HAL

archives-ouvertes.fr

Estimation of the hazard function in a semiparametric model with covariate measurement error

Marie-Laure Martin-Magniette, Marie-Luce Taupin

▶ **To cite this version:**

HAL Id: hal-00480192

https://hal.archives-ouvertes.fr/hal-00480192

Submitted on 3 May 2010

ESTIMATION OF THE HAZARD FUNCTION IN A SEMIPARAMETRIC MODEL WITH COVARIATE MEASUREMENT ERROR

Marie-Laure Martin-Magniette[1, 2] and Marie-Luce Taupin[3]

Abstract. We consider a failure hazard function, conditional on a time-independent covariate Z, given by $\eta_{\gamma^0}(t)f_{\beta^0}(Z)$. The baseline hazard function η_{γ^0} and the relative risk f_{β^0} both belong to parametric families with $\theta^0 = (\beta^0, \gamma^0)^\top \in \mathbb{R}^{m+p}$. The covariate Z has an unknown density and is measured with an error through an additive error model $U = Z + \varepsilon$ where ε is a random variable, independent from Z, with known density f_ε. We observe a n-sample (X_i, D_i, U_i), $i = 1, \ldots, n$, where X_i is the minimum between the failure time and the censoring time, and D_i is the censoring indicator. Using least square criterion and deconvolution methods, we propose a consistent estimator of θ^0 using the observations (X_i, D_i, U_i), $i = 1, \ldots, n$. We give an upper bound for its risk which depends on the smoothness properties of f_ε and $f_\beta(z)$ as a function of z, and we derive sufficient conditions for the $\sqrt{n}$-consistency. We give detailed examples considering various type of relative risks f_β and various types of error density f_ε. In particular, in the Cox model and in the excess risk model, the estimator of θ^0 is $\sqrt{n}$-consistent and asymptotically Gaussian regardless of the form of f_ε.

1991 Mathematics Subject Classification. 62G05, 62F12,62G99, 62J02.

The dates will be set by the publisher.

1. Introduction

We are interested in the relationship between a survival time T and a covariate Z described by the conditional hazard function of T given $Z = z$ intuitively defined by

$$R(t|z) = \lim_{\Delta_t \downarrow 0} \frac{1}{\Delta_t} \mathbb{P}(t \leq T < t + \Delta_t | T \geq t, Z = z).$$

In this paper we consider a parametric proportional hazard model, $R(t, \theta^0 | Z) = \eta_{\gamma^0}(t) f_{\beta^0}(Z)$, conditional on a time-independent covariate Z with unknown density g. The proportional hazard model is often used to describe a covariate effect on a survival time. Under the condition $f_\beta(0) = 1$, $\eta_{\gamma^0}(t)$ is the baseline hazard function, that is the conditional hazard function of T given $Z = 0$. The function f_{β^0} is the relative risk, and the conditional failure rates associated with any two values of the covariate Z is proportional. Here we assume that the functions η_γ and f_β both belong to parametric families and $\theta^0 = (\beta^0, \gamma^0)^\top$ belongs to the interior of a compact set $\Theta = \mathbb{B} \times \Gamma \subset \mathbb{R}^{m+p}$. To ensure that the hazard function is a nonnegative function, we assume that $\eta_\gamma(t) \geq 0$ for all $\gamma \in \Gamma$ and for all $t \in [0, \tau]$, $\tau < \infty$, and also that $f_\beta(Z) \geq 0$ for all $\beta \in \mathbb{B}$ and Z with density g. The parametric modelling of the hazard function has some advantages. In particular, the coefficients can be clinically meaningful and fitted values from the model can provide estimates of survival time.

Among the best known parametric models are exponential models where $R(t, \theta | Z) = \gamma f(\beta^\top Z)$, Weibull models with $R(t, \theta | Z) = \gamma_1 t^{\gamma_2} f(\beta^\top Z)$, models with a piecewise constant baseline function and Gomperz-Makeham models with $R(t, \theta | Z) = (\gamma_1 + \gamma_2(\gamma_3)^t) f(\beta^\top Z)$. This latter is commonly used in analysis of mortality data (see [37]). We refer to [2, 10, 17] for discussions on parametric survival time models and their advantages.

Suppose we observe the covariate Z in a cohort of n individuals. For each individual, we would observe a triplet (X_i, D_i, Z_i), where $X_i = \min(T_i, C_i)$ is the minimum between the failure time T_i and the censoring time C_i, $D_i = \mathbb{1}_{T_i \leq C_i}$ denotes the failure indicator, and Z_i is the value of the covariate. In this context, one usual way is to estimate $\theta^0 = (\beta^0, \gamma^0)^\top$ by the maximum likelihood estimator. We refer for instance to [1, 4, 5, 16] for related results.

In this paper we consider that the covariate Z is mismeasured. For example, the covariate Z could be a stage of a disease, which may be misdiagnosed, or Z could be a dose of ingested pathogenic agent not correctly evaluated, such that the error range between the unknown dose and the evaluated dose is sizeable. We observe a cohort of n individuals during a fixed time interval $[0, \tau]$ with $\tau < \infty$. For each individual, the available observation is thus the triplet $\Delta_i = (X_i, D_i, U_i)$ where $U_i = Z_i + \varepsilon_i$, and where the sequences of random variables $(\varepsilon_i)_{i=1,\cdots,n}$

and $(Z_i, T_i, C_i)_{i=1,\cdots,n}$ are independent. The density of ε is known and is denoted by f_ε. Our aim is to estimate the parameter $\theta^0 = (\beta^0, \gamma^0)^\top$ from the n-sample of independent and identically distributed random variables $(\Delta_1, \ldots, \Delta_n)$, in the presence of the completely unknown density g of the unobservable covariate Z, where g is viewed as a nuisance parameter belonging to a functional space $\mathcal{G}$. Hence this model belongs to the class of the so-called semiparametric models.

1.1. Our results

We propose an estimation procedure based on the least square criterion estimation using deconvolution methods. The least square criterion is defined by

$$S_{\theta^0,g}(\theta) = \mathbb{E}\Big((f_\beta^2 W)(Z) \int_0^\tau Y(t)\eta_\gamma^2(t)dt\Big) - 2\mathbb{E}\Big((f_\beta W)(Z) \int_0^\tau \eta_\gamma(t)dN(t)\Big) \tag{1.1}$$

where W is a nonnegative weight function to be chosen, $N(t) = \mathbb{1}_{X \le t, D=1}$ and $Y(t) = \mathbb{1}_{X \ge t}$. Since the intensity of the censored process $N(t)$ with respect to $\mathcal{F}_t = \sigma\{Z, U, N(s), \mathbb{1}_{X \ge s}, 0 \le s \le t \le \tau\}$ is equal to $\lambda(t, \theta^0, Z) = \eta_{\gamma^0}(t)Y(t)f_{\beta^0}(Z)$, we can rewrite $S_{\theta^0,g}(\theta)$ as

$$S_{\theta^0,g}(\theta) = \int_0^\tau \Big\{ \mathbb{E}\Big[\big(\eta_\gamma(t)f_\beta(Z) - \eta_{\gamma^0}(t)f_{\beta^0}(Z)\big)^2 Y(t)W(Z)\Big] - \mathbb{E}\Big[\big(\eta_{\gamma^0}(t)f_{\beta^0}(Z)\big)^2 Y(t)W(Z)\Big]\Big\}dt. \tag{1.2}$$

This shows that $S_{\theta^0,g}(\theta)$ is minimum if $\theta = \theta^0$ as soon as W is a nonnegative function. We propose to estimate $S_{\theta^0,g}(\theta)$ for all $\theta \in \Theta$ by a quantity $S_{n,1}(\theta)$, which depends on the error density f_ε, on the observations $\Delta_1, \cdots, \Delta_n$, with $\Delta_i = (X_i, D_i, U_i)$, and where g is replaced with a deconvolution kernel estimator. The parameter θ^0 is then estimated by minimizing $S_{n,1}(\theta)$. We refer to [38] for properties of M-estimators and to [33] for other results on the estimation of intensity processes using the least square criterion.

Under classical smoothness and identifiability assumptions and for a W suitably chosen such that $f_\beta W$ has the best smoothness properties, the estimator $\widehat{\theta}_1 = \underset{\theta \in \Theta}{\arg\min}\, S_{n,1}(\theta)$ converges to $\underset{\theta \in \Theta}{\arg\min}\, S_{\theta^0,g}(\theta) = \theta^0$ which ensures the consistency of $\widehat{\theta}_1$. Its rate of convergence depends on the smoothness of f_ε and on the smoothness of $(f_\beta W)(z)$ as a function of z. More precisely, if we denote by $\varphi^\star(t) = \int e^{itx}\varphi(x)dx$ the Fourier transform of an integrable function φ, the rate of convergence of $\widehat{\theta}_1$ depends on the behavior of the ratios of the Fourier transforms $(f_\beta W)^*(t)/f_\varepsilon^*(-t)$ and $(f_\beta^2 W)^*(t)/f_\varepsilon^*(-t)$ as t tends to infinity. We give an upper bound of the quadratic risk $\mathbb{E} \parallel \widehat{\theta}_1 - \theta^0 \parallel_{\ell^2}^2$ where $\parallel \theta \parallel_{\ell^2}^2 = \sum_{k=1}^{m+p} \theta_k^2$, for various relative risks and various types of error density, and we derive sufficient conditions ensuring $\sqrt{n}$-consistency and asymptotic normality. Through these examples, we also show that a simple choice of W can significantly improve the rate of convergence of $\widehat{\theta}_1$,

up to the parametric rate in many cases. Specific choices for W ensure that $\widehat{\theta}_1$ from the Cox model and the excess relative risk model is $\sqrt{n}$-consistent and asymptotically Gaussian. No consistent estimators for the excess relative risk model with mismeasured covariate have been found previously. Moreover, the previous methods developped for the parameter estimation in the Cox model do not apply in the excess relative risk model. Furthermore, from an application point of view, this result is promising since the excess relative risk model is commonly used in radioprotection research to investigate the relationship between cancer occurence and radiation dose which is always mismeasured (see [27,30]).

By construction our estimation procedure is related to the problem of the estimation of $S_{\theta^0,g}(\theta)$. For some particular relative risks, $S_{\theta^0,g}(\theta)$ can be estimated at the parametric rate without deconvolution methods. In these special cases we propose a second estimator $\widehat{\theta}_2$ which is always a $\sqrt{n}$-consistent and asymptotically Gaussian estimator of θ^0.

1.2. **Previous results and ideas**

Models with measurement errors have been thoroughly studied since the 1950s with the first papers studying regression models with errors-in-variables (see, e.g. [21,32]). We refer to [8,13] for a presentation of such models and results related to measurement error models. It is well known that in regression models, the measurement errors on the explanatory variables make the estimation of the regression function much more difficult, even if the regression function has a parametric form. For recent results and illustration of such difficulties, we refer to [6,35] for parametric regression functions, and to [9,11] for nonparametric regression functions.

Interest in survival models when covariates are subject to measurement errors is more recent. To our knowledge, there are no results on consistent estimation of the hazard function (for any type of modelling) in the case of general parametric relative risk with a mismeasured covariate. All the previously known results are obtained in the semiparametric Cox model. Let us specify the related methods: to take into account the fact that the covariate Z is measured with error, one idea is simply to replace Z with the observation U in the partial likelihood. This idea, usually called naive method, provides biased estimators (see for instance [28,29]). One can also cite [25] who study the resulting bias in case of heterogeneous covariate measurement error. A second idea, related to calibration methods, is to propose corrections of the estimation criterion, for instance by replacing Z_i with an approximation of $\mathbb{E}(Z_i|U_i)$. The third idea, proposed by [30,36,39] is to approximate the partial log-likelihood related to the filtration generated by the observations $\mathcal{E}_t = \sigma\{U, N(s), \mathbb{1}_{X>s}, 0 \leq s \leq t\}$. All these methods provide biased estimators of the parameters in the semiparametric Cox model and also in

general proportional hazard models. The last idea, extensively used in the semiparametric Cox model, is to correct the partial score function where the mismeasured Z_i's have been replaced with the observations U_i's. By definition, the partial score function is (see [14])

$$L_n^{(1)}(\beta, Z^{(n)}) = \frac{1}{n} \sum_{i=1}^{n} \int_0^\tau \left(f_\beta^{(1)}(Z_i)/f_\beta(Z_i) - \left[\sum_{j=1}^{n} Y_j(t) f_\beta^{(1)}(Z_j) \right] / \left[\sum_{j=1}^{n} Y_j(t) f_\beta(Z_j) \right] \right) dN_i(t),$$

with $N_i(t) = \mathbb{1}_{X_i \le t, D_i = 1}$, $Y_i(t) = \mathbb{1}_{X_i \ge t}$, $Z^{(n)} = (Z_1, \ldots, Z_n)$, and with $f_\beta^{(1)}$ the first derivative of f_β with respect to β. The $\tilde{\beta}$ such that $L_n^{(1)}(\tilde{\beta}, U^{(n)}) = 0$ is not consistent but, in the Cox model, one can exhibit corrections of $L_n^{(1)}(\beta, Z^{(n)})$ ensuring the consistency. Among those who have used related methods, one can cite [7, 18–20, 22, 23, 28, 29, 34] and [3], and more recently [26]. These corrections strongly depend on the exponential form of the relative risk of the Cox model. Indeed, if $U = Z + \varepsilon$, with Z independent of ε, then $\lim_{n \to \infty} \mathbb{E}[L_n^{(1)}(\beta, Z^{(n)})]$ only depends on $\mathbb{E}(Z)$ and on $\mathbb{E}[\exp(\beta Z)]$, with $\mathbb{E}(Z) = \mathbb{E}(U)$ and $\mathbb{E}[\exp(\beta U)] = \mathbb{E}[\exp(\beta Z)]\mathbb{E}[\exp(\beta \varepsilon)]$. Extension of such methods to other relative risks have not been conclusive. For instance, in the semiparametric model of excess relative risk where $f_\beta(Z) = (1 + \beta Z)$, easy calculations give that $\lim_{n \to \infty} \mathbb{E}[L_n^{(1)}(\beta, Z^{(n)})]$ depends on $\mathbb{E}[Z/(1 + \beta Z)]$ whereas $\lim_{n \to \infty} \mathbb{E}[L_n^{(1)}(\beta, U^{(n)})]$ depends on $\mathbb{E}[U/(1 + \beta U)]$. Since the error model $U = Z + \varepsilon$ does not provide any expression of $\mathbb{E}[Z/(1 + \beta Z)]$ in terms of $\mathbb{E}[U/(1 + \beta U)]$, a correction analogous to the ones proposed in the Cox model cannot be exhibited. In other words, it seems impossible to find a function $\Psi_n(\beta, U)$ that is independent of the unknown density g and that satisfies $\lim_{n \to \infty} \mathbb{E}(\Psi_n(\beta, U)) = \mathbb{E}[Z/(1 + \beta Z)]$.

The paper is organized as follows. Section 2 presents the model and the assumptions. In Sections 3 and 4 we present the main estimator and its asymptotic properties. In Section 5 we extend our estimation procedure and propose a second estimator. In Section 6 we give detailed examples. The proofs are given in Sections 7 and 8.

2. Model, assumptions and notations

Notation: For two complex-valued functions u and v in $\mathbb{L}_2(\mathbb{R}) \cap \mathbb{L}_1(\mathbb{R})$, we define $u^*(x) = \int e^{itx} u(t) dt$, $u \star v(x) = \int u(y) v(x - y) dy$, and $< u, v > = \int u(x) \overline{v}(x) dx$ with $\overline{z}$ the conjugate of a complex number z. We also use $\|u\|_1 = \int |u(x)| dx$, $\|u\|_2^2 = \int |u(x)|^2 dx$, $\|u\|_\infty = \sup_{x \in \mathbb{R}} |u(x)|$, and for $\theta \in \mathbb{R}^d$, $\| \theta \|_{\ell^2}^2 = \sum_{k=1}^{d} \theta_k^2$. For a map $(\theta, u) \mapsto \varphi_\theta(u)$ from $\Theta \times \mathbb{R}$ to $\mathbb{R}$, the first and second derivatives with respect to θ are denoted by

$$\varphi_\theta^{(1)}(\cdot) = \left(\varphi_{\theta,j}^{(1)}(\cdot) \right)_j \text{ with } \varphi_{\theta,j}^{(1)}(\cdot) = \frac{\partial \varphi_\theta(\cdot)}{\partial \theta_j} \text{ for } j \in \{1, \cdots, m+p\}$$

$$\text{and} \quad \varphi_\theta^{(2)}(\cdot) = \left(\varphi_{\theta,j,k}^{(2)}(\cdot) \right)_{j,k} \text{ with } \varphi_{\theta,j,k}^{(2)}(\cdot) = \frac{\partial^2 \varphi_\theta(\cdot)}{\partial \theta_j \theta_k}, \text{ for } j, k \in \{1, \cdots, m+p\}.$$

Throughout the paper $\mathbb{P}$, $\mathbb{E}$ and Var denote respectively the probability, the expectation, and the variance when the underlying and unknown true parameters are θ^0 and g. Finally, a_- denotes the negative part of a, which is equal to a if $a \leq 0$ and 0 otherwise.

Model assumptions:

$$\text{For all } \gamma \in \Gamma, \ \eta_\gamma \text{ is nonnegative and } \int_0^\tau \eta_\gamma^2(t)dt < \infty. \tag{$\mathbf{A_1}$}$$

$$\text{For all } \beta \in \mathbb{B} \text{ and for all } g \in \mathcal{G}, \text{ if } Z \text{ has density } g, \ f_\beta(Z) \geq 0. \tag{$\mathbf{A_2}$}$$

$$\text{Conditional on } Z \text{ and } U, \text{ the failure time } T \text{ and the censoring time } C \text{ are independent.} \tag{$\mathbf{A_3}$}$$

$$\text{The conditional distribution of the failure time } T \text{ given } (Z,U) \text{ does not depend on } U. \tag{$\mathbf{A_4}$}$$

$$\text{The conditional distribution of the censoring time } C \text{ given } (Z,U) \text{ does not \ depend on } U. \tag{$\mathbf{A_5}$}$$

These assumptions are common in most of the frameworks dealing with survival data analysis and covariates measured with error (see [2, 15, 30, 31, 36]). Concerning $(\mathbf{A_2})$, as is mentioned in [31], a sufficient requirement would be to assume that $f_\beta(z) \geq 0$ for all $z \in \mathbb{R}$. But this condition is too strong in general, and does not allow one to consider regression forms of particular interest, such as linear form $f_\beta(z) = 1 + \beta z$. We only assume in $(\mathbf{A_2})$ that $f_\beta(z) \geq 0$ for all z in the support of the density g of the covariate Z and for all $\beta \in \mathbb{B}$. Assumption $(\mathbf{A_3})$ states that a general censorship model is considered. Assumption $(\mathbf{A_4})$ and $(\mathbf{A_5})$ state that both the failure time and the censoring time are independent of the observed covariate when the observed and true covariates are both given, i.e. the measurement error is not prognostic.

Smoothness assumptions

$$\text{The functions } \beta \mapsto f_\beta \text{ and } \gamma \mapsto \eta_\gamma \text{ admit continuous derivatives up to order 3} \tag{$\mathbf{A_6}$}$$
$$\text{with respect to } \beta \text{ and } \gamma \text{ respectively.}$$

We denote by $S_{\theta^0,g}^{(1)}(\theta)$ and $S_{\theta^0,g}^{(2)}(\theta)$ the first and second derivatives of $S_{\theta^0,g}(\theta)$ with respect to θ. For all t in $[0,\tau]$, let $S_{\theta^0,g}^{(2)}(\theta,t)$ be the second derivative of $S_{\theta^0,g}$ when the integral is taken over $[0,t]$ in (1.2), with the convention that $S_{\theta^0,g}^{(2)}(\theta) = S_{\theta^0,g}^{(2)}(\theta,\tau)$.

Identifiability and moment assumptions

$$S_{\theta^0,g}^{(1)}(\theta) = 0 \text{ if and only if } \theta = \theta^0. \tag{$\mathbf{A_7}$}$$

For all $t \in]0,\tau]$, the matrix $S_{\theta^0,g}^{(2)}(\theta^0,t)$ exists and is positive definite. $\qquad\qquad$ ($\mathbf{A_8}$)

For all $j = 1,\cdots,p$, $\int_0^\tau \eta_{\gamma^0}^4(t)dt < \infty$, and $\int_0^\tau |\eta_{\gamma^0,j}^{(1)}|^3(t)\eta_{\gamma^0}(t)dt < \infty$. $\qquad$ ($\mathbf{A_9}$)

For $\gamma \in \Gamma$ and $j = 1,\cdots,p$, $\mathbb{E}(\int_0^\tau \eta_\gamma(t)dN(t))^2 < \infty$ and $\mathbb{E}(\int_0^\tau \eta_{\gamma^0,j}^{(1)}(t)dN(t))^2 < \infty$. $\qquad$ ($\mathbf{A_{10}}$)

$\sup_{g\in\mathcal{G}} \| f_{\beta^0}g \|_2^2 \le C_1(f_{\beta^0})$, $\sup_{g\in\mathcal{G}} \| f_{\beta^0}^2 g \|_2^2 \le C_2(f_{\beta^0}^2)$. $\qquad\qquad$ ($\mathbf{A_{11}}$)

The function W is such that for all $\beta \in \mathbb{B}$ and $g \in \mathcal{G}$, $\mathbb{E}[(f_\beta^2 W)(Z)]$ is finite . $\qquad$ ($\mathbf{A_{12}}$)

The function W is such that for all $g \in \mathcal{G}$ and for $j = 1,\cdots,m$, $\mathbb{E}[(f_{\beta^0}^6 W)(Z)]$ $\qquad$ ($\mathbf{A_{13}}$)

and $\mathbb{E}[(|f_{\beta^0,j}^{(1)}|^3 |f_{\beta^0}W)(Z)|]$ are finite.

We can use the equality (1.2), to see see that $S_{\theta^0,g}(\theta)$ is minimum at $\theta = \theta^0$. Assumptions ($\mathbf{A_7}$) and ($\mathbf{A_8}$) ensure that θ^0 is the unique minimum. The density g and the parameter β vary over sets $\mathcal{G}$ and $\mathbb{B}$, such that ($\mathbf{A_2}$), ($\mathbf{A_7}$), ($\mathbf{A_8}$), ($\mathbf{A_{10}}$) and ($\mathbf{A_{11}}$) hold.

3. Estimation procedure

If the Z_i's were observed, $S_{\theta^0,g}(\theta)$ would be estimated by

$$\tilde{S}_n(\theta) = -\frac{2}{n}\sum_{i=1}^n (f_\beta W)(Z_i) \int_0^\tau \eta_\gamma(t)dN_i(t) + \frac{1}{n}\sum_{i=1}^n (f_\beta^2 W)(Z_i) \int_0^\tau \eta_\gamma^2(t)Y_i(t)dt.$$

Since the Z_i's are not observable we estimate $S_{\theta^0,g}$ by

$$S_{n,1}(\theta) = \frac{1}{n}\sum_{i=1}^n \left[(f_\beta^2 W) \star K_{n,C_n}(U_i) \int_0^\tau \eta_\gamma^2(t)Y_i(t)dt - 2(f_\beta W) \star K_{n,C_n}(U_i) \int_0^\tau \eta_\gamma(t)dN_i(t) \right]$$

where $K_{n,C_n}(\cdot) = C_n K_n(C_n \cdot)$ is a deconvolution kernel defined through a kernel K, f_ε and a sequence C_n *via*: $K_n^*(t) = K^*(t)/f_\varepsilon^*(-tC_n)$. The kernel K is such that K^* is compactly supported satisfying $|1 - K^*(t)| \le \mathbb{1}_{|t|\ge 1}$, $\int K(t)dt = 1$, and $C_n \to \infty$ as $n \to \infty$. By construction, the deconvolution kernel K_{n,C_n} also satisfies $K_{n,C_n}^*(t) = K_{C_n}^*(t)/f_\varepsilon^*(-t) = K^*(t/C_n)/\overline{f_\varepsilon^*}(t)$.

For the construction of $S_{n,1}(\theta)$ we require that

$$\text{the density } f_\varepsilon \text{ belongs to } \mathbb{L}_2(\mathbb{R}) \cap \mathbb{L}_\infty(\mathbb{R}) \text{ and for all } x \in \mathbb{R}, f_\varepsilon^*(x) \ne 0. \qquad (\mathbf{N_1})$$

The key ideas for this construction are the following: For any integrable function Φ, $\lim_{n\to\infty} n^{-1}\sum_{i=1}^n \Phi \star K_{n,C_n}(U_i) = \mathbb{E}(\Phi(Z))$. Hence we estimate $\mathbb{E}(\Phi(Z))$ by $n^{-1}\sum_{i=1}^n \Phi \star K_{n,C_n}(U_i)$ instead of $n^{-1}\sum_{i=1}^n \Phi(Z_i)$

which is not available. Similarly, for any $\psi \in \mathbb{L}_1(\mathbb{R})$ and Φ such that $\mathbb{E}(\Phi(Z)) < \infty$,

$$\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \psi(X_i) \Phi \star K_{n,C_n}(U_i) = \mathbb{E}(\psi(X)\Phi(Z)).$$

Indeed, if $f_{X,U,Z}$ is the joint distribution of (X, U, Z), Assumptions $(\mathbf{A_4})$-$(\mathbf{A_5})$ and the independence between Z and ε imply that $f_U = g \star f_\varepsilon$ and $f_{X,U,Z}(x, u, z) = f_{X,Z}(x, z) f_\varepsilon(u - z)$. Consequently, under mild conditions $S_{n,1}(\theta) \xrightarrow[n\to\infty]{\mathbb{P}} S_{\theta^0,g}(\theta)$ for all $\theta \in \Theta$ and we propose to estimate θ^0 by

$$\widehat{\theta}_1 = (\widehat{\beta}_1, \widehat{\gamma}_1)^\top = \arg\min_{\theta=(\beta,\gamma)^\top \in \Theta} S_{n,1}(\theta). \tag{3.1}$$

4. Asymptotic properties

4.1. General results for the risk of $\widehat{\theta}_1$

The weight function W is chosen such that

$$\sup_{\beta\in\mathbb{B}}(Wf_\beta), \ W \text{ and } \sup_{\beta\in\mathbb{B}}(Wf_\beta^2) \text{ belong to } \mathbb{L}_1(\mathbb{R}). \tag{$\mathbf{A_{14}}$}$$

$$\sup_{\beta\in\mathbb{B}}(Wf_\beta^{(1)}) \text{ and } \sup_{\beta\in\mathbb{B}}(Wf_\beta f_\beta^{(1)}) \text{ belong to } \mathbb{L}_1(\mathbb{R}). \tag{$\mathbf{A_{15}}$}$$

We say that a function $\psi \in \mathbb{L}_1(\mathbb{R})$ satisfies (4.2) if for a sequence C_n we have

$$\min_{q=1,2} \| \psi^*(K_{C_n}^* - 1) \|_q^2 + n^{-1} \min_{q=1,2} \Big\| \frac{\psi^* K_{C_n}^*}{f_\varepsilon^*} \Big\|_q^2 = o(1). \tag{4.2}$$

We note that for any integrable function ψ, one can always find C_n such that (4.2) hold.

Theorem 4.1. *Let $(\mathbf{A_1})$-$(\mathbf{A_{15}})$ and $(\mathbf{N_1})$ hold. Let $\widehat{\theta}_1 = \widehat{\theta}_1(C_n)$ be defined by (3.1).*

1) For all the sequences C_n such that Wf_β and Wf_β^2 and their first derivatives with respect to β satisfy (4.2), $\mathbb{E}(\|\widehat{\theta}_1(C_n) - \theta^0\|_{\ell^2}^2) = o(1)$ as $n \to \infty$, and $\widehat{\theta}_1 = \widehat{\theta}_1(C_n)$ is consistent.

2) Assume moreover that for all $\beta \in \mathbb{B}$, $f_\beta W$ and $f_\beta^2 W$ and their derivatives up to order 3 with respect to β satisfy (4.2). Then $\mathbb{E}(\| \widehat{\theta}_1 - \theta^0 \|_{\ell^2}^2) = O(\varphi_n^2)$ with $\varphi_n^2 = \|(\varphi_{n,j})\|_{\ell^2}^2$, $\varphi_{n,j}^2 = B_{n,j}^2(\theta^0) + V_{n,j}(\theta^0)/n$, $B_{n,j}^2(\theta^0) = \min\{B_{n,j}^{[1]}(\theta^0), B_{n,j}^{[2]}(\theta^0)\}$, $V_{n,j}(\theta^0) = \min\{V_{n,j}^{[1]}(\theta^0), V_{n,j}^{[2]}(\theta^0)\}$, $j = 1, \cdots, m$ where

$$B_{n,j}^{[q]}(\theta^0) = \big\|(f_{\beta^0}^2 W)^*(K_{C_n}^* - 1)\big\|_q^2 + \big\|(f_{\beta^0} W)^*(K_{C_n}^* - 1)\big\|_q^2 + \big\|(f_{\beta^0,j}^{(1)} W)^*(K_{C_n}^* - 1)\big\|_q^2$$

$$+ \big\|(f_{\beta^0,j}^{(1)} f_{\beta^0} W)^*(K_{C_n}^* - 1)\big\|_q^2,$$

and

$$V_{n,j}^{[q]}(\theta^0) = \big\|(f_{\beta^0}^2 W)^* \frac{K_{C_n}^*}{f_\varepsilon^*}\big\|_q^2 + \big\|(f_{\beta^0} W)^* \frac{K_{C_n}^*}{f_\varepsilon^*}\big\|_q^2 + \big\|\left(f_{\beta^0,j}^{(1)} W\right)^* \frac{K_{C_n}^*}{f_\varepsilon^*}\big\|_q^2 + \big\|(f_{\beta^0,j}^{(1)} f_{\beta^0} W)^* \frac{K_{C_n}^*}{f_\varepsilon^*}\big\|_q^2.$$

The terms $B_{n,j}^2$ and $V_{n,j}$ are the squared bias and variance terms, respectively. As usual, the bias is the smallest for the smoothest functions $(W f_\beta)(z)$ and $\partial(f_\beta W)(z)/\partial\beta$, as functions of z. As in density deconvolution, or for regression function estimation in errors-in-variables models, the biggest variances are obtained for the smoothest error density f_ε. Hence, the slowest rates are obtained for the smoothest error density f_ε, for instance for Gaussian ε's. Consequently, a good choice of W can improve $\widehat\theta_1$'s rate of convergence by smoothing $W f_\beta$. The rate for estimating β^0 depends on the smoothness properties of $\partial(f_\beta W)(z)/\partial\beta$ and $\partial(f_\beta^2 W^2)(z)/\partial\beta$ as functions of z, whereas the rate for estimating γ^0 depends on the smoothness properties of $(f_\beta W)(z)$ and $(f_\beta^2 W)(z)$ as functions of z. In both cases, the smoothness properties of η_γ, as a function of t, do not influence the rate of convergence. The parametric rate of convergence is achieved as soon as $(f_\beta W)$ and $(f_\beta^2 W)$ and their derivatives, as functions of z, are smoother than the error density f_ε.

4.2. Sufficient conditions for $\sqrt{n}$-consistency

We say that $(\mathbf{C_1})$-$(\mathbf{C_3})$ hold if

there exists a weight function W such that the functions $\quad(\mathbf{C_1})$

$\sup_{\beta\in\mathbb{B}}(f_\beta W)^*/f_\varepsilon^*$ and $\sup_{\beta\in\mathbb{B}}(f_\beta^2 W)^*/f_\varepsilon^*$ belong to $\mathbb{L}_1(\mathbb{R})\cap\mathbb{L}_2(\mathbb{R})$;

the functions $\sup_{\beta\in\mathbb{B}}\left(f_\beta^{(1)} W\right)^*/f_\varepsilon^*$ and $\sup_{\beta\in\mathbb{B}}\left(f_\beta^{(1)} f_\beta W\right)^*/f_\varepsilon^*$ belong to $\mathbb{L}_1(\mathbb{R})\cap\mathbb{L}_2(\mathbb{R})$; $\quad(\mathbf{C_2})$

for all $\beta\in\mathbb{B}$, $\left(f_\beta^{(2)} W\right)^*/f_\varepsilon^*$ and $\left(\frac{\partial^2(f_\beta^2 W)}{\partial\beta^2}\right)^*/f_\varepsilon^*$ belong to $\mathbb{L}_1(\mathbb{R})\cap\mathbb{L}_2(\mathbb{R})$. $\quad(\mathbf{C_3})$

Theorem 4.2. *Under the assumptions of Theorem 4.1 and under $(\mathbf{C_1})$-$(\mathbf{C_3})$, one can find C_n such that $\widehat\theta_1 = \widehat\theta_1(C_n)$ defined by (3.1) is a $\sqrt{n}$-consistent estimator of θ^0. Moreover $\sqrt{n}(\widehat\theta_1 - \theta^0) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,\Sigma_1)$, where Σ_1 equals*

$$\left(\mathbb{E}\big[-2\int_0^\tau \frac{\partial^2((f_\beta W)(Z)\eta_\gamma(s))}{\partial\theta^2}\big|_{\theta=\theta^0} dN(s) + \int_0^\tau \frac{\partial^2((f_\beta^2 W)(Z)\eta_\gamma^2(s))}{\partial\theta^2}\big|_{\theta=\theta^0} Y(s)ds\big]\right)^{-1} \times$$

$$\Sigma_{0,1} \times \left(\mathbb{E}\big[-2\int_0^\tau \frac{\partial^2((f_\beta W)(Z)\eta_\gamma(s))}{\partial\theta^2}\big|_{\theta=\theta^0} dN(s) + \int_0^\tau \frac{\partial^2((f_\beta^2 W)(Z)\eta_\gamma^2(s))}{\partial\theta^2}\big|_{\theta=\theta^0} Y(s)ds\big]\right)^{-1}$$

with

$$\Sigma_{0,1} = \mathbb{E}\left\{ \left[-2\int_0^\tau \frac{\partial(R_{\beta,f_\varepsilon,1}(U)\eta_\gamma(s))}{\partial\theta} \mid_{\theta=\theta^0} dN(s) + \int_0^\tau \frac{\partial(R_{\beta,f_\varepsilon,2}(U)\eta_\gamma^2(s))}{\partial\theta} \mid_{\theta=\theta^0} Y(s)ds \right] \right.$$

$$\left. \times \left[-2\int_0^\tau \frac{\partial(R_{\beta,f_\varepsilon,1}(U)\eta_\gamma(s))}{\partial\theta} \mid_{\theta=\theta^0} dN(s) + \int_0^\tau \frac{\partial(R_{\beta,f_\varepsilon,2}(U)\eta_\gamma^2(s))}{\partial\theta} \mid_{\theta=\theta^0} Y(s)ds \right]^\top \right\},$$

$$R_{\beta,f_\varepsilon,1}(U) = (2\pi)^{-1}\int (Wf_\beta)^*(t)\frac{e^{-itU}}{f_\varepsilon^*(t)}dt \quad and \quad R_{\beta,f_\varepsilon,2}(U) = (2\pi)^{-1}\int (Wf_\beta^2)^*(t)\frac{e^{-itU}}{f_\varepsilon^*(t)}dt.$$

Conditions $(\mathbf{C_1})$-$(\mathbf{C_3})$ ensure the existence of the functions $R_{\beta,f_\varepsilon,j}$ for $j = 1, 2$ and hence the $\sqrt{n}$-consistency. Nevertheless it is not always possible to find W such that $(\mathbf{C_1})$-$(\mathbf{C_3})$ hold.

4.3. **Rates for general smoothness classes**

We now specify the asymptotic properties of $\widehat{\theta}_1$ when f_ε^* and $(Wf_\beta)^*$ satisfy assumptions $(\mathbf{N_2})$ and $(\mathbf{R_1})$ given below.

> There exist positive constants $\underline{C}(f_\varepsilon), \overline{C}(f_\varepsilon)$, and nonnegative δ, α, u_0 and $\rho \leq 2$ $(\mathbf{N_2})$
>
> such that $\underline{C}(f_\varepsilon) \leq |f_\varepsilon^*(u)| |u|^\alpha \exp(\delta |u|^\rho) \leq \overline{C}(f_\varepsilon)$ for all $|u| \geq u_0$.

When $\rho = 0 = \delta$ in $(\mathbf{N_2})$, with the convention that $\delta = 0$ if and only if $\rho = 0$, f_ε is called " ordinary smooth ". An example of an ordinary smooth density is the double exponential (also called Laplace) distribution with $\rho = 0 = \delta$ and $\alpha = 2$. The square integrability of f_ε in $(\mathbf{N_1})$ requires that $\alpha > 1/2$ when $\rho = 0$ in $(\mathbf{N_2})$. When $\delta > 0$ and $\rho > 0$, f_ε is infinitely differentiable and is called " super smooth ". The standard examples for super smooth densities are the Gaussian and Cauchy distributions. The smoothness of $f_\beta W$ is described by :

> A function f satisfies $(\mathbf{R_1})$ if f belongs to $\mathbb{L}_1(\mathbb{R} \cap \mathbb{L}_2(\mathbb{R})$ and if there exist $a, d,$ $(\mathbf{R_1})$
>
> $u_0', r \geq 0$ such that $0 < \underline{L}(f) \leq |f^*(u)||u|^a \exp(d|u|^r) \leq \overline{L}(f) < \infty$ for all $|u| \geq u_0',$

with the convention that $d = 0$ if and only if $r = 0$.

Corollary 4.1. *Under the assumptions of Theorem 4.1, assume that f_ε satisfies $(\mathbf{N_2})$ and that for all $\beta \in \mathbb{B}$, $(f_\beta W)$, $(f_\beta^2 W)$ and their derivatives with respect to β_j, $j = 1, \ldots, m$ up to order 3, satisfy $(\mathbf{R_1})$. Consider the sequences C_n such that*

$$C_n^{(2\alpha-2a+1-\rho)+(1-\rho)-} \exp\{-2dC_n^r + 2\delta C_n^\rho\}/n = o(1) \ as \ n \to +\infty. \tag{4.3}$$

Then $\widehat{\theta}_1$ is a consistent estimator of θ^0 and $\mathbb{E}(\|\widehat{\theta}_1 - \theta^0\|^2_{\ell^2}) = O(\varphi_n^2)$ with φ_n^2 as in Table 1.

		f_ε	
		$\rho = 0$ in $(\mathbf{N_2})$ ordinary smooth	$\rho > 0$ in $(\mathbf{N_2})$ super smooth
Wf_{β^0}	$d = r = 0$ in $(\mathbf{R_1})$ Sobolev	$a < \alpha + 1/2 \quad n^{-\frac{2a-1}{2\alpha}}$ $a \geq \alpha + 1/2 \quad n^{-1}$	$(\log n)^{-\frac{2a-1}{\rho}}$
	$r > 0$ in $(\mathbf{R_1})$ $\mathcal{C}^\infty$	n^{-1}	$r < \rho \quad (\log n)^{A(a,r,\rho)} \exp\left\{-2d\left(\frac{\log n}{2\delta}\right)^{r/\rho}\right\}$ $r = \rho$: $d < \delta \quad (\log n)^{A(a,r,\rho)+2\alpha d/(\delta r)} n^{-d/\delta}$ $d = \delta, a < \alpha+1/2 \quad (\log n)^{(2\alpha-2a+1)/r} n^{-1}$ $d = \delta, a \geq \alpha+1/2 \quad n^{-1}$ $d > \delta \quad n^{-1}$ $r > \rho \quad n^{-1}$

where $A(a,r,\rho) = (-2a + 1 - r + (1-r)_-)/\rho$.

TABLE 1. Rates of convergence φ_n^2 of $\widehat{\theta}_1$

5. EXTENSION OF THE ESTIMATION PROCEDURE : A SECOND ESTIMATOR $\widehat{\theta}_2$

Our estimation procedure requires the estimation of the two following linear functionals of the density g, $\mathbb{E}[\int_0^\tau (f_\beta W)(Z) dN(t)]$ and $\mathbb{E}[\int_0^\tau (f_\beta^2 W)(Z) Y(t) dt]$. We now study the particular cases in which these linear functionals can be directly estimated without using a kernel deconvolution plug-in.

We say that conditions $(\mathbf{C_4})$-$(\mathbf{C_6})$ hold if there exist a weight function W and two functions $\Phi_{\beta,f_\varepsilon,1}$ and $\Phi_{\beta,f_\varepsilon,2}$ not depending on g, such that for all $\beta \in \mathbb{B}$ and for all g

$$\mathbb{E}\Big[\int_0^\tau (f_\beta W)(Z) dN(t)\Big] = \mathbb{E}\Big[\int_0^\tau \Phi_{\beta,f_\varepsilon,1}(U) dN(t)\Big], \mathbb{E}\Big[\int_0^\tau (f_\beta^2 W)(Z) Y(t) dt\Big] = \mathbb{E}\Big[\int_0^\tau \Phi_{\beta,f_\varepsilon,2}(U) Y(t) dt\Big]; \quad (\mathbf{C_4})$$

$$\text{for } k = 0,1,2 \text{ and for } j = 1,2, \quad \mathbb{E}[\sup_{\beta \in \mathbb{B}} \|\Phi_{\beta,f_\varepsilon,j}^{(k)}(U)\|_{\ell^2}] < \infty; \quad (\mathbf{C_5})$$

$$\text{for } j = 1,2 \text{ and for all } \beta \in \mathbb{B}, \quad \mathbb{E}\left[\|\Phi_{\beta,f_\varepsilon,j}^{(1)}(U)\|_{\ell^2}^2\right] < \infty. \quad (\mathbf{C_6})$$

Under $(\mathbf{C_4})$-$(\mathbf{C_6})$, we estimate θ^0 by $\widehat{\theta}_2 = \underset{\theta \in \Theta}{\arg\min}\, S_{n,2}(\theta)$ where

$$S_{n,2}(\theta) = -\frac{2}{n}\sum_{i=1}^n \int_0^\tau \Phi_{\beta,f\varepsilon,1}(U_i)\eta_\gamma(t)dN_i(t) + \frac{1}{n}\sum_{i=1}^n \int_0^\tau \Phi_{\beta,f\varepsilon,2}(U_i)\eta_\gamma^2(t)Y_i(t)dt. \tag{5.1}$$

The main difficulty for finding such functions $\Phi_{\beta,f_\varepsilon,1}$ and $\Phi_{\beta,f_\varepsilon,2}$ lies in the constraint that they must not depend on the unknown density g. We refer to Section 5.2 for the construction of such functions.

5.1. Asymptotic properties of $\widehat{\theta}_2$

Theorem 5.1. *Let $(\mathbf{A_1})$ -$(\mathbf{A_{13}})$, and conditions $(\mathbf{C_4})$-$(\mathbf{C_6})$ hold. Then $\widehat{\theta}_2$, defined by (5.1) is a $\sqrt{n}$-consistent estimator of θ^0. Moreover $\sqrt{n}(\widehat{\theta}_2 - \theta^0) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,\Sigma_2)$, where Σ_2 is equal to*

$$\left[\mathbb{E}\Big(-2\int_0^\tau \frac{\partial^2(\Phi_{\beta,f_\varepsilon,1}(U)\eta_\gamma(s))}{\partial\theta^2}\,|_{\theta=\theta^0}\,dN(s) + \int_0^\tau \frac{\partial^2(\Phi_{\beta,f_\varepsilon,2}(U)\eta_\gamma^2(s))}{\partial\theta^2}\,|_{\theta=\theta^0}\,Y(s)ds\Big)\right]^{-1}$$

$$\times \Sigma_{0,2}\left[\mathbb{E}\Big(-2\int_0^\tau \frac{\partial^2(\Phi_{\beta,f_\varepsilon,1}(U)\eta_\gamma(s))}{\partial\theta^2}\,|_{\theta=\theta^0}\,dN(s) + \int_0^\tau \frac{\partial^2(\Phi_{\beta,f_\varepsilon,2}(U)\eta_\gamma^2(s))}{\partial\theta^2}\,|_{\theta=\theta^0}\,Y(s)ds\Big)\right]^{-1}$$

with

$$\Sigma_{0,2} = \mathbb{E}\Big[\Big(-2\int_0^\tau \frac{\partial(\Phi_{\beta,f_\varepsilon,1}(U)\eta_\gamma(s))}{\partial\theta}\,|_{\theta=\theta^0}\,dN(s) + \int_0^\tau \frac{\partial(\Phi_{\beta,f_\varepsilon,2}(U)\eta_\gamma^2(s))}{\partial\theta}\,|_{\theta=\theta^0}\,Y(s)ds\Big)$$

$$\times \Big(-2\int_0^\tau \frac{\partial(\Phi_{\beta,f_\varepsilon,1}(U)\eta_\gamma(s))}{\partial\theta}\,|_{\theta=\theta^0}\,dN(s) + \int_0^\tau \frac{\partial(\Phi_{\beta,f_\varepsilon,2}(U)\eta_\gamma^2(s))}{\partial\theta}\,|_{\theta=\theta^0}\,Y(s)ds\Big)^\top\Big].$$

5.2. Comments:

Let us briefly compare conditions $(\mathbf{C_1})$-$(\mathbf{C_3})$ and $(\mathbf{C_4})$-$(\mathbf{C_6})$. It is noteworthy that conditions $(\mathbf{C_4})$-$(\mathbf{C_6})$ are more general. For instance, condition $(\mathbf{C_4})$ does not require that $f_\beta W$, $f_\beta^2 W$ belong to $\mathbb{L}_1(\mathbb{R})$ (as for instance in the Cox Model with $W \equiv 1$). Moreover $(\mathbf{C_1})$ implies $(\mathbf{C_4})$, with $\Phi_{\beta,f_\varepsilon,j} = R_{\beta,f_\varepsilon,j}$. Indeed, under $(\mathbf{C_1})$-$(\mathbf{C_3})$, if we define $\Phi_{\beta,f_\varepsilon,1}^* = (Wf_\beta)^*/\overline{f_\varepsilon^*}$ and $\Phi_{\beta,f_\varepsilon,2}^* = (Wf_\beta^2)^*/\overline{f_\varepsilon^*}$, we have

$$\mathbb{E}[Y(t)\Phi_{\beta,f_\varepsilon,2}(U)] = \iint \mathbb{1}_{x\geq t}f_{X,Z}(x,z)\frac{1}{2\pi}\int \frac{(Wf_\beta^2)^*(s)}{\overline{f_\varepsilon^*}(s)}e^{-isz}\overline{f_\varepsilon^*}(s)ds\,dx\,dz = \mathbb{E}[Y(t)(Wf_\beta^2)(Z)].$$

Consequently, $\mathbb{E}\big[\int_0^\tau \Phi_{\beta,f_\varepsilon,2}(U)\eta_\gamma^2(t)Y(t)dt\big] = \mathbb{E}\big[\int_0^\tau (f_\beta^2 W)(Z)\eta_\gamma^2(t)dt\big]$, and analogously $\mathbb{E}\big[\int_0^\tau \Phi_{\beta,f_\varepsilon,1}(U)\eta_\gamma(t)dN(t)\big] = \mathbb{E}\big[\int_0^\tau (f_\beta W)(Z)\eta_\gamma(t)dN(t)\big]$. Hence Condition $(\mathbf{C_4})$ holds and $\Sigma_{0,1} = \Sigma_{0,2}$ with $\Sigma_{0,1}$ defined in Theorem 4.2.

The choice of W is very important, as illustrated in Example 4 where we consider $f_\beta(z) = 1 - \beta + \beta/(1+z^2)$ and f_ε is the Gaussian density. In this case, when $W \equiv 1$ it seems impossible to find a function $\Phi_{\beta,f_\varepsilon,2}$ such that $\mathbb{E}[Y(t)\Phi_{\beta,f_\varepsilon,2}(U)] = \mathbb{E}[Y(t)f_\beta^2(Z)]$. Whereas $(\mathbf{C_1})$-$(\mathbf{C_3})$ and $(\mathbf{C_4})$-$(\mathbf{C_6})$ hold by taking $W(z) = (1 + z^2)^4 \exp(-z^2/(4\delta))$ where δ is as $(\mathbf{N_2})$.

To summarize: $\widehat{\theta}_1$ always exists and is consistent under mild regularity conditions, though $\widehat{\theta}_1$'s rate of convergence is not always $\sqrt{n}$. By judicious choice of W the parametric rate of convergence can be achieved in some cases. In contrast, the computation of $\widehat{\theta}_2$ is more straightforward than the computation of $\widehat{\theta}_1$ since it does not require deconvolution estimators, but $\widehat{\theta}_2$ does not always exist.

5.3. Case without errors

If the covariates are measured without errors, that is $U_i = Z_i$ and $\varepsilon_i = 0$ for $i = 1, \cdots, n$, then the procedure still works by taking $f_\varepsilon^* \equiv 1$ in the previous formulae. More precisely, in this context $S_{n,1}(\theta)$ becomes

$$S_{n,1}^{we}(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left[(f_\beta^2 W) \star K_{C_n}(U_i)\int_0^\tau \eta_\gamma^2(t)Y_i(t)dt - 2(f_\beta W) \star K_{C_n}(U_i)\int_0^\tau \eta_\gamma(t)dN_i(t)\right]$$

where $K_{C_n}(\cdot) = C_n K(C_n \cdot)$, $K_{C_n}^*(t) = K^*(t/C_n)$ and the kernel K is as in Section 3. Under the previous assumptions, easy calculations show that $S_{n,1}^{we}(\theta)$ converges to $S_{\theta^0,g}(\theta)$ for all $\theta \in \Theta$ since for any integrable function Φ we have

$$n^{-1}\sum_{i=1}^{n}\Phi \star K_{C_n}(U_i) = n^{-1}\sum_{i=1}^{n}\Phi \star K_{C_n}(Z_i) \xrightarrow[n\to\infty]{\mathbb{P}} \mathbb{E}\big[(\Phi(Z)\big].$$

In this context $\widehat{\theta}_1^{we} = \underset{\theta \in \Theta}{\arg\min}\, S_{n,1}^{we}(\theta)$ is a consistent estimator of θ^0. Moreover it is easy to find W such that conditions $(\mathbf{C_1})$-$(\mathbf{C_3})$ hold when $f_\varepsilon^* \equiv 1$. It follows that $\widehat{\theta}_1^{we}$ is a $\sqrt{n}$-consistent and asymptotically Gaussian estimator of θ^0. We can analogously define $\widehat{\theta}_2^{we}$ and show that it is a $\sqrt{n}$-consistent and asymptotically Gaussian estimator of θ^0.

6. Examples

In this section, we illustrate the asymptotic properties of $\widehat{\theta}_1$ and $\widehat{\theta}_2$ for various relative risks. In all of these examples, $K^*(t) = \mathbb{1}_{|t|\leq 1}$, the baseline function has a nonspecified parametric form and f_ε satisfies $(\mathbf{N_1})$ and $(\mathbf{N_2})$ with $0 \leq \rho \leq 2$.

The first example considers the Cox model. We show that our estimation procedure, based on a nonparametric deconvolution method, provides a $\sqrt{n}$-consistent and asymptotically Gaussian estimator of β^0. The aim of

this example is to show that we recover the $\sqrt{n}$-consistency in a slightly different model (parametric baseline function), and using estimators quite different from the ones proposed by [3, 23, 29]. The other examples show that our estimation procedure provides consistent estimators, and even sometimes $\sqrt{n}$-consistent estimators for general parametric hazard functions.

Example 1. Exponential relative risk (Cox model).

Let f_β be of the form $f_\beta(z) = \exp(\beta z)$ and assume that $\mathbb{E}(\exp(\beta U)) < \infty$ for all β in $\mathbb{B}$. Let $W(z) = \exp\{-z^2/(4\delta)\}$ where δ is as in $(\mathbf{N_2})$. Then conditions $(\mathbf{C_1})$-$(\mathbf{C_3})$ as well as conditions $(\mathbf{C_4})$-$(\mathbf{C_6})$ are satisfied. Hence the estimators $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are $\sqrt{n}$-consistent and asymptotically Gaussian estimators of θ^0, with the same asymptotic variance.

One could also choose $W \equiv 1$ and use the independence between Z and ε to find that $\mathbb{E}[\exp(\beta Z)] = \mathbb{E}[\exp(\beta U)]/\mathbb{E}[\exp(\beta \varepsilon)]$. Consequently, if we choose $\Phi_{\beta,f_\varepsilon,1}(U) = \exp(2\beta U)/\mathbb{E}[\exp(2\beta \varepsilon)]$ and $\Phi_{\beta,f_\varepsilon,2}(U) = \exp(\beta U)/\mathbb{E}[\exp(\beta \varepsilon)]$ we get that $\mathbb{E}[\Phi_{\beta,f_\varepsilon,1}(U)] = \mathbb{E}[f_\beta^2(Z)]$ and also that $\mathbb{E}[Y(t)\Phi_{\beta,f_\varepsilon,2}(U)] = \mathbb{E}[Y(t)f_\beta(Z)]$. It follows that conditions $(\mathbf{C_4})$-$(\mathbf{C_6})$ hold and the criterion $S_{n,2}$ defined by (5.1) exists. In this case $\widehat{\theta}_2$ is still a $\sqrt{n}$-consistent and asymptotically Gaussian estimator of θ^0.

Example 2. Excess relative risk model. Let f_β be of the form $f_\beta(z) = 1 + \beta z$ with β and g in $\mathbb{B}$ and $\mathcal{G}$ such that for $Z \sim g$, $1 + \beta Z$ is nonnegative. For example $\mathcal{G}$ could consist of densities having support included in $[0, +\infty)$ and $\mathbb{B}$ could be a compact set included in $[0, \infty)$, or $\mathcal{G}$ could consist of densities having support included in $[-1, +1]$ and $\mathbb{B}$ could be a compact set included in $[-1, +1]$. Let $W(z) = \exp\{-z^2/(4\delta)\}$ where δ is as in $(\mathbf{N_2})$. Then conditions $(\mathbf{C_1})$-$(\mathbf{C_3})$ as well as conditions $(\mathbf{C_4})$-$(\mathbf{C_6})$ are satisfied. Hence the estimators $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are $\sqrt{n}$-consistent and asymptotically Gaussian estimators of θ^0, with the same asymptotic variance.

Example 3. Polynomial relative risk 1. Let f_β be of the form $f_\beta(z) = 1 + \sum_{k=1}^{m} \beta_k z^k$ with $m \geq 1$, and with β and g in $\mathbb{B}$ and $\mathcal{G}$ such that $f_\beta(Z)$ is nonnegative for $Z \sim g$. For example $\mathcal{G}$ could consist of densities having support included in $[0, +\infty)$ and $\mathbb{B}$ could be a compact set included in $[0, \infty)^m$, or $\mathcal{G}$ could consist of densities having support included in $[-1, +1]$ and $\mathbb{B}$ could be a compact set included in the set $\{\beta = (\beta_k)_{1 \leq k \leq m} \in \mathbb{R}^m$ such that $\sum_{k=1}^{m} |\beta_k| \leq 1\}$. Let $W(z) = \exp\{-z^2/(4\delta)\}$ where δ is as in $(\mathbf{N_2})$. Then conditions $(\mathbf{C_1})$-$(\mathbf{C_3})$ as well as conditions $(\mathbf{C_4})$-$(\mathbf{C_6})$ are satisfied. Hence the estimators $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are $\sqrt{n}$-consistent and asymptotically Gaussian estimators of θ^0, with the same asymptotic variance.

One can also choose $W \equiv 1$, provided that $\mathbb{E}(|U|^m) < \infty$ and that the kernel K has finite absolute moments of order m and satisfies $\int u^r K(u) du = 0$, for $r = 1, \ldots, m$. With this choice of W, $\widehat{\theta}_1$ remains a $\sqrt{n}$-consistent and asymptotically Gaussian estimator of θ^0.

Example 4. Cauchy relative risk 1. Consider f_β of the form $f_\beta(z) = 1 - \beta + \beta/(1+z^2)$ with β and g in $\mathbb{B}$ and $\mathcal{G}$ such that $f_\beta(Z)$ is nonnegative for $Z \sim g$. For example $\mathcal{G}$ could consist of all densities and $\mathbb{B}$ could be a compact set included in $]0,1[$ or $(-\infty, 0[$. The relative risk f_β has the regularity of $z \mapsto 1/(1+z^2)$ which satisfies $(\mathbf{R_1})$ with $a = 0$, $d = 1/2$ and $r = 1$. Let $W(z) = (1+z^2)^4 \exp\{-z^2/(4\delta)\}$ where δ is as in $(\mathbf{N_2})$. The functions $f_\beta W$, $f_\beta^2 W$ and their derivatives with respect to β up to order 3 satisfy $(\mathbf{R_1})$ with $\rho < r = 2$ or $\rho = r = 2$ and $d > \delta$. Consequently, conditions $(\mathbf{C_1})$-$(\mathbf{C_3})$ as well as conditions $(\mathbf{C_4})$-$(\mathbf{C_6})$ are satisfied and the estimators $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are $\sqrt{n}$-consistent and asymptotically Gaussian estimators of θ^0, with the same asymptotic variance.

This example underlines the importance of the choice of smoothing weight function W in the construction of $\widehat{\theta}_1$ or $\widehat{\theta}_2$. Indeed, with $W \equiv 1$, Theorem 4.1 predicts a much slower rate of convergence. For example, if ε is a Gaussian random variable and $W \equiv 1$, the predicted rate is of order $\exp(-2\sqrt{\log n})$.

Example 5. Laplace relative risk. Consider f_β of the form $f_\beta(z) = 1 + \beta f(z)$ with $f(z) = \exp(-|z|/2) - 1$ and with β and g in $\mathbb{B}$ and $\mathcal{G}$ such that $f_\beta(Z)$ is nonnegative for $Z \sim g$. For example $\mathcal{G}$ could consist of all densities and $\mathbb{B}$ could be a compact set included in $]0,1[$ or $(-\infty, 0[$. The Fourier transform of $z \mapsto \ell(z) = \exp(-|z|/2)$ equals $\ell^*(t) = 4/(1+4t^2)$ decays slowly. Hence, if we choose $W \equiv 1$, the estimator $\widehat{\theta}_1$ is not $\sqrt{n}$-consistent as soon as $|f_\varepsilon^*(u)| \leq o(|u|^{-2})$ with $|u| \to \infty$. A closer look shows that f_β and its derivative with respect to β are C^∞ except at the point $z = 0$. Therefore, a proper choice of W can smooth out at 0 and make $W f_\beta$, $W f_\beta^2$ and their derivatives with respect to β infinitely differentiable functions in z. This choice of W ensures the $\sqrt{n}$-consistency of $\widehat{\theta}_1$ whenever f_ε satisfies $(\mathbf{N_2})$ with $0 < \rho < 1$. Even if $\rho \geq 1$, the rate of $\widehat{\theta}_1$ is much faster with our choice of W than it would be for $W \equiv 1$. Let us specify the choice of W. Set

$$\Psi_{A,B,R}(z) = \exp\left(-(z-A)^{-R}(B-z)^{-R}\right) I_{[A,B]}(z), \tag{6.1}$$

where $-\infty < A < B < \infty$ are fixed and $R > 0$. According to [12, 24], p. 346, Theorem 7.3, $|\Psi_{A,B,R}^*(u)| \leq c \exp(-C|u|^{R/(R+1)})$, as $|u| \to \infty$ and c, C are positive constants. We propose to take W equal to $\Psi_{0,100,R}$ or $\Psi_{-100,0,R}$ or their sum. This choice of W ensures that $f_\beta W$, $f_\beta^2 W$ and their derivatives with respect to β up to order 3 satisfy $(\mathbf{R_1})$ with $d > 0$ and $r = R/(R+1)$ closer to 1 as R comes larger.

If f_ε satisfies $(\mathbf{N_2})$ with $0 \leq \rho < 1$, we choose R large enough such that $r = R/(R+1) > \rho$. Hence, conditions $(\mathbf{C_1})$-$(\mathbf{C_3})$ as well as conditions $(\mathbf{C_4})$-$(\mathbf{C_6})$ are satisfied. Consequently, the estimators $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are $\sqrt{n}$-consistent and asymptotically Gaussian estimators of θ^0, with the same asymptotic variance.

If $\rho \geq 1$, for this choice of W, the functions Wf_β and Wf_β^2 and their derivatives with respect to β up to order 3, satisfy $(\mathbf{R_1})$ with $r = R/(R+1)$. Hence, according to Table 1,

$$\mathbb{E} \parallel \widehat{\theta}_1 - \theta^0 \parallel_{\ell^2}^2 = O(1)\big(\log n\big)^{\frac{1-2a-r}{\rho}} \exp\{-2d(\log n/(2\delta))^{r/\rho}\}.$$

Example 6. Irregular relative risk. Consider f_β with $f_\beta(z) = 1 + \beta_2 a_- + \beta_1 z + \beta_2(z-a)\mathbb{I}_{z \geq a}$ with β and g in $\mathbb{B}$ and $\mathcal{G}$ such that $f_\beta(Z)$ is nonnegative for $Z \sim g$. For example $\mathcal{G}$ could consist of all densities with support included in $[0, +\infty)$ and $\mathbb{B}$ could be a compact set included in $[0, +\infty)$. This relative risk is $\mathcal{C}^\infty$ except at point a where it is not differentiable. We suggest to use the smoothing weight function in (6.1) as follows. For $R > 0$, let $W(z) = \Psi_{a-100,a,R}(z)$. In that way, Wf_β and Wf_β^2 and their derivatives with respect to β satisfy $(\mathbf{R_1})$ with $0 < r = R/(R+1) < 1$ as close to 1 as needed.

If f_ε satisfies $(\mathbf{N_2})$ with $0 \leq \rho < 1$, then we take R large enough such that $r = R/(R+1) > \rho$ and thus conditions $(\mathbf{C_1})$-$(\mathbf{C_3})$ as well as conditions $(\mathbf{C_4})$-$(\mathbf{C_6})$ are satisfied. Consequently, the estimators $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are $\sqrt{n}$-consistent and asymptotically Gaussian estimators of θ^0, with the same asymptotic variance.

If $\rho \geq 1$ in $(\mathbf{N_2})$, the functions Wf_β and Wf_β^2 and their derivatives with respect to β up to order 3, satisfy $(\mathbf{R_1})$ with $0 < r = R/(R+1) < 1$ as close to 1 as needed. According to Table 1

$$\mathbb{E} \parallel \widehat{\theta}_1 - \theta^0 \parallel_{\ell^2}^2 = O(1)\big(\log n\big)^{\frac{1-2a-r}{\rho}} \exp\{-2d(\log n/(2\delta))^{r/\rho}\}.$$

Comments on Examples 5 and 6.

In these examples, $f_\beta W$ satisfies $(\mathbf{R_1})$ with $r < 1$. Hence $\widehat{\theta}_1$ is $\sqrt{n}$-consistent provided that f_ε is ordinary smooth or super smooth with an exponent $\rho < 1$. For example, when the ε is Gaussian, it seems impossible to find W such that $(Wf_\beta)^*/f_\varepsilon^*$ belongs to $\mathbb{L}_1(\mathbb{R})$. This comes from the fact that for these relative risks, the least square criterion $S_{\theta^0,g}(\theta)$ cannot be estimated with the parametric rate of convergence and hence probably cannot provide a $\sqrt{n}$-consistent estimator of θ^0. Nevertheless, even in cases where $\sqrt{n}$-consistency does not seem achievable, the resulting rate of the risk of $\widehat{\theta}_1$ is clearly much faster than the predicted logarithmic rate that we could have with $W \equiv 1$ (see Table 1).

In survival data analysis the relative risks f_β are often of the form $f_\beta(z) = f(\beta z)$ (see for instance [31]). Let us present some examples of this type.

Example 7. Polynomial relative risk 2. Let f_β be of the form $f(\beta z)$ with $f(z) = 1 + \sum_{k=1}^{m} a_k z^k$ with known a_k's and with β and g in $\mathbb{B}$ and $\mathcal{G}$ such that for $Z \sim g$, $f(\beta Z)$ is nonnegative. For example $\mathcal{G}$ could consist of all densities having support included in $\mathbb{R}^+$ and $\mathbb{B}$ could be a compact set included in $(\mathbb{R}^+)^m$. Let $W(z) = \exp\{-z^2/(4\delta)\}$ where δ is as in $(\mathbf{N_2})$. Then conditions $(\mathbf{C_1})$-$(\mathbf{C_3})$ as well as conditions $(\mathbf{C_4})$-$(\mathbf{C_6})$ are satisfied. Hence the estimators $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are $\sqrt{n}$-consistent and asymptotically Gaussian estimators of θ^0, with the same asymptotic variance.

Example 8. Cauchy relative risk 2. Consider f_β of the form $f(\beta z)$ with $f(z) = 1/(1 + z^2)$. in this example $\mathcal{G}$ could consist of all densities and $\mathbb{B}$ could be a compact set of $\mathbb{R}^m$. Let $W(z) = (1 + z^2)^4 \exp\{-z^2/(4\delta)\}$ with δ as in $(\mathbf{N_2})$ or $W \equiv 1$. With these choices of W, the functions $f_\beta W$, $f_\beta^2 W$ and their derivatives with respect to β up to order 3 satisfy $(\mathbf{R_1})$ with $a = 0$, $d = 1/\beta$ and $r = 1$. According to Table 1, if f_ε satisfies $(\mathbf{N_2})$ with $0 \leq \rho < 1$, then $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are $\sqrt{n}$-consistent and asymptotically Gaussian estimators of θ^0, with the same asymptotic variance. If f_ε satisfies $(\mathbf{N_2})$ with $\rho \geq 1$, then $\widehat{\theta}_1$ is consistent with a rate that depends on β^0. Let us be more precise. The proof of Theorem 4.1 implies that the terms $B_{n,j}^2(\theta^0)$ are of order $\exp(-2C_n/\beta^0)$ and the terms $V_{n,j}(\theta^0)/n$ are of order $C_n^{2\alpha+(1-\rho)+(1-\rho_-)} \exp(-2C_n/\beta^0 + 2\delta C_n^\rho)/n$, for $j = 1, \ldots, m + p$.

Choose C_n^* that provides the best compromise between the squared bias and the variance terms. It is independent of β^0 and is given by

$$C_n^* = \left[(\log n)/(2\delta) - (2\alpha + (1 - \rho)_-)/(2\delta\rho) \log\left(\log n/(2\delta)\right) \right]^{1/\rho}.$$

This choice yields the rate

$$\varphi_n^2 = \max\left\{ n^{-1}, \exp\left[-\frac{2}{\beta^0}\left(\frac{\log n}{2\delta} - \frac{2\alpha + (1-\rho)_-}{2\delta\rho} \log\left(\frac{\log n}{2\delta}\right)\right)^{1/\rho}\right](\log n)^{(1-\rho)/\rho} \right\}.$$

In other words, if $\rho = 1$, then $\mathbb{E} \parallel \widehat{\theta}_1 - \theta^0 \parallel_{\ell^2}^2 = O(1) \max\left\{ n^{-1}, n^{-1/(\beta^0\delta)}(\log n)^{2\alpha/(\beta^0\delta)} \right\}$ and if $\rho > 1$, then $\mathbb{E} \parallel \widehat{\theta}_1 - \theta^0 \parallel_{\ell^2}^2 = O(1) \exp\left[-2(\beta^0)^{-1}\left(\log n/(2\delta)\right)^{1/\rho}\right]$.

7. Proofs

From now C denotes any numerical constant and C_A indicates that it depends on A. We point out that the value of C may vary from a line to the other.

7.1. **Proof of Theorem 4.1**

7.1.1. *Consistency*

It follows from $(\mathbf{A_7})$ and $(\mathbf{A_8})$ and the two points :

1- For all $\theta \in \Theta$, $\mathbb{E}[S_{n,1}(\theta) - S_{\theta^0,g}(\theta)]^2 = o(1)$ as $n \to \infty$, with $S_{\theta^0,g}(\theta)$ defined in (1.2),

2- If $\omega(n,\rho)$ denotes $\omega(n,\rho) = \sup\left\{|S_{n,1}(\theta) - S_{n,1}(\theta')| : \|\theta - \theta'\|_{\ell^2} \leq \rho\right\}$, there exists ρ_k going to 0, such that $\mathbb{E}[\omega(n,\rho_k)]^2 = O(\rho_k^2)$ as $n \to \infty$ $\forall k \in \mathbb{N}$.

Proof of 1- We will break $\mathbb{E}[S_{n,1}(\theta) - S_{\theta^0,g}(\theta)]^2$ into its corresponding bias $|\mathbb{E}[S_{n,1}(\theta)] - S_{\theta^0,g}(\theta)|^2$ and variance $\mathbb{E}[\{S_{n,1}(\theta) - \mathbb{E}[S_{n,1}(\theta)]\}^2]$ components and then show each term is asymptotically $o_p(1)$. For the study of both terms, we repeatedly use the two following versions of Hölder's Inequality

$$| < \varphi_1, \varphi_2 > | \quad \leq \quad \| \varphi_1 \|_2 \| \varphi_2 \|_2, \tag{7.1}$$

$$\text{and } | < \varphi_1, \varphi_2 > | \quad \leq \quad \| \varphi_1 \|_\infty \| \varphi_2 \|_1 . \tag{7.2}$$

Study of the bias.

Under Assumptions $(\mathbf{A_1})$-$(\mathbf{A_5})$ and using that $(Z_i, U_i, N_i(s), D_i, Y_i(s))_{i=1,\cdots,n}$ are independent, the intensity of the censored process $N_i(t) = \mathbb{I}_{X_i \leq t, D_i = 1}$ with respect to $\mathcal{F}_t = \sigma\{Z_i, U_i, N_i(s), \mathbb{I}_{X_i \geq s}, 0 \leq s \leq t \leq \tau, i = 1, \cdots, n\}$ is equal to

$$\lambda_i(t, \theta^0, Z_i) = \eta_{\gamma^0}(t) Y_i(t) f_{\beta^0}(Z_i). \tag{7.3}$$

We use (7.3) and Lemma 8.1 to write

$$\mathbb{E}\big[S_{n,1}(\theta)\big] = \int_0^\tau \mathbb{E}\Big[(f_\beta^2 W) \star K_{C_n}(Z)\eta_\gamma^2(t)Y(t) - 2(f_\beta W) \star K_{C_n}(Z)\eta_\gamma(t)f_{\beta^0}(Z)\eta_{\gamma^0}(t)Y(t)\Big]dt,$$

and hence

$$
\begin{aligned}
\mathbb{E}\big[S_{n,1}(\theta)\big] - S_{\theta^0,g}(\theta) &= \int\int_0^\tau \eta_\gamma^2(t)\mathbb{I}_{x \geq t}\big\langle f_{X,Z}(x,\cdot), (f_\beta^2 W) \star K_{C_n} - f_\beta^2 W\big\rangle dx\, dt \\
&\quad -2\int\int_0^\tau \eta_{\gamma^0}(t)\eta_\gamma(t)\mathbb{I}_{x \geq t}\big\langle f_{\beta^0}(\cdot)f_{X,Z}(x,\cdot), (f_\beta W) \star K_{C_n} - f_\beta W\big\rangle dx\, dt.
\end{aligned}
$$

If we apply (7.1) we obtain the first bound

$$\left| \mathbb{E}(S_{n,1}(\theta)) - S_{\theta^0,g}(\theta) \right| \leq \left(\int_0^\tau \eta_\gamma^2(t)dt \right) \parallel f_{X,Z} \parallel_2 \parallel (f_\beta^2 W) \star K_{C_n} - (f_\beta^2 W) \parallel_2$$

$$+ \left(2 \int_0^\tau \eta_\gamma(t)\eta_{\gamma^0}(t)dt \right) \int \parallel f_{\beta^0}(\cdot)f_{X,Z}(x,\cdot) \parallel_2 dx \parallel (f_\beta W) \star K_{C_n} - f_\beta W \parallel_2 .$$

Now, Parseval's formula gives

$$\left| \mathbb{E}(S_{n,1}(\theta)) - S_{\theta^0,g}(\theta) \right| \leq (2\pi)^{-1} \left(\int_0^\tau \eta_\gamma^2(t)dt \right) \parallel f_{X,Z} \parallel_2 \parallel (f_\beta^2 W)^*(K_{C_n}^* - 1) \parallel_2$$

$$+ (\pi)^{-1} \left(\int_0^\tau \eta_\gamma(t)\eta_{\gamma^0}(t)dt \right) \int \parallel f_{\beta^0}(\cdot)f_{X,Z}(x,\cdot) \parallel_2 dx \parallel (f_\beta W)^*(K_{C_n}^* - 1) \parallel_2$$

that is

$$\left| \mathbb{E}(S_{n,1}(\theta)) - S_{\theta^0,g}(\theta) \right| \leq C_{\gamma,\gamma^0,f_{\beta^0}} \left[\parallel (f_\beta^2 W)^*(K_{C_n}^* - 1) \parallel_2 + \parallel (f_\beta W)^*(K_{C_n}^* - 1) \parallel_2 \right]. \qquad (7.4)$$

We apply (7.2), to get that $\left| \mathbb{E}(S_{n,1}(\theta)) - S_{\theta^0,g}(\theta) \right|$ is also bounded by

$$\parallel (f_\beta^2 W) \star K_{C_n} - (f_\beta^2 W) \parallel_\infty \left(\int \parallel f_{X,Z}(x,\cdot) \parallel_1 dx \right) \left(\int_0^\tau \eta_\gamma^2(t)dt \right)$$

$$+ \parallel (f_\beta W) \star K_{C_n} - (f_\beta W) \parallel_\infty \left(2 \int \parallel f_{\beta^0}(\cdot)f_{X,Z}(x,\cdot) \parallel_1 dx \right) \left(\int_0^\tau \eta_\gamma(t)\eta_{\gamma^0}(t)dt \right)$$

which is also less than

$$\parallel (f_\beta^2 W)^*(K_{C_n}^* - 1) \parallel_1 \left((2\pi)^{-1} \parallel f_{X,Z} \parallel_1 \right) \left(\int_0^\tau \eta_\gamma^2(t)dt \right)$$

$$+ \parallel (f_\beta W)^*(K_{C_n}^* - 1) \parallel_1 \left(\pi^{-1} \int \parallel f_{\beta^0}(\cdot)f_{X,Z}(x,\cdot) \parallel_1 dx \right) \left(\int_0^\tau \eta_\gamma(t)f_{\beta^0}\eta_{\gamma^0}(t)dt \right).$$

This implies

$$\left| \mathbb{E}(S_{n,1}(\theta)) - S_{\theta^0,g}(\theta) \right| \leq C_{\gamma,\gamma^0,f_{\beta^0}} \left[\parallel (f_\beta W)^*(K_{C_n}^* - 1) \parallel_1 + \parallel (f_\beta^2 W)^*(K_{C_n}^* - 1) \parallel_1 \right]. \qquad (7.5)$$

If we combine bounds (7.4) and (7.5) we get

$$\left| \mathbb{E}(S_{n,1}(\theta)) - S_{\theta^0,g}(\theta) \right| \leq C_{\gamma,\gamma^0,f_{\beta^0}}$$

$$\times \min \left\{ \parallel (f_\beta W)^*(K_{C_n}^* - 1) \parallel_2 + \parallel (f_\beta^2 W)^*(K_{C_n}^* - 1) \parallel_2, \right.$$

$$\left. \parallel (f_\beta W)^*(K_{C_n}^* - 1) \parallel_1 + \parallel (f_\beta^2 W)^*(K_{C_n}^* - 1) \parallel_1 \right\}. \qquad (7.6)$$

We apply Lemma 8.2 and obtain $\left|\mathbb{E}(S_{n,1}(\theta)) - S_{\theta^0,g}(\theta)\right|^2 = O\left(C_n^{-2a+1-r+(1-r)_-}\exp(-2dC_n^r)\right) = o(1)$.

Study of the variance. Since we consider independent and identically distributed random variables, we obtain the variance $\mathrm{Var}[S_{n,1}(\theta)] = (2 + o(1))n^{-1}(A_1 + A_2)$, with

$$
\begin{aligned}
A_1 &= \mathbb{E}\Big[(f_\beta^2 W) \star K_{n,C_n}(U) \int_0^\tau \eta_\gamma^2(t)Y(t)dt\Big]^2, \\
A_2 &= 4\mathbb{E}\Big[(f_\beta W) \star K_{n,C_n}(U) \int_0^\tau \eta_\gamma(t)dN(t)\Big]^2.
\end{aligned}
$$

We apply (7.2) and Lemma 8.1 to obtain that A_1 is less than

$$
\big(\int_0^\tau \eta_\gamma^2(t)dt\big)^2 \int \Big|\big\langle f_{X,Z}(x,\cdot) \star f_\varepsilon, ((f_\beta^2 W) \star K_{n,C_n})^2\big\rangle\Big|dx
$$

$$
\leq \big(\int_0^\tau \eta_\gamma^2(t)dt\big)^2 \big(\int \parallel f_{X,Z}(x,\cdot) \star f_\varepsilon \parallel_\infty dx\big) \parallel (f_\beta^2 W) \star K_{n,C_n} \parallel_2^2
$$

and hence

$$
A_1 \leq (2\pi)^{-1}\big(\int_0^\tau \eta_\gamma^2(t)dt\big)^2 \parallel f_\varepsilon \parallel_\infty \parallel f_{X,Z} \parallel_1 \Big\|\frac{(f_\beta^2 W)^* K_{C_n}^*}{f_\varepsilon^*}\Big\|_2^2.
$$

We now give a first bound for A_2. If we denote $\varphi(X,Z) = \int_0^\tau \eta_\gamma(t)dN(t)$ and apply Lemma 8.1 and (7.2), we obtain that A_2 is bounded by

$$
4\int \big\langle (\varphi^2(x,\cdot)f_{X,Z}(x,\cdot)) \star f_\varepsilon, ((f_\beta W) \star K_{n,C_n})^2\big\rangle dx
$$

$$
\leq 4\big(\int \parallel (\varphi^2(x,\cdot)f_{X,Z}(x,\cdot)) \star f_\varepsilon \parallel_\infty dx\big) \parallel (f_\beta W) \star K_{n,C_n} \parallel_2^2.
$$

Since $\int \parallel \varphi^2(x,\cdot)f_{X,Z}(x,\cdot) \parallel_1 dx = \mathbb{E}\big[\int_0^\tau \eta_\gamma(t)dN(t)\big]^2$, we get

$$
\int \parallel \varphi^2(x,\cdot)f_{X,Z}(x,\cdot)) \star f_\varepsilon \parallel_\infty dx \leq \parallel f_\varepsilon \parallel_\infty \mathbb{E}\big[\int_0^\tau \eta_\gamma(t)dN(t)\big]^2.
$$

Consequently,

$$
A_2 \leq 4(2\pi)^{-1}\mathbb{E}\big(\int_0^\tau \eta_\gamma(t)dN(t)\big)^2 \parallel f_\varepsilon \parallel_\infty \Big\|\frac{(f_\beta W)^* K_{C_n}^*}{f_\varepsilon^*}\Big\|_2^2,
$$

$$
\text{and} \quad \mathrm{Var}[S_{n,1}(\theta)] \leq C_{\theta^0,\parallel f_\varepsilon\parallel_\infty} n^{-1}\Big[\big\|(f_\beta W)^* \frac{K_{C_n}^*}{f_\varepsilon^*}\big\|_2^2 + \big\|(f_\beta^2 W)^* \frac{K_{C_n}^*}{f_\varepsilon^*}\big\|_2^2\Big]. \tag{7.7}
$$

We apply (7.2) and obtain that A_1 is also less than

$$\left(\int_0^\tau \eta_\gamma^2(t)dt\right)^2 \int \left|\left\langle f_{X,Z}(x,\cdot) \star f_\varepsilon, ((f_\beta^2 W) \star K_{n,C_n})^2\right\rangle\right| dx$$

$$\leq \left(\int_0^\tau \eta_\gamma^2(t)dt\right)^2 \left(\parallel \int f_{X,Z}(x,\cdot) \parallel_1 dx\right) \parallel (f_\beta^2 W) \star K_{n,C_n} \parallel_\infty^2 .$$

Similarly, A_2 is less than

$$4 \int \left\langle (\varphi^2(x,\cdot)f_{X,Z}(x,\cdot)) \star f_\varepsilon, ((f_\beta W) \star K_{n,C_n})^2\right\rangle dx$$

$$\leq 4\left(\int \parallel (\varphi^2(x,\cdot)f_{X,Z}(x,\cdot)) \star f_\varepsilon \parallel_1 dx\right) \parallel (f_\beta W) \star K_{n,C_n} \parallel_\infty^2$$

where $\varphi(X,Z)$ is still defined by $\varphi(X,Z) = \int_0^\tau \eta_\gamma(t)dN(t)$. Once again, since

$$\int \parallel (\varphi^2(x,\cdot)f_{X,Z}(x,\cdot)) \star f_\varepsilon \parallel_1 dx = \mathbb{E}\left(\int_0^\tau \eta_\gamma(t)dN(t)\right)^2,$$

$$\mathrm{Var}[S_{n,1}(\theta)] \leq C_{\theta^0} n^{-1}\left[\left\|(f_\beta W)^* \frac{K_{C_n}^*}{f_\varepsilon^*}\right\|_1^2 + \left\|(f_\beta^2 W)^* \frac{K_{C_n}^*}{f_\varepsilon^*}\right\|_1^2\right]. \tag{7.8}$$

We combine (7.7) and (7.8) and have

$$\mathrm{Var}[S_{n,1}(\theta)] \leq \frac{C_{\theta^0, \|f_\varepsilon\|_\infty}}{n} \min\left\{\left\|(f_\beta W)^* \frac{K_{C_n}^*}{f_\varepsilon^*}\right\|_2^2 + \left\|(f_\beta^2 W)^* \frac{K_{C_n}^*}{f_\varepsilon^*}\right\|_2^2,\right.$$

$$\left.\left\|(f_\beta W)^* \frac{K_{C_n}^*}{f_\varepsilon^*}\right\|_1^2 + \left\|(f_\beta^2 W)^* \frac{K_{C_n}^*}{f_\varepsilon^*}\right\|_1^2\right\}. \tag{7.9}$$

We apply Lemma 8.2 and obtain $\mathrm{Var}[S_{n,1}(\theta)] = O\left(C_n^{2(\alpha-a)+1-\rho+(1-\rho)-} \exp(-2dC_n^r + 2\delta C_n^\rho)/n\right)$, and hence under (4.3), $\mathbb{E}\left[S_{n,1}(\theta) - S_{\theta^0,g}(\theta)\right]^2 = o(1)$, as $n \to \infty$.

Proof of 2- By definition $S_{n,1}(\theta) - S_{n,1}(\theta')$ is equal to

$$-\frac{2}{n}\int_0^\tau \left[(f_\beta W) \star K_{n,C_n}(U_i)\eta_\gamma(t) - (f_{\beta'} W) \star K_{n,C_n}(U_i)\eta_{\gamma'}(t)\right]dN_i(t)$$

$$+\frac{1}{n}\int_0^\tau \left[(f_\beta^2 W) \star K_{n,C_n}(U_i)\eta_\gamma^2(t) - (f_{\beta'}^2 W) \star K_{n,C_n}(U_i)\eta_{\gamma'}^2(t)\right]Y_i(t)dt.$$

Under $(\mathbf{A_6})$, $(\mathbf{A_{14}})$, $(\mathbf{A_{15}})$, $(\mathbf{N_2})$ and $(\mathbf{R_1})$, for C_n satisfying (4.3), and since $\|\theta - \theta'\|_{\ell^2} \leq \rho_k$, we get $\mathbb{E}(|S_{n,1}(\theta) - S_{n,1}(\theta')|^2) = O(\rho_k^2)$. Hence **2-** follows. $\qquad\square$

7.1.2. Rate of convergence

Denote by $S_{n,1}^{(1)}(\theta)$ and $S_{n,1}^{(2)}(\theta)$ the first and second derivatives of $S_{n,1}(\theta)$ with respect to θ. We use classical Taylor expansion and the consistency of $\widehat{\theta}_1$ to get $0 = S_{n,1}^{(1)}(\widehat{\theta}_1) = S_{n,1}^{(1)}(\theta^0) + S_{n,1}^{(2)}(\theta^0)(\widehat{\theta}_1 - \theta^0) + R_n(\widehat{\theta}_1 - \theta^0)$, with R_n defined by

$$R_n = \int_0^1 [S_{n,1}^{(2)}(\theta^0 + s(\widehat{\theta}_1 - \theta^0)) - S_{n,1}^{(2)}(\theta^0)]ds. \tag{7.10}$$

This implies

$$\widehat{\theta}_1 - \theta^0 = -[S_{n,1}^{(2)}(\theta^0) + R_n]^{-1} S_{n,1}^{(1)}(\theta^0). \tag{7.11}$$

Consequently, we have to verify the four following points

i) $\mathbb{E}\big[\big(S_{n,1}^{(1)}(\theta^0)) - S_{\theta^0,g}^{(1)}(\theta^0)\big)\big(S_{n,1}^{(1)}(\theta^0)) - S_{\theta^0,g}^{(1)}(\theta^0)\big)^\top\big] = O[\varphi_n \varphi_n^\top]$, with φ_n defined in Theorem 4.1.

ii) $\mathbb{E}\big[S_{n,1}^{(2)}(\theta^0) - S_{\theta^0,g}^{(2)}(\theta^0)\big]^2 = o(1)$,

iii) R_n defined in (7.10) satisfies $\mathbb{E}(\| R_n \|_{\ell^2}^2) = o(1)$ as $n \to \infty$.

iv) $\mathbb{E}\|\widehat{\theta}_1 - \theta^0\|_{\ell^2}^2 \leq C_{m,p}\mathbb{E}\big[(S_{n,1}^{(1)}(\theta^0))^\top S_{n,1}^{(1)}(\theta^0)\big] + o(\varphi_n^2)$.

The rate of convergence of $\widehat{\theta}_1$ is thus given by the order of $S_{n,1}^{(1)}(\theta^0)$.

Proof of i)

Once again we decompose $\mathbb{E}\big[\big(S_{n,1}^{(1)}(\theta^0)) - S_{\theta^0,g}^{(1)}(\theta^0)\big)\big(S_{n,1}^{(1)}(\theta^0)) - S_{\theta^0,g}^{(1)}(\theta^0)\big)^\top\big]$ in its bias and variance components and study the order of each component. To be specific, we first show that for $j = 1, \cdots m$

$$\big|\mathbb{E}\big(\partial S_{n,1}(\theta)/(\partial \beta_j) \mid_{\theta=\theta^0}\big)\big|^2 \leq C_{\theta^0} \times \min_{q=1,2} \Big\{ \big\|(f_{\beta^0,j}^{(1)} W)^*(K_{C_n}^* - 1)\big\|_q^2 + \big\|(f_{\beta^0,j}^{(1)} f_{\beta^0} W)^*(K_{C_n}^* - 1)\big\|_q^2 \Big\}, \tag{7.12}$$

and for $j = 1, \cdots p$

$$\big|\mathbb{E}\big(\partial S_{n,1}(\theta)/(\partial \gamma_j) \mid_{\theta=\theta^0}\big)\big|^2 \leq C_{\theta^0} \times \min_{q=1,2} \Big\{ \big\|(f_{\beta^0} W)^*(K_{C_n}^* - 1)\big\|_q^2 + \big\|(f_{\beta^0}^2 W)^*(K_{C_n}^* - 1)\big\|_q^2 \Big\}. \tag{7.13}$$

Secondly, we will show that for $j = 1, \cdots, m$

$$\begin{aligned}
\mathrm{Var}(\partial S_{n,1}(\theta)/(\partial \beta_j) \mid_{\theta=\theta^0}) \quad &\leq \quad C_{\theta^0,j}\left[\min_{q=1,2} \big\|(f_{\beta^0}^2 W)^* \frac{K_{C_n}^*}{f_\varepsilon^*}\big\|_q^2 + \min_{q=1,2} \big\|(f_{\beta^0}^2 W)^* \frac{K_{C_n}^*}{f_\varepsilon^*}\big\|_q^2 \right. \\
&\quad + \quad \left. \min_{q=1,2} \big\|(f_{\beta^0,j}^{(1)} W)^* \frac{K_{C_n}^*}{f_\varepsilon^*}\big\|_q^2 + \min_{q=1,2} \big\|(f_{\beta^0} W)^* \frac{K_{C_n}^*}{f_\varepsilon^*}\big\|_q^2 \right]. \tag{7.14}
\end{aligned}$$

Study of the bias By definition $S_{n,1}^{(1)}(\theta^0)$ is equal to

$$\frac{2}{n}\sum_{i=1}^{n}\left(\begin{array}{l}-\int_0^\tau (f_{\beta^0}^{(1)}W)\star K_{n,C_n}(U_i)\eta_{\gamma^0}(t)dN_i(t) + \int_0^\tau (f_{\beta^0}f_{\beta^0}^{(1)}W)\star K_{n,C_n}(U_i)\eta_{\gamma^0}^2(t)Y_i(t)dt\\ -\int_0^\tau (f_{\beta^0}W)\star K_{n,C_n}(U_i)\eta_{\gamma^0}^{(1)}(t)dN_i(t) + \int_0^\tau (f_{\beta^0}^2W)\star K_{n,C_n}(U_i)\eta_{\gamma^0}(t)\eta_{\gamma^0}^{(1)}(t)Y_i(t)dt\end{array}\right).$$

Easy calculations give that $\mathbb{E}(\partial S_{n,1}(\theta)/\partial\beta)_{\theta=\theta^0}$ equals

$$-2\mathbb{E}\big[\int_0^\tau (f_{\beta^0}^{(1)}W)\star K_{n,C_n}(U_1)\eta_{\gamma^0}(t)dN_1(t)\big] + 2\mathbb{E}\big[\int_0^\tau (f_{\beta^0}f_{\beta^0}^{(1)}W)\star K_{n,C_n}(U_1)\eta_{\gamma^0}^2(t)Y_1(t)dt\big].$$

Hence, Lemma 8.1 implies

$$\begin{aligned}\mathbb{E}\big(\partial S_{n,1}(\theta)/(\partial\beta)\mid_{\theta=\theta^0}\big) &=& -2\mathbb{E}\big[f_{\beta^0}(Z_1)(f_{\beta^0}^{(1)}W)\star K_{C_n}(Z_1)\int_0^\tau \eta_{\gamma^0}^2(t)Y_1(t)dt\big]\\ && +2\mathbb{E}\big[(f_{\beta^0}f_{\beta^0}^{(1)}W)\star K_{C_n}(Z_1)\int_0^\tau \eta_{\gamma^0}^2(t)Y_1(t)dt\big].\end{aligned}$$

Since $\partial S_{\theta^0,g}(\theta)/(\partial\beta)\mid_{\theta=\theta^0}$ is equal to

$$-2\mathbb{E}\big[f_{\beta^0}(Z_1)(f_{\beta^0}^{(1)}W)(Z_1)\int_0^\tau \eta_{\gamma^0}^2(t)Y_1(t)dt\big] + 2\mathbb{E}\big[(f_{\beta^0}f_{\beta^0}^{(1)}W)(Z_1)\int_0^\tau \eta_{\gamma^0}^2(t)Y_1(t)dt\big] = 0,$$

we get $\mathbb{E}\big(\partial S_{n,1}(\theta)/\partial\beta\mid_{\theta=\theta^0}\big) = \mathbb{E}\big(\partial S_{n,1}(\theta)/\partial\beta\mid_{\theta=\theta^0}\big) - \partial S_{\theta^0,g}(\theta)/(\partial\beta)\mid_{\theta=\theta^0}$ which also equals

$$\begin{aligned}&-2\int \Big\langle f_{\beta^0}(\cdot)f_{X,Z}(x,\cdot), [(f_{\beta^0}^{(1)}W)\star K_{C_n} - (f_{\beta^0}^{(1)}W)]\Big\rangle \int_0^\tau \eta_{\gamma^0}^2(t)\mathbb{1}_{x\geq t}dt\, dx\\ &\qquad +2\int \Big\langle f_{X,Z}(x,\cdot), [(f_{\beta^0}f_{\beta^0}^{(1)}W)\star K_{C_n} - (f_{\beta^0}f_{\beta^0}^{(1)}W)]\Big\rangle \int_0^\tau \eta_{\gamma^0}^2(t)\mathbb{1}_{x\geq t}dt\, dx.\end{aligned}$$

Similarly, Lemma 8.1 implies

$$\begin{aligned}\mathbb{E}\big(\partial S_{n,1}(\theta)/(\partial\gamma)\mid_{\theta=\theta^0}\big) &=& -2\mathbb{E}\big[f_{\beta^0}(Z_1)(f_{\beta^0}W)\star K_{C_n}(Z_1)\int_0^\tau \eta_{\gamma^0}^{(1)}(t)\eta_{\gamma^0}(t)Y_1(t)dt\big]\\ && +2\mathbb{E}\big[(f_{\beta^0}^2W)\star K_{C_n}(Z_1)\int_0^\tau \eta_{\gamma^0}^{(1)}(t)\eta_{\gamma^0}(t)Y_1(t)dt\big].\end{aligned}$$

We use that $\partial S_{\theta^0,g}(\theta)/(\partial\gamma)\mid_{\theta=\theta^0}$ equals

$$-2\mathbb{E}\big[f_{\beta^0}(Z_1)(f_{\beta^0}W)(Z_1)\int_0^\tau \eta_{\gamma^0}^{(1)}\eta_{\gamma^0}(t)Y_1(t)dt\big] + 2\mathbb{E}\big[(f_{\beta^0}^2W)(Z_1)\int_0^\tau \eta_{\gamma^0}^{(1)}(t)\eta_{\gamma^0}(t)Y_1(t)dt\big] = 0,$$

we obtain $\mathbb{E}\big(\partial S_{n,1}(\theta)/(\partial\gamma)\,|_{\theta=\theta^0}\big) = \mathbb{E}\big(\partial S_{n,1}(\theta)/\partial\gamma\,|_{\theta=\theta^0}\big) - \partial S_{\theta^0,g}(\theta)/(\partial\gamma)\,|_{\theta=\theta^0}$ equals

$$-2\int\big\langle f_{\beta^0}(\cdot)f_{X,Z}(x,\cdot),[(f_{\beta^0}W)\star K_{C_n} - (f_{\beta^0}W)]\big\rangle \int_0^\tau \eta_{\gamma^0}^{(1)}(t)\eta_{\gamma^0}(t)\mathbb{1}_{x\geq t}dt\,dx$$

$$+2\int\big\langle f_{X,Z}(x,\cdot),[(f_{\beta^0}^2 W)\star K_{C_n} - (f_{\beta^0}^2 W)]\big\rangle \int_0^\tau \eta_{\gamma^0}^{(1)}(t)\eta_{\gamma^0}(t)\mathbb{1}_{x\geq t}dt\,dx.$$

Hence we obtain that for $j = 1,\cdots,m$ $(1/2)\mathbb{E}\big(\partial S_{n,1}(\theta)/\partial\beta_j|_{\theta=\theta^0}\big)$ is less than

$$\|\,(f_{\beta^0,j}^{(1)}W)\star K_{C_n} - (f_{\beta^0,j}^{(1)}W)\,\|_2 \,\big(\int \|\,f_{\beta^0}(\cdot)f_{X,Z}(x,\cdot)\,\|_2\,dx\big)\big(\int_0^\tau \eta_{\gamma^0}^2(t)dt\big)$$

$$+\big\|(f_{\beta^0,j}^{(1)}f_{\beta^0}W)\star K_{C_n} - (f_{\beta^0,j}^{(1)}f_{\beta^0}W)\big\|_2\big(\int \|\,f_{X,Z}(x,\cdot)\,\|_2\,dx\big)\big(\int_0^\tau \eta_{\gamma^0}^2(t)dt\big)$$

$$\leq \big\|(f_{\beta^0,j}^{(1)}W)^*(K_{C_n}^* - 1)\big\|_2\big((2\pi)^{-1}\int \|\,f_{\beta^0}(\cdot)f_{X,Z}(x,\cdot)\,\|_2\,dx\big)\big(\int_0^\tau \eta_{\gamma^0}^2(t)dt\big)$$

$$+\big\|(f_{\beta^0,j}^{(1)}f_{\beta^0}W)^*(K_{C_n}^* - 1)\big\|_2\big((2\pi)^{-1}\int \|\,f_{X,Z}(x,\cdot)\,\|_2\,dx\big)\big(\int_0^\tau \eta_{\gamma^0}^2(t)dt\big).$$

Similarly, $(1/2)\mathbb{E}\big(\partial S_{n,1}(\theta)/\partial\gamma_j|_{\theta=\theta^0}\big)$ is less than

$$\|\,(f_{\beta^0}W)^*(K_{C_n}^* - 1)\,\|_2\,\big((2\pi)^{-1}\int \|\,f_{\beta^0}(\cdot)f_{X,Z}(x,\cdot)\,\|_2\,dx\big)\big(\int_0^\tau |\eta_{\gamma^0,j}^{(1)}(t)|\eta_{\gamma^0}(t)dt\big)$$

$$+\|\,(f_{\beta^0}^2 W)^*(K_{C_n}^* - 1)\,\|_2\,\big((2\pi)^{-1}\int \|\,f_{X,Z}(x,\cdot)\,\|_2\,dx\big)\big(\int_0^\tau |\eta_{\gamma^0,j}^{(1)}(t)|\eta_{\gamma^0}(t)dt\big).$$

Consequently,

$$\big|\mathbb{E}\big(\partial S_{n,1}(\theta)/(\partial\beta_j)\,|_{\theta=\theta^0}\big) - \big(\partial S_{\theta^0,g}(\theta)/(\partial\beta_j)\,|_{\theta=\theta^0}\big)\big|$$

$$\leq C_{\theta^0}\big[\big\|(f_{\beta^0,j}^{(1)}W)^*(K_{C_n}^* - 1)\big\|_2 + \big\|(f_{\beta^0,j}^{(1)}f_{\beta^0}W)^*(K_{C_n}^* - 1)\big\|_2\big], \quad (7.15)$$

and

$$\big|\mathbb{E}\big(\partial S_{n,1}(\theta)/(\partial\gamma_j)\,|_{\theta=\theta^0}\big) - \big(\partial S_{\theta^0,g}(\theta)/(\partial\gamma_j)\,|_{\theta=\theta^0}\big)\big|$$

$$\leq C_{\theta^0}\big[\big\|(f_{\beta^0}W)^*(K_{C_n}^* - 1)\big\|_2 + \big\|(f_{\beta^0}^2 W)^*(K_{C_n}^* - 1)\big\|_2\big]. \quad (7.16)$$

We can also write that $(1/2)\mathbb{E}\big(\partial S_{n,1}(\theta)/\partial\beta_j\,|_{\theta=\theta^0}\big)$ is bounded by

$$\|\,(f_{\beta^0,j}^{(1)}W)\star K_{C_n} - (f_{\beta^0,j}^{(1)}W)\,\|_\infty\,\big(\int \|\,f_{\beta^0}(\cdot)f_{X,Z}(x,\cdot)\,\|_1\,dx\big)\big(\int_0^\tau \eta_{\gamma^0}^2(t)dt\big)$$

$$+\big\|(f_{\beta^0,j}^{(1)}f_{\beta^0}W)\star K_{C_n} - (f_{\beta^0,j}^{(1)}f_{\beta^0}W)\big\|_\infty\big(\int \|\,f_{X,Z}(x,\cdot)\,\|_1\,dx\big)\big(\int_0^\tau \eta_{\gamma^0}^2(t)dt\big)$$

which is less than

$$\left\|(f^{(1)}_{\beta^0,j}W)^*(K^*_{C_n}-1)\right\|_1\left((2\pi)^{-1}\int\parallel f_{\beta^0}(\cdot)f_{X,Z}(x,\cdot)\parallel_1 dx\right)\left(\int_0^\tau \eta^2_{\gamma^0}(t)dt\right)$$
$$+\left\|(f^{(1)}_{\beta^0,j}f_{\beta^0}W)^*(K^*_{C_n}-1)\right\|_1\left((2\pi)^{-1}\int\parallel f_{X,Z}(x,\cdot)\parallel_1 dx\right)\left(\int_0^\tau \eta^2_{\gamma^0}(t)dt\right).$$

Similarly, $(1/2)\left|\mathbb{E}\left(\partial S_{n,1}(\theta)/\partial\gamma_j\mid_{\theta=\theta^0}\right)\right|$ is bounded by

$$\parallel (f_{\beta^0}W)^*(K^*_{C_n}-1)\parallel_1\left((2\pi)^{-1}\int\parallel f_{\beta^0}(\cdot)f_{X,Z}(x,\cdot)\parallel_1 dx\right)\left(\int_0^\tau\left|\eta^{(1)}_{\gamma^0,j}(t)\right|\eta_{\gamma^0}(t)dt\right)$$
$$+\parallel (f^2_{\beta^0}W)^*(K^*_{C_n}-1)\parallel_1\left((2\pi)^{-1}\int\parallel f_{X,Z}(x,\cdot)\parallel_1 dx\right)\left(\int_0^\tau\left|\eta^{(1)}_{\gamma^0,j}(t)\right|\eta_{\gamma^0}(t)dt\right).$$

Consequently,

$$\left|\mathbb{E}\left(\partial S_{n,1}(\theta)/(\partial\beta_j)\mid_{\theta=\theta^0}\right)-\left(\partial S_{\theta^0,g}(\theta)/(\partial\beta_j)\mid_{\theta=\theta^0}\right)\right|$$
$$\leq C_{\theta^0}\left[\left\|(f^{(1)}_{\beta^0,j}W)^*(K^*_{C_n}-1)\right\|_1+\left\|(f^{(1)}_{\beta^0,j}f_{\beta^0}W)^*(K^*_{C_n}-1)\right\|_1\right],\quad(7.17)$$

and

$$\left|\mathbb{E}\left(\partial S_{n,1}(\theta)/(\partial\gamma_j)\mid_{\theta=\theta^0}\right)-\left(\partial S_{\theta^0,g}(\theta)/(\partial\gamma_j)\mid_{\theta=\theta^0}\right|$$
$$\leq C_{\theta^0,j}\left[\left\|(f_{\beta^0}W)^*(K^*_{C_n}-1)\right\|_1+\left\|(f^2_{\beta^0}W)^*(K^*_{C_n}-1)\right\|_1\right].\quad(7.18)$$

We combine (7.15), (7.16), (7.17) and (7.18) to obtain (7.12) and (7.13).

Study of the variance We proceed as in the proof of the consistency and write $\mathrm{Var}\left(\partial S_{n,1}(\theta)/(\partial\beta_j)\mid_{\theta=\theta^0}\right)=(8+o(1))n^{-1}[V_{1,j}+V_{2,j}]$, with

$$V_{1,j}\quad=\quad\mathbb{E}[(f^{(1)}_{\beta^0,j}f_{\beta^0}W)^\star K_{n,C_n}(U_1)\int_0^\tau\eta^2_{\gamma^0}(t)Y_1(t)dt]^2$$
$$\text{and}\qquad V_{2,j}\quad=\quad\mathbb{E}[(f^{(1)}_{\beta^0,j}W)\star K_{n,C_n}(U_1)\int_0^\tau\eta_{\gamma^0}(t)dN_1(t)]^2.$$

Similarly, $\mathrm{Var}(\partial S_{n,1}(\theta)/(\partial\gamma_j))=(8+o(1))n^{-1}[V_{3,j}+V_{4,j}]$, with

$$V_{3,j}\quad=\quad\mathbb{E}[(f^2_{\beta^0}W)\star K_{n,C_n}(U_1)\int_0^\tau\eta^{(1)}_{\gamma^0,j}(t)\eta_{\gamma^0}(t)Y_1(t)dt]^2$$
$$\text{and}\qquad V_{4,j}\quad=\quad\mathbb{E}[(f_{\beta^0}W)\star K_{n,C_n}(U_1)\int_0^\tau\eta^{(1)}_{\gamma^0,j}(t)dN_1(t)]^2.$$

Lemma 8.1 implies that

$$V_{1,j} \leq \left(\int_0^\tau \eta_{\gamma^0}^2(t)dt \right)^2 \int \left| \left\langle f_{X,Z}(x,\cdot) \star f_\varepsilon, \left((f_{\beta^0,j}^{(1)} f_{\beta^0} W) \star K_{n,C_n} \right)^2 \right\rangle \right| dx$$

$$\text{and} \qquad V_{3,j} \leq \left(\int_0^\tau \eta_{\gamma^0,j}^{(1)}(t)\eta_{\gamma^0}(t)dt \right)^2 \int \left| \left\langle f_{X,Z}(x,\cdot) \star f_\varepsilon, \left((f_{\beta^0}^2 W) \star K_{n,C_n} \right)^2 \right\rangle \right| dx.$$

We apply inequalities (7.1) and (7.2) and obtain

$$V_{1,j} \leq \left(\int_0^\tau \eta_{\gamma^0}^2(t)dt \right)^2 \max(\| f_\varepsilon \|_\infty, 1) \min_{q=1,2} \left\| (f_{\beta^0,j}^{(1)} f_{\beta^0} W)^* \frac{K_{C_n}^*}{f_\varepsilon^*} \right\|_q^2, \qquad (7.19)$$

$$\text{and} \qquad V_{3,j} \leq \left(\int_0^\tau \eta_{\gamma^0,j}^{(1)}(t)\eta_{\gamma^0}(t)dt \right)^2 \max(\| f_\varepsilon \|_\infty, 1) \min_{q=1,2} \left\| (f_{\beta^0}^2 W)^* \frac{K_{C_n}^*}{f_\varepsilon^*} \right\|_q^2. \qquad (7.20)$$

Now, we apply Lemma 8.1 and have

$$V_{2,j} \leq \int \left| \left\langle \varphi_2^2(x,\cdot) f_{X,Z}(x,\cdot) \star f_\varepsilon, \left((f_{\beta^0,j}^{(1)} W) \star K_{n,C_n} \right)^2 \right\rangle \right| dx$$

$$\text{and} \qquad V_{4,j} \leq \int \left| \left\langle \varphi_{4,j}^2(x,\cdot) f_{X,Z}(x,\cdot) \star f_\varepsilon, \left((f_{\beta^0} W) \star K_{n,C_n} \right)^2 \right\rangle \right| dx,$$

where $\varphi_2(X,Z) = \int_0^\tau \eta_{\gamma^0}(t)dN(t)$ and $\varphi_{4,j}(X,Z) = \int_0^\tau \eta_{\gamma^0,j}^{(1)}(t)dN(t)$. We apply inequalities (7.1) and (7.2) to get

$$V_{2,j} \leq \mathbb{E}\left(\int_0^\tau \eta_{\gamma^0}(t)dN(t) \right)^2 \max(\| f_\varepsilon \|_\infty, 1) \min_{q=1,2} \left\| (f_{\beta^0,j}^{(1)} W)^* \frac{K_{C_n}^*}{f_\varepsilon^*} \right\|_q^2, \qquad (7.21)$$

$$\text{and} \qquad V_{4,j} \leq \mathbb{E}\left(\int_0^\tau \eta_{\gamma^0,j}^{(1)}(t)dN(t) \right)^2 \max(\| f_\varepsilon \|_\infty, 1) \min_{q=1,2} \left\| (f_{\beta^0} W)^* \frac{K_{C_n}^*}{f_\varepsilon^*} \right\|_q^2. \qquad (7.22)$$

The bound (7.14) follows by combining the bounds (7.19), (7.20), (7.21) and (7.22) on the $V_{k,j}$'s for $k = 1, \ldots, 4$.

Proof of ii)

By definition of $S_{n,1}$, $S_{n,1}^{(2)}(\theta^0) = \dfrac{\partial^2 S_{n,1}(\theta^0)}{\partial \theta^2} = \begin{pmatrix} (S_{n,1}^{(2)})_{1,1} & (S_{n,1}^{(2)})_{1,2} \\ (S_{n,1}^{(2)})_{1,2}^\top & (S_{n,1}^{(2)})_{2,2} \end{pmatrix}$, with

$$(S_{n,1}^{(2)})_{1,2} = -\frac{2}{n} \sum_{i=1}^n (f_{\beta^0}^{(1)} W) \star K_{n,C_n}(U_i) \int_0^\tau (\eta_{\gamma^0}^{(1)}(t))^\top dN_i(t)$$

$$+\frac{1}{n} \sum_{i=1}^n \left(f_{\beta^0}^{(1)} f_{\beta^0} W \right) \star K_{n,C_n}(U_i) \int_0^\tau (\eta_{\gamma^0}^{(1)}(t))^\top \eta_{\gamma^0}(t) Y_i(t) dt,$$

$$(S_{n,1}^{(2)})_{1,1} = -\frac{2}{n} \sum_{i=1}^n (f_{\beta^0}^{(2)} W) \star K_{n,C_n}(U_i) \int_0^\tau \eta_{\gamma^0}(t) dN_i(t)$$

$$+\frac{1}{n} \sum_{i=1}^n \left(\frac{\partial^2 (f_\beta^2 W)}{\partial \beta^2} \Big|_{\theta=\theta^0} \right) \star K_{n,C_n}(U_i) \int_0^\tau \eta_{\gamma^0}^2(t) Y_i(t) dt$$

and

$$
\begin{aligned}
(S_{n,1}^{(2)})_{2,2}(\theta) \;=\; & -\frac{2}{n}\sum_{i=1}^{n}(f_{\beta^0}W)\star K_{n,C_n}(U_i)\int_0^\tau \eta_{\gamma^0}^{(2)}(t)dN_i(t)\\
& +\frac{1}{n}\sum_{i=1}^{n}(f_{\beta^0}^2 W)\star K_{n,C_n}(U_i)\int_0^\tau\left(\frac{\partial^2\eta_\gamma^2(t)}{\partial\gamma^2}\Big|_{\theta=\theta^0}\right)Y_i(t)dt.
\end{aligned}
$$

Under $(\mathbf{R_1})$, for C_n satisfying (4.3), $\mathbb{E}[S_{n,1}^{(2)}(\theta^0)-S_{\theta^0,g}^{(2)}(\theta^0)]^2=o(1)$ and **ii)** is proved.

Proof of iii)

The proof of **iii)** follows by using the smoothness of $\beta\mapsto Wf_\beta$ and $\beta\mapsto Wf_\beta^2$ up to order 3, the smoothness of $\gamma\mapsto\eta_\gamma$ and $\gamma\mapsto\eta_\gamma^2$ and by using the consistency of $\widehat\theta_1$.

Proof of iv)

Let us introduce the random event $E_n=\cap_{j,k}E_{n,j,k}$, where

$$
E_{n,j,k}=\left\{\omega \text{ such that } \Big|\frac{\partial^2 S_{\theta^0,g}(\theta)}{\partial\theta_j\partial\theta_k}-\frac{\partial^2 S_{n,1}(\theta,\omega)}{\partial\theta_j\partial\theta_k}+(R_n)_{j,k}(\omega)\Big|_{\theta=\theta^0}\le\frac{1}{2}\frac{\partial^2 S_{\theta^0,g}(\theta)}{\partial\theta_j\partial\theta_k}\Big|_{\theta=\theta^0}\right\}.
$$

Now, we decompose $\mathbb{E}\|\widehat\theta_1-\theta^0\|_{\ell^2}^2$ on the event E_n and its complementary event in the following way

$$
\mathbb{E}\|\widehat\theta_1-\theta^0\|_{\ell^2}^2=\mathbb{E}[\|\widehat\theta_1-\theta^0\|_{\ell^2}^2\mathbf{1}_{E_n}]+\mathbb{E}[\|\widehat\theta_1-\theta^0\|_{\ell^2}^2\mathbf{1}_{E_n^c}].
$$

We use that $\widehat\theta_1$ and θ^0 belong both to a compact set and get

$$
\mathbb{E}\|\widehat\theta_1-\theta^0\|_{\ell^2}^2\le\mathbb{E}[\|\widehat\theta_1-\theta^0\|_{\ell^2}^2\mathbf{1}_{E_n}]+2\sup_{\theta\in\Theta}\|\theta\|_{\ell^2}^2\mathbb{P}(E_n^c).
$$

Hence, the main part of the proof lies in proving that

$$
\mathbb{E}\|\widehat\theta_1-\theta^0\mathbf{1}_{E_n}\|_{\ell^2}^2\le C_{m,p,\theta^0}\mathbb{E}\left[(S_{n,1}^{(1)}(\theta^0))^\top S_{n,1}^{(1)}(\theta^0)\right]\quad\text{and}\quad\mathbb{P}(E_n^c)=o(\varphi_n^2).\tag{7.23}
$$

We use (7.10) and (7.11) to write

$$
\begin{aligned}
\mathbb{E}[\|\widehat\theta_1-\theta^0\|_{\ell^2}^2\mathbf{1}_{E_n}]\;\le\;&\mathbb{E}\left[(S_{n,1}^{(1)}(\theta^0))^\top[(S_{n,1}^{(2)}(\theta^0)+R_n)^{-1}]^\top(S_{n,1}^{(2)}(\theta^0)+R_n)^{-1}S_{n,1}^{(1)}(\theta^0)\mathbf{1}_{E_n}\right]\\
\le\;&C_{m,p}\sup_{j,k}\Big|\frac{\partial^2 S_{\theta^0,g}(\theta)}{\partial\theta_j\partial\theta_k}\Big|_{\theta=\theta^0}\Big|^{-2}\mathbb{E}\left[(S_{n,1}^{(1)}(\theta^0))^\top S_{n,1}^{(1)}(\theta^0)\right].
\end{aligned}
$$

It thus remains to show that $\mathbb{P}(E_n^c) = o(\varphi_n^2)$. We start by writing that $\mathbb{P}(E_n^c) \leq \sum_{j=1}^{m+p} \sum_{k=1}^{m+p} \mathbb{P}(E_{n,j,k}^c)$, and then apply Markov's inequality for $q > 2$, to obtain

$$\mathbb{P}(E_{n,j,k}^c) \leq \left(\left| \frac{1}{2} \frac{\partial^2 S_{\theta^0,g}(\theta)}{\partial \theta_j \partial \theta_k} \big|_{\theta=\theta^0} \right|^q \right)^{-1} \mathbb{E}\left[\left| \left(\frac{\partial^2 (S_{\theta^0,g}(\theta) - S_{n,1}(\theta))}{\partial \theta_j \partial \theta_k} \big|_{\theta=\theta^0} \right) + (R_n)_{j,k} \right|^q \right].$$

Since $|a+b|^q \leq 2^{q-1}(|a|^q + |b|^q)$, $\left(\left| 0.5 \, \partial^2 S_{\theta^0,g}(\theta)/(\partial \theta_j \partial \theta_k) \big|_{\theta=\theta^0} \right|^q \right) \mathbb{P}(E_{n,j,k}^c)$ is less than

$$2^{q-1} \left| \left(\frac{\partial^2 S_{\theta^0,g}(\theta)}{\partial \theta_j \partial \theta_k} \big|_{\theta=\theta^0} \right) - \mathbb{E}\left[\left(\frac{\partial^2 S_{n,1}(\theta)}{\partial \theta_j \partial \theta_k} \big|_{\theta=\theta^0} \right) \right] \right|^q$$

$$+ 2^{2q-2} \mathbb{E}\left[\left| \mathbb{E}\left(\frac{\partial^2 S_{n,1}(\theta)}{\partial \theta_j \partial \theta_k} \big|_{\theta=\theta^0} \right) - \left(\frac{\partial^2 S_{n,1}(\theta)}{\partial \theta_j \partial \theta_k} \big|_{\theta=\theta^0} \right) \right|^q \right] + 2^{2q-2} \mathbb{E}|(R_n)_{j,k}|^q.$$

Now, we apply Rosenthal's inequality (see Rosenthal (1970), Petrov (1995)) to the sum of variables

$$\partial^2 S_{n,1}(\theta)/(\partial \theta_j \partial \theta_k) \big|_{\theta=\theta^0} - \mathbb{E}\left[\left(\partial^2 S_{n,1}(\theta)/(\partial \theta_j \partial \theta_k) \big|_{\theta=\theta^0} \right) \right] := n^{-1} \sum_{i=1}^{n} W_{n,i,j,k}$$

and write

$$\mathbb{E}\left[\left| n^{-1} \sum_{i=1}^{n} W_{n,i,j,k} \right|^q \right] \leq C_q \left[n^{1-r} \mathbb{E}|W_{n,1,j,k}|^q + n^{-q/2} \mathbb{E}^{q/2}|W_{n,1,j,k}|^2 \right].$$

Take $q = 4$ to get

$$\mathbb{E}\left[\left| \frac{\partial^2 S_{n,1}(\theta)}{\partial \theta_j \partial \theta_k} \big|_{\theta=\theta^0} - \mathbb{E}\left(\frac{\partial^2 S_{n,1}(\theta)}{\partial \theta_j \partial \theta_k} \big|_{\theta=\theta^0} \right) \right|^4 \right] \leq C_4 \left[n^{-3} \mathbb{E}|W_{n,1,j,k}|^4 + n^{-2} \mathbb{E}^2|W_{n,1,j,k}|^2 \right].$$

Therefore under the conditions ensuring that

$$\mathbb{E}\left[\frac{\partial^2 S_{\theta^0,g}(\theta)}{\partial \theta_j \theta_k} \big|_{\theta=\theta^0} - \frac{\partial^2 S_{n,1}(\theta)}{\theta_j \theta_k} \big|_{\theta=\theta^0} \right]^2 = o(1),$$

we have

$$\mathbb{E}\left[\frac{\partial^2 S_{\theta^0,g}(\theta)}{\partial \theta_j \theta_k} \big|_{\theta=\theta^0} - \frac{\partial^2 S_{n,1}(\theta)}{\theta_j \theta_k} \big|_{\theta=\theta^0} \right]^4 = O(\varphi_n^4) = o(\varphi_n^2).$$

Now, we use the definition of R_n and the smoothness properties of the derivatives of (Wf_β) and (Wf_β^2) with respect to β, up to order 3 and get $\mathbb{E}((R_n)_{j,k}^4) = o(\|\widehat{\theta}_1 - \theta^0\|_{\ell^2}^4)$. Thus $\mathbb{P}(E_n^c) = o(\varphi_n^2) + o(\mathbb{E}[\|\widehat{\theta}_1 - \theta^0\|_{\ell^2}^4]) = o(\varphi_n^2)$, and (7.23) is proved. □

7.2. **Proof of Theorem 4.2 : asymptotic normality**

Theorem 4.1, its proof, and conditions $(\mathbf{C_1})$-$(\mathbf{C_3})$ imply that $V_{n,j}(\theta^0) = O(1)$ and the asymptotic normality of $\widehat{\theta}_1$ follows if we verify that $\sqrt{n} \, S_{n,1}^{(1)}(\theta^0) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, \Sigma_1)$, with Σ_1 defined in Theorem 4.2.

Let $H_{n,i}$, $\widehat{H}_{n,i}$, $G_{n,i}$, and $\widehat{G}_{n,i}$ be the processes defined for all $t \in [0,\tau]$ by

$$
\widehat{H}_{n,i}(s) = \begin{pmatrix} \frac{-2}{\sqrt{n}} (f_{\beta^0}^{(1)} W) \star K_{n,C_n}(U_i) \eta_{\gamma^0}(s) \\ \frac{-2}{\sqrt{n}} - (f_{\beta^0} W) \star K_{n,C_n}(U_i) \eta_{\gamma^0}^{(1)}(s) \end{pmatrix}, \ H_{n,i}(s) = \begin{pmatrix} \frac{-2}{\sqrt{n}} (f_{\beta^0}^{(1)} W)(Z_i) \eta_{\gamma^0}(s) \\ \frac{-2}{\sqrt{n}} (f_{\beta^0} W)(Z_i) \eta_{\gamma^0}^{(1)}(s) \end{pmatrix},
$$

$$
\widehat{G}_{n,i}(s) = \begin{pmatrix} \frac{2}{\sqrt{n}} (f_{\beta^0} f_{\beta^0}^{(1)} W) \star K_{n,C_n}(U_i) \eta_{\gamma^0}^2(s) \\ \frac{2}{\sqrt{n}} (f_{\beta^0}^2 W) \star K_{n,C_n}(U_i) \eta_{\gamma^0}^{(1)}(s) \eta_{\gamma^0}(s) \end{pmatrix}, \ G_{n,i}(s) = \begin{pmatrix} \frac{2}{\sqrt{n}} (f_{\beta^0} f_{\beta^0}^{(1)} W)(Z_i) \eta_{\gamma^0}^2(s) \\ \frac{2}{\sqrt{n}} (f_{\beta^0}^2 W)(Z_i) \eta_{\gamma^0}^{(1)}(s) \eta_{\gamma^0}(s) \end{pmatrix}.
$$

Since $\lambda_i(t,\theta^0, Z_i)$ defined in (7.3) is the intensity of the process $N_i(t)$ with respect to the filtration $\mathcal{F}_t$, the associated compensator of the process $N(t)$ is $\Lambda_i(t) = \int_0^t \lambda_i(s,\theta^0, Z_i)ds$ and the process $M_i(t) = N_i(t) - \Lambda_i(t,\theta^0, Z_i)$ is a local square integrable martingale. Consequently, we get

$$
\sqrt{n}\, S_{n,1}^{(1)}(\theta^0) \quad = \quad \sum_{i=1}^n \int_0^\tau \widehat{H}_{n,i}(s) dN_i(s) + \sum_{i=1}^n \int_0^\tau \widehat{G}_{n,i}(s) Y_i(s) ds = A_1 + A_2 + A_3 + A_4
$$

with

$$
A_1 \quad = \quad \sum_{i=1}^n \int_0^\tau H_{n,i}(s) dM_i(s), \quad A_2 = \sum_{i=1}^n \int_0^\tau [\widehat{H}_{n,i}(s) - H_{n,i}(s)] dM_i(s),
$$

$$
A_3 \quad = \quad \sum_{i=1}^n \int_0^\tau [\widehat{H}_{n,i}(s) - H_{n,i}(s)] d\Lambda_i(s,\theta^0, Z_i) \text{ and } A_4 = \sum_{i=1}^n \int_0^\tau [\widehat{G}_{n,i}(s) - G_{n,i}(s)] Y_i(s) ds.
$$

Study of A_1

The term A_1 is a linear combination of stochastic integrals of locally bounded and predictable processes, $H_{n,i}$, with respect to finite variation and local square integrable martingales, $M_i(\cdot)$. Consequently, $\mathbb{E}(A_1) = 0$. Denoting by $< M >$ the predictable variation process of M, we have to satisfy the two following conditions for all t in $[0,\tau]$ (see [2] page 68) :

L1) $\sum_{i=1}^n \int_0^t H_{n,i}(s)(H_{n,i}(s))^\top d < M_i > (s) \xrightarrow[n \to \infty]{\mathbb{P}} \widetilde{\Sigma}_1^2(t)$, with $\widetilde{\Sigma}_1^2(t)$ a positive covariance matrix defined by

$$
\tilde{\Sigma}_1^2(t) = 4\mathbb{E}\left[\int_0^t \begin{pmatrix} (f_{\beta^0}^{(1)} W)(Z_i) \eta_{\gamma^0}(s) \\ (f_{\beta^0} W)(Z_i) \eta_{\gamma^0}^{(1)}(s) \end{pmatrix} \begin{pmatrix} (f_{\beta^0}^{(1)} W)(Z_i) \eta_{\gamma^0}(s) \\ (f_{\beta^0} W)(Z_i) \eta_{\gamma^0}^{(1)}(s) \end{pmatrix}^\top \eta_{\gamma^0}(s) Y_i(s) ds \right]
$$

L2) For all $\epsilon > 0$, $\sum_{i=1}^n \int_0^t H_{n,i}(s)(H_{n,i}(s))^\top \mathbb{1}_{\|H_{n,i}(s)\|_{\ell^2} \geq \epsilon}\, d < M_i > (s) = o_p(1)$.

Proof of **L1)**. Since $< M_i > (\cdot) = \Lambda_i(\cdot, \theta^0, Z_i)$ with $\Lambda_i(s,\theta^0, Z_i)$ defined in (7.3), we have to prove that for all

$t \in [0, \tau]$,

$$\sum_{i=1}^{n} \int_0^t H_{n,i}(s)(H_{n,i}(s))^\top Y_i(s) f_{\beta^0}(Z_i) \eta_{\gamma^0}(s) ds \xrightarrow[n\to\infty]{\mathbb{P}} \widetilde{\Sigma}_1^2(t).$$

We apply the following Lemma, which is a straightforward consequence of the fact that the set of functions $\mathcal{I}_t = \{x \mapsto \mathbb{1}_{x \geq t}\}$ is a $\mathbb{P}$-Glivenko Cantelli class (see [38]).

Lemma 7.1. *For $j = 1, \cdots, m$*

$$\sup_{0 \leq t \leq \tau} \left| \frac{1}{n} \sum_{i=1}^{n} Y_i(t) f_{\beta^0,j}(Z_i)(f_{\beta^0,j}^{(1)} W)(Z_i) - \mathbb{E}[Y(t) f_{\beta^0}(Z)(f_{\beta^0,j}^{(1)} W)(Z)] \right| \xrightarrow[n\to\infty]{a.s.} 0,$$

$$\sup_{0 \leq t \leq \tau} \left| \frac{1}{n} \sum_{i=1}^{n} Y_i(t)(f_{\beta^0}^2 W)(Z_i) - \mathbb{E}[Y(t)(f_{\beta^0}^2 W)(Z)] \right| \xrightarrow[n\to\infty]{a.s.} 0$$

$$\sup_{0 \leq t \leq \tau} \left| \frac{1}{n} \sum_{i=1}^{n} Y_i(t) f_{\beta^0}(Z_i)|(f_{\beta^0,j}^{(1)} W)(Z_i)|^3 - \mathbb{E}[Y(t) f_{\beta^0}(Z)|(f_{\beta^0,j}^{(1)} W)(Z)|^3] \right| \xrightarrow[n\to\infty]{a.s.} 0,$$

$$and \quad \sup_{0 \leq t \leq \tau} \left| \frac{1}{n} \sum_{i=1}^{n} Y_i(t)|(f_{\beta^0} W)(Z_i)|^3 - \mathbb{E}[Y(t)|(f_{\beta^0} W)(Z)|^3] \right| \xrightarrow[n\to\infty]{a.s.} 0.$$

Thus **L1)** is verified.

Proof of **L2)**. We have to verify that for all $j = 1, \ldots, m$

$$\frac{1}{n} \mathbb{E} \Big[\sum_{i=1}^{n} \int_0^t \left((f_{\beta^0,j}^{(1)} W)(Z_i) \eta_{\gamma^0}(s) \right)^2 \mathbb{1}_{|(f_{\beta^0,j}^{(1)} W)(Z_i) \eta_{\gamma^0}(s)| \geq \epsilon \sqrt{n}} f_{\beta^0}(Z_i) \eta_{\gamma^0}(s) Y_i(s) ds \Big] = o(1)$$

and that for all $j = 1, \ldots, p$

$$\frac{1}{n} \mathbb{E} \Big[\sum_{i=1}^{n} \int_0^t \left((f_{\beta^0}^2 W)(Z_i) \eta_{\gamma^0,j}^{(1)}(s) \right)^2 \mathbb{1}_{|(f_{\beta^0}^2 W)(Z_i) \eta_{\gamma^0,j}^{(1)}(s)| \geq \epsilon \sqrt{n}} f_{\beta^0}(Z_i) \eta_{\gamma^0}(s) Y_i(s) ds \Big] = o(1).$$

This is a straightforward consequence of Lemma 7.2 by writing that for $j = 1, \cdots, m$

$$\frac{1}{n} \mathbb{E} \Big[\sum_{i=1}^{n} \int_0^t \left((f_{\beta^0,j}^{(1)} W)(Z_i) \eta_{\gamma^0}(s) \right)^2 \mathbb{1}_{|(f_{\beta^0,j}^{(1)} W)(Z_i) \eta_{\gamma^0}(s)| \geq \epsilon \sqrt{n}} f_{\beta^0}(Z_i) \eta_{\gamma^0}(s) Y_i(s) ds \Big]$$

$$\leq \frac{1}{n\sqrt{n}\epsilon} \mathbb{E} \Big[\sum_{i=1}^{n} \int_0^t |(f_{\beta^0,j}^{(1)} W)(Z_i) \eta_{\gamma^0}(s)|^3 f_{\beta^0}(Z_i) \eta_{\gamma^0}(s) Y_i(s) ds \Big] = o(1)$$

and for $j = 1, \cdots, p$

$$\frac{1}{n}\mathbb{E}\Big[\sum_{i=1}^{n}\int_{0}^{t}\big((f_{\beta^0}^2 W)(Z_i)(\eta_{\gamma^0,j}^{(1)}(s))^2\mathbb{1}_{|(f_{\beta^0}^2 W)(Z_i)\eta_{\gamma^0,j}^{(1)}(s)|\geq \epsilon\sqrt{n}}f_{\beta^0}(Z_i)\eta_{\gamma^0}(s)Y_i(s)ds\Big]$$

$$\leq \frac{1}{\epsilon n\sqrt{n}}\mathbb{E}\Big[\sum_{i=1}^{n}\int_{0}^{t}|(f_{\beta^0}^2 W)(Z_i)|^3|\eta_{\gamma^0,j}^{(1)}(s)|^3 f_{\beta^0}(Z_i)\eta_{\gamma^0}(s)Y_i(s)ds\Big] = o(1).$$

Thus **L2)** is verified.

<div align="center">Study of A_2</div>

Since $\mathbb{E}(A_2) = 0$, we use the following lemma and conclude that $A_2 = o_p(1)$.

Lemma 7.2. *Under (N$_2$) and (R$_1$), for C_n satisfying (4.3) then for $j = 1, \cdots, m$*

$$\sup_{0\leq t\leq \tau}\Big|\frac{1}{n}\sum_{i=1}^{n}Y_i(t)f_{\beta^0}(Z_i)(f_{\beta^0,j}^{(1)}W)\star K_{n,C_n}(U_i) - \mathbb{E}[Y(t)f_{\beta^0}(Z)(f_{\beta^0,j}^{(1)}W)(Z)]\Big| \xrightarrow[n\to\infty]{a.s.} 0,$$

$$and \quad \sup_{0\leq t\leq \tau}\Big|\frac{1}{n}\sum_{i=1}^{n}Y_i(t)f_{\beta^0}(Z_i)(f_{\beta^0}W)\star K_{n,C_n}(U_i) - \mathbb{E}[Y(t)(f_{\beta^0}^2 W)(Z)]\Big| \xrightarrow[n\to\infty]{a.s.} 0.$$

<div align="center">Study of A_3</div>

The term A_3 can be viewed as triangular arrays of row-wise independent centered random, by writing $A_3 = \sum_{i=1}^{n} V_{n,i} + \mathbb{E}(A_3)$, with $\sum_{i=1}^{n} V_{n,i} = A_3 - \mathbb{E}(A_3)$. Consequently, the asymptotic normality follows if we verify that

v-a) $\sum_{i=1}^{n}\mathbb{E}[(V_{n,i})^2] \xrightarrow[n\to\infty]{} \Sigma_3^2$;

v-b) $\mathbb{E}(A_3) = o_p(1)$;

v-c) for all $\epsilon > 0$, $\sum_{i=1}^{n}\mathbb{E}[(V_{n,i})^2\mathbb{1}_{\|V_{n,i}\|_{\ell^2}\geq \epsilon}] \xrightarrow[n\to\infty]{} 0$ (Lindeberg Condition).

By definition, A_3 is equal to

$$-\frac{2}{\sqrt{n}}\sum_{i=1}^{n}\int_{0}^{\tau}\begin{pmatrix}(f_{\beta^0}^{(1)}W)\star K_{n,C_n}(U_i) - (f_{\beta^0}^{(1)}W)(Z_i)\eta_{\gamma^0}(s)\\(f_{\beta^0}W)\star K_{n,C_n}(U_i) - (f_{\beta^0}W)(Z_i)\eta_{\gamma^0}^{(1)}(s)\end{pmatrix}Y_i(s)f_{\beta^0}(Z_i)\eta_{\gamma^0}(s)ds.$$

Let us start with the study of the variance (**v-a**). Under (**C$_1$**)-(**C$_3$**)

$$\mathrm{Var}\Big[-\frac{2}{\sqrt{n}}\sum_{i=1}^{n}\int_{0}^{\tau}(f_{\beta^0}^{(1)}W)\star K_{n,C_n}(U_i) - (f_{\beta^0}^{(1)}W)(Z_i)\eta_{\gamma^0}(s)Y_i(s)f_{\beta^0}(Z_i)\eta_{\gamma^0}(s)ds\Big] = O(1),$$

and

$$\mathrm{Var}\Big[-\frac{2}{\sqrt{n}}\sum_{i=1}^{n}\int_{0}^{\tau}(f_{\beta^0}^{(1)}W)\star K_{n,C_n}(U_i) - (f_{\beta^0}W)(Z_i)\eta_{\gamma^0}^{(1)}(s)Y_i(s)f_{\beta^0}(Z_i)\eta_{\gamma^0}(s)ds\Big] = O(1).$$

It follows that **v-a)** is verified.

We now come to the bias term and write that $\mathbb{E}(A_3)$ is equal to

$$-2\sqrt{n}\begin{pmatrix} \mathbb{E}\Big[\big((f_{\beta^0}^{(1)}W)\star K_{n,C_n}(U) - (f_{\beta^0}^{(1)}W)(Z)\big)f_{\beta^0}(Z_i)\int_0^\tau \eta_{\gamma^0}^2(s)Y(s)ds\Big] \\ \mathbb{E}\Big[\big((f_{\beta^0}W)\star K_{n,C_n}(U_i) - (f_{\beta^0}W)(Z_i)\big)f_{\beta^0}(Z_i)\int_0^\tau \eta_{\gamma^0}^{(1)}(s)\eta_{\gamma^0}(s)Y(s)ds\Big] \end{pmatrix}.$$

Lemma 8.1 implies

$$\mathbb{E}(A_3) = -2\sqrt{n}\begin{pmatrix} \mathbb{E}\Big[\big((f_{\beta^0}^{(1)}W)\star K_{C_n}(Z) - (f_{\beta^0}^{(1)}W)(Z)\big)f_{\beta^0}(Z)\int_0^\tau Y(s)\eta_{\gamma^0}^2(s)ds\Big] \\ \mathbb{E}\Big[\big((f_{\beta^0}W)\star K_{C_n}(Z) - (f_{\beta^0}W)(Z)\big)f_{\beta^0}(Z)\int_0^\tau Y(s)\eta_{\gamma^0}^{(1)}(s)\eta_{\gamma^0}(s)ds\Big] \end{pmatrix}$$

that is $\mathbb{E}(A_3)$ is equal to

$$-2\sqrt{n}\begin{pmatrix} \int \big\langle (f_{\beta^0}^{(1)}W)\star K_{C_n}(z) - (f_{\beta^0}^{(1)}W)(z), f_{\beta^0}(z)f_{X,Z}(x,z)\big\rangle \big(\int_0^\tau \mathbb{1}_{x\geq s}\,\eta_{\gamma^0}^2(s)ds\big)\,dx \\ \int \big\langle (f_{\beta^0}W)\star K_{C_n}(z) - (f_{\beta^0}W)(z), f_{\beta^0}(z)\big\rangle \big(\int_0^\tau \mathbb{1}_{x\geq s}\eta_{\gamma^0}^{(1)}(s)\eta_{\gamma^0}(s)ds\big)dx \end{pmatrix}.$$

For $j = 1,\ldots,m$

$$\Big|\int \big\langle (f_{\beta^0,j}^{(1)}W)\star K_{C_n}(z) - (f_{\beta^0,j}^{(1)}W)(z), f_{\beta^0}(z)f_{X,Z}(x,z)\big\rangle \Big(\int_0^\tau \mathbb{1}_{x\geq s}\eta_{\gamma^0}^2(s)ds\Big)dx\Big|$$

is less than

$$\Big(\int_0^\tau \eta_{\gamma^0}^2(s)ds\Big)\int \big|\big\langle (f_{\beta^0,j}^{(1)}W)\star K_{C_n}(z) - (f_{\beta^0,j}^{(1)}W)(z), f_{\beta^0}(z)f_{X,Z}(x,z)\big\rangle\big|dx.$$

We apply (7.1) and get that it is also less than

$$\Big(\int_0^\tau \eta_{\gamma^0}^2(s)ds\Big)\min\Big\{\int \|(f_{\beta^0,j}^{(1)}W)\star K_{C_n} - (f_{\beta^0,j}^{(1)}W)\|_2\|f_{\beta^0}(\cdot)f_{X,Z}(x,\cdot)\|_2 dx,$$

$$\int \|(f_{\beta^0,j}^{(1)}W)\star K_{C_n} - (f_{\beta^0,j}^{(1)}W)\|_\infty\|f_{\beta^0}(\cdot)f_{X,Z}(x,\cdot)\|_1 dx\Big\}$$

which is less than

$$(2\pi)^{-1}\Big(\int_0^\tau \eta_{\gamma^0}^2(s)ds\Big)\min\Big\{\|(f_{\beta^0,j}^{(1)}W)^*(K_{C_n}^* - 1)\|_2\int \|f_{\beta^0}(\cdot)f_{X,Z}(x,\cdot)\|_2 dx,$$

$$\|(f_{\beta^0,j}^{(1)}W)^*(K_{C_n}^* - 1)\|_\infty\int \|f_{\beta^0}(\cdot)f_{X,Z}(x,\cdot)\|_1 dx\Big\}.$$

Similarly, we obtain that for $j = 1,\ldots,p$

$$\Big|\int \big\langle (f_{\beta^0}W)\star K_{C_n}(z) - (f_{\beta^0}W)(z), f_{\beta^0}(z)\big\rangle \Big(\int_0^\tau \mathbb{1}_{x\geq s}\eta_{\gamma^0}^{(1)}(s)\eta_{\gamma^0}(s)ds\Big)dx\Big|$$

is less than

$$(2\pi)^{-1}\Big(\int_0^\tau |\eta_{\gamma^0}(s)\eta^{(1)}_{\gamma^0,j}(s)|ds\Big)\min\Big\{\|(f_{\beta^0}W)^*(K^*_{C_n}-1)\|_2\int\|f_{\beta^0}(\cdot)f_{X,Z}(x,\cdot)\|_2 dx,$$

$$\|(f_{\beta^0}W)^*(K^*_{C_n}-1)\|_\infty\int\|f_{\beta^0}(\cdot)f_{X,Z}(x,\cdot)\|_1 dx\Big\}.$$

Consequently, under $(\mathbf{R_1})$, $\mathbb{E}(A_3)=O(\sqrt{n}C_n^{-a+(1-r)/2+(1-r)_-/2}\exp(-dC_n^r))$. Under $(\mathbf{C_1})$-$(\mathbf{C_3})$, $\mathrm{Var}(A_3)=O(1)$ and hence C_n can be chosen such that $\mathbb{E}(A_3)=o(1)$ and **v-b)** is verified.

In order to verify **v-c)** (Lindeberg condition) we write that for $j=1,\cdots,m$

$$\frac{1}{n}\mathbb{E}\Big[\sum_{i=1}^n\int_0^t\big((f^{(1)}_{\beta^0,j}W)\star K_{n,C_n}(U_i)\eta_{\gamma^0}(s)\big)^2\mathbb{1}_{|(f^{(1)}_{\beta^0,j}W)\star K_{n,C_n}(U_i)\eta_{\gamma^0}(s)|\geq\epsilon\sqrt{n}}f_{\beta^0}(Z_i)\eta_{\gamma^0}(s)Y_i(s)ds\Big]$$

$$\leq\frac{1}{n\sqrt{n}\epsilon}\mathbb{E}\Big[\sum_{i=1}^n\int_0^t|(f^{(1)}_{\beta^0,j}W)\star K_{n,C_n}(U_i)\eta_{\gamma^0}(s)|^3 f_{\beta^0}(Z_i)\eta_{\gamma^0}(s)Y_i(s)ds\Big]=o(1)$$

and for $j=1,\cdots,p$

$$\frac{1}{n}\mathbb{E}\Big[\sum_{i=1}^n\int_0^t\big((f^2_{\beta^0}W)\star K_{n,C_n}(U_i)\eta^{(1)}_{\gamma^0,j}(s)\big)^2\mathbb{1}_{|(f^2_{\beta^0}W)\star K_{n,C_n}(U_i)\eta^{(1)}_{\gamma^0,j}(s)|\geq\epsilon\sqrt{n}}f_{\beta^0}(Z_i)\eta_{\gamma^0}(s)Y_i(s)ds\Big]$$

$$\leq\frac{1}{\epsilon n\sqrt{n}}\mathbb{E}\Big[\sum_{i=1}^n\int_0^t|(f^2_{\beta^0}W)\star K_{n,C_n}(U_i)|^3|\eta^{(1)}_{\gamma^0,j}(s)|^3 f_{\beta^0}(Z_i)\eta_{\gamma^0}(s)Y_i(s)ds\Big]=o(1).$$

$$\text{Study of }A_4$$

The study of A_4, quite similar to the study of A_3 is omitted. $\qquad\square$

7.3. **Proof of Corollary 4.1 :**

For the proof of Corollary 4.1, we apply Lemma 8.2 to the bias bounds (7.12) and (7.13) and obtain

$$\big|\mathbb{E}\big(\partial S_{n,1}(\theta)/(\partial\beta_j)\mid_{\theta=\theta^0}\big)\big|^2=O\big(C_n^{-2a+1-r+(1-r)_-}\exp(-2dC_n^r)\big),$$

$$\text{and}\quad\big|\mathbb{E}\big(\partial S_{n,1}(\theta)/(\partial\gamma_j)\mid_{\theta=\theta^0}\big)\big|^2=O\big(C_n^{-2a+1-r+(1-r)_-}\exp(-2dC_n^r)\big).$$

We apply Lemma 8.2 to the variance bounds (7.19), (7.20), (7.21), (7.22) and get

$$\mathrm{Var}\big(\partial S_{n,1}(\theta)/(\partial\gamma_j)\big)=O\big(C_n^{2(\alpha-a)+1-\rho+(1-\rho)_-}\exp(-2dC_n^r+2\delta C_n^\rho))/n\big).$$

The end of the proof follows by choosing C_n, that provides the best trade-off between the squared bias and the variance. We refer to Butucea and Taupin [6] for details on a such trade-off.

7.4. **Proof of Theorem 5.1 :**

The proof of Theorem 5.1, which is quite classical, is omitted.

8. Appendix

Lemma 8.1. *Let φ be such that $\mathbb{E}(|\varphi(X,Z)|) < \infty$ and let Φ such that $\mathbb{E}(|\Phi(U)|) < \infty$. Under assumptions $(\mathbf{A_5})$ and $(\mathbf{A_4})$, then $\mathbb{E}[\varphi(X,Z)\Phi \star K_{n,C_n}(U)] = \mathbb{E}[\varphi(X,Z)\Phi \star K_{C_n}(Z)]$,*

$$and \quad \mathbb{E}\left[\varphi(X,Z)\Phi \star K_{n,C_n}(U)\right]^2 = \int \left\langle \left(\varphi^2(x,\cdot)f_{X,Z}(x,\cdot)\right) \star f_\varepsilon, (\Phi \star K_{n,C_n})^2 \right\rangle dx.$$

Proof of Lemma 8.1 : We apply Parseval's formula and write

$$\mathbb{E}\left[\varphi(X,Z)\Phi \star K_{n,C_n}(U)\right] = (2\pi)^{-1} \iint \varphi(x,z)f_{X,Z}(x,z) \int \Phi^*(y)K_{C_n}^*(y)e^{-iyz} dy\, dx\, dz$$

$$= \iint \varphi(x,z)\Phi \star K_{C_n}(z)f_{X,Z}(x,z)\, dx\, dz.$$

Similarly, $\mathbb{E}\left[\varphi(X,Z)\Phi \star K_{n,C_n}(U)\right]^2 = \int \left\langle \left(\varphi^2(x,\cdot)f_{X,Z}(x,\cdot)\right) \star f_\varepsilon, (\Phi \star K_{n,C_n})^2 \right\rangle dx.$ □

Lemma 8.2. *For ν, λ and ρ nonnegative numbers, then*

$$\int_{|u|\geq C_n} |u|^{-a} \exp(-d|u|^r)du \leq \frac{1}{C(a,d,r)}C_n^{-a+1-r}\exp\{-dC_n^r\}. \tag{8.1}$$

Furthermore, if f_ε satisfies $(\mathbf{N_2})$, then

$$\int_{|u|\leq C_n} \frac{|u|^{-a}\exp(-d|u|^r)}{|f_\varepsilon^*(u)|}du \leq \frac{1}{C(\alpha,\delta,\rho,a,d,r)\underline{C}(f_\varepsilon)}\max[1,C_n^{(\alpha-a+1-\rho)}\exp\{-dC_n^r+\delta C_n^\rho\}].$$

References

[1] M. Aitkin and D. Clayton, The fitting of exponential, Weibull and extreme value distributions to complex censored survival data using GLIM. *J. R. Stat. Soc., Ser. C*, **29** (1980) 156-163.

[2] P. K. Andersen, O. Borgan, R. D. Gill, and N. Keiding, *Statistical models based on counting processes*. Springer Series in Statistics (1993).

[3] T. Augustin, An exact corrected log-likelihood function for Cox's proportional hazards model under measurement error and some extensions. *Scand. J. Stat.*, **31** (2004) 43-50.

[4] Ø. Borgan, Correction to: Maximum likelihood estimation in parametric counting process models, with applications to censored failure time data. *Scand. J. Statist.*, **11** (1984) 275.

[5] Ø. Borgan, Maximum likelihood estimation in parametric counting process models, with applications to censored failure time data. *Scand. J. Stat., Theory Appl.*, **11** (1984) 1-16.

[6] C. Butucea and M.-L. Taupin, New M-estimators in semiparametric regression with errors in variables. *Ann. Inst. Henri Poincaré: Probab. Stat.* (to appear).

[7] J. S. Buzas, Unbiased scores in proportional hazards regression with covariate measurement error. *J. Statist. Plann. Inference*, **67** (1998) 247-257.

[8] R. J. Carroll, D. Ruppert, and L. A. Stefanski, *Measurement error in nonlinear models.* Chapman and Hall (1995).

[9] F. Comte and M.-L. Taupin, Nonparametric estimation of the regression function in an errors-in-variables model. *Statistica Sinica*, **17** (2007) 1065-1090.

[10] D. R. Cox and D. Oakes, *Analysis of survival data.* Monographs on Statistics and Applied Probability. edited by Chapman and Hall (1984).

[11] J. Fan and Y. K. Truong, Nonparametric regression with errors in variables. *Ann. Statist.* **21** (1993) 1900-1925.

[12] M. V. Fedoryuk, *Asimptotika: integraly i ryady.* Asymptotics: Integrals and Series (1987).

[13] W. A. Fuller, *Measurement error models.* Wiley Series in Probability and Mathematical Statistics (1987).

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

[15] G. Gong, A. S. Whittemore, and S. Grosser, Censored survival data with misclassified covariates: A case study of breast-cancer mortality. *J. Amer. Statist. Assoc.*, **85** (1990) 20-28.

[16] N. L. Hjort, On inference in parametric survival data models. *Int. Stat. Rev.* 60 (1992) 355-387.

[17] D. W. J. Hosmer and S. Lemeshow, *Applied survival analysis. Regression modeling of time to event data.* Wiley Series in Probability and Mathematical Statistics (1999).

[18] C. Hu and D. Y. Lin, Semiparametric failure time regression with replicates of mismeasured covariates. *J. Am. Stat. Assoc.* **99** (2004) 105-118.

[19] C. Hu and D.Y. Lin, Cox regression with covariate measurement error. *Scand. J. Stat.* **29** (2002) 637-655.

[20] Y. Huang and C. Y. Wang, Cox regression with accurate covariates unascertainable: A nonparametric-correction approach. *J. Am. Stat. Assoc.* **95** (2000) 1209-1219.

[21] J. Kiefer and J. Wolfowitz, Consistency of the maximum likelihood estimator in the presence of infinitely many nuisance parameters. *Annals math. Statist.* **27** (1956) 887-906.

[22] F. H. Kong, Adjusting regression attenuation in the Cox proportional hazards model. *J. Statist. Plann. Inference* **79** (1999) 31-44.

[23] F. H. Kong and M. Gu, Consistent estimation in Cox proportional hazards model with covariate measurement errors. *Statistica Sinica* **9** (1999) 953-969.

[24] O. V. Lepski and B. Y. Levit, Adaptive minimax estimation of infinitely differentiable functions. *Math. Methods Statist.* **7** (1998) 123-156.

[25] Y. Li and L. Ryan, Survival analysis with heterogeneous covariate measurement error. *J. Amer. Statist. Assoc.* **99** (2004) 724-735.

[26] Y. Li and L. Ryan, Inference on survival data with covariate measurement error - An imputation-based approach. *Scand. J. Stat.*, **33** (2006) 169-190.

[27] M.-L. Martin-Magniette. Nonparametric estimation of the hazard function by using a model selection method: estimation of cancer deaths in Hiroshima atomic bomb survivors. *J. Roy. Statist. Soc. Ser. C* **54** (2005) 317-331.

[28] T. Nakamura, Corrected score function for errors-in-variables models: methodology and application to generalized linear models. *Biometrika* **77** (1990) 127-137.

[29] T. Nakamura, Proportional hazards model with covariates subject to measurement error. *Biometrics* **48** (1992) 829-838.

[30] R. L. Prentice, Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika* **69** (1982) 331-342.

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

[32] O. Reiersøl, Identifiability of a linear relation between variables which are subject to error. *Econometrica.* **18** (1950) 375-389.

[33] P. Reynaud-Bouret, Adaptive estimation of the intensity of inhomogeneous Poisson processes via concentration inequalities. *Probab. Theory Relat. Fields* **126** (2003) 103-153.

[34] L. A. Stefanski, Unbiaised estimation of a nonlinear function of a normal mean with application to measurement error models. *Commun. Stat. -Theory Meth.* **18** (1989) 4335-4358.

[35] M.-L. Taupin, Semi-parametric estimation in the nonlinear structural errors-in-variables model. *Ann. Statist.* **29** (2001) 66-93.

[36] T. T. Tsiatis, V. DeGruttola, and M. S. Wulfsohn, Modeling the relationship of survival to longitudinal data measured with error. Application to survival and cd4 counts in patients with aids. *J. Amer. Statist. Assoc.* **90** (1995) 27-37.

[37] Y. N. Tyurin, A. Yakovlev, J. Shi, and L. Bass, Testing a model of aging in animal experiments. *Biometrics* **51** (1995) 363-372.

[38] A. W. van der Vaart and J. A. Wellner, *Weak convergences and empirical processes. With applications to statistics.* Springer Series in Statistics (1996).

[39] S. X. Xie, C. Y. Wang, and R. L. Prentice, A risk set calibration method for failure time regression by using a covariate reliability sample. *J. R. Stat. Soc., Ser. B, Stat. Methodol.*, **63** (2001) 855-870.

MARIE-LAURE MARTIN-MAGNIETTE

UMR AgroParisTech / INRA MIA 518   Unité de Recherche en Génomique Végétale,

16, rue Claude Bernard   UMR INRA 1165- CNRS 8114 -UEVE

75231 Paris cedex 05, France,   2,rue Gaston Crémieux- CP 5708

e-mail : marie_laure.martin@agroparistech.fr   91057 Evry Cedex, France

Marie-Luce TAUPIN

Laboratoire MAP5, UMR 8145,

Université Paris Descartes

45, rue des Saint Pères,

75006 Paris, France

e-mail : marie-luce.taupin@univ-paris5.fr