = \text{graph} H_n = \bigcup_{l \geq 0} \text{graph}(H_{n\ell}) \in \mathscr{S}_n \otimes \mathscr{B}(\mathscr{T}).$$

A similar argument can be used to show measurability of the process $\dot{\Gamma}_{n\theta}$ in part (ii). Using arguments analogous to Dabrowska (2006), $\|\Gamma_{n\theta_0+h_n} - \Gamma_{n\theta_0} - h_n\dot{\Gamma}_{n\theta_0}\| = O_P(\|h_n\|_1^2)$ and $\|\dot{\Gamma}_{n\theta_0+h_n} - \dot{\Gamma}_{n\theta_0}\| = O_P(\|h_n\|_1) = o_P(1)$ for any deterministic sequence $h_n \to 0$ or a random $\mathscr{S}_n^P$- measurable sequence $h_n \to_P 0$. Therefore if $\widehat{\theta}$ is an $\mathscr{S}_n^P$- measurable $\sqrt{n}$- consistent estimator of θ_0, then setting $h_n = \widehat{\theta} - \theta_0$, we have $\widehat{W}_0(x) = \widehat{W}(x, \theta_0) + \text{rem}_n(x)$, where $\text{rem}_n = \sqrt{n}[\Gamma_{n\widehat{\theta}} - \Gamma_{n\theta_0} - (\widehat{\theta} - \theta_0)\dot{\Gamma}_{n\widehat{\theta}}] = o_P(1)$. For non-measurable h_n and $\widehat{\theta}_n$, convergence is in outer probability. □

Let us assume now that $f_j(y, \theta, z), j \in \mathscr{J}_0$ is a scalar Carathéodory integrands such that $|f_j(y, \theta, z)| \leq \widetilde{\psi}(\|y\|_1)$ and $|f_j(y, \theta', z) - f_j(y', \theta', z)| \leq [|\theta - \theta'| + \|y - y'\|_1] \max(\widetilde{\psi}'(\|y\|_1), \widetilde{\psi}'(\|y'\|_1)$, where $\widetilde{\psi} = \psi, \psi_1, \psi_2$ and $\widetilde{\psi}' = \psi_3$ satisfy conditions 5.1. Put $S_j[f_j](u, \theta) = n^{-1}\Sigma_{i=1}^n S_{j.i}[f_j](u, \theta)$, where $S_{j.i}[f_j](u, \theta) = \sum_m Y_{jmi}(u)(f_j\alpha_j)(\Gamma_\theta(u), \theta, Z_{jmi})$, and let $s_j[f_j] = ES_j[f_j]$. We write $S_j[1]$ and $s_j[1]$ when $f_j \equiv 1$, and set $\widehat{e}_j[f_j] = S_j[f_j]/S_j[1]$ and $e_j[f_j] = s_j[f_j]/s_j[1]$.

Lemma 5.3 We have $\|S_j[f_j]/s_j[1] - s_j[f_j]/s_j[1]\| \to 0$ a.s. for all $j \in \mathscr{J}_0$.

Proof. We have $([S_j[f_j]/s_j[1]])(x, \theta) = \mathbb{P}_n g_{x\theta}$, where

$$g_{x\theta}(D_i) = \frac{\sum_m Y_{jmi}(x)(f_j\alpha_j)(\Gamma_\theta(x), \theta, Z_{jmi})}{E\sum_m Y_{jmi}(x)\alpha_j(\Gamma_\theta(x), \theta, Z_{jmi})}.$$

The conditions 5.1 imply that there exist constants C_1 and C_2 (dependent on the functions $\widetilde{\psi}, \widetilde{\psi}'$) such that

$$|g_{x\theta}(D_i) - g_{x'\theta}(D_i)| \leq C_1[|Y_{j.i}(x') - Y_{j.i}(x)| +$$
$$Y_{j.i}(0)(|EN_{j.i}(x) - EN_{j.i}(x')| + |EY_{j.i}(x) - EY_{j.i}(x')|)],$$
$$|g_{x\theta}(D_i) - g_{x\theta'}(D_i)| \leq |\theta - \theta'|C_2 Y_{j.i}(0)[1 + EN_{j.i}(\tau) + EY_{j.i}(0)].$$

Define $G(D_i) = Y_{j.i}(0)[C_2\text{diam}\,\Theta + C_1][1 + EN_{j.i}(\tau) + EY_{j.i}(0)] + g_{x_0\theta_0}(D_i)$, where (x_0, θ_0) is an arbitrary point in $\Theta \times [0, \tau]$. Let $\theta_p, p = 1, \ldots, \ell = O(\text{diam}\,\Theta/\varepsilon)^d$ be centers of balls $B(\theta_p, \varepsilon)$ of radius ε covering the set Θ. By noting that $EN_{j.i}$ is an increasing continuous function and $EY_{j.i}$ is a decreasing cáglád function, we can construct a finite partition $0 = x_0 < x_1 < \ldots < x_k = \tau$ such that the intervals $I_r = [x_{r-1}, x_r], r = 1, \ldots, k$ satisfy $EN_{j.i}(I_q) \le \varepsilon EN_{j.i}(\tau)$ and $E|Y_{j.i}(I_r)| \le \varepsilon EY_{j.i}(0)$. Let x_q be the center of the interval I_r. Then for each $x \in I_r$ and $\theta \in B(\theta_p, \varepsilon)$, we have $\|g_{x\theta}(D_i) - g_{x_r\theta_p}(D_i)\|_{P,1} \le \varepsilon \|G(D_i)\|_{P,1}$. It follows that the class of functions $\mathscr{G} = \{g_{x\theta} : x \in [0, \tau], \theta \in \Theta\}$ is Euclidean for the envelope $G(D_i)$ and Glivenko-Cantelli. $\square$

Lemma 5.4 For $j \in \mathscr{J}_0$, define $\text{rem}_j(x, \theta) = [V_{jn} - V_{1jn}](x, \theta)$ and

$$B_j(x, \theta) = \int_0^x [\widehat{e}_j[f_j] - e_j[f_j]](u, \theta) M_{j..}(du, \theta),$$

where f_j satisfies assumptions of Lemma 5.3. Then $\|\sqrt{n}\text{rem}_j\| = o_P(1)$ and $\|\sqrt{n}B_j\| = o_P(1)$.

Proof . For the sake of convenience write rem $=$ rem$_j$ and $B = B_j$. Put $\eta_j(u, \theta) = [S_j/s_j](\Gamma_\theta(u), \theta, u) - 1$. A little algebra shows that

$$\begin{aligned}\text{rem}(x, \theta) &= -\int_0^x \eta_j(u, \theta) \frac{[N_{j..} - EN_{j..}](du)}{s_j[1](u, \theta)} + \int_0^x \eta_j^2(u, \theta) \frac{N_{j..}(du)}{S_j[1](u, \theta)} \\ &= \text{rem}_1(x, \theta) + \text{rem}_2(x, \theta).\end{aligned}$$

We have $\text{rem}_2(x, \theta) = O_P(1)\text{rem}_3(\tau, \theta)$, where

$$\text{rem}_3(x, \theta) = \int_0^x \eta_j^2(u, \theta) \frac{[N_{j..} - EN_{j..}](du)}{s_j[1](u, \theta)} + \int_0^x \eta_j^2(u, \theta) \frac{EN_{j..}(du)}{s_j[1](u, \theta)}.$$

In addition,

$$\begin{aligned}B(x, \theta) = &\int_0^x \left(\frac{S_j[f_j] - S_j[1]e_j[f_j]}{s_j[1]} \right)(u, \theta)[N_{j..} - EN_{j..}](du) \\ &- \int_0^x \left[\left(\frac{S_j[f_j] - s_j[f_j]}{s_j[1]} \right) \eta_j \right](u, \theta)[N_{j..} - EN_{j..}](du) \\ &- \int_0^x \left[\left(\frac{S_j[f_j] - s_j[f_j]}{s_j[1]} \right) \eta_j \right](u, \theta) EN_{j..}(du)\end{aligned}$$

$$+\int_0^x S_j[f_j](u,\theta)\mathrm{rem}_2(du,\theta) = \sum_{p=1}^4 B_p(x,\theta).$$

We have $B_4(x,\theta) = O_P(1)B_5(\theta)$,

$$B_5(\theta) = \int_0^\tau (S_j[|f_j|] - s_j[|f_j|])(u,\theta)\mathrm{rem}_3(du,\theta) + \int_0^\tau s_j[|f_j|](u,\theta)\mathrm{rem}_3(du,\theta).$$

These expressions can be rewritten as V processes of degree $r+1$, $r \leq 3$

$$\mathbb{V}_{n,r+1}(g) = \frac{1}{n^{r+1}}\sum_{\mathbf{i_{r+1}}} g(\mathbf{D}_{\mathbf{i_{r+1}}}), g \in \mathscr{G},$$

where the sum extends over sequences $r+1$-tuplets $\mathbf{D}_{\mathbf{i_{r+1}}} = (D_{i_1},\ldots,D_{i_{r+1}})$ $\mathbf{i}_{\mathbf{r}+\mathbf{1}} = (i_{r_1},\ldots,i_{r+1}), i_j \in 1,\ldots,n$. The kernels g vary over the class $\mathscr{G} = \{g_t : t \in \mathscr{T}\}$, where for $t = (x,\theta)$ we have

$$g_t(\mathbf{D}_{\mathbf{i_{r+1}}}) = \tag{5.4}$$
$$= \int_0^x \prod_{\ell=1}^r [h_\ell(D_{i_\ell},\theta,u) - Eh_\ell(D_{i_\ell},\theta,u)][N_{j.i_{r+1}} - EN_{j.i_{r+1}}](du)$$

or

$$g_t(\mathbf{D}_{\mathbf{i_{r+1}}}) = \int_0^x \prod_{\ell=1}^{r+1}[h_\ell(D_{i_\ell},\theta,u) - Eh_\ell(D_{i_\ell},\theta,u)]EN_{j.}(du). \tag{5.5}$$

Here $h_\ell(D_{i\ell})$ are functions of the form $S_j[f_j]/s_j[1]$, $S_j[1]/s[1]$ and $(\sqrt{s_j[|f_j|]})S_j[1]/s[1]$. In all cases, there exists a constant C such that $h_\ell(D_i,\theta,u) \leq CY_{ji}(u)$ and $|h_\ell(D_i,\theta,u) - h_\ell(D_i,\theta',u)| \leq |\theta-\theta'|CY_{j.i}(u)$. Therefore, for any sequence $\mathbf{D}_{\mathbf{i_{r+1}}} = (D_{i_1},\ldots,D_{i_{r+1}})$, we also have

$$|g_{x\theta} - g_{x'\theta}|(\mathbf{D}_{\mathbf{i_{r+1}}}) \leq |G(\mathbf{D}_{\mathbf{i_{r+1}}},x) - G(\mathbf{D}_{\mathbf{i_{r+1}}},x')|,$$
$$|g_{x\theta} - g_{x\theta'}|(\mathbf{D}_{\mathbf{i_{r+1}}}) \leq |\theta-\theta'|G(\mathbf{D}_{\mathbf{i_{r+1}}},\tau),$$

where

$$G(\mathbf{D}_{\mathbf{i_{r+1}}},x) = \int_0^x \prod_{\ell=1}^r [H_\ell(D_{i_\ell},u) + EH_\ell(D_{i_\ell},u)][N_{j.i_{r+1}} + EN_{j.i_{r+1}}](du)$$

and $H_\ell(D_i,u) = CY_{j.i_\ell}(u), \ell = 1,\ldots,r$ for some constant C.

Let $\{\mathbb{U}_{r+1,n}(g_t) : t \in \mathscr{T}\}$ denote the U process associated with the kernels (5.4-5.5). It is easy to see that $\mathbb{U}_{r+1,n}(g_t)$ forms a canonical process. For

$\mathbf{D_{r+1}} = (D_1, \ldots, D_{r+1})$, we have $EG^p(\mathbf{D_{r+1}}) < \infty$ for $p = 1 + 1/(2r+1)$. Therefore, by Marcinkiewicz-Zygmund law in Teicher (1998) and Lemma A.1 in Dabrowska (2009), $\sqrt{n}\sup_t |\mathbb{U}_{r+1,n}(g_t)| \to_P 0$. By Marcinkiewicz-Zygmund theorem in de la Peña and Giné (1999), we also have $\sqrt{n}\sup_t |\mathbb{V}_{r+1,n}(g_t) - \mathbb{U}_{r+1,n}(g_t)| \to 0$ a.s. because

$$EG(\mathbf{D_{i_{r+1}}})^{2d(i_{r+1})/(2r+1)} < \infty,$$

where $\mathbf{i_{r+1}} = (i_1, \ldots, i_{r+1})$ and $d = d(\mathbf{i_{r+1}})$ is the number of distinct coefficients among $\{i_1, \ldots, i_{r+1}\}$, $d = 1, \ldots, r, r \leq 3$. $\square$

We denote now by $\|B\|_{\mathrm{v}}$ the variation norm of a $d \times q$-matrix of functions $B(x) = [b_{kl}(x)], x \in [0, \tau]$. For any interval $I \subseteq [0, \tau]$, $\|B\|_{\mathrm{v}}(I) = \sup \sum_{i=1}^m \|B(x_j) - B(x_{j-1})\|_1$, where the supremum is taken over finite partitions of I such that $x_i < x_j$.

Further, let $\mathscr{B}(\theta_0, \varepsilon_n)$ be a ball centered at θ_0 of radius $\varepsilon_n, \varepsilon_n \downarrow 0, \sqrt{n}\varepsilon_n \uparrow \infty$. Suppose that $\varphi_\theta(x)$ is a $d \times q$ matrix of functions, with columns of the form $\int_0^x g_{j\theta} d\Gamma_{\theta,j}$ such that $\|\varphi_{\theta_0}\|_{\mathrm{v}} = O(1)$. Let $\varphi_{n\theta}$ be a sequence of consistent estimators such that

- (i) $\varphi_{n\theta}(x)$ is a càdlàg or càglàd function (jointly in (x, θ)), continuous with respect to θ;
- (ii) $\limsup_n \sup\{\|\varphi_{n\theta}\|_{\mathrm{v}} : \theta \in \mathscr{B}(\theta_0, \varepsilon_n)\} = O_P(1)$;
- (iii) $\sup\{\|\varphi_{n\theta} - \varphi_{\theta_0}\|_\infty : \theta \in \mathscr{B}(\theta_0, \varepsilon_n)\} = o_P(1)$ or
- (iii') $\varphi_{n\theta} - \varphi_{n\theta'} = (\theta - \theta')\psi_{n\theta,\theta'}$ where $\limsup_n \sup\{\|\psi_{n\theta\theta'}\|_{\mathrm{v}} : \theta, \theta' \in \mathscr{B}(\theta_0, \varepsilon_n)\} = O_P(1)$.

If $\varphi_{n\theta}$ is a jointly $\mathscr{S}_n^P \otimes \mathscr{B}(\mathscr{T})$ measurable estimator then conditions (ii)-(iii) are assumed to hold in probability. If this is not the case then the conditions (ii)-(iii) are taken to hold in outer probability.

Lemma 5.5 (i) If $\varphi_{n\theta}(x)$ is a measurable process satisfying (i)-(ii) and (iii) or (iii') then with probability tending to 1, the equation $U_{n\varphi_n}(\theta) = 0$ has a consistent root $\widehat{\theta}$ in the ball $\mathscr{B}(\theta_0, \varepsilon_n)$. In addition, under the condition (iii'), the score equation has a unique root in $\mathscr{B}(\theta_0, \varepsilon_n)$, with probability tending to 1.

(ii) If $\varphi_{n\theta}$ is not measurable, then statements in part (1) hold with inner probability tending to 1.

(iii) If $\widetilde{\theta}$ is an arbitrary consistent estimator of θ_0, then the equation $U_{n\widetilde{\varphi}_n}(\theta) = 0$, where $\widetilde{\varphi}_n(x) = \varphi_{n\widetilde{\theta}}(x)$ has a unique solution $\widehat{\theta}$, with (inner) probability tending to 1, and $U_{n\varphi_n}(\widehat{\theta}) = o_{p^*}(n^{-1/2})$.

In all three cases, $\widehat{\Xi} = \sqrt{n}(\widehat{\theta} - \theta_0)$ and the process $\widehat{W}_0 = \{\sqrt{n}[\Gamma_{n\widehat{\theta}} - \Gamma_{\theta_0} - (\widehat{\theta} - \theta_0)^T \dot{\Gamma}_{n\widehat{\theta}}](x) : x \leq \tau\}$ converge weakly in $R^d \times \ell^\infty([0, \tau] \times \mathscr{J}_0)$ to a mean zero Gaussian process defined in the statement of Proposition 3.1.

Proof . Case (1). Write $U_n(\theta) = U_{n\varphi_n}(\theta)$ for short. Set $\widetilde{b}_{jmi}(\Gamma_\theta(u), \theta, u) = = \widetilde{b}_{jmi1}(\Gamma_\theta(u), \theta, u) - \varphi_{\theta_0}(u)\widetilde{b}_{jmi2}(\Gamma_\theta(u), \theta, u)$ where

$$\widetilde{b}_{jmi1}(\Gamma_\theta(u), \theta, u) = \dot{\ell}_j(\Gamma_\theta(u), \theta, Z_{jmi}) - e_j[\dot{\ell}_j](u, \theta),$$
$$\widetilde{b}_{jmi2}(\Gamma_\theta(u), \theta, u) = \ell'_j(\Gamma_\theta(u), \theta, Z_{jmi}) - e_j[\ell'_j](u, \theta).$$

Define $\overline{b}_{jmi}(\Gamma_\theta(u), \theta, u), \overline{b}_{jmi1}(\Gamma_\theta(u)\theta, u)$ and $\overline{b}_{jmi2}(\Gamma_\theta(u), \theta, u)$ using similar expressions with $e_j[\dot{\ell}_j]$ and $e_j[\ell'_j]$ replaced by $\widehat{e}_j[\dot{\ell}_j]$ and $\widehat{e}_j[\ell'_j]$. We have $U_n(\theta) = \sum_{p=1}^4 U_{np}(\theta)$, where

$$U_{1n}(\theta) = \frac{1}{n}\sum_{i=1}^n \sum_j \sum_m \int_0^x \widetilde{b}_{jmi}(\Gamma_\theta(u), \theta, u) M_{jmi}(du, \theta),$$
$$U_{2n}(\theta) = \sum_{q=1}^2 \int_0^\tau r_{nq}(du, \theta)[\Gamma_{n\theta} - \Gamma_\theta]^T(u) = \sum_{q=1}^2 U_{2n;q}(\theta),$$
$$U_{3n}(\theta) =$$
$$-\sum_j \int_0^\tau \left[(\widehat{e}_j[\dot{\ell}_j] - e_j[\dot{\ell}_j])(u, \theta) - \varphi_{\theta_0}(u)(\widehat{e}_j[\ell'_j] - e_j[\ell'_j])(u, \theta)\right] M_{j\cdot\cdot}(du, \theta),$$
$$U_{4n}(\theta) = -\frac{1}{n}\sum_{i=1}^n \sum_j \sum_m \int_0^\tau [\varphi_{n\theta} - \varphi_{\theta_0}](u)]\widehat{b}_{jmi2}(\Gamma_{n\theta}(u), \theta, u) N_{jmi}(du),$$

and

$$r_{n1}(x, \theta) = \frac{1}{n}\sum_{i=1}^n \sum_j \sum_m \int_0^x \overline{b}'_{jmi}(\Gamma_\theta(u), \theta, u) N_{jmi}(du),$$
$$r_{n2}(x, \theta) = \frac{1}{n}\sum_{i=1}^n \sum_j \sum_m \int_0^1 \int_0^x [\overline{b}'_{jmi}(\Gamma^\lambda_{n\theta}(u), \theta, u) N_{jmi}(du) d\lambda - r_{n1}(x, \theta).$$

Here $\Gamma^\lambda_{n\theta} = \Gamma_\theta + \lambda(\Gamma_{n\theta} - \Gamma_\theta)$ for $\lambda \in (0,1)$. We have $U_{2n;2}(\theta_0) = \int_0^\tau O_P(\|\Gamma_{n\theta_0} - \Gamma_{\theta_0}\|^2)\sum_j N_{j\cdot\cdot}(du) = o_P(n^{-1/2})$. Moreover, $r_{1n}(x, \theta_0)$ converges almost surely to

$$r(x, \theta_0) = \sum_j \int_0^x [\mathrm{cov}_j(\ell'_j, \dot{\ell}_j)(u, \theta_0) - \varphi_{\theta_0}(u)\mathrm{cov}_j(\ell'_j, \ell'_j)(u, \theta_0)] EN_{j\cdot\cdot}(du)$$

uniformly in $x, x \leq \tau$. Lemma 5.2 and integration by parts imply that the terms $[\sqrt{n}U_{1n}(\theta_0), \sqrt{n}U_{2n;1}(\theta_0)]$ converge weakly to a pair of independent normal variables with mean zero and covariances $\Sigma_0(\theta_0)$ and $\Sigma_2(\theta_0) - \Sigma_0(\theta_0)$, respectively.

By Lemma 5.3-4, we also have $U_{3n}(\theta_0) = o_P(n^{-1/2})$. Finally,

$$U_{4n}(\theta) = -\sum_{p=1}^{3}\int_0^{\tau}[\varphi_{n\theta} - \varphi_{\theta_0}](u)]B_{pn}(du,\theta) = \sum_{p=1}^{3} U_{4n;p}(\theta),$$

where

$$B_{1n}(x,\theta) = \frac{1}{n}\sum_{i=1}^{n}\sum_j\sum_m\int_0^x[\widehat{b}_{2jmi}(\Gamma_{n\theta}(u),\theta,u) - \widehat{b}_{2jmi}(\Gamma_\theta(u),\theta,u)]N_{jmi}(du),$$

$$B_{2n}(x,\theta) = -\sum_j\int_0^x(\widehat{e}_j[\ell'_j] - e_j[\ell'_j])(u,\theta)M_{j..}(du,\theta),$$

$$B_{3n}(x,\theta) = \frac{1}{n}\sum_{i=1}^{n}\sum_j\sum_m\int_0^x\widetilde{b}_{2jmi}(\Gamma_\theta(u),\theta,u)M_{jmi}(du,\theta).$$

By Lemmas 5.2-5.4, we have $\sqrt{n}U_{4n;2}(\theta) = o_P(1)$ and $\sqrt{n}U_{4n;1}(\theta) = \sum_j\int_0^\tau O_P(\sqrt{n}\|\Gamma_{n\theta} - \Gamma_\theta\|_1(u)\|\varphi_{n\theta} - \varphi_{\theta_0}\|_1(u)N_{j..}(du) = o_P(1)$, uniformly in $\theta \in \mathscr{B}(\theta_0,\varepsilon_n)$. On the other hand, at $\theta = \theta_0$, $\{\sqrt{n}B_{3n}(x,\theta_0) : x \le \tau\}$ is a sum of iid mean zero processes. The finite dimensional distributions are mean zero variables with finite variance-covariance matrix and converge weakly to mean zero Gaussian variables. Each component of $B_{3n}(x,\theta_0)$ is a measurable process which can be represented as a finite linear combination of càdlàg monotone functions of x with a square integrable envelope satisfying (5.2). The same argument as in Lemma 5.2 implies that the process is $\sqrt{n}B_{3n}(x,\theta_0)$ converges weakly to a mean zero Gaussian process with sample paths continuous with respect to the variance semi-metric. The space of functions continuous with respect to the variance semi-metric is isometric to the space $C([0,\tau])^q$. By almost sure representation theorem and a similar integration by parts argument as in Bilias et al (1997) we have $\sqrt{n}U_{4n;3}(\theta_0) = o_P(1)$.

Set $\widehat{U}_n(\theta) = \sum_{j=1}^3 U_{jn}(\theta)$. Some elementary algebra shows that for θ, $\theta' \in \mathscr{B}(\theta_0,\varepsilon_n)$, we have $\widehat{U}_n(\theta) = \widehat{U}_n(\theta') + (\Sigma_n(\theta_0) + \text{rem}_{0n}(\theta,\theta'))(\theta - \theta')$, where $\Sigma_n(\theta_0)$ is a matrix which converges in probability $-\Sigma_1(\theta_0)$. The matrix $\Sigma_1(\theta)$ is defined in Section 3 and is non-singular. Further, $U_{4n}(\theta) - U_{4n}(\theta') = \text{rem}_{2n}(\theta,\theta')$ $(\theta - \theta') + \text{rem}_{3n}(\theta,\theta')$ $+O(|\theta - \theta_0| \vee |\theta' - \theta_0|)\text{rem}_{4n}(\theta,\theta')$. Setting $\text{rem}_{1n}(\theta,\theta')$ $= I + \Sigma_1^{-1}(\theta_0)[\Sigma_n(\theta_0) + \text{rem}_{0n}(\theta,\theta')]$, and $b_{qn} = \sup\{|\text{rem}_{qn}(\theta,\theta')| : \theta, \theta' \in \mathscr{B}(\theta_0,\varepsilon_n)\}$, q = 1, ...,4, we have $b_{1n} = o_P(1)$, $b_{2n} = o_P(1)$. Under the condition (iii'), $\text{rem}_{nq} \equiv 0 \equiv b_{qn}, q = 3,4$, while under the condition (iii), $b_{3n} = o_P(n^{-1/2})$ and $b_{4n} = o_P(1)$.

Put $a_n = b_{1n} + b_{2n} + b_{4n}$ and $A_n = b_{5n} + b_{3n}$, where $b_{5n} = |\Sigma(\theta_0)^{-1}\widehat{U}_n(\theta_0)| = O_P(n^{-1/2})$. Let $0 < \eta < 1/2$ and $0 < \eta' < 1$ be given. By asymptotic tightness of A_n, we can find a compact set $K = K(\eta)$ and n_0 such that for all $n \ge n_0$ and all open sets G containing K, we have $P_n(\sqrt{n}A_n \notin G) < \eta$ and $P_n(a_n > \eta') < \eta$. Therefore, we also have $P_n(\sqrt{n}A_n > M(1-\eta')) < \eta$ for all finite $M \geq M_0$, where $M_0 = M_0(\eta)$ is a large enough finite nonnegative constant. Since $\sqrt{n}\varepsilon_n \uparrow \infty$ and $\varepsilon_n \downarrow 0$, by eventually increasing n_0, we can assume that for $n \geq n_0$, we have $\mathscr{B}(\theta_0, \varepsilon_n) \subset \Theta$ and $M < \sqrt{n}\varepsilon_n$. Consequently, the set $E_n \subset \mathbb{S}_n$ given by $E_n = \{\omega_0 : A_n(\omega_0)/(1-a_n(\omega_0) < \varepsilon_n, a_n(\omega_0) \leq \eta'\}$ satisfies $P_n(E_n) \geq 1-2\eta$ for all $n \geq n_0$.

For $n \geq n_0$, consider the set-valued mapping $H_n : \mathbb{S}_n \hookrightarrow R^d$ given by

$$\begin{aligned} H_n(\omega_0) &= \overline{\mathscr{B}}(\theta_0, \frac{A_n(\omega_0)}{1-a_n(\omega_0)}) = \{\theta : |\theta-\theta_0| \leq \frac{A_n(\omega_0)}{1-a_n(\omega_0}\} \quad \text{if} \quad \omega_0 \in E_n, \\ &= \emptyset \qquad \text{if} \quad \omega_0 \notin E_n. \end{aligned}$$

The graph of H_n, graph$H_n = \{(\omega_0, \theta) : \theta \in H_n(\omega_0)\}$ is $\mathscr{S}_n^P \otimes \mathscr{B}(\Theta)$-measurable and dom$H_n = E_n \in \mathscr{S}_n^P$. Further, let $g_n(\omega_0, \theta) = \theta + \Sigma_1^{-1}(\theta_0)U_n(\omega_0, \theta)$. Then g_n is $\mathscr{S}_n^P \otimes \mathscr{B}(\Theta)$ measurable, because it is continuous with respect to θ for fixed ω_0 and $\mathscr{S}_n^P$-measurable for fixed θ. It follows that the set valued mapping

$$\begin{aligned} C_n(\omega_0) &= \{\theta : g_n(\omega_0, \theta) = 0 \quad \text{and} \quad \theta \in H_n(\omega_0)\} \quad \text{for} \quad \omega_0 \in E_n, \\ &= \emptyset \qquad \text{for} \quad \omega_0 \notin E_n \end{aligned}$$

is closed-valued and has an $\mathscr{S}_n^P \otimes \mathscr{B}(\Theta)$- measurable graph. We have dom$C_n = E_n$: for fixed $\omega_0 \in E_n$, $H_n(\omega_0)$ is a closed ball, $g_n(\omega_0, \theta)$ is continuous and maps $H_n(\omega_0)$ into itself. By Brouwer's fixed point theorem, $C_n(\omega_0) \neq \emptyset$. Thus $E_n \subseteq$ domC_n, while the reversed inclusion is obvious.

Further, for any root $\widehat{\theta}$ in domC_n, we have $|\sqrt{n}(\widehat{\theta}-\theta_0)|^* \leq A_n/(1-a_n) = O_P(1)$, and $\sqrt{n}(\widehat{\theta}-\theta_0) = \Sigma(\theta_0)^{-1}\sqrt{n}\widehat{U}_n(\theta_0) + o_{P^*}(n^{-1/2})$ so that $\sqrt{n}(\widehat{\theta}-\theta_0)$ converges in law to the normal distribution given in Section 3. An argument similar to Bickel *et al.* (1993, p.517) shows also that under the condition (iii'), $g_n(\omega_0, \theta)$ is a contraction on $H_n(\omega_0), \omega_0 \in E_n$, with contraction coefficient $a_n(\omega_0)$. Thus in this case, the root is unique: $C_n(\omega_0) = \{\widehat{\theta}(\omega_0)\}$ for $\omega_0 \in E_n$ and $n \geq n_0$.

Case (2). If $\varphi_{n\theta}$ estimators are not $\mathscr{S}_n^P \otimes \mathscr{B}(\mathscr{T})$ measurable, then the score function splits into two parts: $U_n(\theta) = \widehat{U}_n(\theta) + U_{4n}(\theta)$. The term $\widehat{U}_n(\theta)$ remains $\mathscr{S}_n^P \otimes \mathscr{B}(\Theta)$ measurable, while the second term is not. However, $b_{3n} = o_{p*}(n^{-1/2})$, $a_n = o_{P*}(1)$ while $b_{5n} = |\Sigma(\theta_0)^{-1}\widehat{U}_n(\theta_0)| = O_p(n^{-1/2})$. In this case, the set E_n satisfies $\liminf_n P_{n,*}(E_n) \geq 1-2\eta$ and the closed ball $\overline{\mathscr{B}}(\theta_0, A_n/1-a_n)$ is contained in $\mathscr{B}(\theta_0, \varepsilon_n)$ with inner probability tending to 1.

Case (3). We write $\widetilde{U}_n(\theta)$ for the modified score function obtained by substituting in $\widetilde{\varphi}_n(x) = \varphi_{n\widetilde{\theta}}(x)$ in place of $\varphi_{n\theta}$. Suppose that $\widetilde{\theta}$ is $\mathscr{S}_n^P$-measurable and $\varphi_{n\theta}(x)$ is $\mathscr{S}_n^P \otimes \mathscr{B}(\mathscr{T})$ measurable. Then the plug-in estimator $\varphi_{n\widetilde{\theta}}(x)$ is $\mathscr{S}_n^P \otimes$

$\mathscr{B}([0,\tau])$ measurable and the modified score process $\widetilde{U}_n(\theta)$ is $\mathscr{S}_n^P \otimes \mathscr{B}(\Theta)$ measurable. Moreover, we have $\widetilde{U}_n(\theta) = \widehat{U}_n(\theta) + \widetilde{U}_{n4}(\theta)$, where the remainder $\widetilde{U}_{4n}(\theta)$ satisfies $\sqrt{n}[U_{n4}(\theta) - \widetilde{U}_{n4}(\theta)] = o_P(1+\sqrt{n}|\theta-\theta_0|)$, uniformly in $\theta \in B(\theta_0, \varepsilon_0)$. We also have $\widetilde{U}_{4n}(\theta) - \widetilde{U}_{4n}(\theta') = (\theta-\theta')\widetilde{\mathrm{rem}}_{2n}(\theta,\theta')$, $\sup\{\widetilde{\mathrm{rem}}_{2n}(\theta,\theta') : \theta,\theta' \in \mathscr{B}(\theta_0,\varepsilon_n)\} = o_P(1)$. With probability tending to 1, the modified equation has a unique root $\widehat{\theta}$ in a compact random ball contained in $B(\theta_0,\varepsilon_n)$ and $U_n(\widehat{\theta}) = o_{P^*}(n^{-1/2})$. On the other hand, if either $\widetilde{\theta}$ or $\varphi_{n\theta}$ are not measurable, then this remains to hold, except that the modified equation has a unique solution with inner probability tending to one. □

Under assumptions of part (1), measurable selection theorems (Wagner, 1976) ensure that there exists at least one function $\widehat{\widehat{\theta}} : \mathbb{S}_n \to R^d$ such that $\widehat{\widehat{\theta}}(\omega_0) \in C_n(\omega_0)$ whenever $\omega_0 \in E_n$ and $\widehat{\widehat{\theta}}$ is measurable with respect to $\mathscr{S}_n^p$. This also applies to part (3), provided $\widetilde{\theta}$ and $\varphi_{n\theta}$ are $\mathscr{S}_n^P$- measurable.

5.4 Proof of Proposition 3.2. With some abuse of notation, set $V = [V_j, j \in \mathscr{J}_0]$ where $V(x) = V(x,\theta_0)$ and $V(x,\theta)$ is the Gaussian process of Lemma 5.1. Under the assumption that θ_0 is the true parameter of the modulated renewal process, the process V corresponds to a vector of independent time-transformed Brownian motions with covariance

$$\mathrm{cov}(V_j(x),V_j(y)) = C_j(x\wedge y) \quad \text{and} \quad \mathrm{cov}(V_j(x),V_\ell(y)) = 0 \quad \text{if} \quad j \neq \ell.$$

Similarly, let $\check{V} = [\check{V}_j : j \in \mathscr{J}_0]$ be equal to $\check{V}(x) = \sqrt{n}V_{1n}(x,\theta_0)$ where $V_{1n}(x,\theta)$ is defined as in Lemma 5.1. Thus the j-th component of $\check{V}$ is

$$\check{V}_j(x) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \sum_m \int_0^x \frac{M_{jmi}(du)}{s_j(\Gamma_{\theta_0}(u-),\theta_0,u)}.$$

Put $\check{V}^\# = [\check{V}_j^\# : j = 1,\ldots,q]$,

$$\check{V}_j^\#(x) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \sum_m G_{mi} \int_0^x \frac{N_{jmi}(du)}{s_j(\Gamma_{\theta_0}(u-),\theta_0,u)}.$$

Finally, let G_0 be a $\mathscr{N}(0,I_{d\times d})$ variable, independent of (D_i,G_i)'s. Set $\Xi_1^\# = \Sigma_1^{-1}(\theta_0)\Sigma_0(\theta_0)^{1/2}G_0$ and $\widehat{\Xi}_1^\# = \widehat{\Sigma}_1^{-1}(\widehat{\theta})\widehat{\Sigma}_0(\widehat{\theta})^{1/2}G_0$. We have $E\check{V}_j^\#(x) = 0 = E\check{V}_j(x)$,

$$\begin{aligned}
&\mathrm{cov}(\check{V}_j^\#(x),\check{V}_\ell^\#(x')) = \mathrm{cov}(\check{V}_j(x),\check{V}_\ell(x')) = \delta_{jl}C_{j\theta_0}(x\wedge x'), \qquad (5.6)\\
&\mathrm{cov}(\check{V}_j^\#(x),\check{V}_\ell(x')) = 0.
\end{aligned}$$

Moreover, $\Xi_1^{\#}$ is independent of $D_1,\ldots,D_n$. This also means that it is independent of $(\check{V}^{\#},\check{V})$.

We consider first unconditional weak convergence. By central limit theorem and strong law of large numbers, the finite dimensional distributions of the processes $(\check{V},\check{V}^{\#})$ converge weakly to finite dimensional distributions of $(V,V^{\#})$, two independent vectors of Brownian motions with variance functions C_{j,θ_0}, $j=1,\ldots,q$.

For each $j=1,\ldots,q$, the process $\check{V}_j^{\#}$ can be represented as $\check{V}_j^{\#}(x)=n^{-1/2}\Sigma_{i=1}^n f_x^{(j)}(G_i,D_i)$, where

$$f_x^{(j)}(G_i,D_i)=\sum_m G_{mi}\int_0^x \frac{N_{jmi}(du)}{s_j(\Gamma_{\theta_0}(u-),\theta_0,u)}.$$

The class of functions $\mathscr{F}_j=\{f_x^{(j)}(G_i,D_i):x\in[0,\tau]\}$ has a square integrable envelope

$$F_j(G_i,D_i)=\sum_{m=1}|G_{mi}|\int_0^\tau \frac{N_{jmi}(du)}{s_j(\Gamma_{\theta_0}(u-),\theta_0,u)}$$

and is Euclidean for this envelope because each $f_x^{(j)}\in\mathscr{F}_j$ is a difference of two functions increasing in x and bounded by $F_j(G_i,D_i)$. Thus $\mathscr{F}_j$ forms a Donsker class of functions. The union of these classes, $\mathscr{F}=\bigcup_j\mathscr{F}_j$ is Donsker as well. From Lemma 1, the process $\check{V}=\{\check{V}_j(x):x\in[0,\tau],j\in\mathscr{J}\}$ can be also represented as an empirical process over a Euclidean class of functions $\mathscr{G}$ and the union $\mathscr{F}\cup\mathscr{G}$ forms a Donsker class. Using consistency of the estimates $(\widehat{\theta},\Gamma_{n\widehat{\theta}})$, Lemma 5.5 and a couple of lines integration by parts yields also $\|\widehat{V}^{\#}-\check{V}^{\#}\|=o_P(1)$ in outer probability.

Write $\check{V}^{\#}$ as the empirical process $\check{V}^{\#}=\mathbb{P}_n f$, $f\in\mathscr{F}$. Further, let BL_1 be the collection of Lipschitz functions h from $R^d\times\ell^\infty(\mathscr{F})$ into $[0,1]$, such that $|h(r,w)-h(r',w')|\le|r-r'|+\|w-w'\|$ for $r,r'\in R^d$ and $w,w'\in\ell^\infty(\mathscr{F})$. The set $\mathscr{F}$ is totally bounded with respect to the variance pseudo-metric d. Therefore, for fixed $\delta>0$, it can be covered by a finite number of d-balls of radius δ, say $\mathscr{B}(f_l,\delta)$ $\ell=1,\ldots,k=k(\delta)$. Set $V^{\#}\circ\pi_\delta=\mathbb{P}_n\pi_\delta(f)$, where $\pi_\delta(f)=f_\ell$ for $f\in\mathscr{B}(f_\ell,\delta)$ (pick one f_ℓ for each $f\in\mathscr{F}$). By triangular inequality, we have

$$\sup_{h\in BL_1}|E_G h(\widehat{\Xi}_1^{\#},\widehat{V}^{\#})-Eh(\Xi_1^{\#},V^{\#})|\le\sum_{r=1}^4 I_4(\delta),$$

$$\begin{aligned}
I_1(\delta) &= \sup_{h\in BL_1} |Eh(\Xi_1^\#, V^\# \circ \pi_\delta) - Eh(\Xi_1^\#, V^\#)|, \\
I_2(\delta) &= \sup_{h\in BL_1} |Eh(\Xi_1^\#, V^\# \circ \pi_\delta) - E_G h(\Xi_1^\#, \check{V}^\# \circ \pi_\delta)|, \\
I_3(\delta) &= \sup_{h\in BL_1} |E_G h(\Xi_1^\#, \check{V}^\# \circ \pi_\delta) - E_G h(\Xi_1^\#, \check{V}^\#)|, \\
I_4(\delta) &= \sup_{h\in BL_1} |E_G h(\widehat{\Xi}_1^\#, \widehat{V}^\#) - E_G h(\Xi_1^\#, \check{V}^\#)|.
\end{aligned}$$

For given $\varepsilon > 0$, we can choose δ_0 so that $I_1(\delta) < \varepsilon$ for all $\delta < \delta_0$. The second term converges in outer probability to 0, for any δ. This follows from weak convergence of finite dimensional distributions of $\check{V}^\#$ and the same argument as in Van der Vaart and Wellner (1996, p. 182), except that in our setting, the Lindeberg condition of their Lemma 2.9.5 is not needed to verify conditional weak convergence of finite dimensional distributions. We also have $I_3(\delta) \le E_G^* \|\check{V}^\# \circ \pi_\delta - \check{V}^\#\|_{\mathscr{F}_\delta} \le \Sigma E_G^* \|\check{V}^\#\|_{\mathscr{F}_\delta}$ where $\mathscr{F}_\delta = \{f - f' : f, f' \in \mathscr{F} : d(f - f') < \delta\}$. Since $\mathscr{F}$ forms a Euclidean class of functions with a square integrable envelope, we have $\lim_{\delta\downarrow 0} \limsup_n E^* I_3(\delta) \le \lim_{\delta\downarrow 0} \limsup_n E^* E_G^* \|\check{V}^\#\|_{\mathscr{F}_\delta} = 0$. Finally, the term $I_4(\delta)$ does not depend on δ, and we have $I_4(\delta) \le 2P_G^*(|\widehat{\Xi}_1 - \Xi_1| + \|\widehat{V}^\# - \check{V}^\#\| > \varepsilon) + \varepsilon$. By unconditional convergence, we have $I_4(\delta) \to 0$ in outer probability.

Finally, set $\Psi(\widehat{\Xi}_1^\#, \widehat{V}^\#) = [\check{\Xi}^\#, \check{W}_0^\#]$, where

$$\begin{aligned}
\check{\Xi}^\# &= \widehat{\Xi}_1^\# - \Sigma_1^{-1}(\theta_0) \sum_j \int_0^\tau \rho_{j,\varphi}(u, \theta_0) EN_{j\cdot\cdot}(du) \check{W}_0^\#(u)^T, \\
\check{W}_0^\#(x) &= \int_0^x \widehat{V}^\#(du) \mathscr{P}_{\theta_0}(u, x) \\
&= \widehat{V}^\#(x) - \int_0^x \widehat{V}^\#(u-) Q_{\theta_0}(du) \mathscr{P}_{\theta_0}(u, x).
\end{aligned}$$

The estimates $[\widehat{\Xi}^\#, \widehat{W}_0^\#]$ defined in Section 4 are $[\widehat{\Xi}^\#, \widehat{W}_0^\#] = \widehat{\Psi}(\widehat{\Xi}_1^\#, \widehat{V}^\#)$, where $\widehat{\Psi}$ is the sample analogue of Ψ obtained by plugging in the estimates $\widehat{\mathscr{P}}_{\widehat{\theta}}, \widehat{Q}_{\widehat{\theta}}$, $\rho_{j,\varphi_n}(\cdot, \widehat{\theta}_0)$. By the continuous mapping theorem, unconditionally, $\Psi(\widehat{\Xi}_1^\#, \widehat{V}^\#) \Rightarrow \Psi(\Xi_1^\#, V^\#) = (\Xi^\#, W_0^\#)$. By triangular inequality one more time, we have $\sup_{h\in BL_1} |E_G h(\widehat{\Xi}^\#, \widehat{W}_0^\#) - Eh(\Xi^\#, W_0^\#)| \le J_1 + J_2$, where

$$J_1 = \sup_{h\in BL_1} |E_G h(\widehat{\Xi}^\#, \widehat{W}_0^\#) - E_G h(\check{\Xi}^\#, \check{W}_0^\#)|,$$

$$J_2 = \sup_{h \in BL_1} |E_G h(\check{\Xi}^\#, \check{W}_0^\#) - Eh(\Xi^\#, W_0^\#)|.$$

For any Lipschitz continuous function $h \in BL_1$, $h \circ \Psi \in BL_c$ for some constant c. Therefore the preceding implies that J_2 tends to 0 in outer probability. This also holds for the term J_1, because $\|\check{\Xi}^\# - \widehat{\Xi}^\#\| \to_{P^*} 0$ and $\|\widehat{W}_0^\# - \check{W}_0^\#\|_1 \to_{P^*} 0$, by consistency of the estimates $(\widehat{\theta}, \Gamma_{n\widehat{\theta}})$ and integration by parts. □

References

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

Arjas, E. and Eerola, M. (1993). On predictive causality in longitudinal studies. *J. Statist. Planning and Inference,* **34** 361-386.

Bagdonovicius, V. and Nikulin, M. (1999). Generalized proportional hazards model based on modified partial likelihood. *Lifetime Data Analysis,* **5** 329-350.

Bagdonovicius, M. Hafdi, M. A. and Nikulin, M. (2004). Analysis of survival data with cross-effects of survival functions. *Biostatistics,* **5** 415-425.

Beesack, P. R. (1973). *Gronwall Inequalities.* Carlton Math. Lecture Notes **11** Carlton University, Ottawa.

Bickel, P. J. (1986). Efficient testing in a class of transformation models. In *Proceedings of the 45th Session of the International Statistical Institute* 23.3.63-23.3.81. ISI, Amsterdam.

Bickel, P. J., Klaassen, C. A. J., Ritov, Y. and Wellner, J. A. (1993) Efficient and adaptive estimation in semi-parametric models. Johns Hopkins University Press, Baltimore.

Bickel, P. J. and Ritov, Y. (1995). Local asymptotic normality ranks and covariates in transformation models. In *Festschrift for L. LeCam* (Pollard, D. and Yang, G.,eds). Springer, New York.

Bilias, Y., Gu, M. and Ying, Z. (1997). Towards a general asymptotic theory for Cox model with staggered entry. *Ann. Statist.,* **25** 662-683.

Chang, I-S. and Hsiung, C. A. (1994). Information and asymptotic efficiency in some generalized proportional hazard models for counting processes. *Ann. Statist.,* **22** 1275-1298.

Chang, I-S, Chuang Y-C and Hsiung, C.A. (1999). A class of nonparametric k-sample tests for semi-Markov processes. *Statistica Sinica,* **9** 211-277.

Chang, I-S, Hsiung, C.A. and Wu, S-M. (2000). Estimation in a proportional hazard model for semi-Markov counting process. *Statistica Sinica,* **10** 1257-1266.

Chen, K., Jin, Z. and Ying, Z. (2002). Semi-parametric analysis of transformation models with censored data. *Biometrika*, **89** 659-668.

Chintagunta, P. and Prasad, A.R. (1998). An empirical investigation of the "Dynamic McFadden" model of purchase timing and brand choice: implications for market structure. *J. Business and Economic Statist.*, **16** 2-12.

Cinlar, E. (1975). *Introduction to Stochastic Processes.* Prentice-Hall, New Jersey.

Cook, R. J. and Lawless, J.F. (2007). *The Statistical Analysis of Recurrent Events*. Springer, New York.

Commenges, D. Semi-Markov and non-homogeneous Markov models in medical studies. (1986) . In *Semi-Markov models.* (Edited by J. Janssen) 411-422. Plenum Press, New York, 411-422.

Commenges, D., Joly, P., Gégout-Petit, A. and Liquet, B. (2007). Choice between semi-parametric estimators for Markov and non-Markov multistate models from coarsened observations. *Scand. J. Statist.,* **34** 33-52.

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

Cutler, C. and Antin, J.H. (2001) Peripheral blood stem cells for allogeneic transplantation: a review. *Stem Cells*, **19** 108-117.

Cutler, C., Giri, S., Jeyapalan, S., Paniagua, D., Viswanathan, A., and Antin, J. H. (2001). Acute and chronic graft-versus-host disease after allogeneic peripheral blood stem-cell and bone marrow transplantation: a meta analysis. *J. Clin. Oncol.* **19** 3685-3691.

Dabrowska, D. M., Sun, G. 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. (1995). Estimation of transition probabilities and bootstrap in a semi–parametric Markov renewal model. *J. Nonparametric Statist.*, **5** 237–259.

Dabrowska, D.M. (2006). Estimation in a class of semi-parametric transformation models. In *"Second Erich L. Lehmann Symposium - Optimality"* (J. Rojo, Ed.) Institute of Mathematical Statistics, Lecture Notes and Monograph Series, **49** 166-216.

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

Dabrowska, D.M. (2009). Estimation in a semi-parametric two-stage renewal regression model. *Statistica Sinica*, **19**, 981-996.

Daley, D.J. and Vere-Jones, D. (1988). *An Introduction to the Theory of Point Processes*. Springer, New York.

Dellacherie, C. and Meyer, P.A. (1975) *Probabilities and Potentiel*, Hermann, Paris.

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

Dudley, R. M. (1999). *Uniform Central Limit Theorems*. Cambridge University Press.

Eerola, M. (1994). *Probabilistic causality in longitudinal studies*. Springer, New York.

Flowers, M. E. D., Parker, P.M., Johnston, L.J., Matos, A.V., Storer, B., Bensinger, W. I., Storb, R., Appelbaum, F. R., Forman, S. J., Blume, K. G. and Martin, P. J. (2002). Comparison of chronic graft-versus-host disease after transplantation of peripheral blood stem cells versus bone marrow in allogeneic recipients: long-term follow-up of a randomized trial. *Blood*, **100** 415-419.

Friedrichs, B., Tichelli, A., Bacigalupo, A. Russel,N.H., Ruutu, T., Shapira, Y.M., Beksac, M., Hasenclever, D., Socié, G. and Schmitz, N. (2001) Long-term outcome and late effects in patients transplanted with mobilised blood or bone marrow: a randomised trial. *Lancet Oncology,* **11**, 331-338.

Gale, R. P., Bortin, M. M., van Bekkum, D. W., Biggs, J.C., Dicke, K.A., Gluckman, E., Good, R.A., Hoffman, R.G., Key, H. E. M., Kersey, J.H., Marmont, A., Masaoka, T., Rimm, A.A., van Rood, J.J.and Zwaan, F.E. (1987). Risk factors for acute graft-versus-host disease. *Br. J. Haematol.,*, **67**, 397-406.

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

Gill, R. D. and Johansen, S. (1990). A survey of product integration with a view toward application in survival analysis. *Ann. Statist.,* **18** 1501–1555.

Greenwood, P. and Wefelmeyer W . Empirical estimators for semi-Markov processes. (1996). *Math. Meth. Statist.,* **5** 299-315.

Greenwood, P., Müller, U.U. and Wefelmeyer W . Semi-Markov processes and their applications. (2004). *Commun. Stat. Theory Methods,* **33** 419-435.

Himmelberg, C. J. (1975). Measurable relations. *Fund. Math.* **87** 53-72.

Hjort, N. L. and Cleaskens, G. (2003) Frequentist model average estimators. *J. Amer.Statist.Assoc.,* **98** 938-945.

Hjort, N. L. and Cleaskens, G. (2006) Focused information criteria and model averaging for Cox's hazard regression model. *J. Amer.Statist. Assoc.,* **101** 1449-1464.

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

Janssen, J. (1999) *Semi-Markov Models: Theory and Applications.* Springer, New York.

Janssen, J. and Manca, R. (2007) *Semi-Markov Risk Models for Finance, Insurance and Reliability*. Springer, New York.

Janssen, J. and Manca, R. (2006) *Applied Semi-Markov Processes* Springer, New York.

Janssen, J., Limnios, N. (2001) *International Symposium on Semi-Markov Models: Theory and Applications*. Kluwer, Academic Press.

Jones, M. P. and Crowley, J.J. (1992). Nonparametric tests of the Markov model for survival data. *Biometrika*, **79** 513-522.

Kalbfleisch, J. D. and Prentice, R. L. (1981). *Statistical Analysis of Failure Time Data*. Wiley

Karr, A.F. (1991). *Point Processes and their Statistical Inference,* Marcel Dekker, New York.

Keiding, N. (1986). Statistical analysis of semi-Markov models based on the theory of counting processes. In *Semi-Markov models. Theory and Applications*, (J. Janssen, Ed.) Plenum Press, 301-315.

Keiding, N., Klein, JP and Horowitz,MM. (2001). Multistate models and outcome prediction in bone marrow transplantation. *Statist. Med.*, **20** 1871-1885.

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 transplantation patients. *Statist. Med.*, **12** 2315-2332.

Kosorok, M. R. , Lee B. L. and Fine J. P. (2004). Robust inference for univariate proportional hazard models. *Ann. Statist.*, **32** 1448-1491.

Kuratowski, K. (1966). *Topology*. Academic Press.

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

Last, G. and Brandt, A. (1995). *Marked Point Processes on the Real Line: the Dynamic Approach.* Springer, New York.

Limnios, N and Oprisan. (2001). *Semi-Markov Processe and Reliability*. Springer.

Lin, D. Y., Fleming, T. R., Wei, L.J. (1994). Confidence bands for survival curves under the proportional hazards model. *Biometrika*, **81** 73-81.

Lo, S.M.S. and Wilke, R.A. (2010). A copula model for dependent competing risks. *Appl. Statist.*, **59**, 359-376.

Martinussen, T. and Scheike,T. (2006). *Dynamic Regression Models for Survival Data*. Springer, New York.

Moore,E. M. and Pyke,R. (1968). Estimation of the transition distributions of a Markov renewal process. *Ann. Inst. Stat. Math.*, **20** 411-468.

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

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

Oakes, D. and Cui, L. (1994) On semi-parametric inference for modulated renewal processes. *Biometrika*, **81** 83-91

Ouhbi, L. and Limnios, N. (1996). Nonparametric estimation for semi-Markov kernels with applications to reliability analysis. *Appl. Stochastic Models and Data Analysis.*, **12** 209-220.

Ouhbi, L. and Limnios, N. (1999). Nonparametric estimation for semi-Markov processes based on its hazard rate functions. *Stat. Inference Stoch. Processes*, **2** 151-173.

Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer Verlag, New York.

Pollard, D. (1990). *Empirical Processes: Theory and Applications*. Inst. Math. Statist., Hayward.

Phelan, M. F. (1999). Bayes estimation from a Markov renewal process. *Ann. Statist.*, **18** 603-616.

Putter, H., Fiocco, M. and Geskus, R.B. (2007). Tutorial in biostatistics: competing risks and multi-state models. *Statist. Med.,* **26** 2389-2430.

Pyke, R. (1961,a). Markov renewal processes: definitions and preliminary properties. *Ann. Math. Statist.*, **32** 1231–1242.

Pyke, R. (1961,b) Markov renewal processes with finitely many states. *Ann. Math. Statist.*, **32** 1243–1259.

Pyke, R. and Schaufele, R. (1964). Limit theorems for Markov renewal processes. *Ann. Math. Statist.*, **35** 1746-1764.

Pyke, R. and Schaufele, R. (1966). The existence and uniqueness of stationary measures for Markov renewal processes. *Ann. Math. Statist.*, **37** 1439-1462.

Ringden,O. , Labopin, M., Bacigalupo, A., Arcese, W., Schaefer, U.W., Willemze, R., Koc, H., Bunjes, D. Gluckman,E., Rocha, V. Schattenberg, A. and Frassoni, F. (2002). Transplantation of peripheral blood stem cell as compared with bone marrow from HLA-identical siblings in adult patients with acute myeloid leukemia and acute lymphoblastic leukemia. *J. Clin. Oncol.* **20(24)** 4655-4664.

Rivest, L. P. and Wells, M. T. (2001). A martingale approach to the copula-graphic estimator for the survival function under dependent censoring. *J. Multiv. Analysis*, **79**, 138-155.

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

van der Vaart, A.W. and Wellner, J.A. (1996). *Weak convergence and Empirical Processes with Applications to Statistics*. Springer, New York.

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.

Wagner, D. H. (1977). Survey of measurable selection theorems. *SIAM, J. Control and Optimization*, **15**, 859-903.

Weiss, G.H. and Zelen, M. (1965). A semi-Markov model for clinical trials. *J. Appl. Probab.,* **2** 269-285.

Zheng, M. and Klein, J. P. (1995). estimates of marginal survival for dependent competing risks based on an assumed copula model. *Biometrika*, **82** 127-138.