To appear in *Statistics: A Journal of Theoretical and Applied Statistics*
Vol. 00, No. 00, Month 201X, 1–9

Supplementary material for Semiparametric inference for the recurrent events process by means of a single-index model

Olivier Bouaziz[a*], Ségolen Geffray[b] and Olivier Lopez[c,d]

[a] *MAP5, Université Paris Descartes, 45 rue des Saints Pères, 75270 Paris Cedex 06*; [b] *IRMA, Université de Strasbourg, 7 rue René-Descartes 67084 Strasbourg Cedex*; [c] *ENSAE Paris-Tech & CREST, Laboratoire de Finance et d'Assurance, 3 avenue Pierre Larousse, 92245 Malakoff Cedex France*; [d] *Sorbonne Universités, UPMC Université Paris VI, EA 3124, LSTA, 4 place Jussieu 75005 Paris*

(*Received 00 Month 201X; final version received 00 Month 201X*)

1. Extended simulation study

In this section, we extend the simulation study performed in Section 4 of the main article. In order to take into account heterogeneity among individuals, we consider a gamma frailty model where $G_i \sim \Gamma(a, 1/a)$, $i = 1, \ldots, n$, is a gamma variable with shape a and scale $1/a$. The process $\tilde{N}_i(\cdot)$ is simulated in such a way that, conditionally on G_i and Z_i, $\tilde{N}_i(\cdot)$ is a Poisson process with intensity $(\theta_0' Z_i + 5)G_i$. This ensures us that, marginally, $\tilde{N}_i(t)|Z_i$ has a negative binomial distribution with mean equal to $(\theta_0' Z_i + 5)t$ and variance equal to $(\theta_0' Z_i + 5)t + (\theta_0' Z_i + 5)^2 t^2/a$. Note also that the process of interest $N^*(\cdot)$ has a conditional expectation equal to

$$E[N_i^*(t)|Z_i] = (\theta_0' Z_i + 5) \int_0^t (1 - F(s-))ds, \quad i = 1, \ldots, n,$$

which is the same as in the simulations section of the main paper, but has a larger variance than its expectation.

We put $a = 2$. The distributions of the variables D_i and C_i, the parameters and the family of weights are all set to the same values as in the main paper. In Tables 1 and 2 we report the results of our estimators $\tilde{\theta}$ and $\hat{\theta}_{\hat{w},h_0}$ over 1 000 simulations of samples of size 100 for two rates of censoring (30% and 70%). We also compare these results with the Cox estimator as previously. The average weights of $\hat{w}$ were also computed. For 30% of censoring, we obtain, $E[\hat{w}(\{0.9\})] = 0.711, E[\hat{w}(\{1\})] = 0.563, E[\hat{w}(\{1.1\})] = 0.420$

Corresponding author*. Email: olivier.bouaziz@parisdescartes.fr

and $E[\hat{w}(\{1.2\})] = 0.337$ and for 70% of censoring, $E[\hat{w}(\{0.9\})] = 0.725, E[\hat{w}(\{1\})] = 0.561, E[\hat{w}(\{1.1\})] = 0.422$ and $E[\hat{w}(\{1.2\})] = 0.329$.

When no weights are used in the estimation procedure, the simulation results are very similar to those obtained in the main paper. However, an increase in the variance estimates can be seen in the negative binomial context compared to the Poisson framework. As previously, the adaptive measure seems to play an important role in the estimation performance of θ_0. However, the improvement in the quality of estimation in the negative binomial context is clearly not as remarkable as in the Poisson context. Finally, our estimators still outperform the Cox estimator. The latter is still biased and has also a greater variance than in the Poisson situation.

All these results emphasize the fact that the recurrent events, in this simulation design, have a greater variance than in the Poisson context. This seems to deteriorate the quality of estimation of all estimators. However, the adaptive choice of the weights can still improve greatly the simulation results in term of MSE, especially in the case of high censoring, where the MSE is almost divided by 3 (and divided by 1.5 for 30% of censored data).

Table 1. Biases, variances and MSE of $\tilde{\theta}, \hat{\theta}_{\hat{w},h_0}$ and $\hat{\theta}_{cox}$ for 30% of censored data

$p = 30\%$	Bias	Variance	MSE
$\tilde{\theta}$	$\begin{pmatrix} 0.0827 \\ 0.0606 \\ 0.0552 \end{pmatrix}$	$\begin{pmatrix} 0.1008 & -0.0176 & -0.0551 \\ -0.0176 & 0.0794 & -0.0433 \\ -0.0551 & -0.0433 & 0.0980 \end{pmatrix}$	0.2880
$\hat{\theta}_{\hat{w},h_0}$	$\begin{pmatrix} 0.0634 \\ 0.0597 \\ 0.0429 \end{pmatrix}$	$\begin{pmatrix} 0.0559 & -0.0202 & -0.0242 \\ -0.0202 & 0.0624 & -0.0227 \\ -0.0242 & -0.0227 & 0.0679 \end{pmatrix}$	0.1956
$\hat{\theta}_{cox}$	$\begin{pmatrix} -1.4975 \\ -1.1696 \\ -0.6608 \end{pmatrix}$	$\begin{pmatrix} 0.0626 & -0.0002 & 0.0011 \\ -0.0002 & 0.0650 & 0.0056 \\ 0.0011 & 0.0056 & 0.0607 \end{pmatrix}$	4.2353

Table 2. Biases, variances and MSE of $\tilde{\theta}, \hat{\theta}_{\hat{w},h_0}$ and $\hat{\theta}_{cox}$ for 70% of censored data

$p = 70\%$	Bias	Variance	MSE
$\tilde{\theta}$	$\begin{pmatrix} 0.0913 \\ 0.0748 \\ 0.0578 \end{pmatrix}$	$\begin{pmatrix} 0.1449 & -0.0312 & 0.0287 \\ -0.0412 & 0.1210 & -0.0265 \\ 0.0287 & -0.0265 & 0.1927 \end{pmatrix}$	0.6417
$\hat{\theta}_{\hat{w},h_0}$	$\begin{pmatrix} 0.0740 \\ 0.0624 \\ 0.0411 \end{pmatrix}$	$\begin{pmatrix} 0.0643 & -0.0242 & -0.0244 \\ -0.0242 & 0.0682 & -0.0242 \\ -0.0244 & -0.0242 & 0.0731 \end{pmatrix}$	0.2167
$\hat{\theta}_{cox}$	$\begin{pmatrix} -1.5005 \\ -1.1744 \\ -0.6449 \end{pmatrix}$	$\begin{pmatrix} 0.0724 & -0.0004 & 0.0002 \\ -0.0004 & 0.07156 & 0.0771 \\ 0.0002 & 0.0771 & 0.0771 \end{pmatrix}$	4.2676

2. Uniform convergence of the nonparametric estimators

In this section, we show that the kernel estimator $\hat{\mu}_{\theta,h}$ defined by (2.10) satisfies the convergence rates required by Assumption 7, under Assumptions 10 and 11. Introduce the quantity

$$\tilde{\mu}_{\theta,h}(t,u) = \sum_{i=1}^{n} \int_0^t \frac{K\left(\frac{\theta' Z_i - u}{h}\right) dN_i(s)}{\sum_{j=1}^{n} K\left(\frac{\theta' Z_j - u}{h}\right)\left(1 - G(s-)\right)}.$$

We first study the convergence rate of the difference between $\tilde{\mu}_{\theta,h}$ and μ_θ and their derivatives. Since no Kaplan-Meier functions are involved in this expression, we can use classical results on uniform convergence of kernel estimators, mainly from Einmahl and Mason [2005].

We also introduce a trimming function. Its purpose is to circumvent problems caused by too small values of the denominator in the definition of $\hat{\mu}_{\theta,h}$. Indeed, to ensure uniform consistency of our estimator, we need to bound this denominator away from zero. We use the same methodology as in Delecroix *et al.* [2006]. Let $f_{\theta_0' Z}$ denote the density of $\theta_0' Z$ and define the "ideal" trimming function $J_{\theta_0}(\theta_0' Z, c) = I(\theta_0' Z \in B_0)$ where $B_0 = \{u : f_{\theta_0' Z}(u) \geq c\}$ for some constant $c > 0$. As in Delecroix *et al.* [2006] (see also Lopez [2009]), we first assume that we know some set B on which $\inf\{f_{\theta' Z}(\theta' z) : z \in B, \theta \in \Theta\} > c$ where c is a strictly positive constant. In a preliminary step, we can use this set B to compute the preliminary trimming $J_B(z) = I(z \in B)$. Using this trimming function and a deterministic sequence of bandwidth h_0 satisfying (4) in Assumption 10 we define a preliminary estimator θ_n of θ_0 as

$$\theta_n(w) = \arg\min_{\theta \in \Theta} M_{n,w}(\theta, \hat{\mu}_\theta) J_B(z).$$

Given this preliminary consistent estimator of θ_0, we use the following trimming $J_n(\theta_n' Z, c) = I(\hat{f}_{\theta_n' Z}(\theta_n' Z) \geq c)$ which appears to be asymptotically equivalent to $J_{\theta_0}(\theta_0' Z, c)$ (see e.g. Lopez [2009]). Then, our final estimator consists of

$$\hat{\theta}(w) = \arg\min_{\theta \in \Theta_n} M_{n,w}(\theta, \hat{\mu}_\theta) J_n(\theta_n' z, c),$$

where Θ_n is a shrinking neighborhood of θ_0 accordingly to our preliminary estimator θ_n.

As announced, the next proposition gives the rates of convergence of $\tilde{\mu}_{\theta,h}$ and its derivatives. Since we need a convergence over $\theta \in \Theta$, the trimming we need to use is $J_\theta(\theta' Z, c) := I(\hat{f}_{\theta' Z}(\theta' Z) \geq c)$. But notice that $J_{\theta_0}(\theta_0' Z, c)$ can be replaced by $J_\theta(\theta' Z, c/2)$ on shrinking neighborhoods of θ_0.

Proposition 1. *Under Assumption 10, for z such that $J_\theta(\theta' z, c) = 1$ almost surely, we* *have*

$$\sup_{t\leq T_{(n)},\theta,z,h}\sqrt{\frac{nh}{\log n}}\left|\frac{\tilde{\mu}_\theta(t,\theta' z)-\mu_\theta(t,\theta' z)}{\bar{\mu}_{\theta_0}(t,\theta_0' z)^{\lambda_1+\lambda_2}}\right| = O_P(1), \tag{2.1}$$

$$\sup_{t\leq T_{(n)},\theta,z,h}\sqrt{\frac{nh^3}{\log n}}\left\|\frac{\nabla_\theta\tilde{\mu}_\theta(t,z)-\nabla_\theta\mu_\theta(t,z)}{\bar{\mu}_{\theta_0}(t,\theta_0' z)^{\lambda_1+\lambda_2}}\right\| = O_P(1), \tag{2.2}$$

$$\sup_{t\leq T_{(n)},\theta,z,h}\sqrt{\frac{nh^5}{\log n}}\left\|\frac{\nabla^2_\theta\tilde{\mu}_\theta(t,z)-\nabla^2_\theta\mu_\theta(t,z)}{\bar{\mu}_{\theta_0}(t,\theta_0' z)^{\lambda_1+\lambda_2}}\right\| = O_P(1). \tag{2.3}$$

Proof. The proofs of (2.1)-(2.3) are all similar. The most delicate term to handle, coming from (2.3), is

$$\hat{A}_\theta^{n,h}(t,z) := \frac{1}{nh^3}\sum_{i=1}^n \frac{(Z_i-z)^2}{\bar{\mu}_{\theta_0}(t,\theta_0' z)^{\lambda_1+\lambda_2}}K''\left(\frac{\theta' Z_i-\theta' z}{h}\right)\int_0^t \frac{dN_i(s)}{1-G(s-)}.$$

Consider the class of functions $\mathcal{K}$ introduced in Assumption 10. From Nolan and Pollard [1987], it can easily be seen that, using a kernel K satisfying Assumption 10, for some $c'>0$ and $\nu>0$, we have $N(\varepsilon,\mathcal{K},\|\cdot\|_\infty)\leq c'\varepsilon^{-\nu}, 0<\varepsilon<1$.

Then, concerning the uniformity with respect to θ, Lemma 22 (ii) of Nolan and Pollard [1987] shows that the family of functions $\left\{(Z,N)\longmapsto \hat{A}_\theta^{n,h}(t,z)\right\}$ satisfies the assumptions of Proposition 1 in Einmahl and Mason [2005].
Define

$$\tilde{A}_\theta^{h}(t,z) := \frac{1}{h^3}E\left[\frac{(Z-z)^2}{\bar{\mu}_{\theta_0}(t,\theta_0' z)^{\lambda_1+\lambda_2}}K''\left(\frac{\theta' Z-\theta' z}{h}\right)\int_0^t \frac{dN(s)}{1-G(s-)}\right],$$

$$A_\theta^{h}(t,z) := \frac{\partial^2}{\partial u^2}\left\{E\left[\frac{(Z-z)^2}{\bar{\mu}_{\theta_0}(t,\theta_0' z)^{\lambda_1+\lambda_2}}\int_0^t \frac{dN(s)}{1-G(s-)}\middle|\theta' Z=u\right]f_{\theta' Z}(u)\right\}\Bigg|_{u=\theta' z}$$

and apply Talagrand's inequality (see Talagrand [1994], see also Einmahl and Mason [2005]) to obtain that

$$\sup_{t\leq T_{(n)},\theta,z,h}|\hat{A}_{n,h}(t,z)-\tilde{A}_{n,h}(t,z)| = O_P\left(n^{-1/2}h^{-5/2}(\log n)^{1/2}\right).$$

For the bias term, classical kernel arguments (see for instance Bosq and Lecoutre [1997]) show that

$$\sup_{t\leq T_{(n)},\theta,z,h}|\tilde{A}_{n,h}(t,z)-A_{n,h}(t,z)| = O(h^2).$$

□

It remains to study $\hat{\mu}_{\theta,h}-\tilde{\mu}_{\theta,h}$. The following lemma gives some precision on the difference between the Kaplan Meier weights of $\hat{\mu}_{\theta,h}$ and the "ideal" weights involving the true function G in $\tilde{\mu}_{\theta,h}$.

Lemma 2. *Let* $\hat{\Lambda}_G(s) = (1 - \hat{G}(s-))^{-1}$, $\tilde{\Lambda}_G(s) = (1 - G(s-))^{-1}$ *and*

$$C_G(t) = \int_0^t \frac{dG(s)}{\big(1 - G(s-)\big)\big(1 - H(s-)\big)}.$$

(1) We have

$$\sup_{t \le T_{(n)}} \frac{1 - G(t)}{1 - \hat{G}(t)} = O_P(1).$$

(2) For all $0 \le \beta \le 1$ *and* $\varepsilon > 0$*, we have*

$$|\hat{\Lambda}_G(s) - \tilde{\Lambda}_G(s)| \le R_n(s)\tilde{\Lambda}_G(s) C_G(s)^{\beta(1/2+\varepsilon)},$$

where $\sup_{s \le T_{(n)}} R_n(s) = O_P(n^{-\beta/2})$.

Proof. (1) This result is a consequence of Lemma 2.6 in Gill [1983].
(2) For $0 \le \beta \le 1$ and $\varepsilon > 0$, write

$$\hat{\Lambda}_G(s) - \tilde{\Lambda}_G(s) = \tilde{\Lambda}_G(s) C_G(s)^{\beta(1/2+\varepsilon)} \big(R_G(s) C_G(s)^{-1/2-\varepsilon}\big)^{\beta} \big(R_G(s)\big)^{1-\beta} \frac{1 - G(s-)}{1 - \hat{G}(s-)},$$

where $R_G(s) = \big(\hat{G}(s-) - G(s-)\big)\big(1 - G(s-)\big)^{-1}$. Since $\int_0^{\tau_H} C_G(s)^{-1-2\varepsilon} dC_G(s) < \infty$, apply Theorem 1 in Gill [1983] and use the first part of the current lemma to conclude the proof.

□

The next proposition gives the convergence rate of $\hat{\mu}_{\theta,h} - \tilde{\mu}_{\theta,h}$. Notice that if w is supported on a compact interval, we only need this result on a compact subset of $[0, T_{(n)}]$ and in this case Assumption 11 is automatically fulfilled.

Proposition 3. *Under Assumptions 10 and 11, for* z *such that* $J_\theta(\theta' z, c) = 1$ *almost surely, we have*

$$\sup_{t \le T_{(n)}, \theta, z, h} \left| \frac{\hat{\mu}_\theta(t, \theta' z) - \tilde{\mu}_\theta(t, \theta' z)}{\bar{\mu}_{\theta_0}(t, \theta_0' z)^{\lambda_1 + \lambda_2}} \right| = O_P\left(n^{-7/20}\right), \tag{2.4}$$

$$\sup_{t \le T_{(n)}, \theta, z, h} h \left\| \frac{\nabla_\theta \hat{\mu}_\theta(t, z) - \nabla_\theta \tilde{\mu}_\theta(t, z)}{\bar{\mu}_{\theta_0}(t, \theta_0' z)^{\lambda_1 + \lambda_2}} \right\| = O_P\left(n^{-7/20}\right), \tag{2.5}$$

$$\sup_{t \le T_{(n)}, \theta, z, h} h^2 \left\| \frac{\nabla_\theta^2 \hat{\mu}_\theta(t, z) - \nabla_\theta^2 \tilde{\mu}_\theta(t, z)}{\bar{\mu}_{\theta_0}(t, \theta_0' z)^{\lambda_1 + \lambda_2}} \right\| = O_P\left(n^{-7/20}\right). \tag{2.6}$$

Proof. We only prove (2.6) since (2.4) and (2.5) can be handled similarly. Let us consider the following term involving the second derivative of K

$$\frac{1}{nh^3} \sum_{i=1}^n (Z_i - z)^2 K''\left(\frac{\theta' Z_i - \theta' z}{h}\right) \left(\bar{\mu}_{\theta_0}(t, \theta_0' z)^{\lambda_1 + \lambda_2} f_{\theta' Z}(\theta' z)\right)^{-1} \int_0^t \big(\hat{\Lambda}(s) - \tilde{\Lambda}(s)\big) dN_i(s).$$

From Lemma 2, this term can be bounded by

$$O_P(n^{-\beta/2}h^{-2})\left|\frac{1}{nh}\sum_{i=1}^{n}K''\left(\frac{\theta' Z_i-\theta' z}{h}\right)\bar{\mu}_{\theta_0}(t,\theta_0' z)^{-(\lambda_1+\lambda_2)}\int_0^t\tilde{\Lambda}(s)C_G(s)^{\beta(1/2+\varepsilon)}dN_i(s)\right| \tag{2.7}$$

where the O_P- rate does not depend on t,θ,z nor h. Now, consider the family of functions indexed by t,θ,z and h,

$$\left\{(Z,N)\longmapsto K''\left(\frac{\theta' Z-\theta' z}{h}\right)\bar{\mu}_{\theta_0}(t,\theta_0' z)^{-(\lambda_1+\lambda_2)}\int_0^t\tilde{\Lambda}(s)C_G(s)^{\beta(1/2+\varepsilon)}dN(s)\right\}.$$

This family is Euclidian (see Nolan and Pollard [1987]) for an envelope

$$\sup_{t,z}\frac{\tilde{\Lambda}(t)C_G^{\beta(1/2+\varepsilon)}(t)N(t)}{\bar{\mu}_{\theta_0}(t,\theta_0' z)^{\lambda_1+\lambda_2}}$$

which is, for $\beta=7/10$, square integrable from Assumption 11. Then, using the results of Sherman [1994], the second part of (2.7) is $O_P(1)$ uniformly in t,θ,z and h.

□

Finally, combination of Propositions 1 and 3 leads to the following result.

Corollary 4. *Under Assumptions 10 and 11, for z such that $J_\theta(\theta' z,c)=1$ almost surely,*

$$\sup_{t\leq T_{(n)},\theta,z,h}\frac{|\hat{\mu}_\theta(t,\theta' z)-\mu_\theta(t,\theta' z)|\cdot\|\nabla_\theta\hat{\mu}_\theta(t,z)-\nabla_\theta\mu_\theta(t,z)\|}{|\bar{\mu}_{\theta_0}(t,\theta_0' z)|^{2(\lambda_1+\lambda_2)}}=o_P(n^{-1/2}).$$

3. Technical lemmas

3.1 *Gradient vector in the single-index model*

Lemma 5. *Let $\mu'_{\theta_0}(t|u)=\frac{\partial}{\partial u}\mu_{\theta_0}(t,u)$ (assuming that $\mu_{\theta_0}(\cdot,\cdot)$ is C^1). Then, for every (t,z), the map $\theta\mapsto\mu_\theta(t,\theta' z)$ is differentiable with respect to θ, with*

$$\nabla_\theta\mu_{\theta_0}(t,Z)=\mu'_{\theta_0}(t|\theta_0' Z)\big(Z-E(Z|\theta_0' Z)\big),$$

where, as a consequence

$$E\big[\nabla_\theta\mu_{\theta_0}(t,Z)|\theta_0' Z\big]=0. \tag{3.1}$$

Proof. Observe that $\mu_\theta(t,\theta' Z)=E[\mu_{\theta_0}(t,\theta_0' Z)|\theta' Z]$ and let $\zeta(Z,\theta)=\theta_0' Z-\theta' Z$ for $\theta\in\Theta$. We have

$$\mu_\theta(t,\theta' Z)=E\big[\mu_{\theta_0}\big(t,\zeta(Z,\theta)+\theta' Z\big)|\theta' Z\big].$$

Defining $\Gamma(\theta_1, \theta_2) = E[\mu_{\theta_0}(t, \zeta(Z, \theta_1) + \theta_2' Z)|\theta_2' Z]$, we have $\Gamma(\theta, \theta) = \mu_\theta(t, \theta' Z)$, which leads to

$$\nabla_{\theta_1}\Gamma(\theta_1, \theta_0)|_{\theta_1=\theta_0} = -\mu'_{\theta_0}(t, \theta_0' Z) E\left[Z|\theta_0' Z\right],$$
$$\nabla_{\theta_2}\Gamma(\theta_0, \theta_2)|_{\theta_2=\theta_0} = Z\mu'_{\theta_0}(t, \theta_0' Z).$$

□

4. Auxiliary lemma for tightness conditions

Lemma 6. *Let $\mathcal{F}$ be a class of functions. Let $P_n(t,f)$ be a process on $[0,\tau_H]\times\mathcal{F}$. Define, for any $\tau \in [0,\tau_H]$, $R_n(\tau, f) = P_n(\tau_H, f) - P_n(\tau, f)$. Assume that for any τ such that $\tau < \tau_H$*

$$P_n(t,f) \Longrightarrow \mathbb{W}(V_f(t)) \in \mathcal{D}([0,\tau]), f \in \mathcal{F},$$

where $\mathbb{W}(V_f(t))$ is a centered Gaussian process with covariance function V_f and $\mathcal{D}$ denotes the set of càdlàg functions.
Assume that, for a sequence of random variables (X_n) and two functions Γ and Γ_n, the following conditions hold

(1) $\lim_{\tau\to\tau_H} V_f(\tau) = V_f(\tau_H)$ *with* $\sup_{f\in\mathcal{F}} |V_f(\tau_H)| < \infty$,
(2) $|R_n(\tau', f)| \le X_n \times \Gamma_n(\tau)$ *for all* $\tau < \tau' < \tau_H$,
(3) $X_n = O_P(1)$,
(4) $\Gamma_n(\tau) \to \Gamma(\tau)$ *in probability,*
(5) $\lim_{\tau\to\tau_H} \Gamma(\tau) = 0$.

Then $P_n(\tau_H, f) \Longrightarrow \mathcal{N}(0, V_f(\tau_H))$.

Proof. From Theorem 13.5 in Billingsley [1999] and condition (1), it suffices to show that, for all $\varepsilon > 0$

$$\lim_{\tau\to\tau_H} \limsup_{n\to\infty} P\left(\sup_{\tau\le t\le\tau_H, f\in\mathcal{F}} |R_n(t,f)| > \varepsilon\right) = 0. \tag{4.1}$$

Using condition (2) , the probability in equation (4.1) is bounded, for all $M > 0$, by

$$P\big(|\Gamma_n(\tau) - \Gamma(\tau)| > \varepsilon/M - \Gamma(\tau)\big) + P(X_n > M). \tag{4.2}$$

Moreover, from condition (4), we can state that

$$\limsup_{n\to\infty} P\big(|\Gamma_n(\tau) - \Gamma(\tau)| > \varepsilon/M - \Gamma(\tau)\big) = I(\varepsilon/M - \Gamma(\tau) \ge 0).$$

Since $\Gamma(\tau) \to 0$ (condition (5)), we can deduce that

$$\lim_{\tau\to\tau_H} \limsup_{n\to\infty} P\big(|\Gamma_n(\tau) - \Gamma(\tau)| > \varepsilon/M - \Gamma(\tau)\big) = 0.$$

As a consequence,

$$\lim_{\tau\to\tau_H}\limsup_{n\to\infty} P\left(\sup_{\tau\le t\le\tau_H, f\in\mathcal{F}} |R_n(t,f)| > \varepsilon\right) \le \lim_{M\to\infty}\limsup_{n\to\infty} P(X_n > M) = 0,$$

using condition (3). □

5. Covering number results

In this section, we determine the covering numbers of some particular classes of functions. From these computations, sufficient conditions can be deduced to check Property 2 and Assumption 9.

Proposition 7. *Let $\mathcal{F}$ be a class of functions $f(t,z)$ with envelope $\bar{F}$ defined on $\mathbb{R}\times\mathbb{R}^d$ with continuous derivative with respect to the first component. Let $\tilde{F}$ be the envelope of the class of functions $\partial f(s,z)/\partial s$. Let $W(t)$ be a positive bounded decreasing function and set $\mathcal{W} = \{w : dw(t) = W(t)d\tilde{w}(t), \tilde{w} \in \tilde{\mathcal{W}}\}$ where $\tilde{\mathcal{W}}$ is a class of monotone positive functions with envelope function $\tilde{W}$.*
Assume that $E[(\int_0^{\tau_H} \bar{F}(t,z)W(t)dY(t))^2] < \infty$, $E[(\int_0^{\tau_H} \bar{F}(t,z)Y(t)dW(t))^2] < \infty$ and $E[(\int_0^{\tau_H} \tilde{F}(t,z)W(t)Y(t)dt)^2] < \infty$.
Then, the class of functions $\mathcal{H} = \{(z,y) \to \int_0^{\tau_H} f(t,z)y(t)dw(t), f \in \mathcal{F}, w \in \mathcal{W}\}$ has a uniform covering number satisfying, for some constant C,

$$N(\varepsilon,\mathcal{H},\|\cdot\|_2) \le CN(\varepsilon, W\mathcal{F},\|\cdot\|_2)N(\varepsilon,\tilde{\mathcal{W}},\|\cdot\|_2).$$

Proof. Let Q be a probability measure and introduce $N_1 = \sup_Q N(\varepsilon\|W\bar{F}\|_Q, W\mathcal{F},\|\cdot\|_{2,Q})$ and $N_2 = \sup_Q N(\varepsilon\|\tilde{W}\|_Q,\tilde{\mathcal{W}},\|\cdot\|_{2,Q})$. Let $\{f_i^W\}_{1\le i\le N_1}$ (respectively $\{\tilde{w}_j\}_{1\le j\le N_2}$) be the center of the $\varepsilon - \|\cdot\|_{2,Q}$ balls needed to cover $W\mathcal{F}$ (respectively $\tilde{\mathcal{W}}$). Writing $dw(t) = W(t)d\tilde{w}(t)$, we have for any $1 \le i \le N_1$ and $1 \le j \le N_2$

$$\begin{aligned}
&\left|\int_0^{\tau_H} Y(t)f(t,z)W(t)d\tilde{w}(t) - \int_0^{\tau_H} Y(t)f_i^W(t,z)d\tilde{w}_j(t)\right| \\
&\quad\le \left|\int_0^{\tau_H} Y(t)\big(f(t,z)W(t) - f_i^W(t,z)\big)d\tilde{w}_j(t)\right| + \left|\int_0^{\tau_H} Y(t)f(t,z)W(t)(d\tilde{w} - d\tilde{w}_j)(t)\right|.
\end{aligned}$$

For any $f \in \mathcal{F}$, there exists some i such that the first term is seen to be less than $c_1\varepsilon$ in $L^2(Q)-$norm. For the second term, there also exists some j such that this is smaller than $c_2\varepsilon$, which can be seen using integration by parts. The result follows. □

References

[1] Billingsley, P. (1999). *Convergence of probability measures* (Second ed.). Wiley Series in Probability and Statistics: Probability and Statistics. New York: John Wiley & Sons Inc. A Wiley-Interscience Publication.

[2] Bosq, D. and J.-P. Lecoutre (1997). *Théorie de l'estimation fonctionnelle*, Volume 3 of *Economie et statistiques avancées.* Paris: Economica.

[3] Delecroix, M., M. Hristache, and V. Patilea (2006). On semiparametric M-estimation in single-index regression. *J. Statist. Plann. Inference 136*(3), 730–769.

[4] Einmahl, U. and D. M. Mason (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Statist. 33*(3), 1380–1403.

[5] Gill, R. (1983). Large sample behavior of the product-limit estimator on the whole line. *Ann. Statist. 11*, 49–58.

[6] Lopez, O. (2009). Single-index regression models with right-censored responses. *J. Statist. Plann. Inference 139*(3), 1082–1097.

[7] Nolan, D. and D. Pollard (1987). U-processes: rates of convergence. *Ann. Statist. 15*(2), 780–799.

[8] Sherman, R. P. (1994). Maximal inequalities for degenerate U-processes with applications to optimization estimators. *Ann. Statist. 22*(1), 439–459.

[9] Talagrand, M. (1994). Sharper bounds for Gaussian and empirical processes. *Ann. Probab. 22*(1), 28–76.