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

MARIE-LAURE MARTIN-MAGNIETTE\textsuperscript{1,2} AND MARIE-LUCE TAUPIN\textsuperscript{3}

Abstract. We consider a failure hazard function, conditional on a time-independent covariate $Z$, given by $\eta_{\theta,0}(t)f_{\beta,0}(Z)$. The baseline hazard function $\eta_{\theta,0}$ and the relative risk $f_{\beta,0}$ both belong to parametric families with $\theta_0 = (\beta_0, \gamma_0)^T \in \mathbb{R}^{m+p}$. The covariate $Z$ has an unknown density and is measured with an error through an additive error model $U = Z + \varepsilon$ where $\varepsilon$ is a random variable, independent from $Z$, with known density $f_\varepsilon$. 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 $\theta_0$ 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 $f_\varepsilon$ and $f_{\beta,0}(z)$ 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 $f_\beta$ and various types of error density $f_\varepsilon$. In particular, in the Cox model and in the excess risk model, the estimator of $\theta_0$ is $\sqrt{n}$-consistent and asymptotically Gaussian regardless of the form of $f_\varepsilon$.

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


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

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

In this paper we consider a parametric proportional hazard model, $R(t, \theta_0|Z) = \eta_{\theta,0}(t)f_{\beta,0}(Z)$, 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 $f_\beta(0) = 1$, $\eta_{\theta,0}(t)$ is the baseline hazard function, that is the conditional hazard function of $T$ given $Z = 0$. The function $f_{\beta,0}$ is the relative risk and the conditional failure rates associated with any two values of the covariate $Z$ is proportional. Here we assume

\textit{Keywords and phrases.} Semiparametric estimation, errors-in-variables model, measurement error, nonparametric estimation, excess risk model, Cox model, censoring, survival analysis, density deconvolution, least square criterion.

\textsuperscript{1} UMR AgroParisTech/INRA MIA 518, Paris, France.
\textsuperscript{2} URGV UMR INRA 1165/CNRS 8114/UEVE, Évry, France.
\textsuperscript{3} Université Paris Descartes, Laboratoire MAP5, Paris; marie-luce.taupin@math-info.univ-paris5.fr

Article published by EDP Sciences © EDP Sciences, SMAI 2009
that the functions $\eta_1$ and $f_\beta$ 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 $\eta_1(t) \geq 0$ 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 $R(t, \theta|Z) = \gamma_t f(\beta^\top Z)$, Weibull models with $R(t, \theta|Z) = \gamma_t t^{\beta} f(\beta^\top Z)$, models with a piecewise constant baseline function and Gompertz-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 $C_i$, $D_i = \mathbb{I}_{T_i \leq C_i}$ 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 $(\varepsilon_i)_{i=1,\ldots,n}$ and $(Z_i, T_i, C_i)_{i=1,\ldots,n}$ are independent. The density of $\varepsilon$ is known and is denoted by $f_\varepsilon$. 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

$$S_{\theta^0, \varphi}(\theta) = \mathbb{E} \left[ (f_\beta W)(Z) \int_0^\tau Y(t) \eta_1(t) \mathrm{d}t \right] - 2\mathbb{E} \left[ (f_\beta W)(Z) \int_0^\tau \eta_1(t) \mathrm{d}N(t) \right]$$

where $W$ is a nonnegative weight function to be chosen, $N(t) = \mathbb{I}_{X \leq t, D=1}$ and $Y(t) = \mathbb{I}_{X \geq t}$. Since the intensity of the censored process $N(t)$ with respect to $\mathcal{F}_t = \sigma\{Z, U, N(s), \mathbb{I}_{X \geq s}, 0 \leq s \leq t \leq \tau\}$ is equal to $\lambda(t, \theta^0, Z) = \eta_\varphi(t)Y(t) f_{\beta_0}(Z)$, we can rewrite $S_{\theta^0, \varphi}(\theta)$ as

$$S_{\theta^0, \varphi}(\theta) = \int_0^\tau \left\{ \mathbb{E} \left[ \left( \eta_1(t) f_\beta(Z) - \eta_\varphi(t) f_{\beta_0}(Z) \right)^2 Y(t) W(Z) \right] - \mathbb{E} \left[ \left( \eta_\varphi(t) f_{\beta_0}(Z) \right)^2 Y(t) W(Z) \right] \right\} \mathrm{d}t.$$

This shows that $S_{\theta^0, \varphi}(\theta)$ is minimum if $\theta = \theta^0$ as soon as $W$ is a nonnegative function. We propose to estimate $S_{\theta^0, \varphi}(\theta)$ for all $\theta \in \Theta$ by a quantity $S_{n,1}(\theta)$, which depends on the error density $f_\varepsilon$, on the observations $\Delta_{i,1}, \ldots, \Delta_{i,n}$, with $\Delta_i = (X_i, D_i, U_i)$, and where $g$ is replaced with a deconvolution kernel estimator. The parameter $\theta^0$ 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 $\hat{\theta}_1 = \arg\min_{\theta \in \Theta} S_{n,1}(\theta)$ converges to $\arg\min_{\theta \in \Theta} S_{\theta^0, \varphi}(\theta) = \theta^0$ which ensures the consistency of $\hat{\theta}_1$. Its rate of convergence depends on the smoothness of $f_\varepsilon$ and on the smoothness of $(f_\beta W)(z)$ as a function of $z$. More precisely, if we denote by $\varphi^*(t) = \int e^{itz} \varphi(x) \mathrm{d}x$ the Fourier transform of an integrable function $\varphi$, the rate of convergence of $\hat{\theta}_1$ depends on the behavior of the ratios of the Fourier
transforms \((f_dW)^*(t)/f_d^*(t)\) and \((f_d^2W)^*(t)/f_d^*(t)\) as \(t\) tends to infinity. We give an upper bound of the quadratic risk \(E \parallel \hat{\theta}_1 - \theta^0 \parallel_c^2\), where \(\parallel \theta \parallel_c^2 = \sum_{k=1}^{n+1} \theta_k^2\), 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 \(\hat{\theta}_1\), up to the parametric rate in many cases. Specific choices for \(W\) ensure that \(\hat{\theta}_1\) 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 developed 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 occurrence and radiation dose which is always mismeasured (see [27,30]).

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

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 an approximation of \(E(Z|U)\). The third idea, proposed by [30,36,39] is to approximate the partial log-likelihood related to the filtration generated by the observations \(E_t = \sigma\{U, N(s), I_{X_{1,\geq s}}, 0 \leq s \leq t\}\). 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])

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

with \(N_i(t) = I_{X_{1,\leq t}, D_i=1}, Y_i(t) = I_{X_{1,\geq t}}, Z^{(n)} = (Z_1, \ldots, Z_n)\), and with \(f_\beta^{(1)}\) the first derivative of \(f_\beta\) with respect to \(\beta\). 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 \(\varepsilon\), then \(\lim_{n \to \infty} E[L_n^{(1)}(\beta, Z^{(n)})] = 0\) only depends on \(E(Z)\) and on \(E[\exp(\beta Z)]\), with \(E(Z) = E(U)\) and \(E[\exp(\beta U)] = E[\exp(\beta Z)]E[\exp(\beta \varepsilon)]\). 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 \( \lim_{n \to \infty} E[L^{(1)}_n(\beta, Z^{(n)})] \) depends on \( E[Z/(1 + \beta Z)] \) whereas \( \lim_{n \to \infty} E[L^{(1)}_n(\beta, U^{(n)})] \) depends on \( E[U/(1 + \beta U)] \). Since the error model \( U = Z + \varepsilon \) does not provide any expression of \( E[Z/(1 + \beta Z)] \) in terms of \( E[U/(1 + \beta U)] \), 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 \( \lim_{n \to \infty} E(\Psi_n(\beta, U)) = E[Z/(1 + \beta Z)] \).

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 \( L_2(\mathbb{R}) \cap L_2(\mathbb{R}) \), we define \( u^*(x) = \int e^{itx} u(t) dt \), \( u^* v(x) = \int u(y) v(x - y) dy \), and \( \langle u, v \rangle = \int u(x) \overline{v(x)} dx \) with \( \overline{v} \) 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 \( \theta \in \mathbb{R}^d, \| \theta \|_2^2 = \sum_{k=1}^d \theta^2_k \). 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 \( \theta \) are denoted by

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

and

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

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

**Model assumptions:**

1. For all \( \gamma \in \Gamma \), \( \eta_\gamma \) is nonnegative and \( \int_0^\gamma \eta^2_\gamma(t) dt < \infty. \quad (A_1) \)
2. For all \( \beta \in B \) and for all \( g \in G \), if \( Z \) has density \( g \), \( f_\beta(Z) \geq 0. \quad (A_2) \)
3. Conditional on \( Z \) and \( U \), the failure time \( T \) and the censoring time \( C \) are independent. \quad (A_3)
4. The conditional distribution of the failure time \( T \) given \((Z, U)\) does not depend on \( U. \quad (A_4) \)
5. The conditional distribution of the censoring time \( C \) given \((Z, U)\) does not depend on \( U. \quad (A_5) \)

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 \((A_2)\), 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 \((A_2)\) that \( f_\beta(z) \geq 0 \) for all \( z \) in the support of the density \( g \) of the covariate \( Z \) and for all \( \beta \in B \).

Assumption \((A_3)\) states that a general censoring model is considered. Assumption \((A_4)\) and \((A_5)\) 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.**

The functions \( \beta \mapsto f_\beta \) and \( \gamma \mapsto \eta_\gamma \) admit continuous derivatives up to order 3 with respect to \( \beta \) and \( \gamma \) respectively. \quad (A_6)
We denote by $S_{p,\theta}^{(1)}(\theta)$ and $S_{p,\theta}^{(2)}(\theta)$ the first and second derivatives of $S_{p,\theta}(\theta)$ with respect to $\theta$. For all $t$ in $[0, \tau]$, let $S_{p,\theta}^{(2)}(\theta, t)$ be the second derivative of $S_{p,\theta}$ when the integral is taken over $[0, t]$ in (1.2), with the convention that $S_{p,\theta}^{(2)}(\theta) = S_{p,\theta}^{(2)}(\theta, \tau)$.

Identifiability and moment assumptions.

\[ S_{p,\theta}^{(1)}(\theta) = 0 \text{ if and only if } \theta = \theta^0. \] (A7)

For all $t \in [0, \tau]$, the matrix $S_{p,\theta}^{(2)}(\theta^0, t)$ exists and is positive definite. \hspace{1cm} (A8)

For all $j = 1, \cdots, p$, \[ \int_0^T \eta_{p,j}^4(t)dt < \infty, \text{ and } \int_0^T |\eta_{p,j}^{(1)}(t)|^3 \eta_{p,j}(t)dt < \infty. \] (A9)

For $\gamma \in \Gamma$ and $j = 1, \cdots, p$, \[ E \left( \int_0^T \eta_{\gamma,j}^4(t)dt \right)^2 < \infty \text{ and } E \left( \int_0^T \eta_{\gamma,j}^{(1)}(t)dt \right)^2 < \infty. \] (A10)

\[ \sup_{g \in G} \| f_{p,\theta} \|^2 \leq C_1(f_{p,\theta}), \sup_{g \in G} \| f_{p,\theta} \|^2 \leq C_2(f_{p,\theta}). \] (A11)

The function $W$ is such that for all $\beta \in B$ and $g \in G$, $E[(f_{p,\theta}^2 W)(Z)]$ is finite. \hspace{1cm} (A12)

The function $W$ is such that for all $g \in G$ and $E[(f_{p,\theta}^2 W)(Z)]$ are finite. \hspace{1cm} (A13)

and for $j = 1, \cdots, m$, \[ E[(f_{p,\theta}^3 |f_{p,\theta} W)(Z)] \] are finite.

We can use the equality (1.2), to see see that $S_{p,\theta}(\theta)$ is minimum at $\theta = \theta^0$. Assumptions (A7) and (A8) ensure that $\theta^0$ is the unique minimum. The density $g$ and the parameter $\beta$ vary over sets $G$ and $B$, such that (A2), (A7), (A8), (A10) and (A11) hold.

3. Estimation Procedure

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

\[ \hat{S}_n(\theta) = -\frac{2}{n} \sum_{i=1}^n (f_{p,\theta}^2 W)(Z_i) \int_0^T \eta_{\gamma,i}(t)dt + \frac{1}{n} \sum_{i=1}^n (f_{p,\theta}^2 W)(Z_i) \int_0^T \eta_{\gamma,i}^2(t)Y_i(t)dt. \]

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

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

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

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

the density $f_\theta$ belongs to $L_2(R) \cap L_\infty(R)$ and for all $x \in R$, $f_\theta^*(x) \neq 0$. \hspace{1cm} (N1)

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

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

Indeed, if $f_{X,U,Z}$ is the joint distribution of $(X, U, Z)$, Assumptions (A₄)–(A₅) and the independence between $Z$ and $\varepsilon$ imply that $f_U = g \ast f_\varepsilon$ and $f_{X,U,Z}(x,u,z) = f_{X,Z}(x,z)f_\varepsilon(u - z)$. Consequently, under mild conditions $S_{n,1}(\theta) \overset{p}{\to} S_{\beta_0,\varrho}(\theta)$ for all $\theta \in \Theta$ and we propose to estimate $\theta^0$ by

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

4. ASYMPTOTIC PROPERTIES

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

The weight function $W$ is chosen such that

$$
\sup_{\beta \in \mathbb{B}} (Wf_\beta), \quad W \text{ and } \sup_{\beta \in \mathbb{B}} (Wf_\beta^1) \text{ belong to } \mathbb{L}_1(\mathbb{R}).
$$

(A₁₄) \sup_{\beta \in \mathbb{B}} (Wf_\beta) \text{ and } \sup_{\beta \in \mathbb{B}} (Wf_\beta^1) \text{ belong to } \mathbb{L}_1(\mathbb{R}).

(A₁₅)

$$
\mathbb{E} \left[ \sup_{\beta \in \mathbb{B}} S_{n,1}^{(1)}(\theta) \right] < \infty.
$$

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

$$
\min_{q=1,2} \| \psi^*(K_{C_n}^* - 1) \|_q^2 + \max_{q=1,2} \| \psi^* K_{C_n}^* \|_q = o(1).
$$

(4.2)

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

**Theorem 4.1.** Let (A₁)–(A₁₅) and (N₁) hold. Let $\hat{\theta}_1 = (\hat{\beta}_1, \hat{\gamma}_1)$ be defined by (3.1).

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

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

$$
B_n^{(q)}(\theta^0) = \| (f_{\beta^0} W)^*(K_{C_n}^* - 1) \|_q^2 + \| (f_{\beta^0} W)^*(K_{C_n}^* - 1) \|_q^2 + \| (f_{\beta^0}^1 W)^*(K_{C_n}^* - 1) \|_q^2 + \| (f_{\beta^0}^1 W)^*(K_{C_n}^* - 1) \|_q^2
$$

and

$$
V_n^{(q)}(\theta^0) = \| (f_{\beta^0} W)^*(K_{C_n}^* - 1) \|_q^2 + \| (f_{\beta^0} W)^*(K_{C_n}^* - 1) \|_q^2 + \| (f_{\beta^0}^1 W)^*(K_{C_n}^* - 1) \|_q^2 + \| (f_{\beta^0}^1 W)^*(K_{C_n}^* - 1) \|_q^2.
$$

The terms $B_n^2$ and $V_{n,j}$ are the squared bias and variance terms, respectively. As usual, the bias is the smallest for the smoothest functions $(Wf_\beta)(z)$ and $(f_\beta W)(z)/\beta$, 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 $f_\varepsilon$. Hence, the slowest rates are obtained for the smoothest error density $f_\varepsilon$, for instance for Gaussian $\varepsilon$'s. Consequently, a good choice of $W$ can improve $\hat{\theta}_1$’s rate of convergence by smoothing $Wf_\beta$.

The rate for estimating $\beta_0$ 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 $\gamma_0$ 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 $\eta_\gamma$, 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 $f_\varepsilon$.

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

We say that (C1)–(C3) hold if

there exists a weight function $W$ such that the functions $sup_{\beta \in \mathbb{B}} (f_\beta W)/f_\varepsilon^*$ and $sup_{\beta \in \mathbb{B}} (f_\beta^2 W)/f_\varepsilon^*$ belong to $L_1(\mathbb{R}) \cap L_2(\mathbb{R})$;

the functions $sup_{\beta \in \mathbb{B}} (f_\beta^{(1)} W)/f_\varepsilon^*$ and $sup_{\beta \in \mathbb{B}} (f_\beta^{(1)} f_\beta W)/f_\varepsilon^*$ belong to $L_1(\mathbb{R}) \cap L_2(\mathbb{R})$;

for all $\beta \in \mathbb{B}$, $(f_\beta^{(2)} W)/f_\varepsilon^*$ and $(\partial^2 (f_\beta^2 W)/\partial \beta^2)/f_\varepsilon^*$ belong to $L_1(\mathbb{R}) \cap L_2(\mathbb{R})$.

Theorem 4.2. Under the assumptions of Theorem 4.1 and under (C1)–(C3), one can find $C_n$ such that $\hat{\theta}_1 = \theta_1(C_n)$ defined by (3.1) is a $\sqrt{n}$-consistent estimator of $\theta_0$. Moreover $\sqrt{n}(\hat{\theta}_1 - \theta_0) \xrightarrow{\mathcal{L}} N(0, \Sigma_1)$, where $\Sigma_1$ equals

$$
(\mathbb{E}[-2 \int_0^\tau \frac{\partial^2 ((f_\beta W)(Z)\eta_\gamma(s))}{\partial \theta^2} |_{\theta = \theta_0} dN(s) + \int_0^\tau \frac{\partial^2 ((f_\beta^2 W)(Z)\eta_\gamma^2(s))}{\partial \theta^2} |_{\theta = \theta_0} Y(s)ds]^{-1} \times \Sigma_{0,1} \times (\mathbb{E}[-2 \int_0^\tau \frac{\partial^2 ((f_\beta W)(Z)\eta_\gamma(s))}{\partial \theta^2} |_{\theta = \theta_0} dN(s) + \int_0^\tau \frac{\partial^2 ((f_\beta^2 W)(Z)\eta_\gamma^2(s))}{\partial \theta^2} |_{\theta = \theta_0} Y(s)ds]^{-1}
$$

with

$$
\Sigma_{0,1} = \mathbb{E} \left\{ -2 \int_0^\tau \frac{\partial (R_{3, f_\beta, 1}(U)\eta_\gamma(s))}{\partial \theta} |_{\theta = \theta_0} dN(s) + \int_0^\tau \frac{\partial (R_{3, f_\beta, 2}(U)\eta_\gamma^2(s))}{\partial \theta} |_{\theta = \theta_0} Y(s)ds \right\} 
\times \left[ -2 \int_0^\tau \frac{\partial (R_{3, f_\beta, 1}(U)\eta_\gamma(s))}{\partial \theta} |_{\theta = \theta_0} dN(s) + \int_0^\tau \frac{\partial (R_{3, f_\beta, 2}(U)\eta_\gamma^2(s))}{\partial \theta} |_{\theta = \theta_0} Y(s)ds \right]^T,
$$

$$
R_{3, f_\beta, 1}(U) = (2\pi)^{-1} \int (Wf_\beta)^*(t) e^{-itU} f_\varepsilon^*(t) dt \quad \text{and} \quad R_{3, f_\beta, 2}(U) = (2\pi)^{-1} \int (Wf_\beta^2)^*(t) e^{-itU} f_\varepsilon^*(t) dt.
$$

Conditions (C1)–(C3) ensure the existence of the functions $R_{3, f_\beta, j}$ for $j = 1, 2$ and hence the $\sqrt{n}$-consistency. Nevertheless it is not always possible to find $W$ such that (C1)–(C3) hold.
Table 1. Rates of convergence $\varphi_n^2$ of $\tilde{\theta}_1$.

<table>
<thead>
<tr>
<th>$W_{f_{\beta}^{*}}$</th>
<th>$f_{\beta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d = r = 0$ in $(\mathbf{R}_1)$ Sobolev</td>
<td>$\rho = 0$ in $(\mathbf{N}_2)$ ordinary smooth</td>
</tr>
<tr>
<td>$\rho &gt; 0$ in $(\mathbf{N}_2)$ super smooth</td>
<td></td>
</tr>
<tr>
<td>$a &lt; \alpha + 1/2$</td>
<td>$n^{-\frac{2a}{\alpha}}$</td>
</tr>
<tr>
<td>$a \geq \alpha + 1/2$</td>
<td>$n^{-1}$</td>
</tr>
<tr>
<td>$r &lt; \rho$</td>
<td>$(\log n)^{\frac{2a-1}{\rho}}$</td>
</tr>
<tr>
<td>$r = \rho$</td>
<td>$(\log n)^{A(a, r, \rho)} \exp \left{-2d \left(\frac{\log n}{n}\right)^{1/\rho}\right}$</td>
</tr>
<tr>
<td>$r &gt; \rho$</td>
<td>$(\log n)^{A(a, r, \rho)} \exp \left{-2d \left(\frac{\log n}{n}\right)^{1/\rho}\right}$</td>
</tr>
</tbody>
</table>

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

4.3. Rates for general smoothness classes

We now specify the asymptotic smoothness properties of $\hat{\theta}_1$ when $f_{\beta}^{*}$ and $(W_{f_{\beta}})^{*}$ satisfy assumptions $(\mathbf{N}_2)$ and $(\mathbf{R}_1)$ given below.

There exist positive constants $C(f_\epsilon, \overline{C}(f_\epsilon))$, and nonnegative $\delta, \alpha, u_0$ and $\rho \leq 2$ \hspace{1cm} (N2)

such that $C(f_\epsilon) \leq |f_\epsilon^{*}(u)| |u|^{\alpha} \exp(\delta |u|^{\rho}) \leq \overline{C}(f_\epsilon)$ for all $|u| \geq u_0$.

When $\rho = 0 = \delta$ in $(\mathbf{N}_2)$, with the convention that $\delta = 0$ if and only if $\rho = 0$, $f_\epsilon$ is called “ordinary smooth”. An example of an ordinary smooth density is the double exponential (also called Laplace) distribution with $\epsilon > 0$.

An example of a super smooth density is the Gaussian and Cauchy distributions. The smoothness of $\hat{\theta}_1$ 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 $L_1(\mathbb{R} \cap L_2(\mathbb{R}))$ and if there exist $a, d, u_0, r \geq 0$ such that $0 < L(f) \leq |f^{*}(u)| |u|^{\alpha} \exp(d |u|^{\rho}) \leq \overline{C}(f) < \infty$ 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 $f_\epsilon$ 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 $\beta_j$, $j = 1, \ldots, m$ up to order $d$, satisfy $(\mathbf{R}_1)$. Consider the sequences $C_n$ such that

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

Then $\hat{\theta}_1$ is a consistent estimator of $\theta_0$ and $\mathbb{E}(||\hat{\theta}_1 - \theta_0||_2^2) = O(\varphi_n^{2})$ with $\varphi_n^2$ as in Table 1.

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

Our estimation procedure requires the estimation of the two following linear functionals of the density $g$, $\mathbb{E}[\int_0^T (f_{\beta} W)(Z) dN(t)]$ and $\mathbb{E}[\int_0^T (f_{\beta}^2 W)(Z) Y(t) dt]$. 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 (C4)–(C6) hold if there exist a weight function $W$ and two functions $\Phi_{\beta,f,1}$ and $\Phi_{\beta,f,2}$ not depending on $g$, such that for all $\beta \in \mathbb{B}$ and for all $g$

$$\mathbb{E}\left[ \int_0^\tau (f_\beta W)(Z) dN(t) \right] = \mathbb{E}\left[ \int_0^\tau \Phi_{\beta,f,1}(U) dN(t) \right],$$

$$\mathbb{E}\left[ \int_0^\tau (f_\beta^* W)(Z) Y(t) dt \right] = \mathbb{E}\left[ \int_0^\tau \Phi_{\beta,f,2}(U) Y(t) dt \right];$$  \hspace{1cm} (C4)

for $k = 0, 1, 2$ and for $j = 1, 2$,

$$\mathbb{E}[\sup_{\beta \in \mathbb{B}} \| \Phi_{\beta,f,1}(U) \|_{\mathbb{L}_2}] < \infty;$$  \hspace{1cm} (C5)

for $j = 1, 2$ and for all $\beta \in \mathbb{B}$,

$$\mathbb{E}\left[ \| \Phi_{\beta,f,1}(U) \|_{\mathbb{L}_2}^2 \right] < \infty.$$  \hspace{1cm} (C6)

Under (C4)–(C6), we estimate $\theta_0$ by $\hat{\theta}_2 = \arg \min_{\theta \in \Theta} S_{n,2}(\theta)$ where

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

The main difficulty for finding such functions $\Phi_{\beta,f,1}$ and $\Phi_{\beta,f,2}$ 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 $\hat{\theta}_2$

**Theorem 5.1.** Let (A1)–(A13), and conditions (C4)–(C6) hold. Then $\hat{\theta}_2$, defined by (5.1), is a $\sqrt{n}$-consistent estimator of $\theta_0$. Moreover $\sqrt{n}(\hat{\theta}_2 - \theta_0) \xrightarrow{n \to \infty} N(0, \Sigma_2)$, where $\Sigma_2$ is equal to

$$\mathbb{E}\left[ -2 \int_0^\tau \frac{\partial^2 (\Phi_{\beta,f,1}(U) \eta_i(s))}{\partial \theta^2} |_{\theta = \theta_0} dN(s) + \int_0^\tau \frac{\partial^2 (\Phi_{\beta,f,2}(U) \eta_i^2(s))}{\partial \theta^2} |_{\theta = \theta_0} Y(s) ds \right]^{-1}$$

$$\times \Sigma_{0,2} \left[ \mathbb{E}\left[ -2 \int_0^\tau \frac{\partial^2 (\Phi_{\beta,f,1}(U) \eta_i(s))}{\partial \theta^2} |_{\theta = \theta_0} dN(s) + \int_0^\tau \frac{\partial^2 (\Phi_{\beta,f,2}(U) \eta_i^2(s))}{\partial \theta^2} |_{\theta = \theta_0} Y(s) ds \right]^{-1} \right]^{-1}$$

with

$$\Sigma_{0,2} = \mathbb{E}\left[ -2 \int_0^\tau \frac{\partial (\Phi_{\beta,f,1}(U) \eta_i(s))}{\partial \theta} |_{\theta = \theta_0} dN(s) + \int_0^\tau \frac{\partial (\Phi_{\beta,f,2}(U) \eta_i^2(s))}{\partial \theta} |_{\theta = \theta_0} Y(s) ds \right]$$

$$\times \left[ -2 \int_0^\tau \frac{\partial (\Phi_{\beta,f,1}(U) \eta_i(s))}{\partial \theta} |_{\theta = \theta_0} dN(s) + \int_0^\tau \frac{\partial (\Phi_{\beta,f,2}(U) \eta_i^2(s))}{\partial \theta} |_{\theta = \theta_0} Y(s) ds \right]\right]^T.$$

5.2. Comments

Let us briefly compare conditions (C1)–(C3) and (C4)–(C6). It is noteworthy that conditions (C4)–(C6) are more general. For instance, condition (C4) does not require that $f_\beta W$, $f_\beta^* W$ belong to $\mathbb{L}_2(\mathbb{R})$ (as for instance in the Cox Model with $W \equiv 1$). Moreover (C4) implies (C4), with $\Phi_{\beta,f,1} = R_{\beta,f,1}$, $\Phi_{\beta,f,2} = R_{\beta,f,2}$, we have

$$\mathbb{E}[Y(t) \Phi_{\beta,f,2}(U)] = \int \mathbb{I}_{x \geq 1} f x(t, x, z) \left( \frac{1}{2\pi} \int \left( W f_\beta^* \right)(s) e^{-isz} f_z(s) ds \right) dx dz = \mathbb{E}[Y(t) (W f_\beta^*)(z)].$$
Consequently, $E \left[ \int_0^T \Phi_{\beta, f_{\beta}, 2}(U)\eta_i^2(t)Y(t)dt \right] = E \left[ \int_0^T (f_{\beta}^2W)(Z)\eta_i^2(t)dt \right]$, and analogously $E \left[ \int_0^T \Phi_{\beta, f_{\beta}, 1}(U)\eta_i(t)dN(t) \right] = E \left[ \int_0^T (f_{\beta}W)(Z)\eta_i(t)dN(t) \right]$. Hence Condition $(C_4)$ 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 $f_{\beta}$ is the Gaussian density. In this case, when $W \equiv 1$ it seems impossible to find a function $\Phi_{\beta, f_{\beta}, 2}$ such that $E[Y(t)\Phi_{\beta, f_{\beta}, 2}(U)] = E[Y(t)f_{\beta}^2(Z)]$. Whereas $(C_1)$–$(C_3)$ and $(C_4)$–$(C_6)$ hold by taking $W(z) = (1 + z^2)^4 \exp(-z^2/(4\delta))$ where $\delta$ is as in $(N_2)$.

To summarize: $\hat{\theta}_1$ always exists and is consistent under mild regularity conditions, though $\hat{\theta}_1$’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 $\hat{\theta}_2$ is more straightforward than the computation of $\hat{\theta}_1$ since it does not require deconvolution estimators, but $\hat{\theta}_2$ 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 $i = 1, \ldots, n$, then the procedure still works by taking $f_{\beta}^* \equiv 1$ in the previous formulae. More precisely, in this context $S_{n,1}(\theta)$ becomes

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

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

$$n^{-1} \sum_{i=1}^n \Phi * K_{C_n}(U_i) = n^{-1} \sum_{i=1}^n \Phi * K_{C_n}(Z_i) \xrightarrow{P} \frac{P}{\sqrt{n}} E[\Phi(Z)].$$

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

6. EXAMPLES

In this section, we illustrate the asymptotic properties of $\hat{\theta}_1$ and $\hat{\theta}_2$ for various relative errors. In all of these examples, $K^*(t) = I_{|t| \leq 1}$, the baseline function has a nonspecific parametric form and $f_{\beta}$ satisfies $(N_1)$ and $(N_2)$ 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 $\beta^0$. 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 $f_{\beta}$ be of the form $f_{\beta}(z) = \exp(\beta z)$ and assume that $E[\exp(\beta U)] < \infty$ for all $\beta$ in $\mathbb{R}$. Let $W(z) = \exp(-z^2/(4\delta))$ where $\delta$ is as in $(N_2)$. Then conditions $(C_1)$–$(C_3)$ as well as conditions $(C_4)$–$(C_6)$ are satisfied. Hence the estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ are $\sqrt{n}$-consistent and asymptotically Gaussian estimators of $\theta^0$, with the same asymptotic variance.

One could also choose $W \equiv 1$ and use the independence between $Z$ and $\varepsilon$ to find that $E[\exp(\beta Z)] = E[\exp(\beta U)]/E[\exp(\beta \varepsilon)]$. Consequently, if we choose $\Phi_{\beta, f_{\beta}, 1}(U) = \exp(2\beta U)/E[\exp(2\beta \varepsilon)]$ and $\Phi_{\beta, f_{\beta}, 2}(U) = \exp(2\beta U)/E[\exp(2\beta \varepsilon)]$. However, it seems impossible to find a function $\Phi_{\beta, f_{\beta}, 2}$ such that $E[Y(t)\Phi_{\beta, f_{\beta}, 2}(U)] = E[Y(t)f_{\beta}^2(Z)]$. Whereas $(C_1)$–$(C_3)$ and $(C_4)$–$(C_6)$ hold by taking $W(z) = (1 + z^2)^4 \exp(-z^2/(4\delta))$ where $\delta$ is as in $(N_2)$.
Example 2. Excess relative risk model. Let $f_\beta$ be of the form $f_\beta(z) = 1 + \beta z$ with $\beta$ and $\gamma$ in $B$ and $G$ such that for $Z \sim g$, $1 + \beta Z$ is nonnegative. For example $G$ could consist of densities having support included in $[0, +\infty)$ and $B$ could be a compact set included in $[0, +\infty)$, or $G$ could consist of densities having support included in $[-1, +1]$ and $B$ could be a compact set included in $[-1, +1]$. Let $W(z) = \exp(-z^2/(4\delta))$ where $\delta$ is as in (N2). Then conditions (C1)–(C3) as well as conditions (C4)–(C6) are satisfied. Hence the estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ are $\sqrt{n}$-consistent and asymptotically Gaussian estimators of $\theta^0$, with the same asymptotic variance.

Example 3. Polynomial relative risk 1. Let $f_\beta$ be of the form $f_\beta(z) = 1 + \sum_{k=1}^{m} \beta_k z^k$ with $m \geq 1$, and with $\beta$ and $\gamma$ in $B$ and $G$ such that $f_\beta(Z)$ is nonnegative for $Z \sim g$. For example $G$ could consist of densities having support included in $[0, +\infty)$ and $B$ could be a compact set included in $[0, +\infty)^m$, or $G$ could consist of densities having support included in $[-1, +1]$ and $B$ could be a compact set included in the set $\{ \beta = (\beta_k)_{1 \leq k \leq m} \in \mathbb{R}^m \text{ such that } \sum_{k=1}^{m} |\beta_k| \leq 1 \}$. Let $W(z) = \exp(-z^2/(4\delta))$ where $\delta$ is as in (N2). Then conditions (C1)–(C3) as well as conditions (C4)–(C6) are satisfied. Hence the estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ are $\sqrt{n}$-consistent and asymptotically Gaussian estimators of $\theta^0$, with the same asymptotic variance.

One can also choose $W \equiv 1$, provided that $E(|U|^m) < \infty$ 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 $W$, $\hat{\theta}_1$ remains a $\sqrt{n}$-consistent and asymptotically Gaussian estimator of $\theta^0$.

Example 4. Cauchy relative risk 1. Consider $f_\beta$ of the form $f_\beta(z) = 1 - \beta + \beta/(1 + z^2)$ with $\beta$ and $\gamma$ in $B$ and $G$ such that $f_\beta(Z)$ is nonnegative for $Z \sim g$. For example $G$ could consist of all densities and $B$ could be a compact set included in $[0, 1]$ or $(-\infty, 0]$. The relative risk $f_\beta$ has the regularity of $z \mapsto 1/(1 + z^2)$ which satisfies (R1) with $a = 0$, $d = 1/2$ and $r = 1$. Let $W(z) = (1 + z^2)^{4/3} \exp(-z^2/(4\delta))$ where $\delta$ is as in (N2). The functions $f_\beta W$, $f_\beta^2 W$ and their derivatives with respect to $\beta$ up to order 3 satisfy (R4) with $\rho < r = 2$ or $\rho = r = 2$ and $d > \delta$. Consequently, conditions (C1)–(C3) as well as conditions (C4)–(C6) are satisfied and the estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ are $\sqrt{n}$-consistent and asymptotically Gaussian estimators of $\theta^0$, with the same asymptotic variance.

This example underlines the importance of the choice of smoothing weight function $W$ in the construction of $\hat{\theta}_1$ or $\hat{\theta}_2$. Indeed, with $W \equiv 1$, Theorem 4.1 predicts a much slower rate of convergence, for example, if $\epsilon$ 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 variable. Consider $f_\beta$ of the form $f_\beta(z) = 1 + \beta f(z)$ with $f(z) = \exp(-|z|/2) - 1$ and with $\beta$ and $\gamma$ in $B$ and $G$ such that $f_\beta(Z)$ is nonnegative for $Z \sim g$. For example $G$ could consist of all densities and $B$ could be a compact set included in $[0, 1]$ or $(-\infty, 0]$. The Fourier transform of $f(z) = \exp(-|z|/2)$ equals $\ell^+ (t) = 4/(1 + 4t^2)$ decays slowly. Hence, if we choose $W \equiv 1$, the estimator $\hat{\theta}_1$ is not $\sqrt{n}$-consistent as soon as $|f_\beta^u (u)| \leq c |u|^{-2}$ with $|u| \to \infty$. A closer look shows that $f_\beta$ and its derivative with respect to $\beta$ are $C^\infty$ 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 $\beta$ infinitely differentiable functions in $z$. This choice of $W$ ensures the $\sqrt{n}$-consistency of $\hat{\theta}_1$ whenever $f_\beta$ satisfies (N2) with $0 < \rho < 1$. Even if $\rho \geq 1$, the rate of $\hat{\theta}_1$ is much faster with our choice of $W$ than it would be for $W \equiv 1$. Let us specify the choice of $W$. Set

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

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 $\beta$ up to order 3 satisfy (R4) with $d > 0$ and $r = R/(R + 1)$ closer to 1 as $R$ comes larger.

Theorem 5.1 (Laplace relative risk variable). Suppose that conditions (C4)–(C6) hold and the criterion $S_{n,2}$ defined by (5.1) exists. In this case $\hat{\theta}_2$ is still a $\sqrt{n}$-consistent and asymptotically Gaussian estimator of $\theta^0$.
If \( f_\varepsilon \) satisfies (N2) with \( 0 \leq \rho < 1 \), we choose \( R \) large enough such that \( r = R/(R + 1) > \rho \). Hence, conditions (C1)-(C3) as well as conditions (C4)-(C6) are satisfied. Consequently, the estimators \( \hat{\theta}_1 \) and \( \hat{\theta}_2 \) are \( \sqrt{n} \)-consistent and asymptotically Gaussian estimators of \( \theta^0 \), with the same asymptotic variance.

If \( \rho \geq 1 \), for this choice of \( W \), the functions \( Wf_\beta \) and \( Wf_\beta^2 \) and their derivatives with respect to \( \beta \) up to order 3, satisfy (R1) with \( r = R/(R + 1) \). Hence, according to Table 1,

\[
\mathbb{E} \| \hat{\theta}_1 - \theta^0 \|^2_{L^2} = O(1) \left( \log n \right)^{1-2\rho_\varepsilon} \exp\{ -2d(\log n)/(2\delta)^{r/\rho} \}.
\]

**Example 6. Irregular relative risk.** Consider \( f_\beta \) with \( f_\beta(z) = 1 + \beta_2 a_{-} + \beta_1 z + \beta_2(z-a) 1_{z \geq a} \) with \( \beta \) and \( g \) in \( \mathcal{B} \) 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 \( \mathcal{B} \) could be a compact set included in \([0, +\infty)\). This relative risk is \( C^\infty \) 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 \( W(z) = \Psi_{R-100,a,n}(z) \). In that way, \( Wf_\beta \) and \( Wf_\beta^2 \) and their derivatives with respect to \( \beta \) satisfy (R1) with \( 0 < r = R/(R + 1) < 1 \) as close to 1 as needed.

If \( f_\varepsilon \) satisfies (N2) with \( 0 \leq \rho < 1 \), then we take \( R \) large enough such that \( r = R/(R + 1) > \rho \) and thus conditions (C1)-(C3) as well as conditions (C4)-(C6) are satisfied. Consequently, the estimators \( \hat{\theta}_1 \) and \( \hat{\theta}_2 \) are \( \sqrt{n} \)-consistent and asymptotically Gaussian estimators of \( \theta^0 \), with the same asymptotic variance.

If \( \rho \geq 1 \) in (N2), the functions \( Wf_\beta \) and \( Wf_\beta^2 \) and their derivatives with respect to \( \beta \) up to order 3, satisfy (R1) with \( 0 < r = R/(R + 1) < 1 \) as close to 1 as needed. According to Table 1

\[
\mathbb{E} \| \hat{\theta}_1 - \theta^0 \|^2_{L^2} = O(1) \left( \log n \right)^{1-2\rho_\varepsilon} \exp\{ -2d(\log n)/(2\delta)^{r/\rho} \}.
\]

**Comments on Examples 5 and 6.**

In these examples, \( f_\beta W \) satisfies (R1) with \( r < 1 \). Hence \( \hat{\theta}_1 \) is \( \sqrt{n} \)-consistent provided that \( f_\varepsilon \) is ordinary smooth or super smooth with an exponent \( \rho < 1 \). For example, when the \( \varepsilon \) is Gaussian, it seems impossible to find \( W \) such that \( (Wf_\beta)^r/f_\varepsilon \) belongs to \( \mathcal{L}_1(\mathbb{R}) \). This comes from the fact that for these relative risks, the least square criterion \( S_{\rho_\varepsilon,0}(\theta) \) cannot be estimated with the parametric rate of convergence and hence probably cannot provide a \( \sqrt{n} \)-consistent estimator of \( \theta^0 \). Nevertheless, even in cases where \( \sqrt{n} \)-consistency does not seem achievable, the resulting rate of the risk of \( \hat{\theta}_1 \) is clearly much faster than the predicted logarithmic rate that we could have with \( W = 1 \) (see Tab. 1).

In survival data analysis the relative risks \( f_\beta \) 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 \( f_\beta \) be of the form \( f(\beta z) = 1 + \sum_{k=1}^m a_k z^k \) with known \( a_k \)'s and with \( \beta \) and \( g \) in \( \mathcal{B} \) 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 \( \mathcal{B} \) could be a compact set included in \( \mathbb{R}^+ \). Let \( W(z) = \exp\{ -z^2/(4\delta) \} \) where \( \delta \) is as in (N2). Then conditions (C1)-(C3) as well as conditions (C4)-(C6) are satisfied. Hence the estimators \( \hat{\theta}_1 \) and \( \hat{\theta}_2 \) are \( \sqrt{n} \)-consistent and asymptotically Gaussian estimators of \( \theta^0 \), with the same asymptotic variance.

**Example 8. Cauchy relative risk 2.** Consider \( f_\beta \) of the form \( f(\beta z) = 1/(1+z^2) \). In this example \( \mathcal{G} \) could consist of all densities and \( \mathcal{B} \) could be a compact set of \( \mathbb{R}^m \). Let \( W(z) = (1 + z^2)^d \exp\{ -z^2/(4\delta) \} \) with \( \delta \) as in (N2) or \( W = 1 \). With these choices of \( W \), the functions \( f_\beta W, f_\beta^2 W \) and their derivatives with respect to \( \beta \) up to order 3 satisfy (R1) with \( a = 0, d = 1/\beta \) and \( r = 1 \). According to Table 1, if \( f_\varepsilon \) satisfies (N2) with \( 0 \leq \rho < 1 \), then \( \hat{\theta}_1 \) and \( \hat{\theta}_2 \) are \( \sqrt{n} \)-consistent and asymptotically Gaussian estimators of \( \theta^0 \), with the same asymptotic variance. If \( f_\varepsilon \) satisfies (N2) with \( \rho \geq 1 \), then \( \hat{\theta}_1 \) is consistent with a rate that depends on \( \beta^0 \). Let us be more precise. The proof of Theorem 4.1 implies that the terms \( B_{n,j}(\theta^0) \) of order \( \exp( -2C_n/\beta^0 ) \) and the terms \( V_{n,j}(\theta^0)/n \) of order \( C_n^{2\rho_\varepsilon+(1-\rho_\varepsilon)+(1-\rho_\varepsilon)} \exp( -2C_n/\beta^0 + 2\delta C_n^\rho )/n \), 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 $\beta^0$ and is given by

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

This choice yields the rate

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

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

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 (A7) and (A8) and the two points:

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

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

**Proof of 1-** We will break $E[S_{n,1}(\theta) - S_{p^0,\theta}(\theta)]^2$ into its corresponding bias $|E[S_{n,1}(\theta)] - S_{p^0,\theta}(\theta)]^2$ and variance $E[(S_{n,1}(\theta) - E[S_{n,1}(\theta)])^2]$ 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

$$
|\langle \varphi_1, \varphi_2 \rangle| \leq \| \varphi_1 \|_2 \| \varphi_2 \|_2, \quad \text{and} \quad |\langle \varphi_1, \varphi_2 \rangle| \leq \| \varphi_1 \|_{\infty} \| \varphi_2 \|_1. \quad (7.1)
$$

**Study of the bias.**

Under Assumptions (A1)–(A5) and using that $(Z_i, U_i, N_i(s), D_i, Y_i(s))_{i=1,\ldots,n}$ are independent, the intensity of the censored process $N_i(t) = I_{X_i \leq t, D_i=1}$ with respect to $\mathcal{F}_t = \sigma(Z_i, U_i, N_i(s)), I_{X_i \geq s}, 0 \leq s \leq t, i = 1, \ldots, n$ is equal to

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

We use (7.3) and Lemma 8.1 to write

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

and hence

$$
E[S_{n,1}(\theta)] - S_{p^0,\theta}(\theta) = \int_0^\tau \eta_{\gamma^0}(t)I_{s \geq t} \left\langle (f_{X,Z}(x, \cdot), (f_{p^0}W) \ast K_{C_n} - f_{p^0}W)dx \right\rangle dt
$$

$$
-2 \int_0^\tau \eta_{\gamma^0}(t)\eta_{\gamma^0}(t)I_{s \geq t} \left\langle (f_{p^0}(\cdot), f_{X,Z}(x, \cdot), (f_{\beta^0}W) \ast K_{C_n} - f_{\beta^0}W) \right\rangle dx dt.
$$
If we apply (7.1) we obtain the first bound
\[
\| \mathbb{E}(S_{n,1}(\theta)) - S_{\theta^o,g}(\theta) \| \leq \left( \int_0^T \eta_1^2(t)dt \right) \| f_{X,Z} \|_2 \| (f_\beta^2 W) \ast K_{C_n} - (f_\beta^2 W) \|_2 \\
+ \left( 2 \int_0^T \eta_1(t)\eta_\gamma(t)dt \right) \int \| f_{\theta^o} \cdot f_{X,Z}(x,\cdot) \|_2 dx \| (f_\beta W) \ast K_{C_n} - f_\beta W \|_2.
\]
Now, Parseval’s formula gives
\[
\| \mathbb{E}(S_{n,1}(\theta)) - S_{\theta^o,g}(\theta) \| \leq (2\pi)^{-1} \left( \int_0^T \eta_1^2(t)dt \right) \| f_{X,Z} \|_2 \| (f_\beta^2 W)^* (K_{C_n}^* - 1) \|_2 \\
+ \pi^{-1} \left( \int_0^T \eta_1(t)\eta_\gamma(t)dt \right) \int \| f_{\theta^o} \cdot f_{X,Z}(x,\cdot) \|_2 dx \| (f_\beta W)^* (K_{C_n}^* - 1) \|_2.
\]
that is
\[
\| \mathbb{E}(S_{n,1}(\theta)) - S_{\theta^o,g}(\theta) \| \leq C_{\gamma,\gamma^o} \left[ \| (f_\beta^2 W)^* (K_{C_n}^* - 1) \|_2 + \| (f_\beta W)^* (K_{C_n}^* - 1) \|_2 \right] \tag{7.4}
\]
We apply (7.2), to get that \( \| \mathbb{E}(S_{n,1}(\theta)) - S_{\theta^o,g}(\theta) \| \) is also bounded by
\[
\| (f_\beta^2 W) \ast K_{C_n} - (f_\beta^2 W) \|_\infty \left( \int \| f_{X,Z}(x,\cdot) \|_1 dx \right) \left( \int_0^T \eta_1^2(t)dt \right) \\
+ \| (f_\beta W) \ast K_{C_n} - (f_\beta W) \|_\infty \left( 2 \int \| f_{\theta^o} \cdot f_{X,Z}(x,\cdot) \|_1 dx \right) \left( \int_0^T \eta_1(t)\eta_\gamma(t)dt \right)
\]
which is also less than
\[
\| (f_\beta^2 W)^* (K_{C_n}^* - 1) \|_1 \left( (2\pi)^{-1} \left( \int \| f_{X,Z} \|_1 \right) \left( \int_0^T \eta_1^2(t)dt \right) \\
+ \| (f_\beta W)^* (K_{C_n}^* - 1) \|_1 \left( \pi^{-1} \left( \int \| f_{\theta^o} \cdot f_{X,Z}(x,\cdot) \|_1 dx \right) \left( \int_0^T \eta_1(t)\eta_\gamma(t)dt \right) \right) \right).
\]
This implies
\[
\| \mathbb{E}(S_{n,1}(\theta)) - S_{\theta^o,g}(\theta) \| \leq C_{\gamma,\gamma^o} \left[ \| (f_\beta W)^* (K_{C_n}^* - 1) \|_1 + \| (f_\beta^2 W)^* (K_{C_n}^* - 1) \|_1 \right] \tag{7.5}
\]
If we combine bounds (7.4) and (7.5) we get
\[
\| \mathbb{E}(S_{n,1}(\theta)) - S_{\theta^o,g}(\theta) \| \leq C_{\gamma,\gamma^o} \left[ \| (f_\beta W)^* (K_{C_n}^* - 1) \|_2 + \| (f_\beta^2 W)^* (K_{C_n}^* - 1) \|_2, \\
\| (f_\beta W)^* (K_{C_n}^* - 1) \|_1 + \| (f_\beta^2 W)^* (K_{C_n}^* - 1) \|_1 \right]. \tag{7.6}
\]
Since \( W f_\beta \) and \( W f_\beta^2 \) satisfy (4.3), for all \( \theta \in \text{theta}, \| \mathbb{E} S_{n,1}(\theta) - S_{\theta^o,g}(\theta) \| = 0(1) \) as \( n \) tends to infinity.

**Study of the variance.** Since we consider independent and identically distributed random variables, we obtain the variance \( \text{Var}[S_{n,1}(\theta)] = (2 + o(1))n^{-1}(A_1 + A_2), \) with
\[
A_1 = \mathbb{E}[(f_\beta^2 W) \ast K_{n,C_n}(U) \int_0^T \eta_1^2(t)Y(t)dt]^2,
A_2 = 4\mathbb{E}[(f_\beta W) \ast K_{n,C_n}(U) \int_0^T \eta_1(t)dN(t)]^2.
\]
We apply (7.2) and Lemma 8.1 to obtain that $A_1$ is less than
\[
\left(\int_0^T \eta_2^2(t) dt\right)^2 \left(\int \langle f_{X,Z}(x, \cdot) \rangle \cdot f_{\epsilon} \cdot ((f_{\beta}^3 W) \ast K_{n,C_n})^2 \right) dx \leq \left(\int_0^T \eta_2^2(t) dt\right)^2 \left(\int \| f_{X,Z}(x, \cdot) \|_{\infty} dx \right) \cdot \| (f_{\beta}^3 W) \ast K_{n,C_n} \|_2^2
\]
and hence
\[
A_1 \leq (2\pi)^{-1} \left(\int_0^T \eta_2^2(t) dt\right)^2 \| f_{\epsilon} \|_{\infty} \| f_{X,Z} \|_1 \cdot \left\| \frac{(f_{\beta}^3 W)^* K_{n,C_n}^*}{f_{\epsilon}^*} \right\|_2^2.
\]
We now give a first bound for $A_2$. If we denote $\varphi(X, Z) = \int_0^T \eta_1(t) dN(t)$ and apply Lemma 8.1 and (7.2), we obtain that $A_2$ is bounded by
\[
4 \int \langle (\varphi^2(x, \cdot)f_{X,Z}(x, \cdot)) \rangle \cdot f_{\epsilon} \cdot ((f_{\beta}^3 W) \ast K_{n,C_n})^2 \rangle dx \leq 4 \left(\int \| (\varphi^2(x, \cdot)f_{X,Z}(x, \cdot)) \rangle \cdot f_{\epsilon} \|_{\infty} dx \right) \cdot \| (f_{\beta}^3 W) \ast K_{n,C_n} \|_2^2.
\]
Since $\int \| \varphi^2(x, \cdot)f_{X,Z}(x, \cdot) \|_1 dx = E[\int_0^T \eta_1(t) dN(t)]^2$, we get
\[
\int \| \varphi^2(x, \cdot)f_{X,Z}(x, \cdot)) \rangle \cdot f_{\epsilon} \|_{\infty} dx \leq \| f_{\epsilon} \|_{\infty} E \left[\int_0^T \eta_1(t) dN(t)\right]^2.
\]
Consequently,
\[
A_2 \leq 4(2\pi)^{-1} E \left(\int_0^T \eta_1(t) dN(t)\right)^2 \| f_{\epsilon} \|_{\infty} \| \frac{(f_{\beta}^3 W)^* K_{n,C_n}^*}{f_{\epsilon}^*} \|_2^2
\]
and
\[
\text{Var}[S_{n,1}(\theta)] \leq C_{\varphi^2, \| f_{\epsilon} \|_{\infty}} n^{-1} \left[ \| (f_{\beta}^3 W)^* K_{n,C_n}^* \|_2^2 + \| (f_{\beta}^3 W)^* K_{n,C_n}^* \|_2^2 \right].
\]
We apply (7.2) and obtain that $A_1$ is also less than
\[
\left(\int_0^T \eta_2^2(t) dt\right)^2 \left(\int \langle f_{X,Z}(x, \cdot) \rangle \cdot f_{\epsilon} \cdot ((f_{\beta}^3 W) \ast K_{n,C_n})^2 \right) dx \leq \left(\int_0^T \eta_2^2(t) dt\right)^2 \left(\int \| f_{X,Z}(x, \cdot) \|_{1} dx \right) \| (f_{\beta}^3 W) \ast K_{n,C_n} \|_2^2
\]
Similarly, $A_2$ is less than
\[
4 \int \langle (\varphi^2(x, \cdot)f_{X,Z}(x, \cdot)) \rangle \cdot f_{\epsilon} \cdot ((f_{\beta}^3 W) \ast K_{n,C_n})^2 \rangle dx \leq 4 \left(\int \| (\varphi^2(x, \cdot)f_{X,Z}(x, \cdot)) \rangle \cdot f_{\epsilon} \|_{1} dx \right) \| (f_{\beta}^3 W) \ast K_{n,C_n} \|_2^2
\]
where \( \varphi(X, Z) \) is still defined by \( \varphi(X, Z) = \int_0^T \eta_s(t) dN(t) \). Once again, since

\[
\int \| \langle \varphi^2(x, \cdot) f_X, \varphi(x, \cdot) \rangle \ast f_x \|_1 \, dx = \mathbb{E} \left( \int_0^T \eta_s(t) dN(t) \right)^2,
\]

\[
\text{Var}[S_{n,1}(\theta)] \leq C_{\varphi, \|f_x\|_{L^\infty}} n^{-1} \left[ \| (f_{\beta} W)^* \frac{K_c^*}{f_z} \|_1^2 + \| (f_{\beta} W)^* \frac{K_c^*}{f_z} \|_2^2 \right].
\]

We combine (7.7) and (7.8) and have

\[
\text{Var}[S_{n,1}(\theta)] \leq \frac{C_{\varphi, \|f_x\|_{L^\infty}}}{n} \min \left\{ \| (f_{\beta} W)^* \frac{K_c^*}{f_z} \|_2^2, \| (f_{\beta} W)^* \frac{K_c^*}{f_z} \|_2^2 \right\},
\]

\[
\| (f_{\beta} W)^* \frac{K_c^*}{f_z} \|_1^2 + \| (f_{\beta} W)^* \frac{K_c^*}{f_z} \|_2^2.
\]

Since \( W_{\beta}, W_{\beta}^2 \) and their derivatives satisfy (4.3), \( \text{var}(S_{n,1}(\theta)) = 0(1) \) and the same holds for \( \mathbb{E} [S_{n,1}(\theta) - S_{\theta^{(\theta')}, \theta}]^2 \).

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

\[
- \frac{2}{n} \int_0^T \left[ (f_{\beta} W) \ast K_{nC_n}(U_i) \eta_{1}(t) - (f_{\beta} W) \ast K_{nC_n}(U_i) \eta_{1}(t) \right] dN_i(t)
+ \frac{1}{n} \int_0^T \left[ (f_{\beta} W) \ast K_{nC_n}(U_i) \eta_{2}(t) - (f_{\beta} W) \ast K_{nC_n}(U_i) \eta_{2}(t) \right] Y_i(t) dt.
\]

Under Assumption (A16), we get \( \mathbb{E}(|S_{n,1}(\theta) - S_{n,1}(\theta')|^2) = 0(\rho^2) \). \( \square \)

### 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 \( \theta \). We use classical Taylor expansion and the consistency of \( \hat{\theta}_1 \) to get

\[
0 = S_{n,1}^{(1)}(\hat{\theta}_1) = S_{n,1}^{(1)}(\theta^0) + S_{n,1}^{(2)}(\theta^0)(\hat{\theta}_1 - \theta^0) + R_n(\hat{\theta}_1 - \theta^0),
\]

with \( R_n \) defined by

\[
R_n = \int_0^\tau [S_{n,1}^{(2)}(\theta^0) - s(\hat{\theta}_1 - \theta^0)] - [S_{n,1}^{(2)}(\theta^0)] \, ds.
\]

This implies

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

Consequently, we have to verify the four following points

i) \( \mathbb{E} \left[ (S_{n,1}^{(1)}(\theta^0) \ast S_{\theta^0, g}(\theta^0)) (S_{n,1}^{(1)}(\theta^0) \ast S_{\theta^0, g}(\theta^0)) \right] = O(\varphi_n^2) \), with \( \varphi_n \) defined in Theorem 4.1.

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

iii) \( R_n \) defined in (7.10) satisfies \( \mathbb{E} \left( \| R_n \|_2^2 \right) = o(1) \) as \( n \rightarrow \infty \).

iv) \( \mathbb{E} \| \hat{\theta}_1 - \theta^0 \|_2^2 \leq C_{\text{m.p}} \mathbb{E} \left[ (S_{n,1}^{(1)}(\theta^0)) ^\top S_{n,1}^{(1)}(\theta^0) \right] + o(\varphi_n^2) \).

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

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

\[
\|\mathbb{E}(\partial S_{n,1}(\theta)/(\partial \beta_j) |_{\theta = \theta^0})\|_q^2 \leq C_{q^0} \times \min_{q=1,2} \left\{ \| (f_{n,j}^{(1)} W)^* (K_{C_n}^* - 1) \|_q^2 + \| (f_{n,j}^{(0)} W)^* (K_{C_n}^* - 1) \|_q^2 \right\},
\]

and for \( j = 1, \ldots, p \)

\[
\|\mathbb{E}(\partial S_{n,1}(\theta)/(\partial \gamma_j) |_{\theta = \theta^0})\|_q^2 \leq C_{q^0} \times \min_{q=1,2} \left\{ \| (f_{n} W)^* (K_{C_n}^* - 1) \|_q^2 + \| (f_n W)^* (K_{C_n}^* - 1) \|_q^2 \right\}.
\]

Secondly, we will show that for \( j = 1, \ldots, m \)

\[
\text{Var}(\partial S_{n,1}(\theta)/(\partial \beta_j) |_{\theta = \theta^0}) \leq C_{q^0,j} \left[ \min_{q=1,2} \| (f_{n} W)^* (K_{C_n}^* - 1) \|_q^2 + \min_{q=1,2} \| (f_{n} W)^* (K_{C_n}^* - 1) \|_q^2 \right] + \min_{q=1,2} \| (f_{n,j}^{(1)} W)^* (K_{C_n}^* - 1) \|_q^2 + \min_{q=1,2} \| (f_{n,j}^{(0)} W)^* (K_{C_n}^* - 1) \|_q^2 \].
\]

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

\[
\frac{1}{n} \sum_{i=1}^{n} \left( - \int_0^\tau (f_{j0}^{(1)} W) * K_{n,C_n}(U_i) \eta_{n1}(t) dN_i(t) + \int_0^\tau (f_{j0}^{(1)} W) * K_{n,C_n}(U_i) \eta_{n2}(t) Y_i(t) dt - \int_0^\tau (f_{j0}^{(1)} W) * K_{n,C_n}(U_i) \eta_{n2}(t) Y_i(t) dt \right).
\]

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

\[
-2\mathbb{E} \left[ \int_0^\tau (f_{j0}^{(1)} W) * K_{n,C_n}(U_i) \eta_{n1}(t) dN_i(t) \right] + 2\mathbb{E} \left[ \int_0^\tau (f_{j0}^{(1)} W) * K_{n,C_n}(U_i) \eta_{n2}(t) Y_i(t) dt \right].
\]

Hence, Lemma 8.1 implies

\[
\mathbb{E}(\partial S_{n,1}(\theta)/(\partial \beta) |_{\theta = \theta^0}) = -2\mathbb{E} \left[ f_{j0} (Z_1) (f_{j0}^{(1)} W) * K_{C_n}(Z_1) \int_0^\tau \eta_{n2}(t) Y_i(t) dt \right] + 2\mathbb{E} \left[ (f_{j0} f_{j0}^{(1)} W) * K_{C_n}(Z_1) \int_0^\tau \eta_{n2}(t) Y_i(t) dt \right] = 0.
\]

Since \( \partial S_{n,1}(\theta)/(\partial \beta) |_{\theta = \theta^0} \) is equal to

\[
-2\mathbb{E} \left[ f_{j0} (Z_1) (f_{j0}^{(1)} W)(Z_1) \int_0^\tau \eta_{n2}(t) Y_i(t) dt \right] + 2\mathbb{E} \left[ (f_{j0} f_{j0}^{(1)} W)(Z_1) \int_0^\tau \eta_{n2}(t) Y_i(t) dt \right] = 0,
\]

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

\[
-2 \int \left( \left[ f_{j0}(x) f_{j0} f_{j0}^{(1)} W)(x) \right] \int_0^\tau \eta_{n2}(t) I_x \geq t \right) dx + 2 \int \left( \left[ f_{j0}(x) f_{j0}^{(1)} W)(x) \right] \int_0^\tau \eta_{n2}(t) I_x \geq t \right) dx.
\]
Similarly, Lemma 8.1 implies
\[
\mathbb{E}(\partial S_{n,1}(\theta)/(\partial \gamma) \mid \theta = \theta^0) = -2\mathbb{E} \left[ f_{\beta_0}(Z_1)(f_{\beta_0}W) \ast K_{C_n}(Z_1) \int_0^\tau \eta_{\nu_0}^{(1)}(t) \eta_{\nu_0}(t) Y_1(t) \, dt \right] + 2\mathbb{E} \left[ (f_{\beta_0}^2W) \ast K_{C_n}(Z_1) \int_0^\tau \eta_{\nu_0}^{(1)}(t) \eta_{\nu_0}(t) Y_1(t) \, dt \right].
\]

We use that \(\partial S_{\theta_0,g}(\theta)/(\partial \gamma) \mid \theta = \theta^0\) equals
\[
-2\mathbb{E} \left[ f_{\beta_0}(Z_1)(f_{\beta_0}W)(Z_1) \int_0^\tau \eta_{\nu_0}^{(1)}(t) \eta_{\nu_0}(t) Y_1(t) \, dt \right] + 2\mathbb{E} \left[ (f_{\beta_0}^2W)(Z_1) \int_0^\tau \eta_{\nu_0}^{(1)}(t) \eta_{\nu_0}(t) Y_1(t) \, dt \right] = 0,
\]
we obtain \(\mathbb{E}(\partial S_{n,1}(\theta)/(\partial \gamma) \mid \theta = \theta^0) = \mathbb{E}(\partial S_{n,1}(\theta)/(\partial \gamma) \mid \theta = \theta^0 \mid \partial S_{\theta_0,g}(\theta)/(\partial \gamma) \mid \theta = \theta^0\) equals
\[
-2 \int \langle f_{\beta_0}(\cdot) f_{X,Z}(x, \cdot) \rangle \cdot [([f_{\beta_0}W] \ast K_{C_n} - (f_{\beta_0}W))] \int_0^\tau \eta_{\nu_0}^{(1)}(t) \eta_{\nu_0}(t) I_{x \geq \tau} \, dt \, dx
\]
\[+ 2 \int \langle f_{X,Z}(x, \cdot) \rangle \cdot [([f_{\beta_0}^2W] \ast K_{C_n} - (f_{\beta_0}^2W))] \int_0^\tau \eta_{\nu_0}^{(1)}(t) \eta_{\nu_0}(t) I_{x \geq \tau} \, dt \, dx.
\]
Hence we obtain that for \(j = 1, \ldots, m\) \((1/2)\mathbb{E}(\partial S_{n,1}(\theta)/(\partial \beta_j) \mid \theta = \theta^0)\) is less than
\[
\| (f_{\beta_0}^{(1)}W) \ast K_{C_n} - (f_{\beta_0}^{(1)}W) \|_2 \left( \int \| f_{\beta_0}(\cdot) f_{X,Z}(x, \cdot) \|_2 \, dx \right) \left( \int_0^\tau \eta_{\nu_0}(t) \, dt \right)
\]
\[+ \| (f_{\beta_0}^{(1)}W) \ast K_{C_n} - (f_{\beta_0}^{(1)}W) \|_2 \left( \int \| f_{X,Z}(x, \cdot) \|_2 \, dx \right) \left( \int_0^\tau \eta_{\nu_0}^{(1)}(t) \, dt \right)
\]
\[\leq \| (f_{\beta_0}^{(1)}W)^\ast(K_{C_n}^\ast - 1) \|_2 \left( 2\pi \right)^{-1} \left( \int \| f_{\beta_0}(\cdot) f_{X,Z}(x, \cdot) \|_2 \, dx \right) \left( \int_0^\tau \eta_{\nu_0}(t) \, dt \right)
\]
\[+ \| (f_{\beta_0}^{(1)}W)^\ast(K_{C_n}^\ast - 1) \|_2 \left( 2\pi \right)^{-1} \left( \int \| f_{X,Z}(x, \cdot) \|_2 \, dx \right) \left( \int_0^\tau \eta_{\nu_0}^{(1)}(t) \, dt \right).
\]
Similarly, \((1/2)\mathbb{E}(\partial S_{n,1}(\theta)/(\partial \gamma_j) \mid \theta = \theta^0)\) is less than
\[
\| (f_{\beta_0}W)^\ast(K_{C_n}^\ast - 1) \|_2 \left( 2\pi \right)^{-1} \left( \int \| f_{\beta_0}(\cdot) f_{X,Z}(x, \cdot) \|_2 \, dx \right) \left( \int_0^\tau \eta_{\nu_0}^{(1)}(t) \eta_{\nu_0}(t) \, dt \right)
\]
\[+ \| (f_{\beta_0}^2W)^\ast(K_{C_n}^\ast - 1) \|_2 \left( 2\pi \right)^{-1} \left( \int \| f_{X,Z}(x, \cdot) \|_2 \, dx \right) \left( \int_0^\tau \eta_{\nu_0}(t) \, dt \right)
\].

Consequently,
\[
\mathbb{E}(\partial S_{n,1}(\theta)/(\partial \beta_j) \mid \theta = \theta^0) - (\partial S_{\theta_0,g}(\theta)/(\partial \beta_j) \mid \theta = \theta^0) \leq C_{\theta^0} \| (f_{\beta_0}^{(1)}W)^\ast(K_{C_n}^\ast - 1) \|_2 + \| (f_{\beta_0}^{(1)}W)^\ast(K_{C_n}^\ast - 1) \|_2, \quad (7.15)
\]
and
\[
\mathbb{E}(\partial S_{n,1}(\theta)/(\partial \gamma_j) \mid \theta = \theta^0) - (\partial S_{\theta_0,g}(\theta)/(\partial \gamma_j) \mid \theta = \theta^0) \leq C_{\theta^0} \| (f_{\beta_0}W)^\ast(K_{C_n}^\ast - 1) \|_2 + \| (f_{\beta_0}^2W)^\ast(K_{C_n}^\ast - 1) \|_2. \quad (7.16)
\]
We can also write that $(1/2)\mathbb{E}(\partial S_{n,1}(\theta)/\partial \beta_j |_{\theta=\theta_0})$ is bounded by

$$
\| (f_{\beta_{i,j}}W) * K_{\beta_i} - (f_{\beta_{i,j}}W) \|_\infty \left( \int \| f_{\beta_{i,j}}(x,z) \|_1 \, dx \right) \left( \int_0^\tau \eta_{\beta_{i,j}}^2(t) \, dt \right)
+ \| (f_{\beta_{i,j}}W) * K_{\beta_i} - (f_{\beta_{i,j}}W) \|_\infty \left( \int \| f_{X,Z}(x,\cdot) \|_1 \, dx \right) \left( \int_0^\tau \eta_{\beta_{i,j}}^2(t) \, dt \right)
$$

which is less than

$$
\| (f_{\beta_{i,j}}W) * (K_{\beta_i} - 1) \|_1 \left( (2\pi)^{-1} \int \| f_{\beta_{i,j}}(x,z) \|_1 \, dx \right) \left( \int_0^\tau \eta_{\beta_{i,j}}^2(t) \, dt \right)
+ \| (f_{\beta_{i,j}}W) * (K_{\beta_i} - 1) \|_1 \left( (2\pi)^{-1} \int \| f_{X,Z}(x,\cdot) \|_1 \, dx \right) \left( \int_0^\tau \eta_{\beta_{i,j}}^2(t) \, dt \right).
$$

Similarly, $(1/2)\mathbb{E}(\partial S_{n,1}(\theta)/\partial \gamma_j |_{\theta=\theta_0})$ is bounded by

$$
\| (f_{\beta_{i,j}}W) * (K_{\beta_i} - 1) \|_1 \left( (2\pi)^{-1} \int \| f_{\beta_{i,j}}(x,z) \|_1 \, dx \right) \left( \int_0^\tau \eta_{\beta_{i,j}}^2(t) \, dt \right)
+ \| (f_{\beta_{i,j}}W) * (K_{\beta_i} - 1) \|_1 \left( (2\pi)^{-1} \int \| f_{X,Z}(x,\cdot) \|_1 \, dx \right) \left( \int_0^\tau \eta_{\beta_{i,j}}^2(t) \, dt \right).
$$

Consequently,

$$
|\mathbb{E}(\partial S_{n,1}(\theta)/\partial \beta_j |_{\theta=\theta_0}) - \mathbb{E}(\partial S_{n,1}(\theta)/\partial \beta_j |_{\theta=\theta_0})| \leq C_{\theta_0} \| (f_{\beta_{i,j}}W) * (K_{\beta_i} - 1) \|_1 + \| (f_{\beta_{i,j}}W) * (K_{\beta_i} - 1) \|_1, \quad (7.17)
$$

and

$$
|\mathbb{E}(\partial S_{n,1}(\theta)/\partial \gamma_j |_{\theta=\theta_0}) - \mathbb{E}(\partial S_{n,1}(\theta)/\partial \gamma_j |_{\theta=\theta_0})| \leq C_{\theta_0} \| (f_{\beta_{i,j}}W) * (K_{\beta_i} - 1) \|_1 + \| (f_{\beta_{i,j}}W) * (K_{\beta_i} - 1) \|_1. \quad (7.18)
$$

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 $\text{Var}(\partial S_{n,1}(\theta)/\partial \beta_j |_{\theta=\theta_0}) = (8 + o(1))n^{-1}[V_{1,j} + V_{2,j}]$, with

$$
V_{1,j} = \mathbb{E} \left[ (f_{\beta_{i,j}}W) K_{\beta_i} (U_1) \int_0^\tau \eta_{\beta_{i,j}}^2(t) Y_1(t) \, dt \right]^2
$$

and

$$
V_{2,j} = \mathbb{E} \left[ (f_{\beta_{i,j}}W) K_{\beta_i} (U_1) \int_0^\tau \eta_{\gamma_{\beta_{i,j}}}^2(t) \, dt \right]^2.
$$

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

$$
V_{3,j} = \mathbb{E} \left[ (f_{\beta_{i,j}}W) \int_0^\tau \eta_{\gamma_{\beta_{i,j}}}^2(t) Y_1(t) \, dt \right]^2
$$

and

$$
V_{4,j} = \mathbb{E} \left[ (f_{\beta_{i,j}}W) \int_0^\tau \eta_{\gamma_{\beta_{i,j}}}^2(t) \, dt \right]^2.
$$
Lemma 8.1 implies that

\[ V_{1,j} \leq \left( \int_0^\tau \eta_{\rho \sigma}(t)dt \right)^2 \int \left\langle f_{X,Z}(x, \cdot) \ast f_\varepsilon, \left( (f_{\beta}^{(1)} f_{\rho}\ast W) \ast K_{n,C_n} \right)^2 \right\rangle dx \]

and

\[ V_{3,j} \leq \left( \int_0^\tau \eta_{\rho \sigma}^{(1)}(t)\eta_{\rho \sigma}(t)\eta_{\rho \sigma}^{(1)}(t)\eta_{\rho \sigma}(t)dt \right)^2 \int \left\langle f_{X,Z}(x, \cdot) \ast f_\varepsilon, \left( (f_{\beta}^{(2)} W) \ast K_{n,C_n} \right)^2 \right\rangle dx. \]

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

\[ V_{1,j} \leq \left( \int_0^\tau \eta_{\rho \sigma}^2(t)dt \right)^2 \max(\| f_\varepsilon \|_{\infty, 1}) \min_{q=1,2} \left\| (f_{\beta}^{(1)} f_{\rho}\ast W) \ast K_{n,C_n} \right\|^2_{q} \tag{7.19} \]

and

\[ V_{3,j} \leq \left( \int_0^\tau \eta_{\rho \sigma}^{(1)}(t)\eta_{\rho \sigma}(t)dt \right)^2 \max(\| f_\varepsilon \|_{\infty, 1}) \min_{q=1,2} \left\| (f_{\beta}^{(2)} W) \ast K_{n,C_n} \right\|^2_{q}, \tag{7.20} \]

Now, we apply Lemma 8.1 and have

\[ V_{2,j} \leq \int \left\langle \varphi_{\rho \sigma}^2(x, \cdot) f_{X,Z}(x, \cdot) \ast f_\varepsilon, \left( (f_{\beta}^{(1)} W) \ast K_{n,C_n} \right)^2 \right\rangle dx, \]

and

\[ V_{4,j} \leq \int \left\langle \varphi_{\rho \sigma}^2(x, \cdot) f_{X,Z}(x, \cdot) \ast f_\varepsilon, \left( (f_{\beta}^{(2)} W) \ast K_{n,C_n} \right)^2 \right\rangle dx, \]

where \( \varphi_{\rho}(X, Z) = \int_0^\tau \eta_{\rho \sigma}(t)dt \) and \( \varphi_{\rho \sigma}(X, Z) = \int_0^\tau \eta_{\rho \sigma}^{(1)}(t) \) dt. We apply inequalities (7.1) and (7.2) to get

\[ V_{2,j} \leq \mathbb{E} \left( \int_0^\tau \eta_{\rho \sigma}(t)dt \right)^2 \max(\| f_\varepsilon \|_{\infty, 1}) \min_{q=1,2} \left\| (f_{\beta}^{(1)} W) \ast K_{n,C_n} \right\|^2_{q} \tag{7.21} \]

and

\[ V_{4,j} \leq \mathbb{E} \left( \int_0^\tau \eta_{\rho \sigma}^{(1)}(t) \right)^2 \max(\| f_\varepsilon \|_{\infty, 1}) \min_{q=1,2} \left\| (f_{\beta}^{(2)} W) \ast K_{n,C_n} \right\|^2_{q}. \tag{7.22} \]

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 \( S_{n,1}, S_{n,1}^{(2)}(\theta^0) = \frac{\partial^2 S_{n,1}(\theta^0)}{\partial \theta^2} = \left( \begin{array}{cc} (S_{n,1}^{(2)})_{1,1} & (S_{n,1}^{(2)})_{1,2} \\ (S_{n,1}^{(2)})_{1,1} & (S_{n,1}^{(2)})_{1,2} \end{array} \right), \)

with

\[ (S_{n,1}^{(2)})_{1,2} = -\frac{2}{n} \sum_{i=1}^n (f_{\beta}^{(1)} W) \ast K_{n,C_n}(U_i) \int_0^\tau (\eta_{\rho \sigma}^{(1)}(t))^\top dN_i(t) \]

\[ + \frac{1}{n} \sum_{i=1}^n (f_{\beta}^{(1)} f_{\rho}\ast W) \ast K_{n,C_n}(U_i) \int_0^\tau (\eta_{\rho \sigma}^{(1)}(t))^\top \eta_{\rho \sigma}(t)Y_{i}(t)dt, \]

\[ (S_{n,1}^{(2)})_{1,1} = -\frac{2}{n} \sum_{i=1}^n (f_{\beta}^{(2)} W) \ast K_{n,C_n}(U_i) \int_0^\tau \eta_{\rho \sigma}(t)dN_i(t) \]

\[ + \frac{1}{n} \sum_{i=1}^n \left( \frac{\partial^2 (f_{\beta}^{(2)} W)}{\partial \theta^2} \right)_{\theta=\theta^0} \ast K_{n,C_n}(U_i) \int_0^\tau \eta_{\rho \sigma}^{(1)}(t)Y_{i}(t)dt \]
and

$$(S_{n,1}^{(2)})_{2,2}(\theta) = \frac{-2}{n} \sum_{i=1}^{n} (f_{\beta \theta} W) * K_{n,C_n}(U_i) \int_0^{\tau_i} \eta_{i,j}^{(2)}(t) dN_i(t)$$

$$+ \frac{1}{n} \sum_{i=1}^{n} (f_{\beta \theta} W) * K_{n,C_n}(U_i) \int_0^{\tau_i} \left( \frac{\partial^2 \eta_{i,j}^{(2)}(t)}{\partial \gamma^2} \right) \left|_{\theta=\theta_0} \right) Y_i(t) dt.$$ 

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

**Proof of iii)**

The proof of iii) follows by using the smoothness of $\beta \mapsto W_{f_\beta}$ and $\beta \mapsto W_{f_\beta}^2$ up to order 3, the smoothness of $\gamma \mapsto \eta_1$ and $\gamma \mapsto \eta_2$ and by using the consistency of $\hat{\theta}_1$.

**Proof of iv)**

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

$$E_{n,j,k} = \left\{ \omega \text{ such that } \left| \frac{\partial^2 S_{\theta_0,g}(\theta)}{\partial \theta_j \partial \theta_k} - \frac{\partial^2 S_{n,1}(\theta_0,\omega)}{\partial \theta_j \partial \theta_k} + (R_n)_{j,k}(\omega) \right|_{\theta=\theta_0} \leq 1/2 \right\}.$$ 

Now, we decompose $\mathbb{E}[\|\hat{\theta}_1 - \theta_0\|^2_{E_n}]$ on the event $E_n$ and its complementary event in the following way

$$\mathbb{E}[\|\hat{\theta}_1 - \theta_0\|^2_{E_n}] = \mathbb{E}[\|\hat{\theta}_1 - \theta_0\|^2_{E_n}] + \mathbb{E}[\|\hat{\theta}_1 - \theta_0\|^2_{E_n^c}].$$

We use that $\hat{\theta}_1$ and $\theta_0$ belong both to a compact set and get

$$\mathbb{E}[\|\hat{\theta}_1 - \theta_0\|^2_{E_n}] \leq \mathbb{E}[\|\hat{\theta}_1 - \theta_0\|^2_{E_n}] + 2 \sup_{\theta \in \Theta} \|\theta\|^2_{E_n} \mathbb{P}(E_n^c).$$

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

$$\mathbb{E}[\|\hat{\theta}_1 - \theta_0\|^2_{E_n}] \leq C_{m,p,\theta_0} \mathbb{E}\left[ (S_{n,1}^{(1)}(\theta_0))^T S_{n,1}^{(1)}(\theta_0) \right] \text{ and } \mathbb{P}(E_n^c) = o(\varphi_n^2). \quad (7.23)$$

We use (7.10) and (7.11) to write

$$\mathbb{E}[\|\hat{\theta}_1 - \theta_0\|^2_{E_n}] \leq C_{m,p} \sup_{j,k} \left| \frac{\partial^2 S_{\theta_0,g}(\theta)}{\partial \theta_j \partial \theta_k} \right|_{\theta=\theta_0} - 2 \mathbb{E}\left[ (S_{n,1}^{(1)}(\theta_0))^T S_{n,1}^{(1)}(\theta_0) \right].$$

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

$$\mathbb{P}(E_{n,j,k}^c) \leq \left( \frac{1}{2} \left| \frac{\partial^2 S_{\theta_0,g}(\theta)}{\partial \theta_j \partial \theta_k} \right|_{\theta=\theta_0} \right)^{-1} \mathbb{E}\left[ \left| \left( \frac{\partial^2 (S_{\theta_0,g}(\theta) - S_{n,1}(\theta))}{\partial \theta_j \partial \theta_k} \right) \left|_{\theta=\theta_0} \right) + (R_n)_{j,k} \right|^q \right].$$
Proof of Theorem 4.2: asymptotic normality

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

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

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

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

and write

\[ \mathbb{E} \left[ n^{-1} \sum_{i=1}^n W_{n,i,j,k} \right]^q \leq C_4 \left[ n^{1-r} \mathbb{E} |W_{n,i,j,k}|^q + n^{-q/2} \mathbb{E}^{q/2} |W_{n,i,j,k}|^2 \right]. \]

Take \( q = 4 \) to get

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

Therefore under the conditions ensuring that

\[ \mathbb{E} \left[ \left( \frac{\partial^2 S_{\beta, \gamma}(\theta)}{\partial \theta_j \partial \theta_k} \bigg|_{\theta=\theta^0} - \frac{\partial^2 S_{\beta, \gamma}(\theta)}{\partial \theta_j \partial \theta_k} \bigg|_{\theta=\theta^0} \right)^2 \right] = o(1), \]

we have

\[ \mathbb{E} \left[ \left( \frac{\partial^2 S_{\beta, \gamma}(\theta)}{\partial \theta_j \partial \theta_k} \bigg|_{\theta=\theta^0} - \frac{\partial^2 S_{\beta, \gamma}(\theta)}{\partial \theta_j \partial \theta_k} \bigg|_{\theta=\theta^0} \right)^4 \right] = O(\varphi_n^4) = o(\varphi_n^2). \]

Now, we use the definition of \( R_n \) and the smoothness properties of the derivatives of \((W f_\beta)\) and \((W f_\beta^n)\) with respect to \( \beta \), up to order 3 and get \( \mathbb{E}((R_n)_{j,k}^4) = o(\|\hat{\theta}_1 - \theta^0\|^4) \). Thus \( \mathbb{P}(E_n^c) = o(\varphi_n^2) + O(\mathbb{E}[\|\hat{\theta}_1 - \theta^0\|_2^4]) = o(\varphi_n^2) \), and (7.23) is proved. \( \Box \)

7.2. Proof of Theorem 4.2: asymptotic normality

Theorem 4.1, its proof, and conditions (C1)–(C3) imply that \( V_{n,j}(\theta^0) = O(1) \) and the asymptotic normality of \( \hat{\theta}_1 \) follows if we verify that \( \sqrt{n} S_{n,1}(\theta^0) \overset{\mathbb{P}}{\to} N(0, \Sigma_1) \), with \( \Sigma_1 \) defined in Theorem 4.2.

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

\[ \hat{H}_{n,i}(s) = \left( \frac{-2}{\sqrt{n}} (f_{j,0}^{(1)} W)(Z_i) \eta_{\gamma,n}(s) \right), \quad H_{n,i}(s) = \left( \frac{-2}{\sqrt{n}} (f_{j,0}^{(1)} W)(Z_i) \eta_{\gamma,n}(s) \right), \]

\[ \hat{G}_{n,i}(s) = \left( \frac{2}{\sqrt{n}} (f_{j,0}^{(1)} W)(Z_i) \eta_{\gamma,n}(s) \right), \quad G_{n,i}(s) = \left( \frac{2}{\sqrt{n}} (f_{j,0}^{(1)} W)(Z_i) \eta_{\gamma,n}(s) \right). \]
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

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

with

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

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

**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] p. 68):

**L1)** $\sum_{i=1}^{n} \int_0^t H_{n,i}(s)(H_{n,i}(s))^\top d\langle M_i \rangle(s) \xrightarrow{n \to \infty} \Sigma_1^2(t)$, with $\Sigma_1^2(t)$ a positive covariance matrix defined by

$$\Sigma_1^2(t) = 4 \mathbb{E} \left[ \int_0^t \left( f^{(1)}_{\beta^0}(W)(Z_i) \eta_{\gamma,\phi}(s) \right) \left( f^{(1)}_{\beta^0}(W)(Z_i) \eta_{\gamma,\phi}(s) \right)^\top \eta_{\gamma,\phi}(s) Y_i(s) ds \right].$$

**L2)** For all $\epsilon > 0$, $\sum_{i=1}^{n} \int_0^t H_{n,i}(s)(H_{n,i}(s))^\top 1_{||H_{n,i}(s)||_{\mathbb{R}^2} \leq \epsilon} d\langle M_i \rangle(s) \xrightarrow{P} 0$.

**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 $[0, \tau]$,

$$\sum_{i=1}^{n} \int_0^t H_{n,i}(s)(H_{n,i}(s))^\top Y_i(s) f^{(1)}_{\beta^0}(Z_i) \eta_{\gamma,\phi}(s) ds \xrightarrow{P} n \to \infty \Sigma_1^2(t).$$

We apply the following Lemma, which is a straightforward consequence of the fact that the set of functions $I_2 = \{ x \mapsto 1_{x \geq 1} \}$ is a $\mathbb{P}$-Glivenko Cantelli class (see [38]).

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

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

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

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

and

$$\sup_{0 \leq t \leq \tau} \left| \frac{1}{n} \sum_{i=1}^{n} Y_{i}(t)(f^{(1)}_{\beta^0,j}(W)(Z_i))^3 - \mathbb{E}[Y(t)(f^{(1)}_{\beta^0,j}(W)(Z))^3] \right| \xrightarrow{a.s.} 0.$$
Thus L1) is verified.

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

$$
\frac{1}{n} E \left[ \sum_{i=1}^{n} \int_{0}^{t} ((f_{j,0,i}^{(1)} W(Z_i),\eta_{i,0}(s)) \cdot I_{[(f_{j,0,i}^{(1)} W(Z_i),\eta_{i,0}(s)) \geq \epsilon_{\sqrt{3}} f_{j,0}^{1}(Z_i) \eta_{i,0}(s)]} Y_i(s) ds \right] = o(1)
$$

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

$$
\frac{1}{n} E \left[ \sum_{i=1}^{n} \int_{0}^{t} ((f_{j,0,i}^{(1)} W(Z_i),\eta_{i,0}(s)) \cdot I_{[(f_{j,0,i}^{(1)} W(Z_i),\eta_{i,0}(s)) \geq \epsilon_{\sqrt{3}} f_{j,0}^{1}(Z_i) \eta_{i,0}(s)]} Y_i(s) ds \right] = o(1).
$$

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

$$
\frac{1}{n} E \left[ \sum_{i=1}^{n} \int_{0}^{t} ((f_{j,0,i}^{(1)} W(Z_i),\eta_{i,0}(s)) \cdot I_{[(f_{j,0,i}^{(1)} W(Z_i),\eta_{i,0}(s)) \geq \epsilon_{\sqrt{3}} f_{j,0}^{1}(Z_i) \eta_{i,0}(s)]} Y_i(s) ds \right] \leq \frac{1}{n \sqrt{n}} E \left[ \sum_{i=1}^{n} \int_{0}^{t} ((f_{j,0,i}^{(1)} W(Z_i),\eta_{i,0}(s)) \cdot I_{[(f_{j,0,i}^{(1)} W(Z_i),\eta_{i,0}(s)) \geq \epsilon_{\sqrt{3}} f_{j,0}^{1}(Z_i) \eta_{i,0}(s)]} Y_i(s) ds \right] = o(1)
$$

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

$$
\frac{1}{n} E \left[ \sum_{i=1}^{n} \int_{0}^{t} ((f_{j,0,i}^{(1)} W(Z_i),\eta_{i,0}(s)) \cdot I_{[(f_{j,0,i}^{(1)} W(Z_i),\eta_{i,0}(s)) \geq \epsilon_{\sqrt{3}} f_{j,0}^{1}(Z_i) \eta_{i,0}(s)]} Y_i(s) ds \right] \leq \frac{1}{n \sqrt{n}} E \left[ \sum_{i=1}^{n} \int_{0}^{t} ((f_{j,0,i}^{(1)} W(Z_i),\eta_{i,0}(s)) \cdot I_{[(f_{j,0,i}^{(1)} W(Z_i),\eta_{i,0}(s)) \geq \epsilon_{\sqrt{3}} f_{j,0}^{1}(Z_i) \eta_{i,0}(s)]} Y_i(s) ds \right] = o(1).
$$

Thus L2) is verified.

Study of $A_2$

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

**Lemma 7.2.** Under (N2) and (R1), for $C_n$ satisfying (4.3) then for $j = 1, \ldots, m$

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

and

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

Study of $A_3$

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

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

**v-b)** $E(A_3) = o_p(1)$;

**v-c)** for all $\epsilon > 0$, $\sum_{i=1}^{n} E[(V_n,i)^2 I_{\|V_n,i\|_{L^2} \geq \epsilon}] \xrightarrow{n\to\infty} 0$ (Lindeberg condition).
By definition, $A_3$ is equal to

$$-\frac{2}{\sqrt{n}} \sum_{i=1}^{n} \int_{0}^{\tau} \left( (f_{j_0}^{(1)} W) \ast K_{n,C_n}(U_i) - (f_{j_0}^{(1)} W)(Z_i) \eta_{j_0}(s) \right) Y_i(s) f_{j_0}(Z_i) \eta_{j_0}(s) ds.$$

Let us start with the study of the variance (v-a). Under $(C_1)-(C_3)$

$$\text{Var} \left[ -\frac{2}{\sqrt{n}} \sum_{i=1}^{n} \int_{0}^{\tau} (f_{j_0}^{(1)} W) \ast K_{n,C_n}(U_i) - (f_{j_0}^{(1)} W)(Z_i) \eta_{j_0}(s) Y_i(s) f_{j_0}(Z_i) \eta_{j_0}(s) ds \right] = O(1),$$

and

$$\text{Var} \left[ -\frac{2}{\sqrt{n}} \sum_{i=1}^{n} \int_{0}^{\tau} (f_{j_0}^{(1)} W) \ast K_{n,C_n}(U_i) - (f_{j_0}^{(1)} W)(Z_i) \eta_{j_0}(s) Y_i(s) f_{j_0}(Z_i) \eta_{j_0}(s) ds \right] = O(1).$$

It follows that v-a is verified.

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

$$-2\sqrt{n} \left( E \left[ \left( (f_{j_0}^{(1)} W) \ast K_{C_n}(z) - (f_{j_0}^{(1)} W)(z) \right) f_{j_0}(Z_i) \int_{0}^{\tau} \eta_{j_0}(s) Y(s) ds \right] \right).$$

Lemma 8.1 implies

$$E(A_3) = \left( E \left[ \left( (f_{j_0}^{(1)} W) \ast K_{C_n}(z) - (f_{j_0}^{(1)} W)(z) \right) f_{j_0}(Z_i) \int_{0}^{\tau} Y(s) \eta_{j_0}(s) \eta_{j_0}(s) ds \right] \right).$$

that is $E(A_3)$ is equal to

$$-2\sqrt{n} \left( \int \left( (f_{j_0}^{(1)} W) \ast K_{C_n}(z) - (f_{j_0}^{(1)} W)(z), f_{j_0}(z)f_{X,Z}(x,z) \right) \int_{0}^{\tau} Y(s) \eta_{j_0}(s) \eta_{j_0}(s) ds dx \right).$$

For $j = 1, \ldots, m$

$$\left| \int \langle (f_{j_0}^{(1)} W) \ast K_{C_n}(z) - (f_{j_0}^{(1)} W)(z), f_{j_0}(z)f_{X,Z}(x,z) \rangle \left( \int_{0}^{\tau} Y(s) \eta_{j_0}(s) \eta_{j_0}(s) ds dx \right) \right|$$

is less than

$$\left( \int_{0}^{\tau} \eta_{j_0}(s) ds \right) \int \left| \langle (f_{j_0}^{(1)} W) \ast K_{C_n}(z) - (f_{j_0}^{(1)} W)(z), f_{j_0}(z)f_{X,Z}(x,z) \rangle \right| dx.$$
which is less than
\[(2\pi)^{-1} \left( \int_0^\tau \eta_\omega^2(s) ds \right) \min \left\{ \| (f_{\beta_1,j}^{(1)} W)^*(K_{C_n}^* - 1) \|_2, \sqrt{\int f_{\beta_1}(\cdot) f_{X,Z}(x, \cdot) dx} \right. \]
\[\left. \| (f_{\beta_1,j}^{(1)} W)^*(K_{C_n}^* - 1) \|_\infty \int f_{\beta_1}(\cdot) f_{X,Z}(x, \cdot) \|_1 dx \right\}.\]
Similarly, we obtain that for \(j = 1, \ldots, p\)
\[| \int \langle f_{\beta_3} W \rangle \star K_{C_n} (z) - (f_{\beta_3} W)(z), f_{\beta_3}(z) \rangle \left( \int_0^\tau |\eta_\omega(s) \eta_{\omega,j}(s)| ds \right) dx | \]
is less than
\[(2\pi)^{-1} \left( \int_0^\tau |\eta_\omega(s) \eta_{\omega,j}(s)| ds \right) \min \left\{ \| (f_{\beta_3} W)^*(K_{C_n}^* - 1) \|_2, \sqrt{\int f_{\beta_3}(\cdot) f_{X,Z}(x, \cdot) dx} \right. \]
\[\left. \| (f_{\beta_3} W)^*(K_{C_n}^* - 1) \|_\infty \int f_{\beta_3}(\cdot) f_{X,Z}(x, \cdot) \|_1 dx \right\}.\]

Consequently, under \(\mathbf{R}_4\), \(\mathbb{E}(A_3) = O(\sqrt{\alpha n^{\alpha - (1-r)/2(1-r) - 1/2}} \exp(-dC_{C_n}^* \alpha))\). Under \(\mathbf{C}_1 - \mathbf{C}_3\), \(\text{Var}(A_3) = O(1)\) and hence \(C_n\) can be chosen such that \(\mathbb{E}(A_3) = o(1)\) and \(\mathbf{v-b}\) is verified.

In order to verify \(\mathbf{v-c}\) (Lindeberg condition) we write that for \(j = 1, \ldots, m\)
\[\frac{1}{n} \left\langle \sum_{i=1}^n \int_0^t (f_{\beta_2,i}^{(1)} W) \star K_{n,C_n}(U_i) \eta_{\omega,j}(s) \right\rangle^2 \mathbf{1}_{|f_{\beta_2,i}^{(1)} W\star K_{n,C_n}(U_i)\eta_{\omega,j}(s)| \geq \varepsilon \sqrt{n} f_{\beta_2}(Z_i) \eta_{\omega,j}(s) Y_i(s) ds} \leq \frac{1}{n \sqrt{n \varepsilon}} \mathbb{E} \left[ \sum_{i=1}^n \int_0^t (f_{\beta_2,i}^{(1)} W) \star K_{n,C_n}(U_i) \eta_{\omega,j}(s) \right\rangle^2 f_{\beta_2}(Z_i) \eta_{\omega,j}(s) Y_i(s) ds] \right] = o(1)\]
and for \(j = 1, \ldots, p\)
\[\frac{1}{n} \left\langle \sum_{i=1}^n \int_0^t (f_{\beta_2,i}^{(1)} W) \star K_{n,C_n}(U_i) \eta_{\omega,j}(s) \right\rangle^2 \mathbf{1}_{|f_{\beta_2,i}^{(1)} W\star K_{n,C_n}(U_i)\eta_{\omega,j}(s)| \geq \varepsilon \sqrt{n} f_{\beta_2}(Z_i) \eta_{\omega,j}(s) Y_i(s) ds} \leq \frac{1}{n \sqrt{n \varepsilon}} \mathbb{E} \left[ \sum_{i=1}^n \int_0^t (f_{\beta_2,i}^{(1)} W) \star K_{n,C_n}(U_i) \eta_{\omega,j}(s) \right\rangle^2 f_{\beta_2}(Z_i) \eta_{\omega,j}(s) Y_i(s) ds] \right] = o(1).\]

**Study of \(A_4\)**
The study of \(A_4\), quite similar to the study of \(A_3\) is omitted.

### 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
\[|\mathbb{E}(\partial S_{n,1}(\theta) / (\partial \beta_j) |_{\theta = \theta_0})|^2 = O(C_n^{-2a_1+r+1}\exp(-2dC_{C_n}^*))\]
and
\[|\mathbb{E}(\partial S_{n,1}(\theta) / (\partial \gamma_j) |_{\theta = \theta_0})|^2 = O(C_n^{-2a_1+r+1}\exp(-2dC_{C_n}^*)\rangle / n).\]

We apply Lemma 8.2 to the variance bounds (7.19), (7.20), (7.21), (7.22) and get
\[\text{Var}(\partial S_{n,1}(\theta) / (\partial \gamma_j)) = O(C_n^{2(\alpha - \rho + 1)^{1/2}} \exp(-2dC_{C_n}^* + 2dC_{C_n}^*) / n).\]
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 $\varphi$ be such that $E(|\varphi(X,Z)|) < \infty$ and let $\Phi$ such that $E(|\Phi(U)|) < \infty$. Under assumptions (A5) and (A4), then $E[\varphi(X,Z)\Phi \ast K_n,C_n(U)] = E[\varphi(X,Z)\Phi \ast K_{C_n}(Z)]$,

and $E[\varphi(X,Z)\Phi \ast K_n,C_n(U)]^2 = \int (\varphi^2(x,\cdot)f_{X,Z}(x,\cdot)) \ast f_\varepsilon, (\Phi \ast K_{n,C_n})^2) \, dx$.

**Proof of Lemma 8.1.** We apply Parseval’s formula and write

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

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

Similarly, $E[\varphi(X,Z)\Phi \ast K_n,C_n(U)]^2 = \int (\varphi^2(x,\cdot)f_{X,Z}(x,\cdot)) \ast f_\varepsilon, (\Phi \ast K_{n,C_n})^2) \, dx$. \hfill $\Box$

**Lemma 8.2.** For $\nu$, $\lambda$ and $\rho$ nonnegative numbers, then

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

Furthermore, if $f_\varepsilon$ satisfies $(N_2)$, then

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

**Acknowledgements.** The authors would like to thank the Associate Editor and the Referee for carefully reading the paper and for their comments who led to substantial improvements of the paper.

**References**


