HAL
open science

New M-estimators in semiparametric regression with errors in variables

Cristina Butucea, Marie-Luce Taupin

To cite this version:

Cristina Butucea, Marie-Luce Taupin. New M-estimators in semiparametric regression with errors in variables. Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques, 2008, 44 (3), pp.393-421. hal-00013229v2

HAL Id: hal-00013229

https://hal.science/hal-00013229v2

Submitted on 4 Nov 2005

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ccsd-00013229, version 2 - 4 Nov 2005

NEW M-ESTIMATORS IN SEMIPARAMETRIC REGRESSION WITH ERRORS IN VARIABLES

CRISTINA BUTUCEA[1,2], MARIE-LUCE TAUPIN[3,4]

Abstract. In the regression model with errors in variables, we observe n i.i.d. copies of (Y, Z) satisfying $Y = f_{\theta^0}(X) + \xi$ and $Z = X + \varepsilon$ involving independent and unobserved random variables X, ξ, ε plus a regression function f_{θ^0}, known up to some finite dimensional θ^0. The common densities of the X_i's and of the ξ_i's are unknown whereas the distribution of ε is completely known. We aim at estimating the parameter θ^0 by using the observations $(Y_1, Z_1), \cdots, (Y_n, Z_n)$. We propose two estimation procedures based on the least square criterion $\tilde{S}_{\theta^0,g}(\theta) = \mathbb{E}_{\theta^0,g}[((Y - f_\theta(X))^2 w(X)]$ where w is some weight function, to be chosen. In the first estimation procedure, w does not depend on θ and the distribution of the ξ's is unknown. The second estimation procedure is based on $S_{\theta^0,g}(\theta) = \mathbb{E}_{\theta^0,g}[((Y - f_\theta(X))^2 - \sigma^2_{\xi,2}) w_\theta(X)]$ where w_θ is positive weight function, to be chosen, and requires the knowledge of $\sigma^2_{\xi,2} = \mathrm{Var}(\xi)$. In both cases, we propose two estimators and derive upper bounds for the risk of those estimators, depending on the smoothness of the errors density p_ε and on the smoothness properties of $w(x)f_\theta(x)$ or $w_\theta(x)f_\theta(x)$ with respect to x. Furthermore we give sufficient conditions that ensure that the parametric rate of convergence is achieved. We provide practical recipes for the choice of w or w_θ in the case of nonlinear regression functions which are smooth on pieces allowing to gain in the order of the rate of convergence, up to the parametric rate in some cases.

[1] Université Paris X, Modal'X, 200, ave. de la République, 92001 Nanterre Cedex, France
[2] Université Paris VI, PMA, 175, rue de Chevaleret, 75013 Paris, France (butucea@ccr.jussieu.fr)
[3] Université Paris Descartes, IUT de Paris, 143 ave. de Versailles, 75016 Paris, France
[4] Université Paris-Sud, Bât. 425, Département de Mathématiques, 91405 Orsay Cedex, France (marie-luce.taupin@math.u-psud.fr)

November 4, 2005

Keywords: Asymptotic normality, consistency, deconvolution kernel estimator, errors-in-variables model, M-estimators, ordinary smooth and super-smooth functions, rates of convergence, semi-parametric nonlinear regression.
AMS 2000 MSC: Primary 62J02, 62F12, Secondary 62G05, 62G20.

1. Introduction

We consider the regression model with errors in variables where one observe n independent and identically distributed (i.i.d.) copies of (Y, Z) satisfying

$$\begin{cases} Y &= f_{\theta^0}(X) + \xi \\ Z &= X + \varepsilon, \end{cases}$$

involving independent and unobserved random variables X, ξ, ε, plus a regression function f_{θ^0} known up to a finite dimensional parameter θ^0, belonging to the interior of a compact set of $\Theta \subset \mathbb{R}^d$. The common densities of the X_i's and of the ξ_i's are unknown, with $\mathbb{E}(\xi) = 0$, whereas the density of the errors ε_i's is completely known.

In this context we aim at estimating the finite dimensional parameter θ^0 in presence of a functional nuisance parameter g, the density of the X.

Previous known results. This model has been widely studied with first results written in the 1950's (see for instance Reiersøl (1950) or Kiefer and Wolfowitz (1956)). Most of results deal with linear models where $\sqrt{n}$-consistency, asymptotic normality and efficiency have been studied. One can cite among the others Bickel and Ritov (1987), Bickel *et al.* (1993), Cheng and van Ness (1994), van der Vaart (1988), (1996), (2002), Murphy and van der Vaart (1996). The nonlinear models have been more recently considered starting with the case of repeated measurement data as in Fuller (1987), Wolter and Fuller (1982b), (1982a), under additional assumptions as in Gleser (1990), Hsiao (1989), Li (2002), Kukush and Schneeweiss (2005a), (2005b) or by simulation (see Carroll (1995), Hsiao *et al.* (2000), (1997), Li (2000)). To our knowledge the first consistent estimator in nonlinear regression models with errors in variables, under nonparametric assumptions on the design density g has been proposed by Taupin (2001) when the errors ε are Gaussian and by Comte, Taupin (2001) in the context of auto-regressive models with errors in variables for various types of errors density p_ε. In those papers, the estimation procedure is based on the estimation of the modified least square criterion $\mathbb{E}\left[(Y - \mathbb{E}(f_\theta(X)|Z))^2 W(Z)\right]$ where the conditional expectation is estimated by using the observations $Z_1, \cdots, Z_n$ and where W is some weight compactly supported function. It is also stated that the rate of convergence, which has not an explicit form, depends on the smoothness of the regression function as well as the smoothness of p_ε through the increase of the ratio $(f_\theta p_\varepsilon(z - \cdot))^*(t)/p_\varepsilon^*(t)$ as t goes to infinity. The parametric rate of convergence is achieved in some specific examples, such as polynomial or exponential regression functions. The main drawback of this estimator is that, besides its complexity, its rate of convergence has not an explicit form.

More recently, Hong and Tamer (2003) propose a consistent estimator in the specific case where p_ε^* is of the form $p_\varepsilon^*(t) = 1/(1 + \sigma^2 t^2)$. Their estimation procedure strongly depends on this particular form of p_ε^* through the ratio $1/p_\varepsilon^*$ always appearing in errors in variables techniques. The extension of the method for other errors density p_ε seems thus not concluding.

Our results. We propose here new estimation procedures, more explicit, natural and tractable than the one proposed by Taupin (2001), that often provide better results, by using the weight function w in order to smooth the regression function by a multiplication. Furthermore, this new estimation procedure allows to provide sufficient conditions to achieve the parametric rate.

Those estimation procedures are based on the least square criterion, both of them allowing to construct two estimators.

The first estimation procedure, is based on the estimation, using the observations (Y_i, Z_i) for $i = 1, \cdots, n$, of the least squares criterion $\tilde{S}_{\theta^0,g} := \mathbb{E}_{\theta^0,g}[(Y - f_\theta(X))^2 w(X)]$ where w is some positive weight function, to be chosen.

In this context, we define two estimators, and the first one is naturally defined by

$$\tilde{\theta}_1 = \arg\min_{\theta\in\Theta} \tilde{S}_{n,1}(\theta) \text{ with } \tilde{S}_{n,1}(\theta) = \frac{1}{n}\sum_{i=1}^{n}\int [(Y_i - f_\theta(x))^2 w(x)] C_n K_n(C_n(x - Z_i))dx,$$

where K_n is a deconvolution kernel such that its Fourier transform verifies $K_n^*(t) = K^*(t)/p_\varepsilon^*(tC_n)$, for K^* compactly supported, C_n a sequence $C_n \to \infty$, and p^*, defined by $p^*(t) = \int e^{itx} f(x)dx$ for p an arbitrary square integrable function.

We show that under classical identifiability, moment and smoothness assumptions, this estimator $\tilde{\theta}_1$ is a consistent estimator of θ^0 with a rate of convergence depending on two factors, the smoothness of (wf_θ) and on the smoothness of the errors density p_ε. This partly comes from the fact that this estimation procedure is based on the estimation of the linear integral functional $\mathbb{E}_{\theta^0,g}(Yw(X)f_\theta(X))$ and $\mathbb{E}_{\theta^0,g}(w(X)f_\theta^2(X))$ using the observations $(Y_1, Z_1), \cdots, (Y_n, Z_n)$, that is by recovering some information on (Y, X) using the observation (Y, Z). More precisely, it depends on the smoothness of $w(x)\partial f_\theta(x)/\partial\theta$ and $w(x)\partial(f_\theta^2(x))/\partial\theta$ and the smoothness of the errors density $p_\varepsilon(x)$, as functions of x through the behavior of $(w\partial f_\theta/\partial\theta)^*(t)/p_\varepsilon^*(t)$, $(w\partial(f_\theta^2)/\partial\theta)^*(t)/p_\varepsilon^*(t)$ as function of $t \to \infty$. From this construction we derive some sufficient conditions that ensure that $\tilde{\theta}_1$ achieves the parametric rate of convergence.

The second estimator, more simple, is based on conditions ensuring that $\mathbb{E}_{\theta^0,g}(Yw(X)f_\theta(X))$ and $\mathbb{E}_{\theta^0,g}(w(X)f_\theta^2(X))$ are estimated at the parametric rate $\sqrt{n}$. In that case the parametric rate of convergence for the estimation of θ^0 can be achieved and the asymptotic normality of this estimator is stated.

The first estimator is more general and applicable in all setups, but for some specific regression functions, the use of deconvolution kernel is not required. In those simple cases, the second estimator is more simple. Nevertheless, the conditions ensuring that $\mathbb{E}_{\theta^0,g}(Yw(X)f_\theta(X))$ and $\mathbb{E}_{\theta^0,g}(w(X)f_\theta^2(X))$ are estimated at the parametric rate $\sqrt{n}$ are not always fullfilled.

In the first estimation procedure, the weight function is used, first in order to make (wf_θ) integrable and second, in order to smooth (wf_θ) and to improve the rate of convergence of the estimators. For a large class of regression function, w not depending on θ suits well, for instance for polynomial, exponential, cosines regression functions as well as for regression functions f_θ of the form $f_\theta(x) = \varphi(\theta)f(x)$. But sometimes, the smoothing properties of w will be improved by taking w depending on θ, e.g. in case where the regression function has to be smoothed at some point related to θ. That can be done, under an additional assumption, which is the knowledge of $\sigma_{\xi,2}^2 = \mathrm{Var}(\xi)$. In this context, the second estimation procedure, is based on the estimation, using the observations $(Y_1, Z_1), \cdots, (Y_n, Z_n)$, of the least squares criterion

$$S_{\theta^0,g} := \mathbb{E}_{\theta^0,g}[(Y - f_\theta(X))^2 w_\theta(X)] - \sigma_{\xi,2}^2 \mathbb{E}_{\theta^0,g}[w_\theta(X)]$$

where w_θ is some positive weight function, depending on θ, to be chosen and where $\sigma_{\xi,2}^2 = \mathrm{Var}(\xi)$ is known. Using this criterion, we propose two other estimators. Analogously to the construction

of $\tilde{\theta}_1$, we propose a first more general estimator $\widehat{\theta}_1$ which is defined by

$$\widehat{\theta}_1 = \arg\min_{\theta\in\Theta} S_{n,1}(\theta) \text{ with } S_{n,1}(\theta) = \frac{1}{n}\sum_{i=1}^{n}\int [(Y_i - f_\theta(x))^2 - \sigma^2_{\xi,2}] w_\theta(x) C_n K_n(C_n(x - Z_i))dx.$$

We show that under classical identifiability, moment and smoothness assumptions, this estimator $\widehat{\theta}_1$ is a consistent estimator of θ^0 with a rate of convergence depending, as $\tilde{\theta}_1$, on the smoothness of $\partial(w_\theta(x)f_\theta(x))/\partial\theta$ and $\partial(w_\theta(x)f^2_\theta(x))/\partial\theta$ and the smoothness of the errors density $p_\varepsilon(x)$, as functions of x, described by the behavior of $(\partial w_\theta/\partial\theta)^*(t)/p^*_\varepsilon(t)$, $(\partial(w_\theta f_\theta)/\partial\theta)^*(t)/p^*_\varepsilon(t)$, $(\partial(f^2_\theta w_\theta)/\partial\theta)^*(t)/p^*_\varepsilon(t)$ as $t \to \infty$. This comes once again from the fact that this estimation procedure requires the estimation of the linear integral functional of the density g, $\mathbb{E}_{\theta^0,g}(w_\theta(X)f^2_\theta(X))$ and $\mathbb{E}_{\theta^0,g}(Yw_\theta(X)f_\theta(X))$ as well as $\mathbb{E}_{\theta^0,g}(Y^2w_\theta(X))$, using the observations $(Y_1, Z_1), \cdots, (Y_n, Z_n)$.

From this construction we derive some sufficient conditions that ensure that $\widehat{\theta}_1$ achieves the parametric rate of convergence.

The second and more simple estimator, is based on conditions ensuring that $\mathbb{E}_{\theta^0,g}(Y^2w_\theta(X))$, $\mathbb{E}_{\theta^0,g}(Yw_\theta(X)f_\theta(X))$ and $\mathbb{E}_{\theta^0,g}(w_\theta(X)f^2_\theta(X))$ can be estimated at the parametric rate $\sqrt{n}$. This induces that the parametric rate of convergence for the estimation of θ^0 can be achieved. Once again, the conditions ensuring that $\mathbb{E}_{\theta^0,g}(Yw_\theta(X)f_\theta(X))$ and $\mathbb{E}_{\theta^0,g}(w_\theta(X)f^2_\theta(X))$ are estimated at the parametric rate $\sqrt{n}$ are not always fullfilled.

The rates of convergence of the proposed estimators are thoroughly studied for various type of smoothness properties of wf_θ, wf^2_θ and their derivatives in θ as functions of x, respectively w_θ, $w_\theta f_\theta$, $w_\theta f^2_\theta$ and their derivatives in θ as functions of x, and for various type of errors density p_ε. It appears that as for the nonparametric estimation of the regression function in errors in variables models, the slower rates of convergence are obtained for smoother density p_ε. In many practical cases we can improve this smoothness a lot by multiplication with a properly chosen weight function.

The conditions ensuring that the $\sqrt{n}$-consistency is achieved are also deeply studied. The main idea is that the actual shape of the regression function matters less than its smoothness (compared to the noise smoothness). Those conditions are illustrated through various examples of regression functions. The point is that these conditions allow us to include setups that were not known before.

Let us now compare our estimation procedure with the one proposed by Hong and Tamer (2003). It is noteworthy that in their special framework of noise densities satisfying $p^*_\varepsilon(t) = c|t|^{-2}(1 + o(1))$ as $|t| \to \infty$ with regression functions f_θ having derivatives in θ up to order 3, twice continuously differentiable functions of x, our estimation procedure also allows, to recover $\sqrt{n}$-consistency.

The drawback of smoothing by multiplication is that we can obtain infinitely differentiable functions but not analytic. In such a case and if the noise has an analytic density, the parametric rate of convergence can not be attained by our methods. Nevertheless the smoothing technique improves significantly the rate when using a clever choice of weight function w.

The main question that remains open is : is it possible to construct a $\sqrt{n}$-consistent estimator of θ^0 for all regression functions and independently of the smoothness of the density of the errors?

The paper is organized as follows. Sections 2 and 3 present the two estimation procedures and the associated estimators as well as their asymptotic properties. Those asymptotic properties

and practical recipes are illustrated through examples in Section 5. The proofs can be found in Section 6 with some technical lemmas presented in Appendix.

2. First estimation procedure

Notations

We denote by x_- the negative part of x, $\| \varphi \|_2^2 = \int \varphi^2(x)dx$, $\| \varphi \|_\infty = \sup_{x\in\mathbb{R}} |\varphi(x)|$. In the same way $p \star q(z) = \int p(z-x)q(x)dx$ denotes the convolution of two functions p and q, $\mathrm{Var}(\xi) = \mathbb{E}(\xi^2)$ is denoted by $\sigma^2_{\xi,2}$ and $\mathbb{E}(\xi^4) = \sigma^4_{\xi,4}$. For some $\theta \in \mathbb{R}^d$, $\| \theta \|_{\ell^2}^2 = \sum_{k=1}^d \theta_k^2$ and $\theta^\top$ is the transpose matrix of θ.

From now, $\mathbb{P}$, $\mathbb{E}$ and Var denote the probability $\mathbb{P}_{\theta^0,g}$, the expected value $\mathbb{E}_{\theta^0,g}$, and respectively, the variance $\mathrm{Var}_{\theta^0,g}$, when the underlying and unknown true parameters are θ^0 and g.

The starting point of the first estimation procedure is to construct an estimator based on the observations (Y_i, Z_i) for $i = 1, \cdots, n$, of the least square contrast

$$\tilde{S}_{\theta^0,g}(\theta) = \mathbb{E}[((Y - f_\theta(X))^2\, w(X)], \tag{2.1}$$

where w is some positive weight function to be suitably chosen.

This estimation procedure requires at least the following assumptions.

Smoothness assumption

($\mathbf{A_1}$) For any θ in Θ, the function $\theta \mapsto f_\theta$ admits continuous derivatives with respect to θ up to the order 3 .

Identifiability and moment assumptions

($\mathbf{I1_1}$) The quantity $\tilde{S}_{\theta^0,g}(\theta) = \sigma^2_{\xi,2}\mathbb{E}(w(X)) + \mathbb{E}[(f_{\theta^0}(X) - f_\theta(X))^2 w(X)]$ admits one unique minimum at $\theta = \theta^0$.

($\mathbf{I1_2}$) For all $\theta \in \Theta$ the matrix $\tilde{S}^{(2)}_{\theta^0,g}(\theta) = \partial^2 S_{\theta^0,g}(\theta)/\partial\theta^2$ exists and

$$\tilde{S}^{(2)}_{\theta^0,g}(\theta^0) = \mathbb{E}\left[w(X) \left(\frac{\partial f_\theta(X)}{\partial\theta} \mid_{\theta=\theta^0} \right) \left(\frac{\partial f_\theta(X)}{\partial\theta} \mid_{\theta=\theta^0} \right)^\top \right] \text{ is positive definite.}$$

($\mathbf{I1_3}$) The quantities $\mathbb{E}[w^2(X)(Y - f_\theta(X))^4]$ and their derivatives up to order 2 with respect to θ are finite.

We denote by $\mathcal{G}$ the set of densities g such that ($\mathbf{I1_3}$) holds.

2.1. Construction and study of the first estimator $\tilde{\theta}_1$.

We start by presenting an estimator, which is general, constructive, and which allows to give upper bounds in a general setting and to deduce some sufficient conditions to achieve the parametric rate of convergence.

2.1.1. *Construction.* Let us denote by $f_{Y,X}$ and by $f_{Y,Z}$ the joint densities of (Y, X) and respectively of (Y, Z) satisfying in this model,

$$f_{Y,X}(y, x) = g(x)p_\xi(y - f_{\theta^0}(x)) \text{ and } f_{Y,Z}(y, z) = f_{Y,X}(y, \cdot) \star p_\varepsilon(z). \tag{2.2}$$

Since $\tilde{S}_{\theta^0,g}(\theta)$ is given by

$$\tilde{S}_{\theta^0,g}(\theta) = \mathbb{E}[(Y - f_\theta(X))^2 w(X)] = \int (y - f_\theta(x))^2 w(x) f_{Y,X}(y,x) dy\, dx,$$

according to (2.2), we naturally propose to estimate $\tilde{S}_{\theta^0,g}(\theta)$ by

$$\tilde{S}_{n,1}(\theta) = \frac{1}{n}\sum_{i=1}^{n} \int (Y_i - f_\theta(x))^2\, w(x)\, K_{n,C_n}(C_n(x - Z_i)) dx, \tag{2.3}$$

where $K_{n,C_n}(\cdot) = C_n K_n(C_n \cdot)$ is a deconvolution kernel defined via its Fourier transform, such that $\int K_n(x)dx = 1$, and

$$K^*_{n,C_n}(t) = \frac{K^*_{C_n}(t)}{p^*_\varepsilon(t)} = \frac{K^*(t/C_n)}{p^*_\varepsilon(t)}, \tag{2.4}$$

with K^* compactly supported satisfying $|1 - K^*(t)| \leq \mathbb{1}_{|t|\geq 1}$.

Using this criterion we propose to estimate θ^0 by

$$\tilde{\theta}_1 = \arg\min_{\theta\in\Theta} \tilde{S}_{n,1}(\theta). \tag{2.5}$$

2.1.2. *Asymptotic properties of the first estimator* $\tilde{\theta}_1$.

$$\sup_{g\in\mathcal{G}} \| f^2_{\theta^0} g \|_2^2 \leq C(f^2_{\theta^0}),\ \sup_{g\in\mathcal{G}} \| f_{\theta^0} g \|_2^2 \leq C(f_{\theta^0}). \tag{$\mathbf{A_2}$}$$

$$\sup_{\theta\in\Theta} |w f_\theta|,\ |w| \text{ and } \sup_{\theta\in\Theta} |w f^2_\theta| \text{ belong to } \mathbb{L}_1(\mathbb{R}) \tag{$\mathbf{A_3}$}$$

As for density deconvolution, or for nonparametric regression function estimation in errors in variables models, the rate of convergence for estimating θ^0 is given by both the smoothness of the errors density p_ε and the smoothness of $f_\theta w$, and more precisely by the smoothness of $\partial(f_\theta w)/\partial\theta$ and $\partial(f^2_\theta w)/\partial\theta$, as functions of x.

Those smoothness properties are described in both cases, by the asymptotic behaviour of the Fourier transforms. We consider p_ε that satisfies the following assumption.

($\mathbf{N_1}$) The density p_ε belongs to $\mathbb{L}_2(\mathbb{R})$ and for all $x \in \mathbb{R}$, $p^*_\varepsilon(x) \neq 0$.

($\mathbf{N_2}$) There exist some positive constants $\underline{C}(p_\varepsilon), \overline{C}(p_\varepsilon), \beta,\ \rho, \alpha$ and u_0 such that $\underline{C}(p_\varepsilon) \leq |p^*_\varepsilon(u)|\, |u|^\alpha \exp(\beta |u|^\rho) \leq \overline{C}(p_\varepsilon)$ for all $|u| \geq u_0$.

The assumption ($\mathbf{N_1}$), quite usual in density deconvolution, ensures the existence of the deconvolution kernel in (2.4). If $\rho > 0$, then the noise is called exponential noise or super smooth noise. And if $\rho = 0$, then by convention $\beta = 0$, $\alpha > 1$ and the noise is called polynomial noise or ordinary smooth.

The smoothness properties of functions involving the regression function are also given by the asymptotic behaviour of the Fourier transforms described as follows.

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

If $r = 0$, by convention $b = 0$, and the function f is called ordinary smooth. If $r > 0$, the function f is called super smooth.

Theorem 2.1. *Let $\tilde{\theta}_1 = \tilde{\theta}_1(C_n)$ be defined by (2.5) under the assumptions ($\mathbf{I1_1}$)-($\mathbf{I1_3}$), ($\mathbf{N_1}$)-($\mathbf{N_2}$), ($\mathbf{A_1}$), ($\mathbf{A_2}$) and ($\mathbf{A_3}$). Assume moreover that for all $\theta \in \Theta$, $f_\theta w$ and $f_\theta^2 w$ and their derivatives up to order 3, with respect to θ satisfy ($\mathbf{R_1}$). Let C_n be a sequence such that*

$$C_n^{(2\alpha-2a+1-\rho)} \exp\{-2bC_n^r + 2\beta C_n^\rho\}/n = o(1) \text{ as } n \to +\infty. \tag{2.6}$$

1) *Then, for all of the sequences satisfying (2.6), $\mathbb{E}(\|\tilde{\theta}_1(C_n) - \theta^0\|_{\ell^2}^2) = o(1)$, as $n \to \infty$ and $\tilde{\theta}_1(C_n)$ is a consistent estimator of θ^0.*

2) *Moreover $\mathbb{E}(\| \tilde{\theta}_1 - \theta^0 \|_{\ell^2}^2) = O(\tilde{\varphi}_n^2)$ with $\tilde{\varphi}_n = \|(\tilde{\varphi}_{n,j})\|_{\ell^2}$ and $\tilde{\varphi}_{n,j}^2 = \tilde{B}_{n,j}^2(\theta^0) + \tilde{V}_{n,j}(\theta^0)/n$, $j = 1 \cdots, d$,*

$$\begin{aligned}\tilde{B}_{n,j}^2(\theta) \;=\; \min\Bigg\{ & \left\| \left(\frac{\partial(wf_\theta)}{\partial\theta_j}\right)^* (K_{C_n}^* - 1) \right\|_2^2 + \left\| \left(\frac{\partial(wf_\theta^2)}{\partial\theta_j}\right)^* (K_{C_n}^* - 1) \right\|_2^2, \\ & \left\| \left(\frac{\partial(wf_\theta)}{\partial\theta_j}\right)^* (K_{C_n}^* - 1) \right\|_1^2 + \left\| \left(\frac{\partial(wf_\theta^2)}{\partial\theta_j}\right)^* (K_{C_n}^* - 1) \right\|_1^2 \Bigg\},\end{aligned}$$

and

$$\begin{aligned}\tilde{V}_{n,j}(\theta) \;=\; \min\Bigg\{ & \left\| \left(\frac{\partial(wf_\theta)}{\partial\theta_j}\right)^* \frac{K_{C_n}^*}{p_\varepsilon^*} \right\|_2^2 + \left\| \left(\frac{\partial(wf_\theta^2)}{\partial\theta_j}\right)^* \frac{K_{C_n}^*}{p_\varepsilon^*} \right\|_2^2, \\ & \left\| \left(\frac{\partial(wf_\theta)}{\partial\theta_j}\right)^* \frac{K_{C_n}^*}{p_\varepsilon^*} \right\|_1^2 + \left\| \left(\frac{\partial(wf_\theta^2)}{\partial\theta_j}\right)^* \frac{K_{C_n}^*}{p_\varepsilon^*} \right\|_1^2 \Bigg\}.\end{aligned}$$

Comment The terms $\tilde{B}_{n,j}$ and $\tilde{V}_{n,j}/n$ are respectively the bias and variance terms with, as in density deconvolution, bigger variance for smoother error density p_ε and smaller bias for smoother (wf_θ). Contrary to density deconvolution, the smoothness of (wf_θ) also appears in the variance term and not only in the bias. The resulting rate when $(f_\theta w)$ and $(f_\theta^2 w)$, as well as their derivatives with respect to θ up to order 3 satisfy ($\mathbf{R_1}$), are given in the following corollary.

Corollary 2.1. *Under the assumptions of Theorem 2.1, assume that p_ε satisfies ($\mathbf{N_2}$) and that for all $\theta \in \Theta$, $(f_\theta w)$, $(f_\theta^2 w)$ and their derivatives with respect to θ_j, $j = 1, \ldots, d$ up to order 3, satisfy ($\mathbf{R_1}$). Then φ_n^2 is given by Table 1, according to values of parameters a, b, r, α, β and ρ.*

Remark 2.1. Let us comment those rates. First note that the slower rates are obtained for the smoother errors density p_ε, for instance for Gaussian ε's. Nevertheless, those rates are improved with the choice of w such that (wf_θ), (wf_θ^2) and their derivatives with respect to θ are smooth enough.

Second, the parametric rate of convergence is achieved as soon as the derivatives with respect to θ, of (wf_θ) and (wf_θ^2) as functions of x are smoother than the errors density p_ε.

Remark 2.2. It is noteworthy that the rate of convergence of the estimators could be improved by assuming some smoothness properties on the design density g. Since g is unkown, we choose not to assume such smoothness conditions.

		p_ε	
		$\rho = 0$ in $(\mathbf{N_2})$ ordinary smooth	$\rho > 0$ in $(\mathbf{N_2})$ super smooth
wf_{θ^0}	$(\mathbf{R_1})$ $b = r = 0$ Sobolev	$a < \alpha + 1/2$ $\quad n^{-\frac{2a-1}{2\alpha}}$ $a \geq \alpha + 1/2$ $\quad n^{-1}$	$(\log n)^{-\frac{2a-1}{\rho}}$
	$(\mathbf{R_1})$ $r > 0$ $\mathcal{C}^\infty$	n^{-1}	$r < \rho$ $\quad (\log n)^{A(a,r,\rho)} \exp\left\{-2b\left(\frac{\log n}{2\beta}\right)^{r/\rho}\right\}$ $r = \rho$: $b < \beta$ $\quad \frac{(\log n)^{A(a,r,\rho)+2\alpha b/(\beta r)}}{n^{b/\beta}}$; $b = \beta,\ a < \alpha + 1/2$ $\quad \frac{(\log n)^{(2\alpha-2a+1)/r}}{n}$; $b = \beta,\ a \geq \alpha + 1/2$ $\quad n^{-1}$; $b > \beta$ $\quad n^{-1}$ $r > \rho$ $\quad n^{-1}$

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

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

2.2. **Consequence : a sufficient condition to obtain the parametric rate of convergence with $\tilde{\theta}_1$.** We say that the conditions $(\mathbf{C_1})$-$(\mathbf{C_3})$ hold if there exists some weight function w such that for all $\theta \in \Theta$,

$(\mathbf{C_1})$ the functions (wf_θ) and (wf_θ^2) belong to $\mathbb{L}_1(\mathbb{R})$, and the functions $\sup_\theta w^*/p_\varepsilon^*$, $\sup_\theta (f_\theta w)^*/p_\varepsilon^*$, $\sup_\theta (f_\theta^2 w)^*/p_\varepsilon^*$ belong to $\mathbb{L}_1(\mathbb{R}) \cap \mathbb{L}_2(\mathbb{R})$;

$(\mathbf{C_2})$ the functions $\sup_{\theta\in\Theta} \left(\frac{\partial(f_\theta w)}{\partial\theta}\right)^* / p_\varepsilon^*$ and $\sup_{\theta\in\Theta} \left(\frac{\partial(f_\theta^2 w)}{\partial\theta})\right)^* / p_\varepsilon^*$ belong to $\mathbb{L}_1(\mathbb{R}) \cap \mathbb{L}_2(\mathbb{R})$;

$(\mathbf{C_3})$ the functions $\left(\frac{\partial^2(f_\theta w)}{\partial\theta^2}\right)^* / p_\varepsilon^*$ and $\left(\frac{\partial^2(f_\theta^2 w)}{\partial\theta^2}\right)^* / p_\varepsilon^*$ belong to $\mathbb{L}_1(\mathbb{R}) \cap \mathbb{L}_2(\mathbb{R})$.

Theorem 2.2. *Consider Model (1.1) under the assumptions* $(\mathbf{A_1})$,$(\mathbf{I1_1})$-$(\mathbf{I1_3})$ *and the conditions* $(\mathbf{C_1})$-$(\mathbf{C_3})$. *Assume that for all* $\theta \in \Theta$, $f_\theta w$, $f_\theta^2 w$ *and their derivatives up to order 3 satisfy* $(\mathbf{R_1})$. *Then* $\tilde{\theta}_1$ *defined by (2.5) is a* $\sqrt{n}$*-consistent estimator of* θ^0 *which satisfies moreover that* $\sqrt{n}(\tilde{\theta}_1 - \theta^0) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, \tilde{\Sigma}_1)$, *with* $\tilde{\Sigma}_1$ *that equals*

$$\left(\mathbb{E}\left[w(X)\left(\frac{\partial f_\theta(X)}{\partial\theta}\right)\left(\frac{\partial f_\theta(X)}{\partial\theta}\right)^\top\right]|_{\theta=\theta^0}\right)^{-1}\tilde{\Sigma}_{0,1}\left(\mathbb{E}\left[w(X)\left(\frac{\partial f_\theta(X)}{\partial\theta}\right)\left(\frac{\partial f_\theta(X)}{\partial\theta}\right)^\top\right]|_{\theta=\theta^0}\right)^{-1}$$

where $\tilde{\Sigma}_{0,1}$ *is given by*

$$\mathbb{E}\left\{\left[\int\left(\frac{\partial(f_\theta^2 w - 2Yf_\theta w)}{\partial\theta}|_{\theta=\theta^0}\right)^*(u)\frac{e^{-iuZ}}{p_\varepsilon^*(u)}du\right]\left[\int\left(\frac{\partial(f_\theta^2 w - 2Yf_\theta w)}{\partial\theta}|_{\theta=\theta^0}\right)^*(u)\frac{e^{-iuZ}}{p_\varepsilon^*(u)}du\right]^\top\right\}.$$

Comments Theorem 2.2 is an immediate consequence of Theorem 2.1 since conditions $(\mathbf{C_1})$-$(\mathbf{C_3})$ ensure precisely that the variance term $V_{n,j} = O(1)$ for $j = 1, \cdots, d$.

2.3. Construction and study of the risk of the second estimator $\tilde{\theta}_2$.

2.3.1. *Construction.* We now propose another estimator of θ^0, based on some sufficient conditions allowing to construct directly a $\sqrt{n}$-consistent estimator of $\tilde{S}_{\theta^0,g}$ defined in (2.1), based on (Y_i, Z_i), $i = 1, \cdots, n$.

We say that the conditions $(\mathbf{C_4})$-$(\mathbf{C_7})$ hold if there exists some weight function w and there exist some functions $\tilde{\Phi}_{\theta,\varepsilon,j}$, $j = 1, 2, 3$ not depending on g, such that for all $\theta \in \Theta$ and for all g

$$(\mathbf{C_4}) \qquad \int y^2 w(x) f_{Y,X}(y,x)dy\,dx = \int y^2\tilde{\Phi}_{\theta,\varepsilon,3}(z)f_{Y,Z}(y,z)dy\,dz,$$
$$\int y f_\theta(x)w(x)f_{Y,X}(y,x)dy\,dx = \int y\tilde{\Phi}_{\theta,\varepsilon,2}(z)f_{Y,Z}dy\,dz$$
$$\text{and} \int w(x)f_\theta^2(x)g(x)dx = \int \tilde{\Phi}_{\theta,\varepsilon,1}(z)h(z)dz;$$

$$(\mathbf{C_5}) \qquad \text{For } j = 1,2,3,\ \mathbb{E}\left[\sup_{\theta\in\Theta}\left|\frac{\partial\tilde{\Phi}_{\theta,\varepsilon,j(Z)}}{\partial\theta}\right|\right] < \infty;$$

$$(\mathbf{C_6}) \qquad \text{For } j = 1,2,3 \text{ and for all } \theta\in\Theta,\ \mathbb{E}\left[\left|\frac{\partial\tilde{\Phi}_{\theta,\varepsilon,j}}{\partial\theta}(Z)\right|^2\right] < \infty;$$

$$(\mathbf{C_7}) \qquad \text{For } j = 1,2,3 \text{ and for all } \theta\in\Theta,\ \mathbb{E}\left[\left|\frac{\partial^2\tilde{\Phi}_{\theta,\varepsilon,j}}{\partial\theta^2}(Z)\right|\right] < \infty.$$

Note that $\tilde{\Phi}_{\theta,\varepsilon,3}$ exists as soon as the chosen weight function w is smoother than p_ε in the way that w^*/p_ε^* belongs to $\mathbb{L}_1(\mathbb{R})$. Furthermore, $\tilde{\Phi}_{\theta,\varepsilon,3} \equiv \tilde{\Phi}_{\varepsilon,3}$ does not depend on θ. We refer to Section 4 for details on how to construct such functions $\tilde{\Phi}_{\theta,\varepsilon,j}$.

Under $(\mathbf{C_4})$-$(\mathbf{C_7})$, we propose to estimate $\tilde{S}_{\theta^0,g}(\theta)$ by

$$\tilde{S}_{n,2}(\theta) = \frac{1}{n}\sum_{i=1}^{n}[(Y_i^2\Phi_{\theta,\varepsilon,3}(Z_i) - 2Y_i\Phi_{\theta,\varepsilon,2}(Z_i) + \Phi_{\theta,\varepsilon,1}(Z_i)], \tag{2.7}$$

and hence θ^0 is estimated by

$$\tilde{\theta}_2 \quad = \quad \arg\min_{\theta\in\Theta} \tilde{S}_{n,2}(\theta). \tag{2.8}$$

Remark 2.3. The main difficulty for finding such function $\tilde{\Phi}_{\theta,\varepsilon,j}$ lies in the constraint that we expect that they do not depend on the unknown density g. If we relax this constraint, obviously there are a lot of solutions.

2.3.2. *Asymptotic properties of the second estimator $\tilde{\theta}_2$.*

Theorem 2.3. *Consider Model (1.1) under the assumptions $(\mathbf{A_1})$,$(\mathbf{I1_1})$-$(\mathbf{I1_3})$, and the conditions $(\mathbf{C_4})$-$(\mathbf{C_7})$. Then $\tilde{\theta}_2$, defined by (2.8) is a $\sqrt{n}$-consistent estimator of θ^0 which satisfies moreover that $\sqrt{n}(\tilde{\theta}_2 - \theta^0) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, \tilde{\Sigma}_2)$, with $\tilde{\Sigma}_2$ that equals*

$$\left(\mathbb{E}\left[w(X)\left(\frac{\partial f_\theta(X)}{\partial\theta}\right)\left(\frac{\partial f_\theta(X)}{\partial\theta}\right)^\top\right]|_{\theta=\theta^0}\right)^{-1}\tilde{\Sigma}_{0,2}\left(\mathbb{E}\left[w(X)\left(\frac{\partial f_\theta(X)}{\partial\theta}\right)\left(\frac{\partial f_\theta(X)}{\partial\theta}\right)^\top\right]|_{\theta=\theta^0}\right)^{-1}$$

where $\tilde{\Sigma}_{0,2}$ equals

$$\mathbb{E}\left[\left(\frac{\partial(\tilde{\Phi}_{\theta,\varepsilon,1}(Z) - 2Y\tilde{\Phi}_{\theta,\varepsilon,2}(Z)}{\partial\theta}\,|_{\theta=\theta^0}\right)\left(\frac{\partial(\tilde{\Phi}_{\theta,\varepsilon,1}(Z) - 2Y\tilde{\Phi}_{\theta,\varepsilon,2}(Z))}{\partial\theta}\,|_{\theta=\theta^0}\right)^\top\right].$$

Note that, by denoting $\tilde{\Phi}_{\theta,\varepsilon,1} = (wf_\theta^2)^*/p_\varepsilon^*$, $\tilde{\Phi}_{\theta,\varepsilon,2} = (wf_\theta)^*/p_\varepsilon^*$ and $\tilde{\Phi}_{\theta,\varepsilon,3} = w^*/p_\varepsilon^*$, we get that $\tilde{\Sigma}_{0,1} = \tilde{\Sigma}_{0,2}$ with $\tilde{\Sigma}_{0,1}$ defined in Theorem 2.2 (See also Comment 1 in Section 4).

3. Second estimation procedure

In the first estimation procedure, the weight function is used in order to make wf_θ, wf_θ^2 integrable and also in order to get smooth wf_θ, wf_θ^2 and derivatives in θ. In a large class of regression functions, the weight function can smooth these functions without depending on θ. Sometimes, the smoothing properties and hence the rate of convergence are improved by making w to depend on θ. It appears in particular when the points where we need to smooth are related to θ.

The second estimation procedure uses an estimator based on the observations (Y_i, Z_i), $i = 1, \cdots, n$, of the least square contrast

$$S_{\theta^0,g}(\theta) = \mathbb{E}[((Y - f_\theta(X))^2 - \sigma_{\xi,2}^2)\, w_\theta(X)], \tag{3.1}$$

where w_θ is some positive weight function to be suitably chosen. This criterion requires thus the knowledge of $\mathrm{Var}(\xi)$.

Subsequently we consider the following assumptions.

Identifiability and moment assumptions

($\mathbf{I2_1}$) The variance $\sigma^2_{\xi,2} = \mathrm{Var}(\xi_i)$ is known.

($\mathbf{I2_2}$) The quantity $S_{\theta^0,g}(\theta) = \mathbb{E}[w_\theta(X)(Y - f_\theta(X))^2] - \sigma^2_{\xi,2}\mathbb{E}(w_\theta(X))$ admits one unique minimum at $\theta = \theta^0$.

($\mathbf{I2_3}$) For all $\theta \in \Theta$ the matrix $S^{(2)}_{\theta^0,g}(\theta) = \partial^2 S_{\theta^0,g}(\theta)/\partial\theta^2$ exists and

$$S^{(2)}_{\theta^0,g}(\theta^0) = \mathbb{E}\left[w_{\theta^0}(X)\left(\frac{\partial(f_\theta(X)}{\partial\theta}\mid_{\theta=\theta^0}\right)\left(\frac{\partial(f_\theta(X)}{\partial\theta}\mid_{\theta=\theta^0}\right)^\top\right] \text{ is positive definite.}$$

($\mathbf{I2_4}$) The quantities $\mathbb{E}[w^2_\theta(X)(Y - f_\theta(X))^4]$ and their derivatives up to order 2 with respect to θ are finite.

Comments on identifiability assumptions : Easy calculations give that

$$\begin{aligned}
\frac{\partial S_{\theta^0,g}(\theta)}{\partial\theta} &= \mathbb{E}_{\theta^0,g}\left(f_{\theta^0}(X) - f_\theta(X))^2\left(\frac{\partial w_\theta(X)}{\partial\theta}\right)\right] - 2\mathbb{E}_{\theta^0,g}\left[w_\theta(X)(f_{\theta^0}(X) - f_\theta(X))\left(\frac{\partial f_\theta(X)}{\partial\theta}\right)\right],\\
S^{(2)}_{\theta^0,g}(\theta) &= \mathbb{E}_{\theta^0,g}\left[(f_{\theta^0}(X) - f_\theta(X))^2\left(\frac{\partial^2 w_\theta(X)}{\partial\theta^2}\right)\right]\\
&\quad -4\mathbb{E}_{\theta^0,g}\left[(f_{\theta^0}(X) - f_\theta(X))\left(\frac{\partial w_\theta(X)}{\partial\theta}\right)\left(\frac{\partial f_\theta(X)}{\partial\theta}\right)^\top\right]\\
&\quad +2\mathbb{E}_{\theta^0,g}\left[w_\theta(X)\left(\frac{\partial f_\theta(X)}{\partial\theta}\right)\left(\frac{\partial f_\theta(X)}{\partial\theta}\right)^\top\right]\\
&\quad -2\mathbb{E}_{\theta^0,g}\left[w_\theta(X)(f_{\theta^0}(X) - f_\theta(X))\left(\frac{\partial^2 f_\theta(X)}{\partial\theta^2}\right)\right].
\end{aligned}$$

Therefore $S_{\theta^0,g}(\theta)$ is minimum if and only if $\theta = \theta^0$ as soon as ($\mathbf{I2_2}$) and ($\mathbf{I2_3}$) hold.

3.1. **Construction and study of the first estimator $\widehat{\theta}_1$.** As in the first estimation procedure, we start by presenting an estimator, which is general, constructive, and which allows to give upper bounds in a general setting and to deduce some sufficient conditions to achieve the parametric rate of convergence.

3.1.1. *Construction.* According to (2.2) $S_{\theta^0,g}(\theta)$ is naturally estimated by

$$S_{n,1}(\theta) = \frac{1}{n}\sum_{i=1}^n \int \left[(Y_i - f_\theta(x))^2 - \sigma^2_{\xi,2}\right] w_\theta(x)\, K_{n,C_n}(C_n(x - Z_i))dx \tag{3.2}$$

where $K_{n,C_n}(\cdot) = C_n K_n(C_n\cdot)$ is a deconvolution kernel satisfying (2.4). Using this criterion, under ($\mathbf{N_1}$), we propose to estimate θ^0 by

$$\widehat{\theta}_1 = \arg\min_{\theta\in\Theta} S_{n,1}(\theta). \tag{3.3}$$

3.1.2. *Asymptotic properties of the first estimator* $\widehat{\theta}_1$.

Theorem 3.1. *Let* $\widehat{\theta}_1 = \widehat{\theta}_1(C_n)$ *be defined by (3.3) under the assumptions (*$\mathbf{I2_1}$*),(*$\mathbf{I2_2}$*), (*$\mathbf{I2_3}$*), (*$\mathbf{I2_4}$*),(*$\mathbf{N_1}$*)-(*$\mathbf{N_2}$*), (*$\mathbf{A_1}$*), (*$\mathbf{A_2}$*) and (*$\mathbf{A_3}$*) with* w *replaced by* w_θ*. Assume moreover that for all* $\theta \in \Theta$*,* $f_\theta w_\theta$ *and* $f_\theta^2 w_\theta$ *and their derivatives up to order 3 with respect to* θ *satisfy (*$\mathbf{R_1}$*). Let* C_n *be a sequence such that (2.6) holds.*

1) *Then, for all of the sequences satisfying (2.6),* $\mathbb{E}(\|\widehat{\theta}_1(C_n) - \theta^0\|_{\ell^2}^2) = o(1)$*, as* $n \to \infty$ *and* $\widehat{\theta}_1(C_n)$ *is a consistent estimator of* θ^0*.*

2) *Moreover* $\mathbb{E}(\| \widehat{\theta}_1 - \theta^0 \|_{\ell^2}^2) = O(\varphi_n^2)$ *with* φ_n *given* $\varphi_n = \|(\varphi_{n,j})\|$ *with* $\varphi_{n,j}^2 = B_{n,j}^2(\theta^0) + V_{n,j}(\theta^0)/n$*,* $j = 1, \cdots, d$*,*

$$B_{n,j}^2(\theta) = \min\left\{ \left\| \left(\frac{\partial(w_\theta)}{\partial\theta_j}\right)^*(K_{C_n}^* - 1)\right\|_2^2 + \left\| \left(\frac{\partial(f_\theta w_\theta)}{\partial\theta_j}\right)^*(K_{C_n}^* - 1)\right\|_2^2 + \left\| \left(\frac{\partial(f_\theta^2 w_\theta)}{\partial\theta_j}\right)^*(K_{C_n}^* - 1)\right\|_2^2, \right.$$
$$\left. \left\| \left(\frac{\partial(w_\theta)}{\partial\theta_j}\right)^*(K_{C_n}^* - 1)\right\|_1^2 + \left\| \left(\frac{\partial(f_\theta w_\theta)}{\partial\theta_j}\right)^*(K_{C_n}^* - 1)\right\|_1^2 + \left\| \left(\frac{\partial(f_\theta^2 w_\theta)}{\partial\theta_j}\right)^*(K_{C_n}^* - 1)\right\|_1^2 \right\},$$

and

$$V_{n,j}(\theta) = \min\left\{ \left\| \left(\frac{\partial(w_\theta)}{\partial\theta_j}\right)^* \frac{K_{C_n}^*}{p_\varepsilon^*}\right\|_2^2 + \left\| \left(\frac{\partial(f_\theta w_\theta)}{\partial\theta_j}\right)^* \frac{K_{C_n}^*}{p_\varepsilon^*}\right\|_2^2 + \left\| \left(\frac{\partial(f_\theta^2 w_\theta)}{\partial\theta_j}\right)^* \frac{K_{C_n}^*}{p_\varepsilon^*}\right\|_2^2, \right.$$
$$\left. \left\| \left(\frac{\partial(w_\theta)}{\partial\theta_j}\right)^* \frac{K_{C_n}^*}{p_\varepsilon^*}\right\|_1^2 + \left\| \left(\frac{\partial(f_\theta w_\theta)}{\partial\theta_j}\right)^* \frac{K_{C_n}^*}{p_\varepsilon^*}\right\|_1^2 + \left\| \left(\frac{\partial(f_\theta^2 w_\theta)}{\partial\theta_j}\right)^* \frac{K_{C_n}^*}{p_\varepsilon^*}\right\|_1^2 \right\}$$

Comments : The terms $B_{n,j}$ and $V_{n,j}/n$ are respectively the bias and variance terms with, as in density deconvolution, bigger variance for smoother error density p_ε and smaller bias for smoother $(w_\theta f_\theta)$ as function of x. The resulting rate when $(f_\theta w_\theta)$ and $(f_\theta^2 w_\theta)$, as well as their derivatives with respect to θ up to order 3 satisfy $(\mathbf{R_1})$, are given in the following corollary.

Corollary 3.1. *Under the assumptions of Theorem 3.1, assume that* p_ε *satisfies (*$\mathbf{N_2}$*) and that for all* $\theta \in \Theta$*,* $(f_\theta w_\theta)$ *and* $(f_\theta^2 w_\theta)$ *and their derivatives with respect to* θ *up to order 3, satisfy (*$\mathbf{R_1}$*). Then* φ_n^2 *is given by the Corollary 2.1 with* $w f_\theta$ *replaced by* $w_\theta f_\theta$*.*

Remark 3.1. The remarks 2.1 and 2.2 are still valid for the Corollary 3.1.

3.1.3. *Consequence : a sufficient condition to obtain the parametric rate of convergence with* $\widehat{\theta}_1$*.* We say that the conditions $(\mathbf{C_8})$-$(\mathbf{C_{10}})$ hold if there exists some weight function w_θ such that

for all $\theta \in \Theta$,

($\mathbf{C_8}$) the functions $(w_\theta f_\theta)$, (w_θ) and $(w_\theta f_\theta^2)$ belong to $\mathbb{L}_1(\mathbb{R})$ and the functions $(w_\theta)^*/p_\varepsilon^*$, $\sup_\theta (f_\theta w_\theta)^*/p_\varepsilon^*$, $\sup_\theta (f_\theta^2 w_\theta)^*/p_\varepsilon^*$ belong to $\mathbb{L}_1(\mathbb{R}) \cap \mathbb{L}_2(\mathbb{R})$;

($\mathbf{C_9}$) the functions $\sup_{\theta\in\Theta} \left(\frac{\partial w_\theta}{\partial \theta}\right)^* /p_\varepsilon^*$, $\sup_{\theta\in\Theta} \left(\frac{\partial (f_\theta w_\theta)}{\partial \theta}\right)^* /p_\varepsilon^*$ and $\sup_{\theta\in\Theta} \left(\frac{\partial (f_\theta^2 w_\theta)}{\partial \theta}\right)^* /p_\varepsilon^*$ belong to $\mathbb{L}_1(\mathbb{R}) \cap \mathbb{L}_2(\mathbb{R})$;

($\mathbf{C_{10}}$) the functions $\left(\frac{\partial^2 w_\theta}{\partial \theta^2}\right)^* /p_\varepsilon^*$, $\left(\frac{\partial^2 (f_\theta w_\theta)}{\partial \theta^2}\right)^* /p_\varepsilon^*$ and $\left(\frac{\partial^2 (f_\theta^2 w_\theta)}{\partial \theta^2}\right)^* /p_\varepsilon^*$ belong to $\mathbb{L}_1(\mathbb{R}) \cap \mathbb{L}_2(\mathbb{R})$.

Theorem 3.2. *Consider Model (1.1) under the assumptions* ($\mathbf{I2_1}$), ($\mathbf{I2_2}$), ($\mathbf{I2_3}$), ($\mathbf{I2_4}$), ($\mathbf{N_1}$)-($\mathbf{N_2}$), ($\mathbf{A_1}$), ($\mathbf{A_2}$), ($\mathbf{A_3}$) *for w replaced by w_θ, and under* ($\mathbf{C_8}$)-($\mathbf{C_{10}}$). *Assume that for all $\theta \in \Theta$, $(f_\theta w_\theta)$ and $(f_\theta^2 w_\theta)$ and their derivatives up to order 3 satisfy* ($\mathbf{R_1}$). *Then $\widehat{\theta}_1$ defined by (3.3) is a $\sqrt{n}$-consistent estimator of θ^0 which satisfies moreover that $\sqrt{n}(\widehat{\theta}_1 - \theta^0) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, \Sigma_1)$, with Σ_1 that equals*

$$\left(\mathbb{E}\left[w_{\theta^0}(X)\left(\frac{\partial f_\theta(X)}{\partial \theta}\right)\left(\frac{\partial f_\theta(X)}{\partial \theta}\right)^\top\right]|_{\theta=\theta^0}\right)^{-1} \Sigma_{0,1} \left(\mathbb{E}\left[w_{\theta^0}(X)\left(\frac{\partial f_\theta(X)}{\partial \theta}\right)\left(\frac{\partial f_\theta(X)}{\partial \theta}\right)^\top\right]|_{\theta=\theta^0}\right)^{-1}$$

where

$$\begin{aligned}\Sigma_{0,1} &= \mathbb{E}\left\{\left[\int \left(\frac{\partial[f_\theta^2 w_\theta - 2Y f_\theta w_\theta + (Y^2 - \sigma_{\xi,2}^2) w_\theta]}{\partial \theta}|_{\theta=\theta^0}\right)^* (u) \frac{e^{-iuZ}}{p_\varepsilon^*(u)} du\right]\right. \\ &\quad \left. \times \left[\int \left(\frac{\partial[f_\theta^2 w_\theta - 2Y f_\theta w_\theta + (Y^2 - \sigma_{\xi,2}^2) w_\theta]}{\partial \theta}|_{\theta=\theta^0}\right)^* (u) \frac{e^{-iuZ}}{p_\varepsilon^*(u)} du\right]^\top\right\}.\end{aligned}$$

Comments The theorem 3.2 is an immediate consequence of Theorem 3.1 since conditions ($\mathbf{C_8}$)-($\mathbf{C_{10}}$) ensure precisely that the variance term $V_{n,j} = O(1)$ for $j = 1, \cdots, d$.

3.2. **Construction and study of the risk of the second estimator $\widehat{\theta}_2$.** We now propose a second estimator whose construction is based on some sufficient conditions allowing to construct a direct and $\sqrt{n}$-consistent estimator of $S_{\theta^0,g}(\theta)$ defined in (3.1), using the observations (Y_i, Z_i), $i = 1, \cdots, n$.

We say that the conditions ($\mathbf{C_{11}}$)-($\mathbf{C_{14}}$) hold if there exists some weight function w_θ and there exist some functions $\Phi_{\theta,\varepsilon,j}$, $j = 1, \ldots, 3$, not depending on g such that for all $\theta \in \Theta$ and

for all g

$$(\mathbf{C_{11}}) \qquad \int w_\theta(x)g(x)dx = \int \Phi_{\theta,\varepsilon,3}(z)h(z)dz$$
$$\int y f_\theta(x) w_\theta(x) f_{Y,X}(y,x)dy\,dx = \int y\Phi_{\theta,\varepsilon,2}(z) f_{Y,Z}(y,z)dy\,dz$$
$$\text{and } \int w_\theta(x) f_\theta^2(x) g(x)dx = \int \Phi_{\theta,\varepsilon,1}(z)h(z)dz;$$

$$(\mathbf{C_{12}}) \qquad \text{For } j=1,2,3,\ \mathbb{E}\left[\sup_{\theta\in\Theta}\left|\frac{\partial \Phi_{\theta,\varepsilon,j}}{\partial\theta}(Z)\right|\right] < \infty;$$

$$(\mathbf{C_{13}}) \qquad \text{For } j=1,2,3, \text{ and for all } \theta\in\Theta,\ \mathbb{E}\left[\left|\frac{\partial \Phi_{\theta,\varepsilon,j}}{\partial\theta}(Z)\right|^2\right] < \infty;$$

$$(\mathbf{C_{14}}) \qquad \text{For } j=1,2,3, \text{ and for all } \theta\in\Theta,\ \mathbb{E}\left[\left|\frac{\partial^2 \Phi_{\theta,\varepsilon,j}}{\partial\theta^2}(Z)\right|\right] < \infty.$$

Under $(\mathbf{C_{11}})$-$(\mathbf{C_{14}})$, we propose to estimate $S_{\theta^0,g}(\theta)$ by

$$S_{n,2}(\theta) = \frac{1}{n}\sum_{i=1}^n [(Y_i^2 - \sigma_{\xi,2}^2)\Phi_{\theta,\varepsilon,3}(Z_i) - 2Y_i\Phi_{\theta,\varepsilon,2}(Z_i) + \Phi_{\theta,\varepsilon,1}(Z_i)]. \tag{3.4}$$

Using this criterion we propose to estimate θ^0 by

$$\widehat{\theta}_2 \;=\; \arg\min_{\theta\in\Theta} S_{n,2}(\theta). \tag{3.5}$$

We refer to Section 4 for details on how to construct such functions $\Phi_{\theta,\varepsilon,j}$.

Remark 3.2. As in the first estimation procedure (see Remark 2.3)), the main difficulty for finding such function $\Phi_{\theta,\varepsilon,j}$ lies in the constraint that we expect that they do not depend on the unknown density g. If we relax this constraint, obviously there are a lot of solutions.

3.2.1. *Asymptotic properties of the second estimator $\widehat{\theta}_2$.*

Theorem 3.3. *Consider Model (1.1) under the assumptions* $(\mathbf{A_1})$, $(\mathbf{I2_1})$, $(\mathbf{I2_2})$, $(\mathbf{I2_3})$, $(\mathbf{I2_4})$ *and the conditions* $(\mathbf{C_{11}})$-$(\mathbf{C_{14}})$. *Then $\widehat{\theta}_2$, defined by (3.5) is a $\sqrt{n}$-consistent estimator of θ^0 which satisfies moreover that $\sqrt{n}(\widehat{\theta}_2 - \theta^0) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,\Sigma_2)$, with Σ_2 that equals*

$$\left(\mathbb{E}\left[w_{\theta^0}(X)\left(\frac{\partial f_\theta(X)}{\partial\theta}\right)\left(\frac{\partial f_\theta(X)}{\partial\theta}\right)^\top\right]|_{\theta=\theta^0}\right)^{-1}\Sigma_{0,2}\left(\mathbb{E}\left[w_{\theta^0}(X)\left(\frac{\partial f_\theta(X)}{\partial\theta}\right)\left(\frac{\partial f_\theta(X)}{\partial\theta}\right)^\top\right]|_{\theta=\theta^0}\right)^{-1}$$

where $\Sigma_{0,2}$ *equals*

$$\mathbb{E}\left[\left(\frac{\partial(\Phi_{\theta,\varepsilon,1}(Z)-2Y\Phi_{\theta,\varepsilon,2}(Z)+(Y^2-\sigma^2_{\xi,2})\Phi_{\theta^0,\varepsilon,3}(Z))}{\partial\theta}\,|_{\theta=\theta^0}\right)\right.$$
$$\left.\times\left(\frac{\partial(\Phi_{\theta,\varepsilon,1}(Z)-2Y\Phi_{\theta,\varepsilon,2}(Z)+(Y^2-\sigma^2_{\xi,2})\Phi_{\theta^0,\varepsilon,3}(Z))}{\partial\theta}\,|_{\theta=\theta^0}\right)^{\top}\right].$$

Also note that, by denoting $\Phi_{\theta,\varepsilon,1}=(w_\theta f_\theta^2)^*/p_\varepsilon^*$, $\Phi_{\theta,\varepsilon,2}=(w_\theta f_\theta)^*/p_\varepsilon^*$ and $\Phi_{\theta,\varepsilon,3}=w_\theta^*/p_\varepsilon^*$, we get that $\Sigma_{0,1}=\Sigma_{0,2}$ with $\Sigma_{0,1}$ defined in Theorem 3.2 (See also Comment 2 in Section 4).

4. Comments on the conditions ensuring the $\sqrt{n}$-consistency

Comment 1 Let us briefly compare the conditions $(\mathbf{C_1})$-$(\mathbf{C_3})$ to the conditions $(\mathbf{C_4})$-$(\mathbf{C_7})$. It is noteworthy that the conditions $(\mathbf{C_4})$-$(\mathbf{C_7})$ are more general. For instance Condition $(\mathbf{C_1})$ implies $(\mathbf{C_4})$, with $\tilde\Phi_{\theta,\varepsilon,j}$ defined by $\tilde\Phi^*_{\theta,\varepsilon,1}=(wf_\theta^2)^*/p_\varepsilon^*$, $\tilde\Phi^*_{\theta,\varepsilon,2}=(wf_\theta)^*/p_\varepsilon^*$, and $\tilde\Phi^*_{\theta,\varepsilon,3}=w^*/p_\varepsilon^*$. This comes from the following equalities $\mathbb{E}[\tilde\Phi_{\theta,\varepsilon,1}(Z)]=\mathbb{E}[(wf_\theta^2)(X)]$,

$$\begin{aligned}\mathbb{E}[Y^2\tilde\Phi_{\theta,\varepsilon,3}(Z)] &= \mathbb{E}[f^2_{\theta^0}(X)\tilde\Phi_{\theta,\varepsilon,3}(Z)]+\sigma^2_{\xi,2}\mathbb{E}[\tilde\Phi_{\theta,\varepsilon,3}(Z)]\\ &= \left\langle f^2_{\theta_0}g,\tilde\Phi_{\theta,\varepsilon,3}\star p_\varepsilon\right\rangle+\sigma^2_{\xi,2}\left\langle g,\tilde\Phi_{\theta,\varepsilon,3}\star p_\varepsilon\right\rangle\\ &= (2\pi)^{-1}\left\langle (f^2_{\theta_0}g)^*,(\tilde\Phi_{\theta,\varepsilon,3})^*p^*_\varepsilon\right\rangle+\sigma^2_{\xi,2}(2\pi)^{-1}\left\langle g^*,(\tilde\Phi_{\theta,\varepsilon,3})^*p^*_\varepsilon\right\rangle\\ &= \left\langle f^2_{\theta^0}g,w\right\rangle+\sigma^2_{\xi,2}\left\langle g,w\right\rangle=\mathbb{E}[Y^2w(X)],\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}[Y\tilde\Phi_{\theta,\varepsilon,2}(Z)] &= \left\langle f_{\theta_0}g,\tilde\Phi_{\theta,\varepsilon,2}\star p_\varepsilon\right\rangle=(2\pi)^{-1}\left\langle (f_{\theta_0}g)^*,(\tilde\Phi_{\theta,\varepsilon,2})^*p^*_\varepsilon\right\rangle\\ &= \langle f_{\theta^0}g,f_\theta w\rangle=\mathbb{E}[Y(wf_\theta)(X)].\end{aligned}$$

But in the condition $(\mathbf{C_4})$, $f_\theta w$, $f_\theta^2 w$, and w are not necessarily in $\mathbb{L}_1(\mathbb{R})$. Nevertheless, the conditions $(\mathbf{C_1})$-$(\mathbf{C_3})$ are more tractable and constructive conditions.

Comment 2 The same comparison between $(\mathbf{C_{11}})$-$(\mathbf{C_{14}})$ and $(\mathbf{C_8})$-$(\mathbf{C_{10}})$ holds with w replaced by w_θ and with the $\Phi_{\theta,\varepsilon,j}$ defined by $\Phi^*_{\theta,\varepsilon,1}=(w_\theta f_\theta^2)^*/p_\varepsilon^*$, $\Phi^*_{\theta,\varepsilon,2}=(w_\theta f_\theta)^*/p_\varepsilon^*$, and $\Phi^*_{\theta,\varepsilon,3}=(f_\theta w_\theta)^*/p_\varepsilon^*$.

Comment 3 The conditions $(\mathbf{C_4})$-$(\mathbf{C_7})$ as well as $(\mathbf{C_{11}})$-$(\mathbf{C_{14}})$ have to be related to the conditions given in Fazekas and Kukush (1998), Fazekas *et al.*(1999) and Baran (2000) in the context of the functional errors in variables model with $X_1,\cdots,X_n$ not random. In both mentionned papers, in order to construct $\sqrt{n}$-consistent estimator, they assume that there exist some functions ϕ_1 and ϕ_2 such that

$$\mathbb{E}(\phi_1(x+\varepsilon,\theta))=f_\theta(x),\ \text{and}\ \mathbb{E}(\phi_2(x+\varepsilon,\theta))=f_\theta^2(x). \tag{4.1}$$

Clearly our conditions $(\mathbf{C_4})$-$(\mathbf{C_7})$ or $(\mathbf{C_{11}})$-$(\mathbf{C_{14}})$ are less restrictive than the condition (4.1), and hence they allow to achieve the parametric rate of convergence for various type of regression functions, by using the possibility of the choice of the weight function w_θ. For instance, if

$f_\theta(x) = \theta/(1+x^2)$, and p_ε is the Gaussian density, then the conditions given in Fazekas and Kukush (1998) and Fazekas *et al.*(1999) are not fulfilled.

Whereas $\mathbb{E}[f_\theta(X)w(X)]$ and $\mathbb{E}[f_\theta^2(X)w(X)]$ can be estimated with the parametric rate of convergence, by taking $w(x) = (1+x^2)^4 \exp(-x^2/(4\beta))$. It follows that the condition $(\mathbf{C_4})$-$(\mathbf{C_7})$ are fulfilled in this special example (see Section 5 for further details on this example). Nevertheless, such weight function are not always available and hence those conditions $(\mathbf{C_4})$-$(\mathbf{C_7})$ are not always fulfilled.

5. Examples and methodological advice

In this section, we propose a deep study of the asymptotic properties of the proposed estimators through various examples of regression functions. Moreover, we show that for many regression functions the practitioner may encounter there are a few simple smoothing weight functions to choose so that the rates improve significantly, up to parametric rate in many cases. These new procedures allow to achieve the parametric rate of convergence in lot of examples, and especially in examples where the previous known estimator proposed in Taupin (2001) does not. In all of these examples, the noise distribution is arbitrary, as far as it satisfies $(\mathbf{N_1})$ and $(\mathbf{N_2})$ with $\rho \leq 2$. The two first examples simply show that these new and more general estimation procedures allow to recover previous known results in simple cases. The others examples provide new results that underlie the improvement due to those methods.

Example 1. Polynomial regression function Let f_θ be of the form $f_\theta(x) = \sum_{k=1}^p \theta_k x^k$ and p_ε satisfying $(\mathbf{N_2})$ with $\rho \leq 2$. Assume that $\mathbb{E}(Y^2) < \infty$ and that $\mathbb{E}(Z^{2p}) < \infty$. Let K be such $K^*(t) = \mathbb{1}_{|t|\leq 1}$ and let $w_\theta(x) = w(x) = \exp\{-x^2/(4\beta)\}$. Then conditions $(\mathbf{C_4})$-$(\mathbf{C_7})$ as well as conditions $(\mathbf{C_1})$-$(\mathbf{C_3})$ are satisfied. Consequently the estimators $\tilde{\theta}_1$ and $\tilde{\theta}_2$ are $\sqrt{n}$-consistent and asymptotically Gaussian estimators of θ^0, with the same asymptotic variance.

Remark 5.1. In this example, one can also choose $w_\theta \equiv 1$, provided that the kernel K has finite absolute moments of order p and satisfies $\int u^r K(u)du = 0$, for $r = 1, \cdots, p$. In this case $\tilde{\theta}_1$ is again a $\sqrt{n}$-consistent and asymptotically Gaussian estimator of θ^0.

The $\sqrt{n}$-consistency as well as the asymptotic normality was already achieved with different estimators, in the linear case (see *e. g.* Bickel (1987), Bickel *et al.* (1993), van der Vaart (1996) and (1988), Murphy and van der Vaart (1996)). In polynomial case, some $\sqrt{n}$-consistent estimator already exists, without proving the asymptotic normality (see *e.g.* Taupin (2001) and Comte and Taupin (2001)) or in the polynomial functional errors in variables model, with fixed and not random X_i's (see *e. g.* Hausman *et al.* (1995), (1991) or Chan and Mak (1985)). It is noteworthy that these new estimation procedures, quite more simple and natural than the one proposed in Taupin (2001), also work in this simple case.

Example 2. Exponential regression function Let f_θ be of the form $f_\theta(x) = \exp(\theta x)$ and p_ε satisfying $(\mathbf{N_2})$ with $\rho \leq 2$. Assume that $\mathbb{E}(Y^2) < \infty$ and that $\mathbb{E}[\exp(2\theta^0 Z)] < \infty$. Let K be such $K^*(t) = \mathbb{1}_{|t|\leq 1}$ and let $w_\theta(x) = w(x) = \exp\{-x^2/(4\beta)\}$. Then the conditions $(\mathbf{C_4})$-$(\mathbf{C_7})$ as well as conditions $(\mathbf{C_1})$-$(\mathbf{C_3})$ are satisfied. Consequently the estimators $\tilde{\theta}_1$ and $\tilde{\theta}_2$ are $\sqrt{n}$-consistent and asymptotically Gaussian estimators of θ^0, with the same asymptotic variance.

Remark 5.2. In this example, one can also choose $w \equiv 1$ and use that

$$\mathbb{E}[\exp(\theta X)] = \mathbb{E}[\exp(\theta Z)]/\mathbb{E}[\exp(\theta\varepsilon)].$$

This implies that if we denote by

$$\tilde{\Phi}_{\theta,\varepsilon,1}(Z) = \frac{\exp(2\theta Z)}{\mathbb{E}[\exp(2\theta\varepsilon)]} \text{ and } \tilde{\Phi}_{\theta,\varepsilon,2}(Z) = \frac{\exp(\theta Z)}{\mathbb{E}[\exp(\theta\varepsilon)]}$$

then

$$\mathbb{E}_h[\tilde{\Phi}_{\theta,\varepsilon,1}(Z)] = \mathbb{E}_g[f_\theta^2(X)] \text{ and } \mathbb{E}_{\theta^0,h}[Y\tilde{\Phi}_{\theta,\varepsilon,2}(Z)] = \mathbb{E}_{\theta^0,g}[Yf_\theta(X)].$$

Consequently, $\tilde{S}_{n,2}$ satisfies

$$\tilde{S}_{n,2}(\theta) = \frac{1}{n}\sum_{i=1}^n \left[Y_i - \frac{\exp(\theta Z_i)}{\mathbb{E}[\exp(\theta\varepsilon)]}\right]^2 = \frac{1}{n}\sum_{i=1}^n [Y_i^2 - 2Y_i\tilde{\Phi}_{\theta,\varepsilon,2}(Z_i) + \tilde{\Phi}_{\theta,\varepsilon,1}(Z_i)]. \tag{5.1}$$

In this case $\tilde{\theta}_2$ is also a $\sqrt{n}$-consistent and asymptotically Gaussian estimator of θ^0.

Once again, these new estimation procedures allow to achieve the $\sqrt{n}$-consistency and the asymptotic normality in a simple example where a $\sqrt{n}$-consistent estimator was already known (see Taupin (2001)).

Example 3. Cosines regression function Let f_θ be of the form $f_\theta(x) = \sum_{j=1}^d \theta_j \cos(jx)$ and p_ε satisfying $(\mathbf{N_2})$ with $\rho \leq 2$. Let K be such $K^*(t) = \mathbb{1}_{|t|\leq 1}$ and let $w_\theta(x) = w(x) = \exp\{-x^2/(4\beta)\}$. Then the conditions $(\mathbf{C_4})$-$(\mathbf{C_7})$ as well as conditions $(\mathbf{C_1})$-$(\mathbf{C_3})$ are satisfied. Consequently the estimators $\tilde{\theta}_1$ and $\tilde{\theta}_2$ are $\sqrt{n}$-consistent and asymptotically Gaussian estimators of θ^0, with the same asymptotic variance.

Remark 5.3. In this example, one can also choose $w_\theta \equiv 1$ and use that $\mathbb{E}[\exp(ijX)] = \mathbb{E}[\exp(ijZ)]/\mathbb{E}[\exp(ij\varepsilon)]$. This implies that if we denote by

$$\tilde{\Phi}_{\theta,\varepsilon,1}(Z) = \frac{1}{4}\left\{1 + \sum_{j=1}^d \theta_j^2 \left[\frac{\exp(2ijZ)}{p_\varepsilon^*(2j)} + \frac{\exp(-2ijZ)}{p_\varepsilon^*(-2j)}\right]\right.$$
$$\left. + \sum_{j=1}^d \sum_{k\neq j} \theta_j\theta_k \left[\frac{\exp(i(j+k)Z)}{p_\varepsilon^*(j+k)} + \frac{\exp(-i(j+k)Z)}{p_\varepsilon^*(-(j+k))} + \frac{\exp(i(j-k)Z)}{p_\varepsilon^*(j-k)} + \frac{\exp(i(-j+k)Z)}{p_\varepsilon^*(-j+k)}\right]\right\}$$

and

$$\tilde{\Phi}_{\theta,\varepsilon,2}(Z) = \frac{1}{2}\left[\frac{\exp(ijZ)}{p_\varepsilon^*(j)} + \frac{\exp(-ijZ)}{p_\varepsilon^*(-j)}\right]$$

then

$$\mathbb{E}[\tilde{\Phi}_{\theta,\varepsilon,1}(Z)] = \mathbb{E}_g[f_\theta^2(X)] \text{ and } \mathbb{E}[Y\tilde{\Phi}_{\theta,\varepsilon,2}(Z)] = \mathbb{E}[Yf_\theta(X)].$$

Consequently, $\tilde{S}_{n,2}$ satisfies

$$\tilde{S}_{n,2}(\theta) = \frac{1}{n}\sum_{i=1}^n [Y_i^2 - 2Y_i\tilde{\Phi}_{\theta,\varepsilon,2}(Z_i) + \tilde{\Phi}_{\theta,\varepsilon,1}(Z_i)]. \tag{5.2}$$

In this case $\tilde{\theta}_2$ with $w \equiv 1$ is again a $\sqrt{n}$-consistent and asymptotically Gaussian estimator of θ^0. In the same way, $\tilde{\theta}_1$ with $w \equiv 1$ is again a $\sqrt{n}$-consistent and asymptotically Gaussian estimator of θ^0.

Remark 5.4. This example has already been considered in Taupin (2001) and Comte and Taupin (2001), but the $\sqrt{n}$-consistency as well as the asymptotic normality is new in this context. More precisely, the estimator constructed in Taupin (2001) has a rate of convergence of order $\exp(\sqrt{\log n})/n$ for Gaussian errors, and it was not clear that this rate can be improved by using the previous method.

Example 4. Cauchy regression function 1 Consider Model (1.1) with $f_\theta(x) = \theta/(1+x^2)$ satisfying $(\mathbf{R_1})$ with $a = 0$, $b = 1/2$ and $r = 1$ and p_ε satisfying $(\mathbf{N_2})$ with $\rho \leq 2$. Let K be such $K^*(t) = \mathbb{1}_{|t|\leq 1}$ and let $w_\theta(x) = w(x) = (1+x^2)^4 \exp\{-x^2/(4\beta)\}$. With our choice of w, the functions $f_\theta w$, $f_\theta^2 w$ and their derivatives in θ up to order 3 satisfy $(\mathbf{R_1})$ with $\rho < r = 2$ or $\rho = r = 2$ and $b > \beta$. Consequently, the conditions $(\mathbf{C_4})$-$(\mathbf{C_7})$ as well as conditions $(\mathbf{C_1})$-$(\mathbf{C_3})$ are satisfied and the estimators $\tilde{\theta}_1$ and $\tilde{\theta}_2$ are $\sqrt{n}$-consistent and asymptotically Gaussian estimators of θ^0, with the same asymptotic variance.

This simple example underlies the importance of the smoothing weight function w in the construction of $\tilde{\theta}_1$ or $\tilde{\theta}_2$, since without a smoothing function w in front of the regression function, Theorem 2.1 predicts for a Gaussian noise ε a rate of convergence of order $\exp(-\sqrt{\log n})$ instead of the parametric rate of convergence.

This has two consequences. First, it seems that the estimator proposed in Taupin (2001) with a rate of order $\exp(-\sqrt{\log n})$ could not achieve the parametric rate of convergence since no weight function appears in front of the regression function, in order to smooth it. Moreover it shows that in this case, the conditions (4.1) given in Fazekas and Kukush (1998) and Fazekas *et al.*(1999) are not satisfied when our conditions $(\mathbf{C_4})$-$(\mathbf{C_7})$ or $(\mathbf{C_{11}})$-$(\mathbf{C_{14}})$ hold and our estimators achieve the parametric rate.

Example 5. Laplace regression function Consider Model (1.1) with $f_\theta(x) = \theta f(x)$ and $f(x) = \exp(-|x|/2)$. The Fourier transform of f and hence of f_θ is slowly decaying, like $|u|^{-2}$ as $|u| \to \infty$. The estimator $\tilde{\theta}_1$ with $w \equiv 1$ would not provide $\sqrt{n}$ consistent estimator for smoother noise densities (as soon as $|p_\varepsilon^*(u)| \leq o(|u|^{-2})$ with $|u| \to \infty$). A closer look tells us that f_θ and its derivative in θ is $\mathcal{C}^\infty$ except at one point $x = 0$. Therefore , a proper choice of w can smooth out at 0 and make wf_θ, wf_θ^2 and their derivatives in θ infinitely differentiable functions in x. With such a choice of the weight function w, the estimator $\tilde{\theta}_1$ attains the parametric rate of convergence for a large set of noise densities, such that p_ε satisfies $(\mathbf{N_2})$ for some $0 < \rho < 1$. Even if the noise is smoother than that (say Gaussian), the rate of $\tilde{\theta}_1$ is much faster when using our choice of w then it would be for $w \equiv 1$ or when using the estimator proposed in Taupin (2001).

Let us be more precise, and define

$$\Psi_{a,b}(x) = \exp\left(-\frac{1}{(x-a)^R(b-x)^R}\right) I_{[a,b]}(x), \tag{5.3}$$

where $-\infty < a < b < \infty$ are fixed and A, B, $R > 0$. Following Lepski and Levit (1998) and Fedoryuk (1987), p. 346, Theorem 7.3, the Fourier transform of this function is such that

$$|\Psi^*_{a,b}(u)| \leq c \exp(-C|u|^{R/(R+1)}), \text{ as } |u| \to \infty$$

and c, $C > 0$ are some constants. Then take w like $\Psi_{0,100}$ or $\Psi_{-100,0}$ or their sum (for some $R > 0$ large enough).

Another way to smooth without restraining to compact support is the following. Let $w(x) = \exp(-1/|x|^{2R})$ a weight function which smoothes at 0 as $R > 0$ is large.

We can vary the coefficient in the exponential in the expression of w and check that $f_\theta w$, $f_\theta^2 w$ and their derivatives up to order 3 satisfy $(\mathbf{R_1})$ with the same $r = R/(R+1)$ closer to 1 as R grows large and $b > 0$.

If the noise satisfies $(\mathbf{N_2})$ with $0 \leq \rho < 1$, then we find R large enough such that $r = R/(R+1) > \rho$ and thus conditions $(\mathbf{C_4})$-$(\mathbf{C_7})$ as well as conditions $(\mathbf{C_1})$-$(\mathbf{C_3})$ are satisfied. Consequently the estimators $\tilde{\theta}_1$ and $\tilde{\theta}_2$ are $\sqrt{n}$-consistent and asymptotically Gaussian estimators of θ^0, with the same asymptotic variance.

If $\rho \geq 1$, for w suitably chosen, then

$$\mathbb{E} \parallel \tilde{\theta}_1 - \theta^0 \parallel^2_{\ell^2} = O(1) (\log n)^{\frac{1-2a-r}{\rho}} \exp\{-2b(\log n/(2\beta))^{r/\rho}\}.$$

Note that for the same regression functions with $f(x) = \exp(|x|)$ we can multiply the previous weight function w by $\exp(-4|x|)$ or by $\exp(-x^2)$ in order to solve integrability problems without changing the previous conclusions.

Example 6. Irregular regression function Consider Model (1.1) with $f_\theta(x) = \theta \mathbb{1}_{[-1,1]}(x)$ and p_ε satisfying $(\mathbf{N_2})$. Let K be such that $K^*(t) = \mathbb{1}_{|t|\leq 1}$, and take $w = \Psi_{-1,1}$ for some $R > 0$ defined by (5.3).

If $\rho = 0$ in $(\mathbf{N_2})$, then the conditions $(\mathbf{C_4})$-$(\mathbf{C_7})$ as well as conditions $(\mathbf{C_1})$-$(\mathbf{C_3})$ are satisfied. Consequently the estimators $\tilde{\theta}_1$ and $\tilde{\theta}_2$ are $\sqrt{n}$-consistent and asymptotically Gaussian estimators of θ^0, with the same asymptotic variance.

If $\rho > 0$, then the best rate for estimating θ^0 is obtained by choosing $w = \Psi_{-1,1}$ with $R > 0$ sufficiently large such that wf_θ and wf_θ^2 satisfy $(\mathbf{R_1})$ with $0 < r = R/(R+1) < 1$ as close to 1 as needed.

It follows that if $0 < \rho < 1$, then we can find $w = \Psi_{-1,1}$ (with R large enough) satisfying $(\mathbf{R_1})$ with $r = R/(R+1) > \rho$ and hence the conditions $(\mathbf{C_4})$-$(\mathbf{C_7})$ as well as conditions $(\mathbf{C_1})$-$(\mathbf{C_3})$ are satisfied. Thus the estimators $\tilde{\theta}_1$ and $\tilde{\theta}_2$ are $\sqrt{n}$-consistent and asymptotically Gaussian estimators of θ^0. Whereas, if $\rho \geq 1$, for a suitably chosen w, then

$$\mathbb{E} \parallel \tilde{\theta}_1 - \theta^0 \parallel^2_{\ell^2} = O(1) (\log n)^{\frac{1-2a-r}{\rho}} \exp\{-2b(\log n/(2\beta))^{r/\rho}\}.$$

Example 7. Polygonal regression function Consider Model (1.1) with $f_\theta(x) = \theta_0 + \theta_1 x + \theta_2(x-a)_+ + \theta_3|x-b|^3$ and p_ε satisfying $(\mathbf{N_2})$. Let K be such $K^*(t) = \mathbb{1}_{|t|\leq 1}$. This regression function is $\mathcal{C}^\infty$ except at points a and b where it is not differentiable. We suggest to use the smoothing weight function in (5.3) as follows. Let for some $R > 0$,

$$w_\theta(x) = w(x) = \Psi_{a-100,a}(x) + \Psi_{a,b}(x) + \Psi_{b,b+100}(x)$$

with Ψ defined in (5.3). The idea is that we can truncate the regression function (say on the interval $[a-100, b+100]$, or larger) in order to smooth out at points a, b and the end points of the support of w.

If the noise satisfies $(\mathbf{N_2})$ with $0 \leq \rho < 1$, then take R large enough such that $r = R/(R+1) > \rho$ and thus conditions $(\mathbf{C_4})$-$(\mathbf{C_7})$ as well as conditions $(\mathbf{C_1})$-$(\mathbf{C_3})$ are satisfied. Consequently the estimators $\tilde{\theta}_1$ and $\tilde{\theta}_2$ are $\sqrt{n}$-consistent and asymptotically Gaussian estimators of θ^0, with the same asymptotic variance.

If $\rho \geq 1$ in $(\mathbf{N_2})$, according to Table 1

$$\mathbb{E} \parallel \tilde{\theta}_1 - \theta^0 \parallel_{\ell^2}^2 = O(1) \, (\log n)^{\frac{1-2a-r}{\rho}} \exp\{-2b(\log n/(2\beta))^{r/\rho}\}.$$

Comments on the examples 5, 6, and 7

In those three examples, $\tilde{\theta}_1$ achieves the $\sqrt{n}$-rate of convergence provided that p_ε is ordinary smooth or super smooth with an exponent $\rho < 1$. But, $f_\theta w$ will satisfy $(\mathbf{R_1})$ with r at most such that $r < 1$ and it seems therefore impossible to have $(w_\theta f_\theta)^*/p_\varepsilon^*$ in $\mathbb{L}_1(\mathbb{R})$ if the ε_i's are Gaussian. For this regression function, if the ε_i's are Gaussian, those least square criterion can not be estimated with the parametric rate of convergence and hence could probably, not provide a $\sqrt{n}$-consistent estimator of θ^0 in this context. Nevertheless, even in cases where the parametric rate of convergence seems not achievable by such estimators, the resulting rate of the risk of $\tilde{\theta}_1$ is clearly infinitely faster than the logarithmic rate we could have without such a choice of w or by using the estimator proposed by Taupin (2001).

Example 8. Growth curves 1 Consider Model (1.1) with $f_\theta(x) = \theta_1/(1+\theta_2 \exp(\theta_3 x))$ and p_ε satisfying $(\mathbf{N_2})$ with $\rho \leq 2$. Let $K^*(t) = \mathbb{1}_{|t|\leq 1}$ and let $w_\theta(x) = (1+\theta_2 \exp(\theta_3 x))^4 \exp\{-x^2/(4\beta)\}$. Then the conditions $(\mathbf{C_{11}})$-$(\mathbf{C_{14}})$ as well as conditions $(\mathbf{C_8})$-$(\mathbf{C_{10}})$ are satisfied. Consequently the estimators $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are $\sqrt{n}$-consistent and asymptotically Gaussian estimators of θ^0, with the same asymptotic variance.

Example 9. Growth curve 2 Consider Model (1.1) with $f_\theta(x) = \theta_2 + (\theta_1 - \theta_2)/(1+\exp(\theta_3 + \theta_4 x))$ and p_ε satisfying $(\mathbf{N_2})$ with $\rho \leq 2$. Let K be such $K^*(t) = \mathbb{1}_{|t|\leq 1}$ and let $w_\theta(x) = (1+\exp(\theta_3+\theta_4 x))^4 \exp\{-x^2/(4\beta)\}$. Then the conditions $(\mathbf{C_{11}})$-$(\mathbf{C_{14}})$ as well as conditions $(\mathbf{C_8})$-$(\mathbf{C_{10}})$ are satisfied. Consequently the estimators $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are $\sqrt{n}$-consistent and asymptotically Gaussian estimators of θ^0, with the same asymptotic variance.

Example 10. Cauchy regression function 2 Consider Model (1.1) with $f_\theta(x) = 1/(1+\theta x^2)$ and p_ε satisfying $(\mathbf{N_2})$ with $\rho \leq 2$. Let K be such $K^*(t) = \mathbb{1}_{|t|\leq 1}$ and let $w_\theta(x) = (1+\theta x^2)^2 \exp\{-x^2/(4\beta)\}$. Then the conditions $(\mathbf{C_{11}})$-$(\mathbf{C_{14}})$ as well as conditions $(\mathbf{C_8})$-$(\mathbf{C_{10}})$ are satisfied. Consequently the estimators $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are $\sqrt{n}$-consistent and asymptotically Gaussian estimators of θ^0, with the same asymptotic variance.

In these three last examples, we see the importance of the weight function w_θ, with the improvement of the rate of convergence by taking w_θ depending on θ. Clearly, for such regression function, the estimator in Taupin (2001) does not achieve the parametric rate of convergence whereas, $\widehat{\theta}_1$ or $\widehat{\theta}_2$ do.

6. Proofs of Theorems

We only detail the proof of Theorems 3.1, and 3.3 as well as Corollary 3.1 since Theorem 3.2 is an immediate consequence of 3.1 and also since the proofs of Theorems 2.1, 2.2 and 2.3 and Corollary 2.1 follow the same lines with w_θ replaced by w and by putting $\sigma_{\xi,2}=0$ in the proofs. In the sequel, we denote by C some absolute constant whose value may vary from one line to the other and we always mention the dependency of some constant C with respect to some parameters. For instance $C(a,b)$ stands for a constant depending on a and b.

6.1. **Proof of 1) of Theorem 3.1:** The main point, of the proof consists in showing that for any θ in Θ, $\mathbb{E}[(S_{n,1}(\theta)-S_{\theta^0,g}(\theta))^2]=o(1)$, with $S_{\theta^0,g}(\theta)$ admitting a unique minimum in $\theta=\theta^0$. The second part of the proof consists in studying $\omega_2(n,\rho)$ defined as

$$\omega_2(n,\rho)=\sup\left\{|S_{n,1}(\theta)-S_{n,1}(\theta')|:\|\theta-\theta'\|_{\ell^2}\leq\rho\right\}.$$

By using the regularity assumptions on the regression function f, we state that there exist two sequences ρ_k and ϵ_k tending to 0, such that for all $k\in\mathbb{N}$

$$\lim_{n\to\infty}\mathbb{P}[\omega_2(n,\rho_k)>\epsilon_k]=0 \text{ and that } \mathbb{E}[(\omega_2(n,\rho_k))^2]=O(\rho_k^2). \tag{6.1}$$

Let us start with the proof of

$$\mathbb{E}[(S_{n,1}(\theta)-S_{\theta^0,g}(\theta))^2]=o(1),\quad \text{for any } \theta\in\Theta \text{ and as } n\to\infty. \tag{6.2}$$

We have to check successively that for any $\theta\in\Theta$,

$$\mathbb{E}[S_{n,1}(\theta)]-S_{\theta^0,g}(\theta)=o(1)\quad \text{and } \mathrm{Var}(S_{n,1}(\theta))=o(1),\quad \text{as } n\to\infty. \tag{6.3}$$

For both the bias and the variance, we give two upper bounds, based on the two following applications of the Hölder's inequality

$$|<\varphi_1,\varphi_2>|\leq\|\varphi_1\|_2\|\varphi_2\|_2 \text{ and } |<\varphi_1,\varphi_2>|\leq\|\varphi_1\|_\infty\|\varphi_2\|_1. \tag{6.4}$$

Proof of the first part of (6.3). According to Lemma 6.2, we have

$$\begin{aligned}\mathbb{E}[S_{n,1}(\theta)] &= \mathbb{E}\left[\left((Y-f_\theta)^2w_\theta-\sigma^2_{\xi,2}w_\theta\right)\star K_{n,C_n}(Z)\right]\\ &= \mathbb{E}\left[(Y^2-\sigma^2_{\xi,2})w_\theta\star K_{C_n}(X)\right]-2\mathbb{E}\left[Y(f_\theta w_\theta)\star K_{C_n}(X)\right]+\mathbb{E}\left[(f_\theta^2w_\theta)\star K_{C_n}(X)\right],\end{aligned}$$

and hence

$$\begin{aligned}\mathbb{E}[S_{n,1}(\theta)]-S_{\theta^0,g}(\theta) &= (2\pi)^{-1}\left\langle(f^2_{\theta^0}g)^*,w^*_\theta(K^*_{C_n}(.)-1)\right\rangle-\pi^{-1}\left\langle(f_{\theta^0}g)^*,(f_\theta w_\theta)^*(K^*_{C_n}-1)\right\rangle\\ &\quad+(2\pi)^{-1}\left\langle g^*,(f^2_\theta w_\theta)^*(K^*_{C_n}-1)\right\rangle.\end{aligned}$$

Consequently, under the assumption $(\mathbf{A_2})$, $|\mathbb{E}[S_{n,1}(\theta)]-S_{\theta^0,g}(\theta)|^2=o(1)$. Indeed, according to (6.4), a first upper bound of $|\mathbb{E}[S_{n,1}(\theta)]-S_{\theta^0,g}(\theta)|$ is given by

$$\begin{aligned}(2\pi)^{-1}\|(f^2_{\theta^0}g)^*\|_2\|w^*_\theta(K^*_{C_n}-1)\|_2+\pi^{-1}\|(f_{\theta^0}g)^*\|_2\|(f_\theta w_\theta)^*(K^*_{C_n}-1)\|_2\\ +(2\pi)^{-1}\|g^*\|_2\|(f^2_\theta w_\theta)^*(K^*_{C_n}-1)\|_2\end{aligned}$$

and hence

$$(6.5)\quad |\mathbb{E}[S_{n,1}(\theta)] - S_{\theta^0,g}(\theta)|^2 \leq C(f_{\theta^0}, w_\theta)\left[\|(w_\theta^*(K_{C_n}^*-1)\|_2^2 + \left\|(w_\theta f_\theta)^*(K_{C_n}^*-1)\right\|_2^2 + \left\|(w_\theta f_\theta^2)^*(K_{C_n}^*-1)\right\|_2^2\right]$$

And we also have that $|\mathbb{E}[S_{n,1}(\theta)] - S_{\theta^0,g}(\theta)|$ is bounded by

$$(2\pi)^{-1} \parallel (f_{\theta^0}^2 g)^* \parallel_\infty \parallel w_\theta^*(K_{C_n}^*-1) \parallel_1 + \pi^{-1} \parallel (f_{\theta^0} g)^* \parallel_\infty \parallel (f_\theta w_\theta)^*(K_{C_n}^*-1) \parallel_1 + (2\pi)^{-1} \parallel g^* \parallel_\infty \parallel (f_\theta^2 w_\theta)^*(K_{C_n}^*-1) \parallel_1$$

and hence

$$(6.6)\quad |\mathbb{E}[S_{n,1}(\theta)] - S_{\theta^0,g}(\theta)|^2 \leq C(f_{\theta^0}, w_\theta)\left[\|w_\theta^*(K_{C_n}^*-1)\|_1^2 + \|(w_\theta f_\theta)^*(K_{C_n}^*-1)\|_1^2 + \|(w_\theta f_\theta^2)^*(K_{C_n}^*-1)\|_1^2\right].$$

Consequently, by combining the two bounds (6.5) and (6.6) we get that $|\mathbb{E}[S_{n,1}(\theta)] - S_{\theta^0,g}(\theta)|^2$ is bounded by

$$C(f_{\theta^0}, w_\theta)\min\left\{\|w_\theta^*(K_{C_n}^*-1)\|_2^2 + \|(w_\theta f_\theta)^*(K_{C_n}^*-1)\|_2^2 + \|(w_\theta f_\theta^2)^*(K_{C_n}^*-1)\|_2^2, \|w_\theta^*(K_{C_n}^*-1)\|_1^2 + \|(w_\theta f_\theta)^*(K_{C_n}^*-1)\|_1^2 + \|(w_\theta f_\theta^2)^*(K_{C_n}^*-1)\|_1^2\right\}$$

that is by applying Lemma 6.1,

$$|\mathbb{E}[S_{n,1}(\theta)] - S_{\theta^0,g}(\theta)|^2 \leq C(f_{\theta^0}, w_\theta, b, r) C_n^{-2a+(1-r)+(1-r)_-} e^{-2bC_n^r},$$

and the first part of (6.3) follows.

Proof of the second part of (6.3). Since the variables are i.i.d. random variables, the stochastic term on the left-hand side in (6.3) has variance

$$\begin{aligned}
\mathrm{Var}[S_{n,1}(\theta)] &\leq n^{-1}\mathbb{E}\left[\left|\left((Y-f_\theta)^2 w_\theta\right) - \sigma_{\xi,2}^2 w_\theta\right) \star K_{n,C_n}(Z)\right|^2\right] \\
&\leq 3n^{-1}\left\{\mathbb{E}[(Y^4+\sigma_{\xi,2}^4)|w_\theta \star K_{n,C_n}(Z)|^2] + 4\mathbb{E}\left[Y^2\left|(f_\theta w_\theta)\star K_{n,C_n}(Z)\right|^2\right]\right. \\
&\qquad \left. +\mathbb{E}\left[\left|(f_\theta^2 w_\theta)\star K_{n,C_n}(Z)\right|^2\right]\right\} \\
&\leq (C/n)\left\{\mathbb{E}\left[f_{\theta^0}^4(X)(w_\theta \star K_{n,C_n})^2(Z)\right] + (\mathbb{E}(\xi^4)+\sigma_{\xi,2}^4)\mathbb{E}\left[(w_\theta\star K_{n,C_n})^2(Z)\right]\right. \\
&\qquad \left. +\mathbb{E}\left[f_{\theta^0}^2(X)((f_\theta w_\theta)\star K_{n,C_n}(Z))^2\right] + \mathbb{E}\left[((f_\theta^2 w_\theta)\star K_{n,C_n}(Z))^2\right]\right\}
\end{aligned}$$

that is, according to Lemma 6.2

$$\begin{aligned}
\mathrm{Var}[S_{n,1}(\theta)] \leq\ & (C/n)\left\{\left\langle((f_{\theta^0}^4 g)\star p_\varepsilon), (w_\theta\star K_{n,C_n})^2\right\rangle + (\mathbb{E}(\xi^4)+\sigma_{\xi,2}^4)\left\langle(g\star p_\varepsilon), (w_\theta\star K_{n,C_n})^2\right\rangle\right. \\
& + \left\langle((f_{\theta^0}^2 g)\star p_\varepsilon), ((f_\theta w_\theta)\star K_{n,C_n})^2\right\rangle + \sigma_{\xi,2}^2\left\langle(g\star p_\varepsilon), ((f_\theta w_\theta)\star K_{n,C_n})^2\right\rangle \\
& \left. + \left\langle(g\star p_\varepsilon), (f_\theta^2 w_\theta)\star K_{n,C_n})^2\right\rangle\right\}.
\end{aligned}$$

On one hand, according to (6.4), under the assumption $(\mathbf{A_2})$,

$$\begin{aligned}\mathrm{Var}[S_{n,1}(\theta)] \;\leq\; & (C/n)\left\{[\| (f_{\theta^0}^4 g)\star p_\varepsilon \|_\infty + (\mathbb{E}(\xi^4)+\sigma_{\xi,2}^4) \| g\star p_\varepsilon \|_\infty] \| w_\theta \star K_{n,C_n} \|_2^2 \right.\\ & +[\| (f_{\theta^0}^2 g)\star p_\varepsilon \|_\infty + \sigma_{\xi,2}^2 \| g\star p_\varepsilon \|_\infty] \| (f_\theta w_\theta)\star K_{n,C_n} \|_2^2 \\ & \left. + \| g\star p_\varepsilon \|_\infty \| (f_\theta^2 w_\theta)\star K_{n,C_n} \|_2^2 \right\},\end{aligned}$$

and we get that

$$\begin{aligned}(6.7)\quad & \mathrm{Var}[S_{n,1}(\theta)] \\ & \leq C(\sigma_{\xi,2}^2,\sigma_{\xi,4},f_\theta,w_\theta,p_\varepsilon)n^{-1}\left[\left\|(w_\theta)^*\frac{K_{C_n}^*}{p_\varepsilon^*}\right\|_2^2+\left\|(f_\theta w_\theta)^*\frac{K_{C_n}^*}{p_\varepsilon^*}\right\|_2^2+\left\|(f_\theta^2 w_\theta)^*\frac{K_{C_n}^*}{p_\varepsilon^*}\right\|_2^2\right].\end{aligned}$$

On the other hand, under the assumption $(\mathbf{A_2})$,

$$\begin{aligned}\mathrm{Var}[S_{n,1}(\theta)] \;\leq\; & \frac{C}{n}\left\{[\| (f_{\theta^0}^4 g)\star p_\varepsilon \|_1 + (\mathbb{E}(\xi^4)+\sigma_{\xi,2}^4) \| g\star p_\varepsilon \|_1] \| w_\theta \star K_{n,C_n} \|_\infty^2 \right.\\ & +[\| (f_{\theta^0}^2 g)\star p_\varepsilon \|_1 + \sigma_{\xi,2}^2 \| g\star p_\varepsilon \|_1] \| (f_\theta w_\theta)\star K_{n,C_n} \|_\infty^2 \\ & \left. + \| g\star p_\varepsilon \|_1 \| (f_\theta^2 w_\theta)\star K_{n,C_n} \|_\infty^2 \right\},\end{aligned}$$

and hence we get that

$$\begin{aligned}(6.8)\qquad & \mathrm{Var}[S_{n,1}(\theta)] \\ & \leq C(\sigma_{\xi,2}^2,\sigma_{\xi,4},f_\theta,w_\theta,p_\varepsilon)n^{-1}\left[\left\|(w_\theta)^*\frac{K_{C_n}^*}{p_\varepsilon^*}\right\|_1^2+\left\|(f_\theta w_\theta)^*\frac{K_{C_n}^*}{p_\varepsilon^*}\right\|_1^2+\left\|(f_\theta^2 w_\theta)^*\frac{K_{C_n}^*}{p_\varepsilon^*}\right\|_1^2\right].\end{aligned}$$

By combining (6.7) and (6.8) and by applying Lemma 6.1 in Appendix, we get that, under assumption $(\mathbf{A_2})$ $\mathrm{Var}[S_{n,1}(\theta)]$ is bounded by

$$\begin{aligned}n^{-1}C(\sigma_\xi^2,\sigma_{\xi,4},f_\theta,w_\theta,p_\varepsilon)\times\min\Bigg\{&\left\|(w_\theta)^*\frac{K_{C_n}^*}{p_\varepsilon^*}\right\|_2^2+\left\|(f_\theta w_\theta)^*\frac{K_{C_n}^*}{p_\varepsilon^*}\right\|_2^2+\left\|(f_\theta^2 w_\theta)^*\frac{K_{C_n}^*}{p_\varepsilon^*}\right\|_2^2,\\ &\left\|(w_\theta)^*\frac{K_{C_n}^*}{p_\varepsilon^*}\right\|_1^2+\left\|(f_\theta w_\theta)^*\frac{K_{C_n}^*}{p_\varepsilon^*}\right\|_1^2+\left\|(f_\theta^2 w_\theta)^*\frac{K_{C_n}^*}{p_\varepsilon^*}\right\|_1^2\Bigg\}\end{aligned}$$

that is

$$\mathrm{Var}[S_{n,1}(\theta)] \leq C(\sigma_\xi^2,\sigma_{\xi,4},f_\theta,w_\theta,p_\varepsilon)\max[1, C_n^{2\alpha-2a+(1-\rho)+(1-\rho)_-}e^{-2bC_n^r+2\beta C_n^\rho}]/n.$$

and under (2.6), then (6.2) is proved.

It remains now to check that there exists two sequences ρ_k and ϵ_k tending to 0, such that (6.1) holds. First write that

$$\begin{aligned}|S_{n,1}(\theta)-S_{n,1}(\theta')| &= \Bigg|\frac{2}{n}\sum_{i=1}^n Y_i[(f_{\theta'}w_{\theta'}-f_\theta w_\theta)]\star K_{n,C_n}(Z_i)-\frac{1}{n}\sum_{i=1}^n[(f_{\theta'}^2w_{\theta'}-f_\theta^2w_\theta)]\star K_{n,C_n}(Z_i)\\ &\quad -\frac{1}{n}\sum_{i=1}^n(Y_i^2-\sigma_{\xi,2}^2)[w_{\theta'}-w_\theta]\star K_{n,C_n}(Z_i)\Bigg|\\ &= \left|\frac{2}{n}\sum_{i=1}^n(\theta'-\theta)^\top\left(\frac{\partial[Y_if_\theta w_\theta+f_\theta^2w_\theta-(Y_i^2-\sigma_{\xi,2}^2)w_\theta]}{\partial\theta}\,|_{\theta=\bar\theta}\right)\star K_{n,C_n}(Z_i)\right|\end{aligned}$$

where $\|\bar\theta\|_{\ell^2}\le\|\theta-\theta'\|_{\ell^2}$, Consequently for $\|\theta-\theta'\|_{\ell^2}\le\rho_k$,

$$|S_{n,1}(\theta)-S_{n,1}(\theta')| \le \frac{2\rho_k}{n}\left\|\sum_{i=1}^n\left(\frac{\partial[Y_if_\theta w_\theta+f_\theta^2w_\theta+(Y_i^2-\sigma_{\xi,2}^2)w_\theta]}{\partial\theta}\,|_{\theta=\bar\theta}\right)\star K_{n,C_n}(Z_i)\right\|_{\ell^2}.$$

It follows that (6.1) holds since for C_n satisfying (2.6), by using the same arguments as for the proof of (6.2), we have that for all $\theta\in\Theta$

$$\mathbb{E}[\|n^{-1}\sum_{i=1}^n\left(\frac{\partial[Y_if_\theta w_\theta+f_\theta^2w_\theta+(Y_i^2-\sigma_{\xi,2}^2)w_\theta]}{\partial\theta}\right)\star K_{n,C_n}(Z_i)\|_{\ell^2}^2]=O(1),$$

and hence for all $k\in\mathbb{N}$, $\mathbb{E}[\sup_{\|\theta'-\theta\|_{\ell^2}\le\rho_k}|S_{n,1}(\theta)-S_{n,1}(\theta')|]=O(\rho_k)$ as $n\to\infty$. □

6.2. **Proofs of 2) of Theorem 3.1 and Corollary 3.1:** If we denote by $S_{n,1}^{(1)}$ and $S_{n,1}^{(2)}$ the first and second derivatives of $S_{n,1}(\theta)$ with respect to θ, by using classical Taylor expansion based on the smoothness properties of $\theta\mapsto w_\theta f_\theta$ and the consistency of $\widehat\theta_1$, we obtain that

$$0=S_{n,1}^{(1)}(\widehat\theta_1)=S_{n,1}^{(1)}(\theta^0)+S_{n,1}^{(2)}(\theta^0)(\widehat\theta_1-\theta^0)+R_{n,1}(\widehat\theta_1-\theta^0),$$

with $R_{n,1}$ defined by

$$R_{n,1}=\int_0^1[S_{n,1}^{(2)}(\theta^0+s(\widehat\theta_1-\theta^0))-S_{n,1}^{(2)}(\theta^0)]ds. \tag{6.9}$$

This implies that

$$\widehat\theta_1-\theta^0=-[S_{n,1}^{(2)}(\theta^0)+R_{n,1}]^{-1}S_{n,1}^{(1)}(\theta^0), \tag{6.10}$$

Consequently, we have to check the four following points.

i) $\mathbb{E}\left[(S_{n,1}^{(1)}(\theta^0)-S_{\theta^0,g}^{(1)}(\theta^0))(S_{n,1}^{(1)}(\theta^0)-S_{\theta^0,g}^{(1)}(\theta^0))^\top\right]=O[\varphi_n\varphi_n^\top]$,

ii) $\mathbb{E}\left[\|S_{n,1}^{(2)}(\theta^0)-S_{\theta^0,g}^{(2)}(\theta^0)\|_{\ell^2}^2\right]=o(1)$.

iii) $R_{n,1}$ defined in (6.9) satisfies $\mathbb{E}(\|R_{n,1}\|_{\ell^2}^2)=o(1)$ as $n\to\infty$.

iv) $\mathbb{E}\|\widehat\theta_1-\theta^0\|_{\ell^2}^2\le 4\mathbb{E}\left[(S_{n,1}^{(1)}(\theta^0))^\top[(S_{\theta^0,g}^{(2)}(\theta^0))^{-1}]^\top(S_{\theta^0,g}^{(2)}(\theta^0))^{-1}S_{n,1}^{(1)}(\theta^0)\right]+o(\varphi_n^2)$.

The rate of convergence of $\widehat\theta_1$ is thus given by the order of $S_{n,1}^{(1)}(\theta^0)-S_{\theta^0,g}^{(1)}(\theta^0)=S_{n,1}^{(1)}(\theta^0)$.

Proof of i)

Write that

$$\begin{aligned} S_{n,1}^{(1)}(\theta) &= \frac{\partial}{\partial\theta}\left(\frac{1}{n}\sum_{i=1}^{n}[((Y_i-f_\theta)^2-\sigma_{\xi,2}^2)w_\theta]\star K_{n,C_n}(Z_i)-\mathbb{E}[((Y-f_\theta(X))^2-\sigma_{\xi,2}^2)w_\theta(X)]\right) \\ &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial[((Y_i-f_\theta)^2-\sigma_{\xi,2}^2)w_\theta]}{\partial\theta}\right)\star K_{n,C_n}(Z_i)-\mathbb{E}\left[\frac{\partial[((Y-f_\theta(X))^2-\sigma_{\xi,2}^2)w_\theta(X)]}{\partial\theta}\right] \end{aligned}$$

Study of the bias

According to Lemma 6.2, $\mathbb{E}[S_{n,1}^{(1)}(\theta^0)]$ is equal to

$$\begin{aligned} -2\mathbb{E}\left[f_{\theta^0}(X)\left(\frac{\partial(f_\theta w_\theta)}{\partial\theta}\star K_{C_n}(X)-\frac{\partial(f_\theta w_\theta)}{\partial\theta}(X)\right)\right]+\mathbb{E}\left[\frac{\partial(f_\theta^2 w_\theta)}{\partial\theta}\star K_{C_n}(X)-\frac{\partial(f_\theta^2 w_\theta)}{\partial\theta}(X)\right] \\ +\mathbb{E}\left[f_{\theta^0}^2(X)\left(\frac{\partial w_\theta}{\partial\theta}\star K_{C_n}(X)-\frac{\partial w_\theta}{\partial\theta}(X)\right)\right] \end{aligned}$$

that is

$$\begin{aligned} \mathbb{E}[S_{n,1}^{(1)}(\theta^0)] &= -2\left\langle (f_{\theta^0}g)^*, \left(\frac{\partial(f_\theta w_\theta)}{\partial\theta}\mid_{\theta=\theta^0}\right)^*(K_{C_n}^*-1)\right\rangle \\ &+ \left\langle g^*, \left(\frac{\partial(f_\theta^2 w_\theta)}{\partial\theta}\mid_{\theta=\theta^0}\right)^*(K_{C_n}^*-1)\right\rangle+\left\langle (f_{\theta^0}^2 g)^*, \left(\frac{\partial w_\theta}{\partial\theta}\mid_{\theta=\theta^0}\right)^*(K_{C_n}^*-1)\right\rangle. \end{aligned}$$

On one hand, according to (6.4), under the assumption $(\mathbf{A_2})$, the bias is bounded in the following way

$$\begin{aligned} \left|\mathbb{E}\left[\frac{\partial S_{n,1}(\theta)}{\partial\theta_j}\right]_{\theta=\theta^0}\right| &\leq \pi^{-1}\parallel (f_{\theta^0}g)^*\parallel_2\left\|\left(\frac{\partial(f_\theta w_\theta)}{\partial\theta_j}\mid_{\theta=\theta^0}\right)^*(K_{C_n}^*-1)\right\|_2 \\ &+(2\pi)^{-1}\parallel g^*\parallel_2\left\|\left(\frac{\partial(f_\theta^2 w_\theta)}{\partial\theta_j}\mid_{\theta=\theta^0}\right)^*(K_{C_n}^*-1)\right\|_2 \\ &+(2\pi)^{-1}\parallel (f_{\theta^0}^2 g)^*\parallel_2\left\|\left(\frac{\partial w_\theta}{\partial\theta_j}\mid_{\theta=\theta^0}\right)^*(K_{C_n}^*-1)\right\|_2 \end{aligned}$$

and consequently,

$$\begin{aligned} \left|\mathbb{E}\left[\left(\frac{\partial S_{n,1}(\theta)}{\partial\theta_j}\mid_{\theta=\theta^0}\right)\right]\right| &\leq C(f_{\theta^0},w_{\theta^0},p_\varepsilon) \\ &\times\left[\left\|\left(\frac{\partial(w_\theta)}{\partial\theta_j}\mid_{\theta=\theta^0}\right)^*(K_{C_n}^*-1)\right\|_2+\left\|\left(\frac{\partial(w_\theta f_\theta)}{\partial\theta_j}\mid_{\theta=\theta^0}\right)^*(K_{C_n}^*-1)\right\|_2\right. \\ &\left.+\left\|\left(\frac{\partial(w_\theta f_\theta^2)}{\partial\theta_j}\mid_{\theta=\theta^0}\right)^*(K_{C_n}^*-1)\right\|_2\right]. \end{aligned} \tag{6.11}$$

And, on the other hand, the bias can also be bounded in the following way

$$\begin{aligned}\left|\mathbb{E}\left[\frac{\partial S_{n,1}(\theta)}{\partial\theta_j}\right]_{\theta=\theta^0}\right| &\leq \pi^{-1}\parallel (f_{\theta^0}g)^*\parallel_\infty \left\|\left(\frac{\partial(f_\theta w_\theta)}{\partial\theta_j}\mid_{\theta=\theta^0}\right)^*(K^*_{C_n}-1)\right\|_1\\ &+(2\pi)^{-1}\parallel g^*\parallel_\infty \left\|\left(\frac{\partial(f_\theta^2 w_\theta)}{\partial\theta_j}\mid_{\theta=\theta^0}\right)^*(K^*_{C_n}-1)\right\|_1\\ &+(2\pi)^{-1}\parallel (f^2_{\theta^0}g)^*\parallel_\infty \left\|\left(\frac{\partial w_\theta}{\partial\theta_j}\mid_{\theta=\theta^0}\right)^*(K^*_{C_n}-1)\right\|_1\end{aligned}$$

and consequently,

$$\begin{aligned}\left|\mathbb{E}\left[\left(\frac{\partial S_{n,1}(\theta)}{\partial\theta_j}\mid_{\theta=\theta^0}\right)\right]\right| &\leq C(f_{\theta^0},w_{\theta^0},p_\varepsilon)\\ &\times\left[\left\|\left(\frac{\partial(w_\theta)}{\partial\theta_j}\mid_{\theta=\theta^0}\right)^*(K^*_{C_n}-1)\right\|_1+\left\|\left(\frac{\partial(w_\theta f_\theta)}{\partial\theta_j}\mid_{\theta=\theta^0}\right)^*(K^*_{C_n}-1)\right\|_1\right.\\ &\left.+\left\|\left(\frac{\partial(w_\theta f_\theta^2)}{\partial\theta_j}\mid_{\theta=\theta^0}\right)^*(K^*_{C_n}-1)\right\|_1\right].\end{aligned}\tag{6.12}$$

By combining (6.11) and (6.12), we get that

$$\left|\mathbb{E}\left[\left(\frac{\partial S_{n,1}(\theta)}{\partial\theta_j}\mid_{\theta=\theta^0}\right)\right]\right|\leq C(f_{\theta^0},w_{\theta^0},p_\varepsilon)\min[B^{(1)}_{n,j}(\theta^0),B^{(2)}_{n,j}(\theta^0)]$$

with $B^{(1)}_{n,j}(\theta^0)$ and $B^{(2)}_{n,j}(\theta^0)$ defined in Theorem 3.1..

Consequently, by applying Lemma we obtain that 6.1,

$$\mathbb{E}^2\left[\left(\frac{\partial S_{n,1}(\theta)}{\partial\theta_j}\mid_{\theta=\theta^0}\right)\right]\leq C^{(1)}_{b_2}(f_{\theta^0},w_{\theta^0},p_\varepsilon,b,r)C_n^{-2a+(1-r)+(1-r)_-}e^{-2bC_n^r}.$$

Study of the variance

For the variance term, it is easy to see that

$$\mathrm{Var}\left(\frac{\partial S_{n,1}(\theta)}{\partial\theta_j}\mid_{\theta=\theta^0}\right)\leq\frac{C}{n}\mathbb{E}\left[\left(\frac{\partial[-2Y_if_\theta w_\theta+f_\theta^2w_\theta+(Y_i^2-\sigma^2_{\xi,2})w_\theta]}{\partial\theta_j}\mid_{\theta=\theta^0}\right)\star K_{n,C_n}(Z_i)\right]^2$$

that is, according to Lemma 6.2, $\mathrm{Var}\,(\partial S_{n,1}(\theta)/\partial\theta_j\mid_{\theta=\theta^0})$ equals

$$\begin{aligned}\frac{C+o(1)}{n}\ \Bigg\{&\left\langle((f^2_{\theta^0}+\sigma^2_{\xi,2})g)\star p_\varepsilon,\left(\left(\frac{\partial(f_\theta w_\theta)}{\partial\theta_j}\mid_{\theta=\theta^0}\right)\star K_{n,C_n}\right)^2\right\rangle\\ &+\left\langle((f^4_{\theta^0}+\sigma_{\xi,4}+4f^2_{\theta^0}\sigma^2_{\xi,2}-\sigma^4_{\xi,2}+4f_{\theta^0}\sigma_{\xi,3})g)\star p_\varepsilon,\left(\left(\frac{\partial w_\theta}{\partial\theta_j}\mid_{\theta=\theta^0}\right)\star K_{n,C_n}\right)^2\right\rangle\\ &+\left\langle g\star p_\varepsilon,\left(\left(\frac{\partial(f_\theta^2w_\theta)}{\partial\theta_j}\mid_{\theta=\theta^0}\right)\star K_{n,C_n}\right)^2\right\rangle\Bigg\}.\end{aligned}$$

It follows that, according to (6.4), $\mathrm{Var}\,(\partial S_{n,1}(\theta)/\partial\theta_j \mid_{\theta=\theta^0})$ is less than

$$\frac{C}{n}\left\{[\sigma^2_{\xi,2} \parallel g\star p_\varepsilon \parallel_\infty + \parallel (f^2_{\theta^0}g)\star p_\varepsilon \parallel_\infty]\left\|\left(\frac{\partial(f_\theta w_\theta)}{\partial\theta_j}\mid_{\theta=\theta^0}\right)\star K_{n,C_n}\right\|_2^2\right.$$
$$+\parallel ((f^4_{\theta^0}+\sigma_{\xi,4}+4f^2_{\theta^0}\sigma^2_{\xi,2}-\sigma^4_{\xi,2}+4f_{\theta^0}\sigma_{\xi,3})g)\star p_\varepsilon \parallel_\infty \left\|\left(\frac{\partial w_\theta}{\partial\theta_j}\mid_{\theta=\theta^0}\right)\star K_{n,C_n}\right\|_2^2$$
$$\left.+\parallel g\star p_\varepsilon \parallel_\infty \left\|\left(\frac{\partial(f^2_\theta w_\theta)}{\partial\theta_j}\mid_{\theta=\theta^0}\right)\star K_{n,C_n}\right\|_2^2\right\},$$

that is

$$\mathrm{Var}\left(\frac{\partial S_{n,1}(\theta)}{\partial\theta_j}\mid_{\theta=\theta^0}\right)\leq\frac{C(\sigma^2_{\xi,2},f_{\theta^0},f^{(1)}_{\theta^0,j},w_{\theta^0},p_\varepsilon)}{n}$$
$$\times\left[\left\|\left(\frac{\partial(w_{\theta)}}{\partial\theta_j}\mid_{\theta=\theta^0}\right)^*\frac{K^*_{C_n}}{p^*_\varepsilon}\right\|_2^2+\left\|\left(\frac{\partial(f_\theta w_\theta)}{\partial\theta_j}\mid_{\theta=\theta^0}\right)^*\frac{K^*_{C_n}}{p^*_\varepsilon}\right\|_2^2+\left\|\left(\frac{\partial(f^2_\theta w_\theta)}{\partial\theta_j}\mid_{\theta=\theta^0}\right)^*\frac{K^*_{C_n}}{p^*_\varepsilon}\right\|_2^2\right]. \tag{6.13}$$

And once again, according to (6.4), another bound the variance term can be obtained to get that $\mathrm{Var}\,(\partial S_{n,1}(\theta)/\partial\theta_j \mid_{\theta=\theta^0})$ is bounded by

$$\frac{C}{n}\left\{[\sigma^2_{\xi,2} \parallel g\star p_\varepsilon \parallel_1 + \parallel (f^2_{\theta^0}g)\star p_\varepsilon \parallel_1]\left\|\left(\frac{\partial(f_\theta w_\theta)}{\partial\theta_j}\mid_{\theta=\theta^0}\right)\star K_{n,C_n}\right\|_\infty^2\right.$$
$$+\parallel g\star p_\varepsilon \parallel_1 \left\|\left(\frac{\partial(f^2_\theta w_\theta)}{\partial\theta_j}\mid_{\theta=\theta^0}\right)\star K_{n,C_n}\right\|_\infty^2$$
$$\left.+\parallel ((f^4_{\theta^0}+\sigma_{\xi,4}+4f^2_{\theta^0}\sigma^2_{\xi,2}-\sigma^4_{\xi,2}+4f_{\theta^0}\sigma_{\xi,3})g)\star p_\varepsilon \parallel_1 \left\|\left(\frac{\partial w_\theta}{\partial\theta_j}\mid_{\theta=\theta^0}\right)\star K_{n,C_n}\right\|_\infty^2\right\},$$

that is

$$\mathrm{Var}\left(\frac{\partial S_{n,1}(\theta)}{\partial\theta_j}\mid_{\theta=\theta^0}\right)\leq\frac{C(\sigma^2_{\xi,2},f_{\theta^0},f^{(1)}_{\theta^0,j},w_{\theta^0},p_\varepsilon)}{n}$$
$$\times\left[\left\|\left(\frac{\partial(w_{\theta)}}{\partial\theta_j}\mid_{\theta=\theta^0}\right)^*\frac{K^*_{C_n}}{p^*_\varepsilon}\right\|_1^2+\left\|\left(\frac{\partial(f_\theta w_\theta)}{\partial\theta_j}\mid_{\theta=\theta^0}\right)^*\frac{K^*_{C_n}}{p^*_\varepsilon}\right\|_1^2+\left\|\left(\frac{\partial(f^2_\theta w_\theta)}{\partial\theta_j}\mid_{\theta=\theta^0}\right)^*\frac{K^*_{C_n}}{p^*_\varepsilon}\right\|_1^2,\right]. \tag{6.14}$$

By combining (6.13) and (6.14), we get that

$$\mathrm{Var}\left(\frac{\partial S_{n,1}(\theta)}{\partial\theta_j}\mid_{\theta=\theta^0}\right)\leq\frac{C(\sigma^2_{\xi,2},f_{\theta^0},f^{(1)}_{\theta^0,j},w_{\theta^0},p_\varepsilon)}{n}$$
$$\min\left\{\left\|\left(\frac{\partial(w_\theta)}{\partial\theta_j}\mid_{\theta=\theta^0}\right)^*\frac{K^*_{C_n}}{p^*_\varepsilon}\right\|_2^2+\left\|\left(\frac{\partial(f_\theta w_\theta)}{\partial\theta_j}\mid_{\theta=\theta^0}\right)^*\frac{K^*_{C_n}}{p^*_\varepsilon}\right\|_2^2+\left\|\left(\frac{\partial(f^2_\theta w_\theta)}{\partial\theta_j}\mid_{\theta=\theta^0}\right)^*\frac{K^*_{C_n}}{p^*_\varepsilon}\right\|_2^2,\right.$$
$$\left.\left\|\left(\frac{\partial(w_\theta)}{\partial\theta_j}\mid_{\theta=\theta^0}\right)^*\frac{K^*_{C_n}}{p^*_\varepsilon}\right\|_1^2+\left\|\left(\frac{\partial(f_\theta w_\theta)}{\partial\theta_j}\mid_{\theta=\theta^0}\right)^*\frac{K^*_{C_n}}{p^*_\varepsilon}\right\|_1^2+\left\|\left(\frac{\partial(f^2_\theta w_\theta)}{\partial\theta_j}\mid_{\theta=\theta^0}\right)^*\frac{K^*_{C_n}}{p^*_\varepsilon}\right\|_1^2\right\}$$

that is, by applying Lemma 6.1

$$\mathrm{Var}\left(\frac{\partial S_{n,1}(\theta)}{\partial\theta_j}\mid_{\theta=\theta^0}\right)\leq\frac{C_{v_2}(\sigma^2_{\xi,2},f_{\theta^0}f^{(1)}_{\theta^0,j},w_{\theta^0},p_\varepsilon)}{n}\max[1,C_n^{2\alpha-2a+(1-\rho)+(1-\rho)_-}\exp\{-2bC_n^r+2\beta C_n^\rho\}].$$

The rate of convergence of $\widehat{\theta}_1$ denoting by $\varphi_n\varphi_n^\top$ corresponds to the best choice for the sequence C_n^*, minimizing the sum of the variance $\mathrm{Var}[S^{(1)}_{n,1}(\theta^0)]$ and the square of the bias $(\mathbb{E}[S^{(1)}_{n,1}(\theta^0)]-S^{(1)}(\theta^0))(\mathbb{E}[S^{(1)}_{n,1}(\theta^0)]-S^{(1)}(\theta^0))^\top=(\mathbb{E}[S^{(1)}_{n,1}(\theta^0)])(\mathbb{E}[S^{(1)}_{n,1}(\theta^0)])^\top$.

Polynomial noise (see $(\mathbf{N_2})$ with $\beta=\rho=0$)

• If for $j=1,\cdots,m$, $\partial(f_\theta w_\theta)/\partial\theta_j|_{\theta=\theta^0}$, $\partial w_\theta/\partial\theta_j|_{\theta=\theta^0}$ and $\partial(f^2_\theta w_\theta)/\partial\theta_j|_{\theta=\theta^0}$ satisfy $(\mathbf{R_1})$ with $r=0$, then

$$|\mathbb{E}[\partial S_{n,1}(\theta)/\partial\theta_j|_{\theta=\theta^0}]|^2\quad\leq\quad C_{b_2}(f_{\theta^0},w_{\theta^0},b,\sigma^2_{\xi,2})C_n^{-2a+1},$$

and

$$\mathrm{Var}(\partial S_{n,1}(\theta)/\partial\theta_j|_{\theta=\theta^0})\quad\leq\quad C_{v_2}(\mathbb{E}(f^2_{\theta^0}(X)),\sigma^2_{\xi,2},f^{(1)}_{\theta^0},f_{\theta^0},w_{\theta^0},p_\varepsilon)\max[1,C_n^{2\alpha-2a+1}]/n$$

It follows that if $r=0$ and $a<\alpha+1/2$ then which implies that

$$C_n^*=n^{1/(2\alpha)}\quad\text{and}\quad\varphi_n^2=O(n^{(1-2a)/(2\alpha)}).\tag{6.15}$$

If $r=0$, and $a\geq\alpha+1/2$, then

$$C_n^*=n^{1/(2a-1)}\quad\text{and}\quad\varphi_n^2=O(n^{-1}).$$

• If for $j=1,\cdots,m$, $\partial(f_\theta w_\theta)/\partial\theta_j|_{\theta=\theta^0}$, $\partial w_\theta/\partial\theta_j|_{\theta=\theta^0}$ and $\partial(f^2_\theta w_\theta)/\partial\theta_j|_{\theta=\theta^0}$ satisfy $(\mathbf{R_1})$ with $r>0$, then

$$|\mathbb{E}[\partial S_{n,1}(\theta)/\partial\theta_j|_{\theta=\theta^0}]|^2\quad\leq\quad C_{b_2}(f_{\theta^0},\sigma^2_{\xi,2},f^{(1)}_{\theta^0},w_{\theta^0})C_n^{-2a+(1-r)+(1-r)_-}\exp\{-2bC_n^r\},$$

and

$$\mathrm{Var}(\partial S_{n,1}(\theta)/\partial\theta_j|_{\theta=\theta^0})\quad\leq\quad C_{v_2}(f_{\theta^0},\sigma^2_{\xi,2},f^{(1)}_{\theta^0},w_{\theta^0},p_\varepsilon)/n$$

which implies that

$$C_n^*=\left[\frac{\log n}{2b}+\frac{-2a+(1-r)+(1-r)_-}{2br}\log\left(\frac{\log n}{2b}\right)\right]^{1/r}\quad\text{and}\quad\varphi_n^2=O(n^{-1}).\tag{6.16}$$

Exponential noise (see $(\mathbf{N_2})$ with $\rho > 0$)

• If for $j = 1, \cdots, m$, $\partial(f_\theta w_\theta)/\partial\theta_j|_{\theta=\theta^0}$, $\partial w_\theta/\partial\theta_j|_{\theta=\theta^0}$ and $\partial(f_\theta^2 w_\theta)/\partial\theta_j|_{\theta=\theta^0}$ satisfy $(\mathbf{R_1})$ with $r = 0$, then

$$|\mathbb{E}[\partial S_{n,1}(\theta)/\partial\theta_j|_{\theta=\theta^0}]|^2 \quad \leq \quad C_{b_2}(f_{\theta^0}, \sigma^2_{\xi,2}, f^{(1)}_{\theta^0}, w_{\theta^0})C_n^{-2a+1},$$

and

$$\mathrm{Var}(\partial S_{n,1}(\theta)/\partial\theta_j|_{\theta=\theta^0}) \quad \leq \quad C_{v_2}(f_{\theta^0}, \sigma^2_{\xi,2}, f^{(1)}_{\theta^0}, w_{\theta^0}, p_\varepsilon)C_n^{2\alpha-2a+(1-\rho)+(1-\rho)_-}\exp\{2\beta C_n^\rho\}/n$$

which implies that

$$C_n^* = \left[\frac{\log n}{2\beta} - \frac{2\alpha + (1-\rho)_-}{2\rho\beta}\log\left(\frac{\log n)}{2\beta}\right)\right]^{1/\rho} \quad \text{and} \quad \varphi_n^2 = O[(\log n/(2\beta))^{(1-2a)/\rho}]. \tag{6.17}$$

• If for $j = 1, \cdots, m$, $\partial(f_\theta w_\theta)/\partial\theta_j|_{\theta=\theta^0}$, $\partial w_\theta/\partial\theta_j|_{\theta=\theta^0}$ and $\partial(f_\theta^2 w_\theta)/\partial\theta_j|_{\theta=\theta^0}$ satisfy $(\mathbf{R_1})$ with $r > 0$ and $\{r > \rho\}$ or $\{r = \rho$ and $b > \beta\}$ or $\{r = \rho,\ b = \beta$ and $a \geq \alpha + 1/2\}$ then

$$|\mathbb{E}[\partial S_{n,1}(\theta)/\partial\theta_j|_{\theta=\theta^0}]|^2 \leq C_{b_2}(\mathbb{E}(f^2_{\theta^0}(X)), \sigma^2_{\xi,2}, L(f^{(1)}_{\theta^0}w_{\theta^0}), L(f_{\theta^0}f^{(1)}_{\theta^0}w_{\theta^0}))$$
$$\times\, C_n^{-2a+(1-r)+(1-r)_-}\exp\{-2bC_n^r\},$$

and

$$\mathrm{Var}(\partial S_{n,1}(\theta)/\partial\theta_j|_{\theta=\theta^0}) \leq C_{v_2}(\mathbb{E}(f^2_{\theta^0}(X)), \sigma^2_{\xi,2}, L(f^{(1)}_{\theta^0}w_{\theta^0}), L(f_{\theta^0}f^{(1)}_{\theta^0}w_{\theta^0}))/n$$

which implies that

$$C_n^* = \left[\frac{\log n}{2b} + \frac{-2a + (1-r) + (1-r)_-}{2br}\log\left(\frac{\log n}{2b}\right)\right]^{1/r} \quad \text{and} \quad \varphi_n^2 = O(n^{-1}). \tag{6.18}$$

• If for $j = 1, \cdots, m$, $\partial(f_\theta w_\theta)/\partial\theta_j|_{\theta=\theta^0}$, $\partial w_\theta/\partial\theta_j|_{\theta=\theta^0}$ and $\partial(f_\theta^2 w_\theta)/\partial\theta_j|_{\theta=\theta^0}$ satisfy $(\mathbf{R_1})$ with $r > 0$, and $\{r \leq \rho$ or $\{r = \rho$ and $b < \beta\}$ then

$$|\mathbb{E}[\partial S_{n,1}(\theta)/\partial\theta_j|_{\theta=\theta^0}]|^2 \leq C_b(\mathbb{E}(f^2_{\theta^0}(X)), \sigma^2_{\xi,2}, L(f^{(1)}_{\theta^0}w_{\theta^0}), L(f_{\theta^0}f^{(1)}_{\theta^0}w_{\theta^0}))$$
$$\times\, C_n^{-2a+(1-r)+(1-r)_-}\exp\{-2bC_n^r\},$$

and

$$\mathrm{Var}(\partial S_{n,1}(\theta)/\partial\theta_j|_{\theta=\theta^0}) \leq C_{v_2}(\mathbb{E}(f^2_{\theta^0}(X)), \sigma^2_{\xi,2}, L(f^{(1)}_{\theta^0}w_{\theta^0}), L(f_{\theta^0}f^{(1)}_{\theta^0}w_{\theta^0}))$$
$$\times\, C_n^{2\alpha-2a+(1-\rho)+(1-\rho)_-}\exp\{-2bC_n^r + 2\beta C_n^\rho\}/n$$

which implies that the bias is by a logarithmic factor larger than the variance

$$\begin{aligned} C_n^* &= \left[\frac{\log n}{2\beta} - \frac{2\alpha + (1-\rho)_- - (1-r)_-}{2\rho\beta}\log\left(\frac{\log n)}{2\beta}\right)\right]^{1/\rho} \\ \text{and } \varphi_n^2 &= O\left[(\log n)^{\frac{-2a+(1-r)+(1-r)_-}{\rho}}\exp\left\{-2b\left(\frac{\log n}{2\beta}\right)^{r/\rho}\right\}\right]. \end{aligned}$$

If $r = \rho$, $b = \beta$ and $a < \alpha + 1/2$ then

$$\mathrm{Var}(\partial S_{n,1}(\theta)/\partial\theta_j|_{\theta=\theta^0}) \quad \leq \quad C_{v_2}(\mathbb{E}(f^2_{\theta^0}(X)), \sigma^2_{\xi,2}, L(f^{(1)}_{\theta^0}w_{\theta^0}), L(f_{\theta^0}f^{(1)}_{\theta^0}w_{\theta^0}))C_n^{2\alpha-2a+1}/n$$

giving the rate $\varphi_n^2 = O[(\log n)^{\frac{2\alpha-2a+1}{r}}/n]$.

Proof of ii)

By using that

(6.19)
$$\left(S_{n,1}^{(2)}(\theta)\right)_{j,k} = -\frac{2}{n}\sum_{i=1}^{n}\left(-2Y_i\frac{\partial^2(f_\theta w_\theta)}{\partial\theta_j\partial\theta_k} + \frac{\partial^2(f_\theta^2 w_\theta)}{\partial\theta_j\partial\theta_k} + (Y_i^2-\sigma_{\xi,2}^2)\frac{\partial^2 w_\theta}{\partial\theta_j\partial\theta_k}\right)|_{\theta=\theta^0}\star K_{n,C_n}(Z_i),$$

we write $S_{n,1}^{(2)}(\theta^0) - S_{\theta^0,g}(\theta^0) = A_1 + A_2 + A_3$ with

$$\begin{aligned}
A_1 &= -\frac{2}{n}\sum_{i=1}^{n} Y_i\left(\frac{\partial^2(f_\theta w_\theta)}{\partial\theta_j\partial\theta_k}|_{\theta=\theta^0}\right)\star K_{n,C_n}(Z_i) - \mathbb{E}\left[Y_i\left(\frac{\partial^2(f_\theta w_\theta)}{\partial\theta_j\partial\theta_k}(X)|_{\theta=\theta^0}\right)\right]\\
A_2 &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial^2(f_\theta^2 w_\theta)}{\partial\theta_j\partial\theta_k}|_{\theta=\theta^0}\right)\star K_{n,C_n}(Z_i) - \mathbb{E}\left[\left(\frac{\partial^2(f_\theta^2 w_\theta)}{\partial\theta_j\partial\theta_k}(X)|_{\theta=\theta^0}\right)\right]\\
A_3 &= \frac{1}{n}\sum_{i=1}^{n}(Y_i^2-\sigma_{\xi,2}^2)\left(\frac{\partial^2 w_\theta}{\partial\theta_j\partial\theta_k}|_{\theta=\theta^0}\right)\star K_{n,C_n}(Z_i) - \mathbb{E}\left[(Y_i^2-\sigma_{\xi,2}^2)\left(\frac{\partial^2 w_\theta}{\partial\theta_j\partial\theta_k}(X)|_{\theta=\theta^0}\right)\right].
\end{aligned}$$

As soon as $w_\theta f_\theta$ and $w_\theta f_\theta^2$ and their derivatives up to order 2, satisfy $(\mathbf{R_1})$, then for C_n satisfying that (2.6), we get that $A_j = o_p(1)$ as $n\to\infty$ for $j = 1,\cdots,3$, and **ii)** is proved.

Proof of iii)

Again using (6.19), the smoothness properties of the derivatives of $w_\theta f_\theta$ and $w_\theta f_\theta^2$ up to order 3 and the consistency of $\widehat{\theta}_1$, we get that $\| R_{n,1}\|_{\ell^2} = o_p(1)$ as $n\to\infty$.

Proof of iv)

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

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

According to (6.10) and (6.9) we have

$$\begin{aligned}
\mathbb{E}\|\widehat{\theta}_1-\theta^0\|_{\ell^2}^2 &= \mathbb{E}[\|\widehat{\theta}_1-\theta^0\|_{\ell^2}^2\mathbb{1}_{E_n}] + \mathbb{E}[\|\widehat{\theta}_1-\theta^0\|_{\ell^2}^2\mathbb{1}_{E_n^c}]\\
&\le \mathbb{E}[\|\widehat{\theta}_1-\theta^0\|_{\ell^2}^2\mathbb{1}_{E_n}] + 2\sup_{\theta\in\Theta}\|\theta\|_{\ell^2}^2\mathbb{P}_{\theta^0,g}(E_n^c)\\
&\le \mathbb{E}\left[(S_{n,1}^{(1)}(\theta^0))^\top[(S_{n,1}^{(2)}(\theta^0)+R_{n,1})^{-1}]^\top(S_{n,1}^{(2)}(\theta^0)+R_{n,1})^{-1}S_{n,1}^{(1)}(\theta^0)\mathbb{1}_{E_n}\right]\\
&\quad +2\sup_{\theta\in\Theta}\|\theta\|_{\ell^2}^2\mathbb{P}_{\theta^0,g}(E_n^c)\\
&\le C4^{2m^2}\sup_{j,k}\left|\frac{\partial^2 S_{\theta^0,g}(\theta)}{\partial\theta_j\partial\theta_k}|_{\theta=\theta^0}\right|^{-2}\mathbb{E}\left[(S_{n,1}^{(1)}(\theta^0))^\top S_{n,1}^{(1)}(\theta^0)\right] + 2\sup_{\theta\in\Theta}\|\theta\|_{\ell^2}^2\mathbb{P}_{\theta^0,g}(E_n^c)\\
&\le C4^{2m^2}\sup_{j,k}\left|\frac{\partial^2 S_{\theta^0,g}(\theta)}{\partial\theta_j\partial\theta_k}|_{\theta=\theta^0}\right|^{-2}\varphi_n^2 + 2\sup_{\theta\in\Theta}\|\theta\|_{\ell^2}^2\mathbb{P}_{\theta^0,g}(E_n^c).
\end{aligned}$$

It remains thus to show that $\mathbb{P}_{\theta^0,g}(E_n^c) = o(\varphi_n^2)$ with

$$\sup_{j,k} \mathbb{E}\left[\left(\frac{\partial^2(S_{n,1}(\theta) - S_{\theta^0,g}(\theta))}{\partial\theta_j\partial\theta_k}\mid_{\theta=\theta^0}\right)^2\right] \leq \varphi_n^2.$$

By Markov's inequality, for some $p > 2$,

$$\mathbb{P}_{\theta^0,g}(E_n^c) \quad \leq \quad \sum_{j=1}^m\sum_{k=1}^m \mathbb{P}_{\theta^0,g}(E_{n,j,k}^c),$$

with

$$\begin{aligned}
\mathbb{P}_{\theta^0,g}(E_{n,j,k}^c) \quad &\leq \quad \mathbb{E}\left[\left|\left(\frac{\partial^2(S_{\theta^0,g}(\theta) - S_{n,1}(\theta))}{\partial\theta_j\partial\theta_k}\mid_{\theta=\theta^0}\right) + (R_{n,1})_{j,k}\right|^p\right] \\
&\leq \quad 2^{p-1}\left|\left(\frac{\partial^2 S_{\theta^0,g}(\theta)}{\partial\theta_j\partial\theta_k}\mid_{\theta=\theta^0}\right) - \mathbb{E}\left[\left(\frac{\partial^2 S_{n,1}(\theta)}{\partial\theta_j\partial\theta_k}\mid_{\theta=\theta^0}\right)\right]\right|^p \\
&\quad +2^{p-1}\mathbb{E}\left[\left|\mathbb{E}\left(\frac{\partial^2 S_{n,1}(\theta)}{\partial\theta_j\partial\theta_k}\mid_{\theta=\theta^0}\right) - \left(\frac{\partial^2 S_{n,1}(\theta)}{\partial\theta_j\partial\theta_k}\mid_{\theta=\theta^0}\right) + (R_{n,1})_{j,k}\right|^p\right] \\
&\leq \quad 2^{p-1}\left|\left(\frac{\partial^2 S_{\theta^0,g}(\theta)}{\partial\theta_j\partial\theta_k}\mid_{\theta=\theta^0}\right) - \mathbb{E}\left[\left(\frac{\partial^2 S_{n,1}(\theta)}{\partial\theta_j\partial\theta_k}\mid_{\theta=\theta^0}\right)\right]\right|^p \\
&\quad +2^{2p-2}\left\{\mathbb{E}\left[\left|\mathbb{E}\left(\frac{\partial^2 S_{n,1}(\theta)}{\partial\theta_j\partial\theta_k}\mid_{\theta=\theta^0}\right) - \left(\frac{\partial^2 S_{n,1}(\theta)}{\partial\theta_j\partial\theta_k}\mid_{\theta=\theta^0}\right)\right|^p\right] + 2^{2p-2}\mathbb{E}|(R_{n,1})_{j,k}|^p\right\}.
\end{aligned}$$

Now we apply the Rosenthal's inequality to the sum of centered variables

$$\left(\frac{\partial^2 S_{n,1}(\theta)}{\partial\theta_j\partial\theta_k}\mid_{\theta=\theta^0}\right) - \mathbb{E}\left[\left(\frac{\partial^2 S_{n,1}(\theta)}{\partial\theta_j\partial\theta_k}\mid_{\theta=\theta^0}\right)\right] \quad = \quad n^{-1}\sum_{i=1}^n W_{n,i,j,k}$$

where $W_{n,i,j,k}$ equals

$$\begin{aligned}
&\left(\frac{\partial^2[-2Y_i f_\theta w_\theta + f_\theta^2 w_\theta + (Y_i^2 - \sigma_{\xi,2}^2)w_\theta]}{\partial\theta_j\partial\theta_k}\mid_{\theta=\theta^0}\right) \star K_{n,C_n}(Z_i) \\
&-2\mathbb{E}\left[\left(\frac{\partial^2[-2Y_i f_\theta w_\theta + f_\theta^2 w_\theta + (Y_i^2 - \sigma_{\xi,2}^2)w_\theta]}{\partial\theta_j\partial\theta_k}\mid_{\theta=\theta^0}\right) \star K_{n,C_n}(Z_i)\right].
\end{aligned}$$

Lemma 6.1. ***Rosenthal's inequality*** *(Rosenthal (1970), Petrov (1995)). For $U_1, \ldots, U_n$, be n independent centered random variables, there exists a constant $C(p)$ such that for $p \geq 1$,*

$$\mathbb{E}[|\sum_{i=1}^n U_i|^p] \leq C(p)[\sum_{i=1}^n \mathbb{E}[|U_i|^p] + (\sum_{i=1}^n \mathbb{E}[U_i^2])^{p/2}]. \tag{6.20}$$

Consequently

$$\mathbb{E}\left[\left|\left(\frac{\partial^2 S_{n,1}(\theta)}{\partial\theta_j\partial\theta_k}\mid_{\theta=\theta^0}\right)-\mathbb{E}\left(\left(\frac{\partial^2 S_{n,1}(\theta)}{\partial\theta_j\partial\theta_k}\mid_{\theta=\theta^0}\right)\right)\right|^p\right] \leq C(p)\left[n^{1-p}\mathbb{E}|W_{n,1,j,k}|^p+n^{-p/2}\mathbb{E}^{p/2}|W_{n,1,j,k}|^2\right].$$

Take $p=4$ to get that

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

Therefore under the conditions ensuring that $(\mathbb{E}[S^{(2)}_{\theta^0,g}(\theta^0)-S^{(2)}_{n,1}(\theta^0)]^2)_{j,k}=o(1)$ we have

$$(\mathbb{E}[(S^{(2)}_{n,1}(\theta^0)-S^{(2)}_{\theta^0,g}(\theta^0)]^4)_{j,k}=O(\varphi_n^4)=o(\varphi_n^2).$$

Now, by using the definition of $R_{n,1}$ combined with (6.19), and the smoothness properties of the derivatives of $(w_\theta f_\theta)$ and $(w_\theta f_\theta^2)$ up to order 3, we get that $\mathbb{E}((R_{n,1})^4_{j,k})=o(\|\widehat{\theta}_1-\theta^0\|^4_{\ell^2})$, and we conclude that

$$\mathbb{E}\|\widehat{\theta}_1-\theta^0\|^2_{\ell^2}\leq 4\mathbb{E}\left[(S^{(1)}_{n,1}(\theta^0))^\top[(S^{(2)}_{\theta^0,g}(\theta^0))^{-1}]^\top(S^{(2)}_{\theta^0,g}(\theta^0))^{-1}S^{(1)}_{n,1}(\theta^0)\right]+o(\varphi_n^2)+o(\mathbb{E}[\|\widehat{\theta}_1-\theta^0\|^4_{\ell^2}]).$$

□

6.3. **Proof of Theorem 3.3:** As for the proof of Theorem 3.1, the main point of the proof consists in showing that for any $\theta\in\Theta$, $\mathbb{E}[(S_{n,2}(\theta)-S_{\theta^0,g}(\theta))^2]=o(1)$, with $S_{\theta^0,g}(\theta)$ admitting a unique minimum in $\theta=\theta^0$. The second part of the proof consists in studying $\omega_1(n,\rho)$ defined as $\omega_1(n,\rho)=\sup\{|S_{n,2}(\theta)-S_{n,2}(\theta')|:\|\theta-\theta'\|_2\leq\rho\}$. By using the regularity assumptions on the regression function f, we state that there exists two sequences ρ_k and ϵ_k tending to 0, such that for all $k\in\mathbb{N}$,

$$\lim_{n\to\infty}\mathbb{P}[\omega_1(n,\rho_k)>\epsilon_k]=0 \text{ and that } \mathbb{E}[(\omega_1(n,\rho_k))^2]=O(\rho_k^2). \tag{6.21}$$

Under the conditions $(\mathbf{C_{11}})$-$(\mathbf{C_{13}})$ and by applying the Law of Large numbers, we get that for any $\theta\in\Theta$, $\mathbb{E}[(S_{n,2}(\theta)-S_{\theta^0,g}(\theta))^2]=o(1)$ as $n\to\infty$ and $\mathbb{E}[(S_{n,2}(\theta)-S_{n,2}(\theta'))^2]=o(1)$ and consequently $\widehat{\theta}_2$ is a consistent estimator of θ^0. By using classical Taylor expansion based on the smoothness properties of the regression function, with respect to θ and the consistency of $\widehat{\theta}_2$, we get that $0=S^{(1)}_{n,2}(\widehat{\theta}_1)=S^{(1)}_{n,2}(\theta^0)+S^{(2)}_{n,2}(\theta^0)(\widehat{\theta}_2-\theta^0)+R_{n,2}(\widehat{\theta}_2-\theta^0)$, with $R_{n,2}$ defined by

$$R_{n,2}=\int_0^1[S^{(2)}_{n,2}(\theta^0+s(\widehat{\theta}_2-\theta^0))\widehat{\theta}-S^{(2)}_{n,2}(\theta^0)]ds. \tag{6.22}$$

This implies that

$$\widehat{\theta}_2-\theta^0=-[S^{(2)}_{n,2}(\theta^0)+R_{n,2}]^{-1}S^{(1)}_{n,2}(\theta^0), \tag{6.23}$$

where $S^{(1)}_{n,2}(\theta)$ and $S^{(2)}_{n,2}(\theta)$ denote the first and second derivatives of $S_{n,2}(\theta)$ with respect to θ. The $\sqrt{n}$-consistency follows by applying the Central limit theorem to $S^{(1)}_{n,2}(\theta^0)$ to get that

$$\sqrt{n}S^{(1)}_{n,2}(\theta^0)\xrightarrow[n\to\infty]{\mathcal{L}}\mathcal{N}(0,\Sigma_{0,2}) \tag{6.24}$$

with

$$\begin{aligned}\Sigma_{0,2} &= \mathbb{E}\Big[\Big(\Phi^{(1)}_{\theta^0,\varepsilon,1}(Z) - 2Y\Phi^{(1)}_{\theta^0,\varepsilon,2}(Z) + (Y^2-\sigma^2_{\xi,2})\Phi^{(1)}_{\theta^0,\varepsilon,3}(Z)\Big) \\ &\quad \times \Big(\Phi^{(1)}_{\theta^0,\varepsilon,1}(Z) - 2Y\Phi^{(1)}_{\theta^0,\varepsilon,2}(Z) + (Y^2-\sigma^2_{\xi,2})\Phi^{(1)}_{\theta^0,\varepsilon,3}(Z)\Big)^\top\Big]. \qquad \square\end{aligned}$$

Appendix : technical lemmas

Lemma 6.1. *Let φ a function such that φ belongs to $\mathbb{L}_2(\mathbb{R})$ satisfying ($\mathbf{R_1}$). Then*

$$\int_{|u|\geq C_n} |\varphi^*(u)|du \leq \frac{L(\varphi)}{R(a,b,r)} C_n^{-a+1-r}\exp\{-bC_n^r\}. \tag{6.1}$$

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

$$\int_{|u|\leq C_n} \frac{|\varphi^*(u)|}{|p^*_\varepsilon(u)|}du \leq \frac{L(\varphi)}{R(\alpha,\beta,\rho,a,b,r)\underline{C}(p_\varepsilon)}\max[1, C_n^{\alpha-a+1-\rho}\exp\{-bC_n^r+\beta C_n^\rho\}].$$

Lemma 6.2. *Let φ such that $\mathbb{E}(|\varphi(Y,X)|)$ is finite and let Φ such that $\mathbb{E}(|\Phi(U)|)$ is finite. Then*

$$\mathbb{E}[\varphi(Y,X)\Phi\star K_{n,C_n}(U)] = \mathbb{E}[\varphi(Y,X)\Phi\star K_{C_n}(X)] = \langle \varphi(y,\cdot)f_{Y,X}(y,\cdot), \Phi\star K_{C_n}\rangle$$

and

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

Proof of Lemma 6.2 : If we denote by $f_{Y,Z,X}$ the joint distribution of (Y,Z,X), then

$$f_{Y,Z,X}(y,z,x) = f_{Y,X}(y,x)f_\varepsilon(z-x) \text{ and } f_{Y,Z}(y,z) = f_{Y,X}(y,\cdot)\star f_\varepsilon(z).$$

Since $K_{n,C_n}\star p_\varepsilon = K_{C_n}$ by construction, we get by the Parseval's formula that

$$\begin{aligned}\mathbb{E}\left[\varphi(Y,X)\Phi\star K_{n,C_n}(Z)\right] &= \iiint \varphi(y,x)\Phi\star K_{n,C_n}(z)f_{Y,Z,X}(y,z,x)dy\,dz\,dx \\ &= \iiint \varphi(y,x)\Phi\star K_{n,C_n}(z)f_{Y,X}(y,x)f_\varepsilon(z-x)dz\,dw\,dx \\ &= \iint \varphi(y,x)f_{Y,X}(y,x)\int \Phi\star K_{n,C_n}(z)f_\varepsilon(z-x)dz\,dy\,dx \\ &= (2\pi)^{-1}\iint \varphi(y,x)f_{Y,X}(y,x)\int \Phi^*(u)K^*_{n,C_n}(u)f^*_\varepsilon(u)e^{-iux}du\,dy\,dx \\ &= (2\pi)^{-1}\iint \varphi(y,x)f_{Y,X}(y,x)\int \Phi^*(u)\frac{K^*_{C_n}(u)}{f^*_\varepsilon(u)}f^*_\varepsilon(u)e^{-iux}du\,dy\,dx \\ &= (2\pi)^{-1}\iint \varphi(y,x)f_{Y,X}(y,x)\int \Phi^*(u)K^*_{C_n}(u)e^{-iux}du\,dy\,dx \\ &= \iint \varphi(y,x)f_{Y,X}(y,x)\int \Phi(z)K_{C_n}(x-z)dz\,dy\,dx \\ &= \iint \varphi(y,x)f_{Y,X}(y,x)\Phi\star K_{C_n}(x)\,dy\,dx.\end{aligned}$$

The second equality follows by writing that

$$\begin{aligned}
\mathbb{E}\left[\varphi(Y,X)\Phi\star K_{n,C_n}(Z)\right]^2 &= \iiint \varphi(y,x)(\Phi\star K_{n,C_n}(z))^2 f_{Y,Z,X}(y,z,x)dy\,dz\,dx \\
&= \iiint \varphi(y,x)(\Phi\star K_{n,C_n}(z))^2 f_{Y,X}(y,x)f_\varepsilon(z-x)dy\,dz\,dx \\
&= \iiint \varphi(y,x)(\Phi\star K_{n,C_n}(z))^2 f_{Y,X}(y,x)f_\varepsilon(z-x)dy\,dz\,dx \\
&= \int \left\langle (\varphi(y,\cdot)f_{Y,X}(y,\cdot))\star f_\varepsilon, (\Phi\star K_{n,C_n})^2\right\rangle dy. \qquad \square
\end{aligned}$$

References

Baran, S. (2000). A consistent estimator in general functional errors-in-variables models. *Metrika 51*(2), 117–132 (electronic).

Bickel, P. J., Klaassen, A. J. C., Ritov, Y., and Wellner, J. A. (1993). *Efficient and adaptative estimation for semiparametric model*. Johns Hopkins Series.

Bickel, P. J. and Ritov, A. J. C. (1987). Efficient estimation in the errors-in-variables model. *Ann. Statist. 15*, 513–540.

Carroll, R. J., Ruppert, D., and Stefanski, L. A. (1995). *Measurement error in Nonlinear Models*. Chapman and Hall, London.

Chan, L. K. and Mak, T. K. (1985). On the polynomial functionnal relationship. *J. R. Statist. Soc. B 47*, 510–518.

Cheng, C. H. and Van Ness, J. W. (1994). On estimating linear relationships when both variables are subject to errors. *J.R. Statist. Soc. B 56*, 167–183.

Comte, F. and Taupin, M.-L. (2001). Semiparametric estimation in the (auto)-regressive β-mixing model with errors-in-variables. *Math. Methods Statist. 10*(2), 121–160.

Fazekas, I., Baran, S., Kukush, A., and Lauridsen, J. (1999). Asymptotic properties in space and time of an estimator in nonlinear functional errors-in-variables models. *Random Oper. Stochastic Equations 7*(4), 389–412.

Fazekas, I. and Kukush, A. G. (1998). Asymptotic properties of estimators in nonlinear functional errors-in-variables with dependent error terms. In *Proceedings of the 18th Seminar on Stability Problems for Stochastic Models, Part I (Hajdúszoboszló, 1997)*, Volume 92, pp. 3890–3895.

Fedoryuk, M. V. (1987). *Asimptotika: integraly i ryady*. Spravochnaya Matematicheskaya Biblioteka. [Mathematical Reference Library]. Moscow: "Nauka".

Fuller, W. A. (1987). *Measurement Error Models*. Wiley.

Gleser, L. J. (1990). Improvements of the naive approach to estimation in nonlinear errors-in-variables regression models. *Contemp. Math. 112*, 99–114.

Hausman, J. A., Newey, W. K., Ichimura, I., and Powell, J. L. . (1991). Identification and estimation of polynomial errors-in-variables models. *J. Econometrics 50*, 273–295.

Hausman, J. A., Newey, W. K., and Powell, J. L. . (1995). Nonlinear errors in variables estimation of some engel curves. *J. Econometrics 65*, 205–233.

Hong, H. and Tamer, E. (2003). A simple estimator for nonlinear error in variable models. *J. Econometrics 117*(1), 1–19.

Hsiao, C. (1989). Consistent estimation for some nonlinear errors-in-variables models. *Journal of Econometrics 41*, 159–185.

Hsiao, C., Wang, L., and Wang, Q. (1997). Estimation of nonlinear errors-in-variables models: an approximate solution. *Statist. Papers 38*(1), 1–25.

Hsiao, C. and Wang, Q. K. (2000). Estimation of structural nonlinear errors-in-variables models by simulated least-squares method. *Internat. Econom. Rev. 41*(2), 523–542.

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

Kukush, A. and Schneeweiss, H. (2005a). Comparing different estimators in a nonlinear measurement error model. I. *Math. Methods Statist. 14*(1), 53–79.

Kukush, A. and Schneeweiss, H. (2005b). Comparing different estimators in a nonlinear measurement error model. II. *Math. Methods Statist. 14*(2), 203–223.

Lepski, O. V. and Levit, B. Y. (1998). Adaptive minimax estimation of infinitely differentiable functions. *Math. Methods Statist. 7*(2), 123–156.

Li, T. (2000). Estimation of nonlinear errors-in-variables models: a simulated minimum distance estimator. *Statist. Probab. Lett. 47*(3), 243–248.

Li, T. (2002). Robust and consistent estimation of nonlinear errors-in-variables models. *J. Econometrics 110*(1), 1–26.

Murphy, S. A. and Van der Vaart, A. W. (1996). Likelihood inference in the errors-in-variables model. *Journal of Multivariate Analysis 59*, 81–108.

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

Taupin, M.-L. (2001). Semi-parametric estimation in the nonlinear structural errors-in-variables model. *Ann. Statist. 29*(1), 66–93.

van der Vaart, A. (2002). Semiparametric statistics. In *Lectures on probability theory and statistics (Saint-Flour, 1999)*, Volume 1781 of *Lecture Notes in Math.*, pp. 331–457. Berlin: Springer.

Van der Vaart, A. W. (1988). Estimating a real parameter in a class of semiparametric models. *Ann. Statist. 16*, 1450–1474.

Van der Vaart, A. W. (1996). Efficient estimation in semi-parametric mixture models. *Ann. Statist. 24*, 862–878.

Wolter, K. M. and Fuller, W. A. (1982a). Estimation of nonlinear errors-in variables models. *Ann. Statist. 10*, 539–548.

Wolter, K. M. and Fuller, W. A. (1982b). Estimation of the quadratic errors-in-variables model. *Ann. Statist. 69*, 175–182.