The Annals of Statistics
2001, Vol. 29, No. 1, 66–93

SEMI-PARAMETRIC ESTIMATION IN THE NONLINEAR STRUCTURAL ERRORS-IN-VARIABLES MODEL

BY MARIE-LUCE TAUPIN

Université Paris-Sud

In the nonlinear structural errors-in-variables model, we propose a consistent estimator of the unknown parameter using a modified least squares criterion. We give an upper bound of its rate of convergence which is strongly related to the regularity of the regression function and is generally slower than the parametric rate of convergence $n^{-1/2}$. Nevertheless, the rate is of order $n^{-1/2}$ for some particular analytic regression functions. For instance, when the regression function is either a polynomial function or an exponential function, we prove that our estimator achieves the parametric rate of convergence.

1. Introduction. Let $(Y_1, Z_1), \dots, (Y_n, Z_n)$ be n independent observations and consider the following model :

$$\begin{cases} Y_i = f_{\beta^0}(X_i) + \xi_i, \\ Z_i = X_i + \varepsilon_i, \end{cases}$$

where the regression function f_{β^0} is known up to a finite dimensional parameter β^0 belonging to $\overset{\circ}{\Theta}$, Θ being a compact subset of R^m. The X_i's are unobservable i.i.d. univariate random variables with an unknown density g with respect to the Lebesgue measure. The errors (ξ_i, ε_i) are also unobservable and are i.i.d. centered Gaussian random variables, independent of the X_i's. Furthermore we assume that the variance of the ε_i's is known and equal to one and we denote by σ^2 the variance of the ξ_i's.

In this semiparametric model, our aim is to estimate the parameter β^0 in the presence of the nuisance parameter g belonging to a functional space.

In the linear case (i.e., $f_{\beta^0}(X_i) = \beta_1^0 + \beta_2^0 X_i$) the first results have been written in the 1950's (see [32], [27]). In such a case $\sqrt{n}$-consistent (see [18], [19]) and efficient estimators have also been constructed (see [3], [36], [2] and [37]). However, the methods used for the construction of the information bound and for the construction of the estimators in the linear case do not seem to extend to the nonlinear case.

The nonlinear case has already been studied by several authors but generally under different or more restrictive assumptions.

Let us start with the functional model, when the unobservable X_i are fixed unknown constants. Wolter and Fuller [38] consider the situation where the error variances tend to zero as n tends to infinity and showed that in this case

Received May 22, 1998.

AMS 2000 *subject classifications*. Primary 62J02, 62F12; secondary 62G05, 62G20.

Key words and phrases. Semi-parametric estimation, analytic function, Fourier transform, errors-in-variables model, nonparametric estimation.

the maximum likelihood estimator is consistent and that its rate depends on the decrease of the error variances to zero. Zwanzig [40] proved that under an entropy condition on the set of nuisance parameters $X_1, \dots, X_n$, the least squares estimator defined by

$$\overline{\beta} = \operatorname{Arg}\min_{\beta\in\Theta} \min_{X_1,\dots,X_n} \frac{1}{n}\sum_{i=1}^{n} \left[(Y_i - f_\beta(X_i))^2 + (Z_i - X_i)^2\right],$$

is consistent with an exponential rate of convergence. Kukush and Zwanzig [28] proved that without assumptions on the decrease of the error variances or without entropy conditions, the least squares estimator defined above is not consistent. Chan and Mak [8] proposed a $\sqrt{n}$-consistent estimator when the function is a polynomial function and the error variances are known.

The model we consider in this paper is the structural model, when the unobservable X_i are i.i.d. random variables with an unknown density g. A natural idea would be to replace the unobservable X_i by the observations Z_i in the usual least squares criterion and to minimize

$$\sum_{i=1}^{n}[Y_i - f_\beta(Z_i)]^2.$$

Unfortunately this estimator is known to be inconsistent even in the linear case. Gleser [20] improved this naive approach by considering the modified least squares criterion

$$\sum_{i=1}^{n}[Y_i - f_\beta(\widetilde{Z}_i)]^2 \qquad \text{where} \qquad \widetilde{Z}_i = \widehat{\Lambda}Z_i + \widehat{\mu}(1-\widehat{\Lambda}),$$

with μ denoting the expectation of X and $\Lambda = \operatorname{Var}(X)/[\operatorname{Var}(X)+1]$. The quantity $\Lambda Z + \mu(1-\Lambda)$ is a linear approximation to $\mathbb{E}(X|Z)$ which equals $\mathbb{E}(X|Z)$ when (X, Z) is normally distributed and $\widehat{\Lambda}$ and $\widehat{\mu}$ are consistent estimators of Λ and μ. This method provides a consistent estimator in the linear case but not in general.

Several other results have been published in the structural case, but under more restrictive assumptions. For instance, Hsiao [24] considered the case where the density g of the X_i's is known up to a finite-dimensional parameter. He proved that under identifiability conditions the estimator obtained by minimizing

$$\sum_{i=1}^{n}[Y_i - \mathbb{E}(f_\beta(X_i)|Z_i)]^2$$

is consistent and asymptotically Gaussian. Other authors studied estimators assuming the presence of additional observations (see [22], [23], [7] and [34]). Moreover [10] and [5] proposed a numerical procedure based on simulations (SIMEX).

Fan and Truong [17], Fan and Masry [15] and Fan, Truong and Wang [16] treated the nonparametric case, when both the regression function and the density of the unobservable X_i are unknown, using methods based on deconvolution.

Now let us motivate our method and then introduce our estimation criterion. If g is known a natural estimation criterion is given by

$$\widetilde{S}_n(\beta, g) = n^{-1} \sum_{i=1}^{n} W(Z_i)[Y_i - \mathbb{E}(f_\beta(X_i)|Z_i)]^2, \tag{1.1}$$

where $W(.)$ is a deterministic weight function with compact support. We can easily prove that under suitable assumptions $\widetilde{\beta}_g$ defined as

$$\widetilde{\beta}_g = \operatorname*{Arg\,min}_{\beta \in \Theta} \widetilde{S}_n(\beta, g), \tag{1.2}$$

is a consistent estimator of β^0 and is asymptotically Gaussian. In the model (1.1), g is unknown and $\mathbb{E}[f_\beta(X_i)|Z_i]$ does not only depend on Z_i but also on the unknown density g. Hence $\widetilde{\beta}_g$ is obviously not an estimator, and our idea is simply to replace in $\widetilde{S}_n(\beta, g)$, the unknown conditional expectation $\mathbb{E}[f_\beta(X_i)|Z_i]$ by a nonparametric estimate based on the sample $Z_1, \dots, Z_n$ and then to estimate β^0 by minimizing this new criterion.

Hence we consider the modified least squares criterion

$$S_n(\beta) = n^{-1} \sum_{i=1}^{n} W(Z_i)[Y_i - \widehat{\Phi}_\beta(Z_i)]^2, \tag{1.3}$$

where $\widehat{\Phi}_\beta(Z_i)$ is an estimate of $\mathbb{E}[f_\beta(X_i)|Z_i]$, defined in Section 2, based on the sample $Z_1, \dots, Z_n$. The estimator $\widehat{\beta}$ of β^0 is then defined by

$$\widehat{\beta} = \operatorname*{Arg\,min}_{\beta \in \Theta} S_n(\beta). \tag{1.4}$$

Under suitable assumptions, the estimator $\widehat{\beta}$ is consistent and its rate of convergence is strongly related to the regularity of the function f_β as a function of X. We give an upper bound of the rate of $\widehat{\beta}$ for general f_β. This upper bound is not explicit and depends on the rate of convergence of the estimator of $\mathbb{E}[f_\beta(X_i)|Z_i]$. More precisely, $\mathbb{E}[f_\beta(X_i)|Z_i]$ is estimated by the ratio of two estimators defined in Section 2. The upper bound of the rate of convergence of $\widehat{\beta}$ depends on the rate of convergence of the numerator of $\widehat{\Phi}_\beta(Z_i)$. As we will see in Section 2, the more regular f_β (as a function of X), the faster the rate of convergence of this numerator. Consequently the same holds for the rate of convergence of $\widehat{\beta}$. This motivates us to evaluate the upper bound of the rate of $\widehat{\beta}$ when f_β admits an analytic continuation in the complex plane.

When the function f_β is such that precise calculations of this upper bound are possible, the rate of $\widehat{\beta}$ is often faster than the rate for estimating the conditional expectation. In particular we establish that the estimator $\widehat{\beta}$ achieves the parametric rate of convergence when f_β is a polynomial function and when f_β is an exponential function. Consequently, our method provides the parametric rate of convergence in the linear case and extends this result to polynomials with arbitrary degree.

In most cases, precise calculations of the rate of $\widehat{\beta}$ being impossible, we bound this rate from above by the rate of convergence of the numerator of $\widehat{\phi}_\beta(Z_i)$.

In Section 2 we construct the estimator of β^0. This construction is composed of two parts. First we build an estimator of the conditional expectation $\mathbb{E}[f_\beta(X)|Z]$ and second we plug this estimator in the criterion $\widetilde{S}_n(\beta, g)$ defined in (1.1). This provides the estimation criterion $S_n(\beta)$. Section 3 is devoted to the statement of the asymptotic properties of $\widehat{\beta}$. As in Section 2, we first give the asymptotic properties of the estimators appearing in the estimate of the conditional expectation and next we establish that $\widehat{\beta}$ is consistent and calculate an upper bound for its rate of convergence for any function f_β. In Section 4 we treat several examples, in particular the case f_β is a polynomial function. In Section 5 we study the variance estimation. Proofs of the theorems can be found in Section 6.

2. Construction of the estimator.

Write X, Z and ε for "generic" observations. Our purpose in the beginning of this section is to construct an estimator of the conditional expectation $\mathbb{E}[f_\beta(X)|Z]$, in order to get the estimation criterion defined in (1.3). We start with some notation.

NOTATION 1. We denote by $\|u\|_1 = \int |u(x)|\,dx$ the $L_1(R)$-norm of the function u, by $\langle u, v\rangle = \int u(x)\overline{v}(x)dx$ the inner product in $L_2(R)$ and by $u^\star$ the Fourier transform of u.

We introduce the functions, $\mathbb{I}_C(t) = \mathbb{I}_{|t|\le C}$, $\overline{\mathbb{I}}_C(t) = \mathbb{I}_{|t|\ge C}$, $T_{f_\beta,z}(x) = f_\beta(x)\eta(z-x)$ and $T^\star_{f_\beta,z}(t) = \int T_{f_\beta,z}(x)\exp\{itx\}\,dx$, where η denotes the standard Gaussian density (density of ε). Furthermore, we denote by $\|u\|_{\infty,W} = \sup_{x\in S_W}|u(x)|$ where S_W denotes the compact support of the weight function $W(.)$.

We assume subsequently that the functions $T_{f_\beta,z}$ and $T^\star_{f_\beta,z}$ belong to $L_1(R)$ for any $z \in S_W$.

Because of the independence between X and ε, the conditional expectation $\mathbb{E}[f_\beta(X)|Z]$ is written as

$$\mathbb{E}[f_\beta(X)|Z] = \frac{\int f_\beta(x)\eta(Z-x)g(x)dx}{\int \eta(Z-x)g(x)dx} \equiv \frac{\Gamma_{f_\beta}(Z)}{h(Z)} \equiv \Phi_\beta(Z), \tag{2.1}$$

where $h = \eta * g$ ($*$ denoting the convolution) is the density of Z.

According to (2.1), we need to estimate the numerator $\Gamma_{f_\beta}(z_0)$ and the denominator $h(z_0)$ at a fixed point z_0.

2.1. *Estimation of the numerator.*

Applying the Parseval–Plancherel formula and using the independence between X and ε, we obtain

$$\Gamma_{f_\beta}(z_0) = (2\pi)^{-1}\langle T^\star_{f_\beta,z_0}, g^\star\rangle = (2\pi)^{-1}\langle T^\star_{f_\beta,z_0}(\eta^\star)^{-1}, h^\star\rangle. \tag{2.2}$$

We replace $h^\star(t)$ by its empirical estimator $n^{-1}\sum_{j=1}^n e^{itZ_j}K_n^\star(t) = n^{-1}\sum_{j=1}^n h_{n,j}^\star(t)$, where $K_n^\star(t)$ is the Fourier transform of the function K_n defined by $K_n(x) = C_n K(C_n x)$, (K being a kernel to be chosen), and C_n tends to infinity. We propose to estimate $\Gamma_{f_\beta}(z_0)$ by

$$\widehat{\Gamma}_{f_\beta}(z_0) = (2\pi n)^{-1}\sum_{j=1}^n \operatorname{Re}\left\langle T^\star_{f_\beta,z_0}(\eta^\star)^{-1}, h^\star_{n,j}\right\rangle, \tag{2.3}$$

where $\operatorname{Re}(z)$ denotes the real part of z.

REMARK 2.1. This estimator depends on C_n and on the kernel K. These two quantities will be chosen in each example. But in all cases the kernel K has to satisfy the following conditions.

ASSUMPTIONS.

(A1) The kernel K belongs to $L_2(\mathbb{R})$ and is an even function.
(A2) Its Fourier transform satisfies $K^\star(t) = 1$ for any t in $[-1, 1]$.
(A3) $|K^\star(t)| \le \mathbb{1}_{[-2,2]}(t)$ for any t in $\mathbb{R}$.

REMARK 2.2. Assumption (A1) ensures that the Fourier transform of the kernel is a real valued function. Assumption (A2) allows us to control the bias term, and Assumption (A3) ensures the existence of the estimator $\widehat{\Gamma}_{f_\beta}(z)$. Note that the so-called naive kernel $K(x) = \sin(x)/(\pi x)$ satisfies these assumptions. In the same way the analogue of the de La Vallée-Poussin kernel V, defined by

$$V(x) = [\cos(x) - \cos(2x)]/[\pi x^2] \tag{2.4}$$

satisfies Assumptions (A1–A3). For some examples we need the kernel to satisfy additional assumptions. For polynomial functions (see Section 4) the kernel K and its derivatives must have all moments finite.

2.2. *Estimation of the denominator.* The denominator $h(z_0)$, which is the value of the density $h = \eta * g$ at the fixed point z_0, is estimated by the kernel estimator

$$\widehat{h}(z_0) = n^{-1}\sum_{j=1}^n V_n(z_0 - Z_j), \tag{2.5}$$

where $V_n(x) = b_n^{-1}V(xb_n^{-1})$, $b_n^{-1} = \sqrt{\log n}$ and V is the analogue of the de La Vallée-Poussin kernel defined by (2.4).

Since $h(z_0) = \Gamma_1(z_0)$, estimating $h(z_0)$ by (2.5) corresponds to choosing the kernel $K = V$ and $C_n = \sqrt{\log n}$ in formula (2.3). The naive kernel $K(x) = \sin(x)/(\pi x)$ provides the same rate of convergence of the quadratic risk as the kernel V. However, the fact that V belongs to $L_1(\mathbb{R})$ is of particular interest when studying the risk with respect to the L_∞-norm.

2.3. *Construction of* $\widehat{\beta}$. Recall that our estimation criterion $S_n(\beta)$, is given by (1.3) where $\widehat{\Phi}_\beta(Z_i)$ is the estimate of $\Phi_\beta(Z_i)$, defined as the ratio

$$\widehat{\Phi}_\beta(Z_i) = \widehat{\Gamma}_{f_\beta}(Z_i)/\widehat{h}(Z_i)$$

with $\widehat{\Gamma}_{f_\beta}(Z_i)$ and $\widehat{h}(Z_i)$ defined as in (2.3) and (2.5) where we sum over $j \neq i$. To be more explicit,

$$\widehat{h}(Z_i) = (n-1)^{-1} \sum_{j=1,\, j\neq i}^{n} V_n(Z_i - Z_j) \equiv (n-1)^{-1} \sum_{j=1,\, j\neq i}^{n} \mathscr{X}^h_{n,\beta}(Z_i, Z_j) \tag{2.6}$$

and

$$\begin{aligned} \widehat{\Gamma}_{f_\beta}(Z_i) &= \frac{1}{2\pi(n-1)} \sum_{j=1,\, j\neq i}^{n} \operatorname{Re}\langle T^\star_{f_\beta, Z_i}(\eta^\star)^{-1}, h^\star_{n,j}\rangle \\ &\equiv \frac{1}{n-1} \sum_{j=1,\, j\neq i}^{n} \mathscr{X}_{n,\beta}(Z_i, Z_j) \end{aligned} \tag{2.7}$$

Putting these estimators in $S_n(\beta)$, we propose to estimate β^0 by $\widehat{\beta} = \operatorname{Argmin}_{\beta\in\Theta} S_n(\beta)$.

REMARK 2.3. The introduction of the weight function W in (1.3) is necessary in order to ensure that $\|\Phi_\beta\|_{\infty,W}$ is finite and to ensure the convergence of $\|\widehat{\Gamma}_{f_\beta}(z) - \Gamma_{f_\beta}(z)\|_{\infty,W}$ to zero.

3. Asymptotic properties. This section consists of three parts. First we give the asymptotic properties of $\widehat{\Gamma}_{f_\beta}(z)$ and the asymptotic properties of $\widehat{h}(z)$, useful in the study of $\widehat{\beta}$. Next we give the asymptotic properties of $\widehat{\beta}$: we state the consistency of $\widehat{\beta}$ and calculate an upper bound of its rate of convergence.

Let us introduce some additional notation and assumptions, needed to characterize the rates of convergence of $\widehat{\Gamma}_{f_\beta}(z)$ and $\widehat{h}(z)$.

ASSUMPTION.

(A4) We say that a function $\overline{f}_\beta$ satisfies Assumption (A4) if the functions

$$T_{\overline{f}_\beta, z},\ T^\star_{\overline{f}_\beta, z},\ \frac{d}{dz} T_{\overline{f}_\beta, z} \text{ and } \frac{d}{dz} T^\star_{\overline{f}_\beta, z} \text{ belong to } L_1(R) \text{ for any } z \in S_W.$$

NOTATION 2. Let $\overline{f}_\beta$ be a function satisfying Assumption (A4). Take C_n and K as in (2.3) and set

$$\lambda_n(\overline{f}_\beta, z) = \mathbb{E}\left|\langle T^\star_{\overline{f}_\beta, z}(\eta^\star)^{-1}, h^\star_{n,j}\rangle\right|^2, \quad \sigma_n^2(\overline{f}_\beta) = \sup_{z\in S_W} \lambda_n(\overline{f}_\beta, z),$$

$$B_n(\overline{f}_\beta) = \sup_{z\in S_W} \|T^\star_{\overline{f}_\beta, z} \mathbb{I}_{C_n}\|_1,\ M_n(\overline{f}_\beta) = C_n \sup_{z\in S_W} \sup_{|t|\le C_n} |T^\star_{\overline{f}_\beta, z}(t) e^{t^2/2}|$$

and

$$D_n(\overline{f}_\beta) = C_n \sup_{z \in S_W} \sup_{|t| \le C_n} |\frac{d}{dz} T^\star_{\overline{f}_\beta, z}(t) e^{t^2/2}|.$$

Using this notation, we consider the following assumptions for any function $\overline{f}_\beta$.

ASSUMPTION.

(A5) We say that a function $\overline{f}_\beta$ satisfies Assumption (A5) if $\sigma_n(\overline{f}_\beta)\sqrt{\log n} = o(\sqrt{n})$, $M_n(\overline{f}_\beta)\log n = o(n)$ and $D_n(\overline{f}_\beta) = o[n^{3/2}\sigma_n(\overline{f}_\beta)\sqrt{\log n}]$.

3.1. *Asymptotic properties of the estimate of the numerator.*

PROPOSITION 3.1. *Let* $\Gamma_f(z_0)$ *be defined by* (2.2) *and* $\widehat{\Gamma}_f(z_0)$ *be the estimator defined by* (2.3) *with a kernel satisfying Assumptions* (A1)–(A3). *Assume that for any* β *in* Θ, f_β *satisfies Assumption* (A4). *Then, there exists a sequence* C_n *such that* f_β *satisfies* (3) *for any* β *in* Θ. *For such a sequence, the following results hold for any* β *in* Θ.

(a) *For any* z_0 *in* $\mathbb{R}$,

$$|\mathbb{E}[\widehat{\Gamma}_{f_\beta}(z_0)] - \Gamma_{f_\beta}(z_0)|^2 \le \|T^\star_{f_\beta, z_0} \mathbb{I}_{C_n}\|_1^2, \text{ and } \operatorname{Var}[\widehat{\Gamma}_{f_\beta}(z_0)] \le n^{-1}\lambda_n(f_\beta, z_0).$$

(b) *For any* $p \ge 1$,

$$\mathbb{E}[\|\widehat{\Gamma}_{f_\beta} - \Gamma_{f_\beta}\|^p_{\infty, W}] \le K(p)\left[\frac{[\sigma_n(f_\beta)\sqrt{\log n}]^p}{\sqrt{n}^p} + \frac{[M_n(f_\beta)\log n]^p}{n^p} + B_n^p(f_\beta)\right].$$

REMARK 3.1. Note that, at worst, choosing $K(x) = \sin(x)/(\pi x)$ provides

$$\sigma_n^2(f_\beta) \le e^{C_n^2}, \quad M_n(f_\beta) = O(C_n e^{C_n^2/2}) \text{ and } D_n(f_\beta) = O(C_n e^{C_n^2/2}).$$

Hence the choice $C_n = \alpha\sqrt{\log n}$ with $\alpha < 1$ ensures that $\widehat{\Gamma}_{f_\beta}(z_0)$ is a consistent estimator of $\Gamma_{f_\beta}(z_0)$ and that Assumption (A5) is satisfied.

REMARK 3.2. For all the examples we consider in Section 6, the quantity $M_n(f_\beta)$ is of order $\sigma_n(f_\beta)\sqrt{\log n}$. Therefore for any $p \ge 1$ we get that

$$\mathbb{E}[\|\widehat{\Gamma}_{f_\beta} - \Gamma_{f_\beta}\|^p_{\infty, W}] \le K'(p)\left[(\sigma_n(f_\beta)\sqrt{\log n}/\sqrt{n})^p + B_n^p(f_\beta)\right].$$

REMARK 3.3. The regularity of the function f_β (as function of X) plays an essential role in the rate of convergence of $\widehat{\Gamma}_{f_\beta}(z_0)$. Let us successively study the bias and variance terms.

It follows from Proposition 3.1 (a) and the assumption that $T^\star_{f_\beta, z}$ belongs to $L_1(\mathbb{R})$, that the bias tends to zero as n tends to infinity. Moreover, due to the properties of Fourier transform, the more regular f_β, the faster the decrease of $T^\star_{f_\beta,z}$ as t tends to infinity.

The upper bound for the variance term depending on the behavior of the function $R(t)$ defined by $R(t) = T^\star_{f_\beta,z_0}(t)e^{t^2/2}$ increases with the sequence C_n. The more regular f_β is, the slower $R(t)$ increases.

Let us study two particular cases. If f_β is regular enough such that $R(t)$ belongs to $L_1(\mathbb{R})$, the variance is of order n^{-1}, the bias term of order $\exp(-C_n^2/2)$ and $\widehat{\Gamma}_f(z_0)$ converges to $\Gamma_{f_\beta}(z_0)$ with the parametric rate of convergence $n^{-1/2}$. If f_β belongs to a Sobolev Class $W_2^m = \{\text{functions } f \in L_1(\mathbb{R}) \text{ such that } \int(1+|t|^2)^m f^\star(t)dt < \infty\}$, $m > 1$, $R(t)$ does not belong to $L_1(\mathbb{R})$. The square of the bias term is of order C_n^{-2m} (see [26] and [31]) and the variance term will be of order $\exp\left\{C_n^2\right\}/n$. Hence we get an upper bound of order $(\log n)^{-m}$ for the quadratic risk.

3.2. *Asymptotic properties of the estimate of the denominator.* Because of the independence between X and ε, $h = \eta * g$ and $|h^\star(t)| \le \exp\{-t^2/2\}$. Therefore, since V satisfies Assumption (A2), it is easy to see that

$$|\mathbb{E}[\widehat{h}(z_0)] - h(z_0)| \le \|h^\star \overline{\mathbb{I}}_{b_n^{-1}}\|_1 \le Ab_n \exp\{-b_n^{-2}/2\}, \tag{3.1}$$

where A is a numerical constant. Furthermore, classical calculations ensure that the variance is bounded from above by $(nb_n)^{-1}$. Taking $b_n = (\log n)^{-1/2}$ provides the result

$$\overline{\lim}_{n\to\infty} \frac{n}{\sqrt{\log n}} \sup_{h\in\mathscr{H}} \mathbb{E}\left[\widehat{h}(z_0) - h(z_0)\right]^2 \le \|V\|_2^2, \tag{3.2}$$

where $\mathscr{H} = \{\eta * g, g \text{ being a density on the real line}\}$. Note that this rate of convergence is the best in the minimax sense (see [35]). We now calculate an upper bound for the risk of $\widehat{h}(z)$ related to the L_∞-norm. It is known from [26] that the rate of convergence for estimating regular densities in L_p-norms, $1 \le p \le \infty$ is related to the distance between the density to be estimated and the class of entire functions of exponential type. More precisely let us define the distance

$$D_\nu^\infty(h) = \inf_{f\in\mathscr{M}_{\nu,\infty}} \|h - f\|_\infty,$$

$\mathscr{M}_{\nu,\infty}$ being the collection of all entire functions of exponential type ν (see [33], page 372 for further references about entire functions of exponential type). Now, nequalities (3.1) ensure that

$$D^\infty_{b_n^{-1}}(h) \le \|h - h * V_n\|_\infty \le \|h^\star(1 - V_n^\star)\|_1 \le Ab_n \exp\{-b_n^{-2}/2\}. \tag{3.3}$$

It follows, applying the result of [26] and according to (3.3), that there exists a constant A' depending on p such that

$$\overline{\lim}_n \left(\frac{n}{\sqrt{\log n}\log\log n}\right)^{p/2} \sup_{h\in\mathscr{H}} \mathbb{E}\,\|\widehat{h} - h\|_\infty^p \le A'. \tag{3.4}$$

The rates of convergence obtained for estimating $h(z)$ are classical results for strongly regular densities. Note that, whatever the density g is, the density $\eta * g$ has obviously strong regularity properties. We immediately see for instance that $h = \eta * g$ is $\mathscr{C}^\infty$, that for any $\gamma > 0$, there exists $C_\gamma > 0$ such that h belongs to the class $\mathscr{A}_\gamma(C_\gamma)$ of densities admitting an analytic continuation to the strip $\{(x + iy), |y| \le \gamma\}$ with $(2\pi)^{-1}\int \cosh^2 \gamma t |h^\star(t)|^2 dt \le C_\gamma$. More precisely, according to (3.1), we see that $\mathscr{H}$ is included in the class $\mathscr{B}_{1,1/2}(2)$ where the classes $\mathscr{B}_{A,\rho}(r)$ are more generally defined as $\mathscr{B}_{A,\rho}(r) = \{\phi \text{ such that } |\phi^\star(t)| \le A\exp\{-\rho|t|^r\}\}$. See [12], [13], [26], [25], [21] for further references about density estimation in classes of densities with strong regularity properties. Note that the problem of estimating the value of h at a fixed point in the convolution model $Z_i = X_i + \varepsilon_i$ is different from the deconvolution problem. In the latter case the aim is to estimate the density g of the X_i's. It is known that the slowest rates of convergence for estimating g are obtained for the smoothest error densities. See [14], [6], [39] for references about deconvolution problems.

3.3. *Asymptotic properties of $\widehat{\beta}$.* Before stating our results we give some preliminary assumptions.

ASSUMPTIONS. For any β in Θ,

(A6) $f_\beta(x)$ admits continuous derivatives up to order 3 with respect to β.

(A7) $\mathbb{E}\left\{W^2(Z)[f_{\beta^0}(X) - \Phi_\beta(Z))]^4\right\} < \infty$.

(A8) $S^{(2)}_{a,b}(\beta) = \mathbb{E}[W(Z)\mathbb{E}[f^{(1)}_{\beta_a}(X)|Z]\mathbb{E}[f^{(1)}_{\beta_b}(X)|Z] < \infty$ for $a, b = 1, \dots, m$, $f^{(1)}_a$ denoting the first derivative of f with respect to β_a.

(A9) The quantity $S(\beta, \beta^0) = \mathbb{E}[W(Z)]\sigma^2 + \mathbb{E}\{W(Z)[f_{\beta^0}(X) - \Phi_\beta(Z)]^2\}$, admits one unique minimum at $\beta = \beta^0$.

(A10) The matrix $S^{(2)}(\beta^0)$ is positive definite.

(A11) For $a, b, c = 1, \dots, m$, denote by $f^{(2)}_{\beta_{a,b}}(x)$ and $f^{(3)}_{\beta_{a,b,c}}(x)$ the second and third derivatives of f with respect to β_a, β_b and β_c respectively. For any z in S_W and for $a, b, c = 1, \dots, m$ the functions

$$\sup_{\beta\in\Theta}\left|f^{(1)}_{\beta_a}(.)\eta(z - .)\right|, \ \sup_{\beta\in\Theta}\left|f^{(2)}_{\beta_{a,b}}(.)\eta(z - .)\right| \text{ and } \sup_{\beta\in\Theta}\left|f^{(3)}_{\beta_{a,b,c}}(.)\eta(z - .)\right|$$

are in $L_1(R)$.

Observe that Assumption (A11) ensures in particular that

$$\frac{\partial}{\partial\beta}\Phi_\beta(z) = \Gamma_{f^{(1)}_\beta}(z)/h(z) \equiv \Phi^{(1)}_\beta(z), \tag{3.5}$$

and therefore

$$S^{(2)}_{a,b}(\beta) = \mathbb{E}[W(Z)\Phi^{(1)}_{\beta_a}(Z)\Phi^{(1)}_{\beta_b}(Z)] = \mathbb{E}[W(Z)\mathbb{E}[f^{(1)}_{\beta_a}(X)|Z]\mathbb{E}[f^{(1)}_{\beta_b}(X)|Z]$$

Consistency of $\widehat{\beta}$.

THEOREM 3.1. *Let* $\widehat{\Gamma}_{f_\beta}(Z_i)$ *be defined by* (2.7) *with the kernel* K *satisfying Assumptions* (A1–A3). *Assume that for any* β *in* Θ, f_β *satisfies Assumptions* (A4), (A6–A10). *Then there exists a sequence* C_n *such that* f_β *satisfies Assumption* (A5) *for any* β *in* Θ. *For this sequence,* $\widehat{\beta} = \widehat{\beta}(C_n)$ *defined by* (1.4), *is a consistent estimator of* β^0.

The proof of the consistency is mainly based on the properties of the criterion $\widetilde{S}_n(\beta, g)$ defined in (1.1). We state in the proof of Lemma 6.1 that $\widetilde{S}_n(\beta, g)$ is a contrast (see [11], Definition 3.2.7. and Theorem 3.2.8, pages 124–126 or [4] for further references about contrasts). Next, using the asymptotic properties of $\widehat{\Gamma}_{f_\beta}(z)$ and $\widehat{h}(z)$ we prove that for any β in Θ, $\widetilde{S}_n(\beta, g) - S_n(\beta)$ converges to zero in probability. To establish this convergence, we use that $\|\widehat{\Gamma}_{f_\beta} - \Gamma_{f_\beta}\|_{\infty,W} = o_p(1)$. This convergence holds provided that Assumption (A4) is satisfied and that C_n is such that Assumption (A5) is satisfied (see Proposition 3.1).

Consequently the assumptions needed to prove the consistency are integrability, differentiability assumptions (A6–A8), and identifiability assumptions (A9) and (A10).

Rate of convergence of $\widehat{\beta}$.

THEOREM 3.2. *Assume that for any* β *in* Θ, f_β *satisfies Assumptions* (A4), (A6)–(A11). *Let* $\widehat{\Gamma}_{f_\beta}(Z_i)$ *be as in Therorem* 3.1, *take* C_n *such that Theorem* 3.1 *holds and such that the first and second derivatives of* f_β *with respect to* β *satisfy Assumption* (A5). *Then the following result holds*:

$$\widehat{\beta} - \beta^0 = O_p\left[\delta_{1,n}(f_{\beta^0}) + n^{-1/2} + \delta_{2,n}(f_{\beta^0})\right],$$

where

$$\delta_{1,n}(f_{\beta^0}) = n^{-1/2}\mathbb{E}^{1/2}\left\{\mathbb{E}^2\left[\frac{W(Z_1)}{h(Z_1)}\Phi^{(1)}_{\beta^0}(Z_1)\mathscr{X}_{n,\beta^0}(Z_1, Z_2)\middle| Z_2\right]\right\} + B_n(f_{\beta^0}),$$

$$\delta_{2,n}(f_{\beta^0}) = M_n(f_{\beta^0})M_n(f^{(1)}_{\beta^0})/n^{3/2} + o\left[M_n(f_{\beta^0})(\log n)^{5/4}\sqrt{\log\log n}/n^{3/2}\right],$$

with $\mathscr{X}_{n,\beta}(z, Z_j)$ *and* $\Phi^{(1)}_\beta(z)$ *defined in* (2.7) *and* (3.5) *respectively*.

REMARK 3.4. The rate of convergence arising from the difference

$$\frac{\partial}{\partial\beta}\widetilde{S}_n\left(\beta^0, g\right) - \frac{\partial}{\partial\beta}S_n\left(\beta^0\right),$$

with $\widetilde{S}_n(\beta)$ defined in (1.1), is given by three terms. The first term $\delta_{1,n}(f_{\beta^0})$ is strongly related to the rate of convergence of $\widehat{\Gamma}_{f_{\beta^0}}(z) - \Gamma_{f_{\beta^0}}(z)$ through quantities of the form

$$n^{-1}\sum_{i=1}^{n} W(Z_i)F(Z_i)[\widehat{\Gamma}_{f_{\beta^0}}(Z_i) - \Gamma_{f_{\beta^0}}(Z_i)], \tag{3.6}$$

where F is some function of Z_i satisfying $\mathbb{E}[W^2(Z_i)F^2(Z_i)] < \infty$. The second term of order $n^{-1/2}$ comes from the study of

$$n^{-1}\sum_{i=1}^{n} W(Z_i)F(Z_i)[\widehat{h}(Z_i) - h(Z_i)] = O_p(n^{-1/2}), \tag{3.7}$$

and from the fact that

$$-\sqrt{n}\frac{\partial}{\partial\beta}\widetilde{S}_n(\beta^0, g) \underset{n\to\infty}{\overset{\mathscr{L}}{\longrightarrow}} \mathscr{N}(0, \Sigma_1),$$

Σ_1 being defined in Lemma 6.1. The last term $\delta_{2,n}(f_{\beta^0})$ comes from the study of quantities of the form

$$A_2 = n^{-1}\sum_{i=1}^{n} W(Z_i)\Big[\widehat{\Phi}_{\beta^0}(Z_i) - \Phi_\beta(Z_i)\Big]\Big[\widehat{\Phi}^{(1)}_{\beta^0}(Z_i) - \Phi^{(1)}_{\beta^0}(Z_i)\Big],$$

where $\widehat{\Phi}^{(1)}_{\beta^0}(Z_i)$ is the estimate of $\mathbb{E}[f^{(1)}_{\beta^0}(X_i)|Z_i] = \Phi^{(1)}_{\beta^0}(Z_i)$ defined in (3.5). This term $\delta_{2,n}(f_{\beta^0})$ will generally be negligible compared to the others. Consequently the rate of convergence of $\widehat{\beta}$ is in most cases governed by the first term $\delta_{1,n}(f_{\beta^0})$ and hence we can expect to achieve the parametric rate of convergence only for functions f_β such that (3.6) is of order $O_p(n^{-1/2})$. When $f_\beta \equiv \beta$ and $K \equiv V$, (3.6) and (3.7) are the same, and in this case $\delta_{1,n}(f_{\beta^0}) = O_p(n^{-1/2})$.

REMARK 3.5. We have seen in Remark 2.1 that $\widehat{\Gamma}_f(z)$ depends on the sequence C_n and on the kernel K. It follows that $\widehat{\beta}$ also depends on these two quantities that we need to choose in each case. According to Remark 3.1, one can always find a sequence C_n such that $\widehat{\beta}$ is consistent. In fact, it is chosen among sequences C_n satisfying Assumption (A5), in order to balance the first and the last term in $\delta_{1,n}(f)$. In the same way the kernel K is chosen in each example.

When f_β is such that precise calculations of $\delta_{1,n}(f_{\beta^0})$ are possible, we can choose a kernel K that provides a rate of convergence of $\widehat{\beta}$ faster than the rate of convergence of $\widehat{\Gamma}_{f_{\beta^0}}(z)$. If not, we only use the fact that $\delta_{1,n}(f_{\beta^0})$ is dominated by $\sigma_n(f_{\beta^0}) + \beta_n(f_{\beta^0})$. In these cases, the kernel K is chosen as the naive kernel $K(x) = \sin(x)/(\pi x)$, and we calculate an upper bound for the rate of convergence of $\widehat{\beta}$ which is directly related to the rate of convergence of the numerator $\widehat{\Gamma}_{f_{\beta^0}}(z)$.

The proof of Theorem 3.2 can be found in Section 6.

4. Applications. Applying Remarks 3.3 and 3.5 when f_β belongs to a Sobolev class W_2^m, $m > 1$, provides that $\delta_{1,n}(f_{\beta^0})$ is of order $(\log n)^{m/2}$ which remains very slow. These remarks motivate us to consider regular functions with respect to x, in particular functions admitting an analytic continuation to a strip in the complex plane containing the real line.

First we consider functions admitting an analytic continuation to the whole complex plane and second we consider functions admitting an analytic continuation in a finite width strip.

4.1. *Functions admitting an analytic continuation in the whole complex plane.*

EXAMPLE (Polynomials). The following result states that if we consider a kernel K that is a fast decreasing function, we get the $\sqrt{n}$-consistency.

ASSUMPTION. For any integer l and p, the kernel K satisfies

(A12) $\int |u^l K^{(p)}| du < \infty$

where $K^{(p)}$ denotes the derivative of order p of the kernel K.

Assume that $f_\beta(x) = \sum_{l=0}^m \beta_l x^l$ with $m \geq 1$. Let $\widehat{\Gamma}_{f_\beta}(Z_i)$ be defined by (2.7) with $C_n = \sqrt{\log n}$ and with the kernel K satisfying Assumptions (A1)–(A3) and (A12). Then we have

$$\widehat{\beta} - \beta^0 = O_p[n^{-1/2}].$$

REMARK 4.1. It is easily shown that there exist kernels satisfying Assumptions (A1)–(A3) and (A12).

REMARK 4.2. Note that when $m = 1$, then $f_\beta(X) = \beta_0 + \beta_1 X$, and the rate is of order $n^{-1/2}$. It follows that $\widehat{\beta}$ achieves the parametric rate of convergence (result already known in that case). Several other methods provide an estimator of β converging with the parametric rate (see [18], [1], [19], [9]). For instance the estimator minimizing

$$\sum_{i=1}^n (Y_i - \beta_0 - \beta_1 Z_i)^2/(1+\beta^2),$$

is known to be $\sqrt{n}$-consistent and asymptotically Gaussian.

Our method provides the parametric rate of convergence for estimating β in the linear case and extends this result to polynomials of arbitrary degree.

SKETCH OF PROOF. In view of (2.7) and after some calculations we find

$$\mathscr{X}_{n,\beta^0}(Z_1, Z_2) = \sum_{l=0}^m \beta_l^0 \sum_{k=0}^l \binom{l}{k} \gamma(l-k) \sum_{q=0}^k \binom{k}{q} Z_1^{k-q} K_n^{(q)}(Z_1 - Z_2), \tag{4.1}$$

where $\gamma(l-k) = (\sqrt{2\pi})^{-1}\int u^{l-k}\exp\{-u^2/2\}dx$. Combining (4.1) with the fact that the kernel K satisfies Assumptions (A1–A3) yield $B_n(f_{\beta^0}) = o_p(n^{-1/2})$.

Let $\varphi(z) = W(z)\Phi^{(1)}_{\beta^0}(z)/h(z)$. Then, according to (4.1), the following equality holds

$$\delta_{1,n}(f_{\beta^0}) = n^{-1/2}\mathbb{E}^{1/2}[D_n^2(Z_2)] + B_n(f_{\beta^0}),$$

where

$$D_n(Z_2) = \sum_{l=0}^{m}\beta_l^0\sum_{k=0}^{l}\binom{l}{k}\gamma(l-k)\sum_{q=0}^{k}\binom{k}{q}\mathbb{E}\left[\varphi(Z_1)Z_1^{k-q}K_n^{(q)}(Z_1-Z_2)\mid Z_2\right].$$

Now, the last expectation can be rewritten in the following way:

$$\mathbb{E}\left[\varphi(Z_1)Z_1^{k-q}K_n^{(q)}(Z_1-Z_2)\mid Z_2\right] = C_n^q\int\varphi_{k,q}\left(Z_2+uC_n^{-1}\right)K^{(q)}(u)du$$

with $\varphi_{k,q}(z) = \varphi(z)h(z)z^{k-q}$. Proceeding to a Taylor expansion yields

$$\begin{aligned} C_n^q\int\varphi_{k,q}\left(Z_2+uC_n^{-1}\right)K^{(q)}(u)du &= \sum_{r=0}^{q-1}\varphi_{k,q}^{(r)}(Z_2)C_n^{q-r}\int u^rK^{(q)}(u)du \\ &\quad+\varphi_{k,q}^{(q)}(Z_2)\int u^qK^{(q)}(u)du \\ &\quad+C_n^{-1}\int u^{q+1}\varphi_{k,q}^{(q+1)}(u_2^\star)K^{(q)}(u)du \end{aligned} \tag{4.2}$$

$u_2^\star$ being a point in the interval $[Z_2, Z_2+u/C_n]$. Integrating by parts provides that the first term in (4.2) is equal to zero, the last term tends to zero with C_n and the second term is bounded from above. This entails that $\delta_{1,n}(f_{\beta^0}) = O_p(n^{-1/2})$. □

EXAMPLE (Exponential). Assume that $f_\beta(x) = \exp\{\beta x\}$. Let $\widehat{\Gamma}_{f_\beta}(Z_i)$ be defined by (2.7) with $C_n = \sqrt{\log n}$ and with the kernel $K \equiv V$ defined in (2.4).

Then we get the following result: $\widehat{\beta} - \beta^0 = O_p(n^{-1/2})$. □

Combining this result and the result on polynomials ensures that $\widehat{\beta}$ achieves the parametric rate of convergence, for any function $f_\beta(x)$ of the form

$$f_\beta(x) = \sum_{l=0}^{m}\beta_l x^l\exp\{\alpha x\}.$$

EXAMPLE (Sum of cosines). Assume that $f_\beta(x) = \sum_{l=0}^m\beta_l\cos(lx)$ with $m \geq 1$. Let $\widehat{\Gamma}_{f_\beta}(Z_i)$ be defined by (2.7) with $C_n = \sqrt{2\log n}$ and with the kernel $K(x) = \sin(x)/(\pi x)$. Then we get $\widehat{\beta} - \beta^0 = O_p[n^{-1/2}\exp\{m\sqrt{\log n}\}]$. □

4.2. *Functions admitting an analytic continuation in a finite width strip.*

EXAMPLE. Let us denote by $B_\gamma = \{(x+iy), (x, y) \in \mathbb{R}^2, |y| \le \gamma\}$ and consider the following assumption.

ASSUMPTION. For any β in Θ the function $f_\beta(.)$ admits an analytic continuation to the strip B_γ, and there exist $c > 2$ and $M > 0$ such that for all $|A| \ge M$,

(A13) $\sup_{|y|\le\gamma} \left|f_\beta(A+iy)\right| \le \exp\left\{A^2/c\right\}$.

COROLLARY 4.1. *Let $\widehat{\beta}$ be defined by* (1.4) *with $C_n = \sqrt{\log n}$ and $K(x) = \sin(x)/(\pi x)$. Assume that there exists $\gamma > 0$ such that f satisfies Assumption* (A13). *If f satisfies Assumptions* (A6)–(A10), *we have the following results*:

(a) *$\widehat{\beta}$ is a consistent estimator of β^0.*

(b) *Furthermore if f and its derivatives with respect to β satisfy Assumptions* (A11) *and* (A13) *then we get*

$$\widehat{\beta} - \beta^0 = O_p\left[\exp\left\{-\gamma\sqrt{\log n}\right\}\right].$$

Easy calculations ensure that $B_n(f^0_\beta) = O[\exp\{-\gamma C_n\}]$, and $\sigma^2_n(f) = O[\exp\{(\gamma - C_n)^2\}]$. Moreover, since $f^{(1)}_\beta$ also admits an analytic continuation to the strip B_γ, $\sigma^2_n(f^{(1)}_{\beta^0})$ is of the same order than $\sigma^2_n(f_{\beta^0})$ which gives the result. □

5. Estimation of σ^2. We consider now the problem of estimating the variance σ^2 of the errors ξ_i's. The main point is that $\widehat{\beta}$ depends on the observations (Y_i, Z_i), on the density η of the errors ε_i and on the regression function f_β, but it does not depend on σ^2. Hence we propose to estimate σ^2 by

$$\widehat{\sigma}^2 = T_n(\widehat{\beta}) - n^{-1}\sum_{i=1}^{n} W_n(Z_i)\left[\widehat{\Phi}_{f^2_{\widehat{\beta}}}(Z_i) - \widehat{\Phi}^2_{\widehat{\beta}}(Z_i)\right], \tag{5.1}$$

where $T_n(\widehat{\beta})$ and $W_n(Z_i)$ are defined by $T_n(\widehat{\beta}) = n^{-1}\sum_{i=1}^n W_n(Z_i)[Y_i - \widehat{\Phi}_{\widehat{\beta}}(Z_i)]^2$, and $W_n(Z_i) = W(Z_i)/[n^{-1}\sum_{i=1}^n W(Z_i)]$. Under suitable assumptions, $\widehat{\sigma}^2$ is a consistent estimator of σ^2 as it is stated in the following corollary.

COROLLARY 5.1. *Let $\widehat{\beta}$ be defined by* (1.4), *$\widehat{\sigma}^2$ be defined by* (5.1) *and assume that Theorem* 3.1 *holds. If furthermore f^2 satisfies Assumption* (A4), *then taking C_n such that f^2 satisfies* (A5) *ensures that $\widehat{\sigma}^2$ is a consistent estimator of σ^2.*

REMARK 5.1. Note that all calculations made with $\mathrm{Var}(\varepsilon_i) = 1$ remain available with $\mathrm{Var}(\varepsilon_i)$ known but different from one. The problem of estimating $\mathrm{Var}(\varepsilon_i)$ is more complicated and needs further calculations, since $\widehat{\beta}$ defined in (1.4) depends explicitly on η (the density of the ε_i's) and hence on $\mathrm{Var}(\varepsilon_i)$.

6. Proofs. We start with the proof of Proposition 3.1, concerning the asymptotic properties of $\widehat{\Gamma}_{f_\beta}(z)$. Second we give the asymptotic properties of $\widetilde{\beta}_g$ defined in (1.2) and then we prove Theorem 3.1 (consistency of $\widehat{\beta}$). Third we proceed to the proof of Theorem 3.2 (upper bound of the rate of convergence of $\widehat{\beta}$) and fourth we prove Corollary 5.1 (consistency of $\widehat{\sigma}^2$).

6.1. *Proof of Proposition* 3.1. Part (a) of Proposition 3.1 is an almost immediate consequence of Definition (2.3).

To prove part (b) we use a chaining method and exponential inequalities.

We start with studying $\mathbb{E}[\|\widehat{\Gamma}_{f_\beta} - \Gamma_{f_\beta}\|_{\infty,W}]$ using a version of a lemma of Pisier (see [30]). For higher order moments, we use a version of Talagrand's inequality that can be found in [29].

Denoting by $\nu_n(z)$ the quantity $\nu_n(z) = \widehat{\Gamma}_{f_\beta}(z) - \mathbb{E}[\widehat{\Gamma}_{f_\beta}(z)]$ and applying the part a) we get that

$$\mathbb{E}[\|\widehat{\Gamma}_{f_\beta} - \Gamma_{f_\beta}\|_{\infty,W}] \le \mathbb{E}[\|\nu_n\|_{\infty,W}] + B_n(f_\beta).$$

Using the definition of $\widehat{\Gamma}_{f_\beta}(z)$ given by (2.3) and (2.7), $\nu_n(z) = n^{-1}\sum_{j=1}^n \psi_{n,j}(z)$ where $\psi_{n,j}(z) = \mathscr{X}_{n,\beta}(z, Z_j) - \mathbb{E}[\mathscr{X}_{n,\beta}(z, Z_j)]$ are i.i.d centered variables and satisfy

$$\|\psi_{n,j}\|_{\infty,W} \le M_n(f_\beta), \quad \sup_{z\in S_W} \mathrm{Var}[\psi_{n,j}(z)] \le \sigma_n^2(f_\beta)$$

and

$$\sup_{z\in S_W}\left|\frac{d}{dz}\psi_{n,j}(z)\right| \le D_n(f_\beta)$$

with $M_n(f_\beta)$ and $D_n(f_\beta)$ defined in Notation 2. To apply Pisier's Lemma (see [30]) we need to consider a grid G_n on S_W. We thus cover the compact S_W by $|G_n|$ intervals $[z_l, z_{l+1}[$, $l = 1, \ldots, |G_n| - 1$. We denote by L the length of the interval S_W and by $\Pi(z)$ the projection of $z \in S_W$ on G_n. The projection $t = \Pi(z)$ is defined by $t = z_l$ for $z \in [z_l, z_{l+1}[$ for $l = 1, \ldots, |G_n|$. Since $\nu_n(z) = \nu_n(t) + \nu_n(z) - \nu_n(t)$ it follows that

$$\mathbb{E}[\|\nu_n\|_{\infty,W}] \le \mathbb{E}_1 + \mathbb{E}_2$$

with $\mathbb{E}_1 = \mathbb{E}[\|\nu_n\|_{\infty,G_n}]$ and $\mathbb{E}_2 = \mathbb{E}[\sup_{z\in S_W} |\nu_n(z) - \nu_n(\Pi(z))|]$. Now, applying Pisier's Lemma we obtain

$$\mathbb{E}_1 \le \sigma_n(f_\beta)\sqrt{2n^{-1}\log(2|G_n|)} + [M_n(f_\beta)\log(2|G_n|)]/3n$$

Finally, taking $|G_n| = n^2$, $\mathbb{E}_1$ is bounded from above in the following way:

$$\mathbb{E}_1 \le \sigma_n(f_\beta)\sqrt{n^{-1}2[\log 2 + 2\log n]} + M_n(f_\beta)[\log 2 + 2\log n]/3n. \tag{6.1}$$

The derivatives of $\widehat{\Gamma}_{f_\beta}(z)$ and $\mathbb{E}[\widehat{\Gamma}_{f_\beta}(z)]$ are continuous and we get that

$$\mathbb{E}_2 \le \left[\sup_{z\in S_W} |z-\Pi(z)| \sup_{z\in S_W} \left|\frac{d}{dz}\nu_n(z)\right|\right] \le 2LD_n(f_\beta)/(|G_n|). \tag{6.2}$$

Combining (6.1) and (6.2) we get by Assumption (A5) that $\mathbb{E}[\|\nu_n\|_{\infty,W}]$ is bounded by

$$\sigma_n(f_\beta)\sqrt{n^{-1}2[\log 2+2\log n]} + KM_n(f_\beta)[\log 2 + 2\log n]/3n, \tag{6.3}$$

which implies that

$$\mathbb{E}\left[\sup_{z\in S_W}\left|\widehat{\Gamma}_{f_\beta}(z)-\Gamma_{f_\beta}(z)\right|\right] \le K'\left[\sigma_n(f_\beta)\sqrt{\log n}/\sqrt{n} + M_n(f_\beta)\log n/n + B_n(f_\beta)\right],$$

and the result holds for $p=1$.

Consider now the case $p>1$. Denote by x_+ the quantity $x_+ = x\mathbb{1}_{x\ge 0}$. For any $\rho>0$ and for any finite family $\mathscr{T}$ we have that

$$\|\nu_n\|_{\infty,\mathscr{T}} \le \left[\|\nu_n\|_{\infty,\mathscr{T}} - (1+\rho)\mathbb{E}(\|\nu_n\|_{\infty,\mathscr{T}})\right]_+ + (1+\rho)\mathbb{E}\left[\|\nu_n\|_{\infty,\mathscr{T}}\right].$$

Consequently, $\mathbb{E}[\|\nu_n\|_{\infty,\mathscr{T}}]^p$ is bounded by

$$2^p\mathbb{E}[\|\nu_n\|_{\infty,\mathscr{T}} - (1+\rho)\mathbb{E}(\|\nu_n\|_{\infty,\mathscr{T}})]_+^p + 2^p(1+\rho)^p\mathbb{E}^p[\|\nu_n\|_{\infty,\mathscr{T}}].$$

Denoting by $Z^{(n)}$, the variable $Z^{(n)} = \|\nu_n\|_{\infty,\mathscr{T}} - (1+\rho)\mathbb{E}[\|\nu_n\|_{\infty,\mathscr{T}}]$, using that $\mathbb{E}(X_+)^p = \int py^{p-1}\mathbb{P}(X>y)\mathbb{1}_{y\ge 0}dy$ and applying Talagrand's inequality (see [29]) we get that

$$\mathbb{E}(Z_+^{(n)})^p \le p\Gamma(p/2)\left[\sigma_n(f_\beta)/\sqrt{n\tau(\rho)}\right]^p + 2p![M_n(f_\beta)/(n\tau(\rho))]^p.$$

It follows from the definition of $\sigma_n(f_\beta)$ and using (6.3), that for any finite subset $\mathscr{T}$ of S_W there exists a constant K_p depending on p such that,

$$\mathbb{E}[\|\nu_n\|_{\infty,\mathscr{T}}]^p \le K_p\mathbb{E}^p[\|\nu_n\|_{\infty,\mathscr{T}}].$$

This entails that $\mathbb{E}[\nu_n\|_{\infty,W}]^p \le K_p\mathbb{E}^p[\|\nu_n\|_{\infty,W}]$ which is bounded by

$$K_1(p)[(\sigma_n(f_\beta)\sqrt{\log n}/\sqrt{n})^p + (M_n(f_\beta)\log n/n)^p]. \qquad \square$$

All the properties of $\widehat{\beta}$ follow from asymptotic properties of $\widetilde{\beta}_g$. More precisely the consistency of $\widehat{\beta}$ comes from the difference $S_n(\beta)-\widehat{S_n}(\beta)$. The upper bound of the rate of convergence of $\widehat{\beta}$ comes from $S_n^{(1)}(\beta^0)-\widetilde{S}_n^{(1)}(\beta^0, g)$. Therefore we begin with the statement of the asymptotic properties of $\widetilde{\beta}_g$, proved at the beginning of the Appendix.

LEMMA 6.1. *Let* $\widetilde{\beta}_g$ *be defined by* (1.2). *If Assumptions* (A6–A10) *are satisfied, we get the following results*:

(a) $\widetilde{\beta}_g$ *converges in probability to* β^0.

(b) *Furthermore* $\sqrt{n}(\widetilde{\beta}_g - \beta^0) \xrightarrow[n\to\infty]{\mathscr{L}} \mathscr{N}(0, \Sigma)$ *with* $\Sigma = \left[\mathbb{E}\,[W(Z)\Phi_{\beta^0}(Z)]^2\right]^{-2}\Sigma_1$ *and*

$$\Sigma_1 = \sigma^2 \mathbb{E}\left[W^2(Z)\Phi^{(1)}_{\beta^0}(Z)\Phi^{(1)}_{\beta^0}(Z)^T\right] + \mathbb{E}\left[W^2(Z)[f_{\beta^0}(X) - \Phi_{\beta^0}(Z)]^2\Phi^{(1)}_{\beta^0}(Z)\Phi^{(1)}_{\beta^0}(Z)^T\right].$$

We repeatedly use the following equalities, needed in the proof of consistency and in the proof of the bound for the rate of convergence.

For any function $\bar{f}$ satisfying Assumption (A4), using both (3.4) and Proposition 3.1, yields the following equalities:

$$\widehat{\Phi}_{\bar{f}}(Z_i) - \Phi_{\bar{f}}(Z_i) = \frac{\widehat{\Gamma}_{\bar{f}}(Z_i) - \Gamma_{\bar{f}}(Z_i)}{h(Z_i)} - \frac{\widehat{\Gamma}_{\bar{f}}(Z_i)}{h(Z_i)}\frac{\widehat{h}(Z_i) - h(Z_i)}{h(Z_i)} + r_n(\bar{f}, Z_i)$$

with $|U_i^\star| \le |\widehat{h}(Z_i) - h(Z_i)|$ and $r_n(\bar{f}, Z_i)$ defined by

$$r_n(\bar{f}, Z_i) = 2\frac{\widehat{\Gamma}_{\bar{f}}(Z_i)}{h(Z_i)}\frac{[\widehat{h}(Z_i) - h(Z_i)]^2}{h^2(Z_i)}\frac{1}{(1+U_i^\star)^3}.$$

We deduce immediately from (3.4) that

$$|r_n(\bar{f}, Z_i)| \le \|\widehat{h} - h\|_\infty^2 = o_p(n^{-1/2}),$$

which entails

$$\widehat{\Phi}_{\bar{f}}(Z_i) - \Phi_{\bar{f}}(Z_i) = \frac{\widehat{\Gamma}_{\bar{f}}(Z_i) - \Gamma_{\bar{f}}(Z_i)}{h(Z_i)} - \frac{\widehat{\Gamma}_{\bar{f}}(Z_i)}{h(Z_i)}\frac{\widehat{h}(Z_i) - h(Z_i)}{h(Z_i)} + o_p(n^{-1/2}). \tag{6.4}$$

6.2. *Proof of Theorem* 3.1. The main point of the proof lies in stating that for any β in Θ, $\widetilde{S}_n(\beta, g) - S_n(\beta) = o_p(1)$ whence $S_n(\beta)$ is a contrast. This implies by Lemma 6.1 that $S_n(\beta)$ converges in probability to $S(\beta, \beta^0)$, admitting an unique minimum at $\beta = \beta^0$. Next denoting by $w(n, \rho)$ the quantity

$$w(n, \rho) = \sup\left\{|S_n(\alpha) - S_n(\beta)| : |\alpha - \beta| \le \rho\right\},$$

and using the regularity of f we obtain, after easy calculations, that there exist two sequence ρ_k and ε_k such that

$$\lim_{n\to\infty} \mathbb{P}[w(n, \rho_k) > \varepsilon_k] = 0. \tag{6.5}$$

The consistency of $\widehat{\beta} = \text{Arg}\ \min_{\beta\in\Theta} S_n(\beta)$ follow thus from [11] (Theorem 3.2.8, page 126). Let us study $S_n(\beta) - \widetilde{S}_n(\beta, g)$. Simple calculations give the decomposition

$$S_n(\beta) - \widetilde{S}_n(\beta, g) = \frac{1}{n}\sum_{i=1}^{n} W(Z_i)\Big\{[\Phi_\beta(Z_i) - \widehat{\Phi}_\beta(Z_i)] \times [2Y_i - (\widehat{\Phi}_\beta(Z_i) + \Phi_\beta(Z_i))]^2\Big\}.$$

Now, combining (6.4), Proposition 3.1 and inequality (3.4), we infer that

$$\|\widehat{\Phi}_\beta - \Phi_\beta\|_{\infty, W} = o_p(1), \tag{6.6}$$

which entails that $S_n(\beta) - \widetilde{S}_n(\beta, g) = o_p(1)$. Finally, applying Lemma 6.1, $\widetilde{S}_n(\beta, g)$ is a contrast which implies that $S_n(\beta)$ is also a contrast and arguing as for $\widetilde{\beta}_g$, we deduce that $\widehat{\beta}$ is consistent. □

6.3. *Proof of Theorem* 3.2. For convenience, we subsequently omit β^0 in the notation and we denote by $\widehat{\Gamma}_f(z)$ and $\Gamma_f(z)$ the quantities $\widehat{\Gamma}_{f_{\beta^0}}(z)$ and $\Gamma_{f_{\beta^0}}(z)$. In the same way denote by $T_{f,z}$ the function $T_{f_{\beta^0},z}$. Applying (3.5), we consider the vector

$$\frac{\partial}{\partial\beta}\widehat{\Gamma}_{f_{\beta^0}}(Z_i) = \widehat{\Gamma}_{f^{(1)}}(Z_i) = \left(\widehat{\Gamma}_{f_1^{(1)}}(Z_i), \ldots, \widehat{\Gamma}_{f_p^{(1)}}(Z_i)\right)^T,$$

defined as the vector of the first derivatives of $\widehat{\Gamma}_{f_\beta}(Z_i)$ with respect to the parameter β and taken at $\beta = \beta^0$. Similarly, we define

$$\frac{\partial^2}{\partial\beta_a\partial\beta_b}\widehat{\Gamma}_{f_{\beta^0}}(Z_i) = \widehat{\Gamma}_{f_{a,b}^{(2)}}(Z_i), \tag{6.7}$$

the matrix of the second derivatives of $\widehat{\Gamma}_{f_\beta}(Z_i)$ with respect to the component β_a and β_b of β and taken at the value $\beta = \beta^0$. By Definition (1.4), of $\widehat{\beta}$ we have

$$\frac{\partial}{\partial\beta}S_n\left(\widehat{\beta}\right) := S_n^{(1)}\left(\widehat{\beta}\right) = 0.$$

Under Assumptions (A6) and (A11), proceeding to a Taylor expansion, we can write

$$\widehat{\beta} - \beta^0 = -[S_n^{(2)}\left(\beta^0\right) + R_n]^{-1} S_n^{(1)}\left(\beta^0\right),$$

with R_n defined by

$$R_n = \int_0^1 \left[S_n^{(2)}\left(\beta^0 + s\left(\widehat{\beta} - \beta^0\right)\right) - S_n^{(2)}\left(\beta^0\right)\right] ds. \tag{6.8}$$

In other words,

$$\widehat{\beta}-\beta^0=-\left[S_n^{(2)}(\beta^0)+R_n\right]^{-1}\widetilde{S}_n^{(1)}\left(\beta^0,g\right)$$
$$+\left[S_n^{(2)}\left(\beta^0\right)+R_n\right]^{-1}\left[\widetilde{S}_n^{(1)}\left(\beta^0,g\right)-S_n^{(1)}\left(\beta^0\right)\right],$$

where $\widetilde{S}_n^{(1)}(\beta,g)$ and $\widetilde{S}_n^{(2)}(\beta,g)$ denote the first and second derivatives of $\widetilde{S}_n(\beta,g)$ with respect to β. It follows from Lemma 6.1 that $-\sqrt{n}\widetilde{S}_n^{(1)}(\beta^0,g)$ converges in distribution to $\mathscr{N}(0,\Sigma_1)$, and therefore

$$\widehat{\beta}-\beta^0=\left[S_n^{(2)}(\beta^0)+R_n\right]^{-1}\left[\widetilde{S}_n^{(1)}\left(\beta^0,g\right)-S_n^{(1)}\left(\beta^0\right)\right]+O_p\left(n^{-1/2}\right).$$

The rate of convergence of $\widehat{\beta}$ is thus governed by the difference $\widetilde{S}_n^{(1)}(\beta^0,g)-S_n^{(1)}(\beta^0)$. and the proof of Theorem 3.2 lies in checking the three following points:

(i) $\widetilde{S}_n^{(1)}(\beta^0,g)-S_n^{(1)}(\beta^0)=O_p[\delta_{1,n}(f)+n^{-1/2}+\delta_{2,n}(f)]$;
(ii) $S_n^{(2)}(\beta^0)-S^{(2)}(\beta^0)=o_p(1)$, with $S^{(2)}(\beta^0)$ defined in (3.3);
(iii) R_n defined by (6.8) satisfies $R_n=o_p(1)$.

PROOF OF (i). Start by writing $\widetilde{S}_n^{(1)}(\beta^0,g)-S_n^{(1)}(\beta^0)=A_1+A_2+A_3$, where

$$A_1=\frac{2}{n}\sum_{i=1}^n W(Z_i)[Y_i-\Phi_{\beta^0}(Z_i)]\left[\widehat{\Phi}_{\beta^0}^{(1)}(Z_i)-\Phi_{\beta^0}^{(1)}(Z_i)\right],$$

$$A_2=\frac{2}{n}\sum_{i=1}^n W(Z_i)\left[\widehat{\Phi}_{\beta^0}(Z_i)-\Phi_{\beta^0}(Z_i)\right]\left[\widehat{\Phi}_{\beta^0}^{(1)}(Z_i)-\Phi_{\beta^0}^{(1)}(Z_i)\right],$$

$$A_3=\frac{2}{n}\sum_{i=1}^n W(Z_i)\left[\widehat{\Phi}_{\beta^0}(Z_i)-\Phi_{\beta^0}(Z_i)\right]\Phi_{\beta^0}^{(1)}(Z_i).$$

Control of A_1. Using (6.4) write $A_1=B_1+B_2+B_3+o_p(n^{-1/2})$ where

$$B_1=\frac{2}{n}\sum_{i=1}^n\frac{W(Z_i)}{h(Z_i)}[Y_i-\Phi_{\beta^0}(Z_i)]\left[\widehat{\Gamma}_{f^{(1)}}(Z_i)-\Gamma_{f^{(1)}}(Z_i)\right],$$

$$B_2=-\frac{2}{n}\sum_{i=1}^n\frac{W(Z_i)}{h^2(Z_i)}[Y_i-\Phi_{\beta^0}(Z_i)]\left[\widehat{h}(Z_i)-h(Z_i)\right]\Gamma_{f^{(1)}}(Z_i),$$

$$B_3=-\frac{2}{n}\sum_{i=1}^n\frac{W(Z_i)}{h^2(Z_i)}[Y_i-\Phi_{\beta^0}(Z_i)]\left[\widehat{h}(Z_i)-h(Z_i)\right]\left[\widehat{\Gamma}_{f^{(1)}}(Z_i)-\Gamma_{f^{(1)}}(Z_i)\right].$$

Applying Lemma A.2 to B_1 with $F(z)=1/h(z)$, and to B_2 with both $F(z)=1/h^2(z)$ and $f_a^{(1)}\equiv 1$, we can conclude that for any $a=1,\dots,m$, $B_{1,a}=o_p(n^{-1/2})$ and $B_{2,a}=o_p(n^{-1/2})$. Under Assumptions (A5), both $\sigma_n^2(f_a^{(1)})\sqrt{\log n}=o(n)$ and

$M_n^2(f_a^{(1)})\log n = o(n^3)$. Consequently, applying Lemma A.4 after changing f in $f_a^{(1)}$ yields $B_3 = o_p(n^{-1/2})$, which entails that $A_1 = o_p(n^{-1/2})$.

Control of A_2. Let us denote by $A_2 = (A_{2,1}, \dots, A_{2,m})^T$ and apply (6.4) to f and to $f^{(1)}$, just as in A_1 to write $A_{2,a} = C_1 + C_2 + C_3 + o_p(n^{-1/2})$ where

$$C_1 = \frac{2}{n}\sum_{i=1}^{n}\frac{W(Z_i)}{h^2(Z_i)}\left[\widehat{\Gamma}_{f_a^{(1)}}(Z_i) - \Gamma_{f_a^{(1)}}(Z_i)\right]\left[\widehat{\Gamma}_f(Z_i) - \Gamma_f(Z_i)\right],$$

$$C_2 = -\frac{4}{n}\sum_{i=1}^{n}\frac{W(Z_i)}{h^2(Z_i)}\left[\widehat{\Gamma}_f(Z_i) - \Gamma_f(Z_i)][\widehat{h}(Z_i) - h(Z_i)\right]\widehat{\Gamma}_{f_a^{(1)}}(Z_i),$$

$$C_3 = -\frac{4}{n}\sum_{i=1}^{n}\frac{W(Z_i)}{h^2(Z_i)}\left[\widehat{\Gamma}_{f_a^{(1)}}(Z_i) - \Gamma_{f_a^{(1)}}(Z_i)\right]\left[\widehat{h}(Z_i) - h(Z_i)\right]\widehat{\Gamma}_f(Z_i).$$

Applying Lemma A.3 to C_1 ensures that

$$C_1 = O_p\left[M_n(f)M_n(f_a^{(1)})/n^{3/2}\right] + o_p\left[\sigma_n(f)/\sqrt{n} + B_n(f)\right].$$

From Assumption (A5), Lemma A.3 and using the fact that $M_n(f)\sqrt{\log n}/n^{3/2} = o(n^{-1/2})$ together with formulas (3.4) and Proposition 3.1 b) we infer that

$$C_2 = o_p\left[M_n(f)(\log n)^{5/4}\sqrt{\log\log n}/n^{3/2}\right] + o_p\left[n^{-1/2}\right].$$

Arguing as for C_2 (replacing f by $f_a^{(1)}$ in C_2) and using that $M_n(f_a^{(1)})\sqrt{\log n}/n^{3/2} = o(n^{-1/2})$ we obtain $C_3 = o_p[M_n(f)(\log n)^{5/4}\sqrt{\log\log n}/n^{3/2}] + o_p[n^{-1/2}]$ and finally

$$\begin{aligned}A_{2,a} = {} & O_p\left[M_n(f)M_n(f_a^{(1)})/n^{3/2}\right] + o_p\left[M_n(f)(\log n)^{5/4}\sqrt{\log\log n}/n^{3/2}\right]\\ & + o_p\left[\sigma_n(f)/\sqrt{n} + B_n(f) + n^{-1/2}\right].\end{aligned}$$

Control of A_3. In most cases, the rate of convergence of $\widehat{\beta}$ is governed by the rate of convergence of this term which does not depends on the observations Y_i. This arises from the fact that the term A_1 depending on the Y_i's, is centered and then is negligible with respect to the last term A_3.

Using again (6.4) write $A_3 = D_1 + D_2 + D_3 + o_p(n^{-1/2})$ where

$$D_1 = \frac{2}{n}\sum_{i=1}^{n}\frac{W(Z_i)}{h(Z_i)}\left[\widehat{\Gamma}_f(Z_i) - \Gamma_f(Z_i)\right]\Phi_{\beta^0}^{(1)}(Z_i),$$

$$D_2 = \frac{2}{n}\sum_{i=1}^{n}\frac{W(Z_i)}{h(Z_i)}\left[\widehat{h}(Z_i) - h(Z_i)\right]\Phi_{\beta^0}(Z_i)\Phi_{\beta^0}^{(1)}(Z_i),$$

$$D_3 = -\frac{2}{n}\sum_{i=1}^{n}\frac{W(Z_i)}{h^2(Z_i)}\left[\widehat{h}(Z_i) - h(Z_i)\right]\left[\Gamma_f(Z_i) - \widehat{\Gamma}_f(Z_i)\right]\Phi_{\beta^0}^{(1)}(Z_i).$$

Applying Lemma A.1 to D_1 and D_2 we infer that

$$D_1 = O_p\left[\delta_{1,n}(f) + B_n(f)\right] \quad \text{and} \quad D_2 = O_p\left[n^{-1/2}\right].$$

Under Assumption (A5), applying Lemma A.3 to D_3 we can conclude that $D_3 = o_p[n^{-1/2}]$, which entails that $A_3 = O_p\left[\delta_{1,n}(f)/\sqrt{n} + B_n(f) + n^{-1/2}\right]$ and (i) holds.

PROOF OF (ii). Since $\widetilde{S}_n^{(2)}(\beta^0, g) - S^{(2)}(\beta^0) = o_p(1)$ (see the proof of Lemma 6.1), it remains to prove that $S_n^{(2)}(\beta^0) - \widetilde{S}_n^{(2)}(\beta^0, g) = o_p(1)$. Proceeding as for the consistency, write $S_n^{(2)}(\beta^0) - \widetilde{S}_n^{(2)}(\beta^0, g) = E_1 + E_2 + E_3$, where

$$E_1 = \frac{2}{n}\sum_{i=1}^n W(Z_i)\left[\widehat{\Phi}_{\beta^0}^{(1)}(Z_i)\widehat{\Phi}_{\beta^0}^{(1)}(Z_i)^T - \Phi_{\beta^0}^{(1)}(Z_i)\Phi_{\beta^0}^{(1)}(Z_i)^T\right],$$

$$E_2 = \frac{2}{n}\sum_{i=1}^n W(Z_i)Y_i\left[\widehat{\Phi}_{\beta^0}^{(2)}(Z_i) - \Phi_{\beta^0}^{(2)}(Z_i)\right],$$

$$E_3 = \frac{2}{n}\sum_{i=1}^n W(Z_i)\left[\widehat{\Phi}_{\beta^0}(Z_i)\widehat{\Phi}_{\beta^0}^{(2)}(Z_i) - \Phi_{\beta^0}(Z_i))\Phi_{\beta^0}^{(2)}(Z_i)\right].$$

Using equality (6.6) we infer that $E_1 = o_p(1)$, $E_2 = o_p(1)$ and denoting by $(E_3)_{a,b}$ the components of the matrix E_3, that $(E_3)_{a,b} = o_p(1)$ for any integer a, b in $[1, m]$ and (ii) follows.

PROOF OF (iii). Under stated assumptions, the regularity of the function f implies that the same regularity holds for $S_n(\beta)$. Let $\beta_n(s) = \beta^0 + s(\widehat{\beta} - \beta^0)$ and write $S_n^{(2)}[\beta_n(s)] - S_n^{(2)}(\beta^0) = F_1 + F_2 + F_3$ where

$$F_1(s) = \frac{2}{n}\sum_{i=1}^n W(Z_i)G_1(Z_i), \qquad F_3(s) = \frac{2}{n}\sum_{i=1}^n W(Z_i)G_2(Z_i),$$

$$F_2(s) = \frac{2}{n}\sum_{i=1}^n W(Z_i)Y_i\left[\widehat{\Phi}_{\beta_n(s)}^{(2)}(Z_i) - \widehat{\Phi}_{\beta^0}^{(2)}(Z_i)\right],$$

with $G_1(Z_i) = \widehat{\Phi}_{\beta_n(s)}^{(2)}(Z_i)\widehat{\Phi}_{\beta_n(s)}^{(2)}(Z_i)^T - \widehat{\Phi}_{\beta^0}^{(2)}(Z_i)\widehat{\Phi}_{\beta^0}^{(2)}(Z_i)^T$ and $G_2(Z_i) = \widehat{\Phi}_{\beta_n(s)}(Z_i)\widehat{\Phi}_{\beta_n(s)}^{(2)}(Z_i) - \widehat{\Phi}_{\beta^0}(Z_i)\widehat{\Phi}_{\beta^0}^{(2)}(Z_i)$. By linearity of the Fourier transform, and the regularity of f, we get $\widehat{\Phi}_{\beta_n(s)}^{(1)}(Z_i) - \widehat{\Phi}_{\beta^0}^{(1)}(Z_i) = s\widehat{\Phi}_{\beta^\star}^{(2)}(Z_i)(\widehat{\beta} - \beta^0)$,with $|\beta_a^\star| \le |\widehat{\beta}_a - \beta_a^0|$, for any integer a in $[1, m]$. Therefore $F_1(s)$ becomes

$$F_1(s) = s\frac{2}{n}\sum_{i=1}^n W(Z_i)\widehat{\Phi}_{\beta^\star}^{(1)}(Z_i)(\widehat{\beta} - \beta^0)\left[\widehat{\Phi}_{\beta_n(s)}^{(1)}(Z_i) + \widehat{\Phi}_{\beta^0}^{(1)}(Z_i)\right]^T$$

whence, using the consistency of $\widehat{\beta}$, the fact that the weight function $W(.)$ has a compact support together with the compactness of Θ and arguing as for (6.5) (see the proof of the consistency), we get $\int_0^1 F_1(s)ds = o_p(1)$ and the same holds for $\int_0^1 F_2(s)ds$ and $\int_0^1 F_3(s)ds$. □

6.4. *Proof of Corollary* 5.1. Write $\widehat{\sigma}^2$ as $\widehat{\sigma}^2 = G_1 + G_2 + G_3 + G_4$, with G_1, G_2, G_3 and G_4 defined by $G_1 = n^{-1}\sum_{i=1}^n W_n(Z_i)[f^2_{\beta^0}(X_i) - \widehat{\Phi}_{f^2_{\widehat{\beta}}}(Z_i)]$,

$$G_2 = \frac{2}{n}\sum_{i=1}^n W_n(Z_i)[\widehat{\Phi}^2_{\widehat{\beta}}(Z_i) - f_{\beta^0}(X_i)\widehat{\Phi}_{\widehat{\beta}}(Z_i)],$$

$$G_3 = \frac{1}{n}\sum_{i=1}^n W_n(Z_i)\xi_i^2 \quad \text{and} \quad G_4 = \frac{2}{n}\sum_{i=1}^n W_n(Z_i)\xi_i[f_{\beta^0}(X_i) - \widehat{\Phi}_{\widehat{\beta}}(Z_i)].$$

The consistency of $\widehat{\beta}$ and the fact that ξ_i's are centered independent from the Z_i's combined with Assumption (A7) imply that $G_3 \xrightarrow[n\to\infty]{\mathbb{P}} \sigma^2$ and $G_4 = o_p(1)$. Under Assumptions (A4) and (A5) for f^2, arguing as in the proof of Theorem 3.1, we infer both that $G_1 = o_p(1)$ and $G_2 = o_p(1)$ which gives the result. □

APPENDIX

We first proceed to the proof of Lemma 6.1 and second we state and prove technical lemmas about U-statistics used in the proofs of Theorem 3.1 and Theorem 3.2.

PROOF OF LEMMA 6.1. We start with proving that $\widetilde{S}_n(\beta, g) - S(\beta, \beta^0) = o_p(1)$ where $S(\beta, \beta^0) = \mathbb{E}[W(Z)]\sigma^2 + \mathbb{E}\{W(Z)\left[f_{\beta^0}(X) - \Phi_\beta(Z)\right]^2\}$. Write $\widetilde{S}_n(\beta, g)$ as

$$\frac{1}{n}\sum_{i=1}^n W(Z_i)\xi_i^2 + \frac{1}{n}\sum_{i=1}^n W(Z_i)[f_{\beta^0}(X_i) - \Phi_\beta(Z_i)]^2 + \frac{2}{n}\sum_{i=1}^n W(Z_i)\xi_i[f_{\beta^0}(X_i) - \Phi_\beta(Z_i)],$$

part (a) arises immediately from the fact that the ξ_i are centered and independent of Z_i's, from the law of large numbers and from Assumption (A9), (A10) and Theorem 3.2.8, page 126 in [11]. Now write

$$\frac{\partial}{\partial\beta}S(\beta, \beta^0) = S^{(1)}(\beta, \beta^0) = -2\mathbb{E}\left\{W(Z)\Phi^{(1)}_\beta(Z)\left[f_{\beta^0}(X) - \Phi_\beta(Z)\right]\right\}.$$

By Definition (1.2), $\widetilde{S}^{(1)}_n(\widehat{\beta}, g) = 0$. The regularity of f with respect to β and a Taylor expansion ensures that

$$\sqrt{n}(\widetilde{\beta}_g - \beta^0) = -\left[\widetilde{S}^{(2)}_n(\beta^0, g) + R_n\right]^{-1}\sqrt{n}\widetilde{S}^{(1)}_n(\beta^0, g),$$

with

$$R_n = \int_0^1\left[\widetilde{S}^{(2)}_n\left(\beta^0 + s(\widetilde{\beta} - \beta^0), g\right) - \widetilde{S}^{(2)}_n(\beta^0, g)\right]ds.$$

Part (b) follows from Assumptions (A7) and (A8), from the consistency of $\widetilde{\beta}_g$ and from the central limit theorem and the law of large numbers which give that $-\sqrt{n}\widetilde{S}_n^{(1)}(\beta^0, g) \xrightarrow[n\to\infty]{\mathscr{L}} \mathscr{N}(0, \Sigma_1)$, $\widetilde{S}_n^{(2)}(\beta^0, g) \xrightarrow[n\to\infty]{\mathbb{P}} \mathbb{E}[\Phi_{\beta^0}^{(1)}(Z)\Phi_{\beta^0}^{(1)}(Z)^T]$ and $R_n = o_p(1)$. □

LEMMA A.1. *Assume that f satisfies Assumptions* (A4) *and let $F(Z_i)$ be a continuous function and C_n be such that f satisfies Assumption* (A5). *Now consider the following quantity:*

$$U_n = n^{-1}\sum_{i=1}^{n} W(Z_i)F(Z_i)\left[\widehat{\Gamma}_f(Z_i) - \Gamma_f(Z_i)\right].$$

Then we have the following results:

(a) $U_n = O_p\left\{n^{-1/2}\mathbb{E}^{1/2}\mathbb{E}^2\left(W(Z_1)F(Z_1)\mathscr{K}_n(Z_1, Z_2) \mid Z_2\right) + B_n(f)\right\}$.

(b) *If $f \equiv 1$ then replacing $\widehat{\Gamma}_f(Z_i)$ in $\widehat{h}(Z_i)$ we have that* $U_n = O_p[n^{-1/2}]$.

PROOF. Part (a) arises from both $\mathbb{E}(U_n) = O[B_n^2(f)]$ and from $\mathbb{E}(U_n^2) = O\{n^{-1}\mathbb{E}[\mathbb{E}^2(W(Z_1)F(Z_1)\mathscr{K}_n(Z_1, Z_2)|Z_2)]\}$.

Let us start with the study of $\mathbb{E}(U_n)$. Denoting by $\psi_n(Z_i, Z_j)$ the function

$$\psi_n(Z_i, Z_j) = W(Z_i)F(Z_i)[\mathscr{K}_n(Z_i, Z_j) - \Gamma_f(Z_i)],$$

U_n is a U-statistic of order 2 which can be written as follows:

$$U_n = [n(n-1)]^{-1}\sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n}\psi_n(Z_i, Z_j).$$

Arguing as for Proposition 3.1 we get $|\mathbb{E}[\mathscr{K}_n(Z_i, Z_j)|Z_i] - \Gamma_f(Z_i)| \leq \|T^{\star}_{f,Z_i}\overline{\mathbb{1}}_{C_n}\|_1$. Hence, from the independence between Z_i and Z_j $(i \neq j)$ we obtain

$$|E(U_n)| = |\mathbb{E}[\psi_n(Z_i, Z_j)]| \leq B_n(f)\mathbb{E}[W(Z_i)|F(Z_i)|].$$

Now, some crude calculations provide the decomposition

$$\mathbb{E}[U_n^2] = [\Sigma_1 + \Sigma_2 + \Sigma_3 + \Sigma_4](1 + O(n^{-1}))$$

where

$$\Sigma_1 = \mathbb{E}^2[\psi_n(Z_1, Z_2)], \qquad \Sigma_2 = n^{-2}\mathbb{E}[\psi_n^2(Z_1, Z_2)],$$
$$\Sigma_3 = n^{-1}\mathbb{E}[\psi_n(Z_1, Z_2)\psi_n(Z_3, Z_2)], \qquad \Sigma_4 = n^{-1}\mathbb{E}[\psi_n(Z_1, Z_2)\psi_n(Z_1, Z_3)].$$

Arguing as for $\mathbb{E}(U_n)$ we easily obtain that $\Sigma_1 \leq B_n^2(f)\mathbb{E}^2[W(Z_1)|F(Z_1)|]$, and applying Proposition 3.1 provides that $\mathrm{Var}[\mathscr{K}_n(Z_i, Z_j) - \Gamma_f(Z_i)|Z_i] \leq \lambda_n(f, Z_i)$ and

$$\mathbb{E}\left[\psi_n^2(Z_1, Z_2)\right] \leq \left[\sigma_n^2(f) + B_n^2(f)\right]\mathbb{E}\left[W^2(Z_1)\Phi^2(Z_1)\right].$$

Since f satisfies Assumption (A4), $B_n(f) = o(1)$ and hence $\Sigma_2 = O[n^{-2}\sigma_n^2(f)]$.

The following step is to study Σ_3. Since the Z_i's are i.i.d. random variables

$$\mathbb{E}\left[\psi_n(Z_1, Z_2)\psi_n(Z_3, Z_2)\right] = \mathbb{E}\left[\mathbb{E}^2(\psi_n(Z_1, Z_2)|Z_2)\right].$$

But $\mathbb{E}[\mathbb{E}^2(\psi_n(Z_1, Z_2))|Z_2]$ equals

$$\mathbb{E}\left[\mathbb{E}\left(W(Z_1)F(Z_1)\mathrm{Re}\langle T^{\star}_{f,Z_1}(\eta^{\star})^{-1}, h^{\star}_{n,2}\rangle \mid Z_2\right) - \mathbb{E}[W(Z_1)F(Z_1)\Gamma_f(Z_1)]\right]^2$$

which is bounded from above by

$$\mathbb{E}\left[\mathbb{E}^2\left(W(Z_1)F(Z_1)\mathscr{X}_n(Z_1, Z_j) \mid Z_2\right)\right] + \mathbb{E}^2[W(Z_1)F(Z_1)\Gamma_f(Z_1)].$$

Consequently,

$$\Sigma_3 = O\left\{n^{-1}\mathbb{E}\left[\mathbb{E}\left(W(Z_1)F(Z_1)\mathscr{X}_n(Z_1, Z_2) \mid Z_2\right)\right]^2\right\}.$$

When precise calculations are impossible we only use the upper bound

$$\begin{aligned}\mathbb{E}\left[\psi_n(Z_1, Z_2)\psi_n(Z_3, Z_2)\right] &\le \mathbb{E}\left[W^2(Z_1)F^2(Z_1)(\mathscr{X}_n(Z_1, Z_2) - \Gamma_f(Z_1))^2\right]\\ &= O\left[\sigma_n^2(f) + B_n^2(f)\right]\mathbb{E}\left[W^2(Z_1)\Phi^2(Z_1)\right],\end{aligned}$$

which implies, in these cases, $\Sigma_3 = O[n^{-1}\sigma_n^2(f)]$. Now,

$$\mathbb{E}[\psi_n(Z_1, Z_2)\psi_n(Z_1, Z_3)] \le B_n^2(f)\mathbb{E}[W^2(Z_1)F^2(Z_1)],$$

which entails that $\Sigma_4 = O[n^{-1}B_n^2(f)]$. Finally, part (a) of Lemma A.1 follows since

$$\mathbb{E}(U_n^2) = O\left\{n^{-1}\mathbb{E}\left[\mathbb{E}\left(W(Z_1)F(Z_1)\mathscr{X}_n(Z_1, Z_2) \mid Z_2\right)\right]^2 + B_n^2(f)\right\}.$$

When $f \equiv 1$ we have

$$\mathbb{E}\left(W(Z_1)F(Z_1)\mathscr{X}_n^h(Z_1, Z_2) \mid Z_2\right) = \mathbb{E}[W(Z_1)F(Z_1)V_n(Z_2 - Z_1)|Z_2].$$

According to the definition of $\widehat{h}(Z_i)$, $\mathbb{E}[W(Z_1)F(Z_1)V_n(Z_2 - Z_1)|Z_2]$ equals

$$\int W(Z_2 + uh_n)F(Z_2 + uh_n)h(Z_2 + uh_n)V(u)du,$$

which is bounded from above by $\|WFh\|_\infty^2\|V\|_1$. Therefore the expectation $\mathbb{E}\{\mathbb{E}^2[\psi_n(Z_1, Z_2)|Z_2]\}$ is bounded. This combined with (3.1), entail that $U_n = O_p(n^{-1/2})$. □

LEMMA A.2. *Assume that for any integer a in $[1, m]$, $f_a^{(1)}$ satisfies Assumption* (A4) *and let $F(Z_i)$ be a continuous function. Let C_n be such that $f_a^{(1)}$ satisfies Assumption* (A5) *and consider the following quantity*:

$$U_n = n^{-1}\sum_{i=1}^{n} W(Z_i)F(Z_i)[Y_i - \Phi_{\beta^0}(Z_i)]\left[\widehat{\Gamma}_{f_a^{(1)}}(Z_i) - \Gamma_{f_a^{(1)}}(Z_i)\right].$$

then, for any integer a in $[1, m]$ we have $U_n = o_p[n^{-1/2}]$.

PROOF. Denoting by $T_i = (Y_i, Z_i)$ and by $\psi_n(T_i, T_j)$ the function

$$\psi_n(T_i, T_j) = W(Z_i)F(Z_i)[Y_i - \Phi_{\beta^0}(Z_i)][\dot{\mathscr{X}}_{n,j}(Z_i) - \Gamma_{f_a^{(1)}}(Z_i)],$$

U_n can be written $U_n = [n(n-1)]^{-1}\sum_{i=1}^n \sum_{j=1, j\neq i}^n \psi_n(T_i, T_j)$. Since $\widehat{\Phi}_{\beta^0}(Z_i) - \Phi_{\beta^0}(Z_i)$ depends only on the observations Z_i, $\mathbb{E}(U_n) = 0$. Now arguing as in the proof of Lemma A.1 we have

$$\mathbb{E}^2[\psi_n(T_1, T_2)] = \mathbb{E}[\psi_n(T_1, T_2)\psi_n(T_3, T_2)] = 0,$$
$$\mathbb{E}\left[\psi_n^2(T_1, T_2)\right] = O\left[n^{-2}\sigma_n^2(f_a^{(1)})\right]$$

and

$$\mathbb{E}[\psi_n(T_1, T_2)\psi_n(T_1, T_3)] = O\left[n^{-1}B_n^2(f_a^{(1)})\right].$$

It follows by Assumption (A4) that $U_n = O_p[B_n(f_a^{(1)})/\sqrt{n}] = o_p(n^{-1/2})$. □

LEMMA A.3. *Assume that f and $f_a^{(1)}$ satisfy Assumption* (A4) *for any integer a in $[1, m]$ and let $F(Z_i)$ be a continuous function. Let C_n be such that f and $f_a^{(1)}$ satisfy Assumption* (A5) *for any integer a in $[1, m]$ and consider the following quantity*:

$$U_n = n^{-1}\sum_{i=1}^n W(Z_i)F(Z_i)[\widehat{\Gamma}_f(Z_i) - \Gamma_f(Z_i)]\left[\widehat{\Gamma}_{f_a^{(1)}}(Z_i) - \Gamma_{f_a^{(1)}}(Z_i)\right].$$

Then:

(a) $U_n = O_p\left[M_n(f)M_n(f_a^{(1)})/n^{3/2}\right] + o_p[\sigma_n(f)/\sqrt{n} + B_n(f)]$.

(b) *If $f_a^{(1)} = 1$, then replacing $\widehat{\Gamma}_{f_a^{(1)}}(Z_i)$ in $\widehat{h}(Z_i)$ we get that $U_n = o_p[n^{-1/2}]$.*

REMARK A.1. Note that we also have

$$U_n = O_p\left[M_n(f)M_n\left(f_a^{(1)}\right)/n^{3/2}\right] + o_p\left[\sigma_n\left(f_a^{(1)}\right)/\sqrt{n} + B_n\left(f_a^{(1)}\right)\right].$$

PROOF OF LEMMA A.3. Denote by $\phi_n(Z_i, Z_j, Z_k)$ the quantity

$$W(Z_i)F(Z_i)[\mathscr{X}_n(Z_i, Z_j) - \Gamma_f(Z_i)]\left[\mathscr{X}_n^{(1)}(Z_i, Z_k) - \Gamma_{f_a^{(1)}}(Z_i)\right]$$

and write U_n as the sum of two U-statistics respectively of order 3 and order 2: $U_n = (n-2)/(n-1)U_n^{(1)} + U_n^{(2)}$, where $U_n^{(1)}$ and $U_n^{(2)}$ are defined by

$$U_n^{(1)} = \frac{1}{n(n-1)(n-2)}\sum_{i=1}^n \sum_{j\neq k\neq i} \phi_n(Z_i, Z_j, Z_k)$$

and

$$U_n^{(2)} = \frac{1}{n(n-1)^2}\sum_{i=1}^n \sum_{j\neq i} \phi_n(Z_i, Z_j, Z_j).$$

Calculations similar to those done in the proof of Lemma A.1, provide

$$\begin{aligned}\mathbb{E}[U_n^{(1)}]^2 = O\Big[&B_n^2(f)B_n^2\left(f_a^{(1)}\right) + n^{-1}B_n^2(f)B_n^2\left(f_a^{(1)}\right)\\ &+n^{-1}B_n^2\left(f_a^{(1)}\right)\left[\sigma_n^2(f)+B_n^2(f)\right]\\ &+n^{-1}B_n^2(f)\left[\sigma_n^2\left(f_a^{(1)}\right)+B_n^2\left(f_a^{(1)}\right)\right]\\ &+n^{-2}B_n^2\left(f_a^{(1)}\right)\left[\sigma_n^2(f)+B_n^2(f)\right]\\ &+n^{-2}B_n^2(f)\left[\sigma_n^2\left(f_a^{(1)}\right)+B_n^2\left(f_a^{(1)}\right)\right]\\ &+n^{-2}\left[\sigma_n^2(f)+B_n^2(f)\right]\left[\sigma_n^2\left(f_a^{(1)}\right)+B_n^2\left(f_a^{(1)}\right)\right]\\ &+n^{-3}\left[\sigma_n^2(f)+B_n^2(f)\right]\left[\sigma_n^2\left(f_a^{(1)}\right)+B_n^2\left(f_a^{(1)}\right)\right]\Big].\end{aligned}$$

Since C_n is such that f and $f_a^{(1)}$ satisfy Assumption (A5), we have $\mathbb{E}[U_n^{(1)}]^2 = o[n^{-1}\sigma_n^2(f) + B_n^2(f)]$. Note that by symmetry we also have that $\mathbb{E}[U_n^{(1)}]^2 = o[n^{-1}\sigma_n^2(f_a^{(1)})+B_n^2(f_a^{(1)})]$ and $\mathbb{E}[U_n^{(1)}]=o[B_n(f)]=o[B_n(f_a^{(1)})]$. Finally, $\mathrm{Var}(U_n^{(1)}) = o[n^{-1}\sigma_n^2(f) + B_n^2(f)]$ and $\mathrm{Var}(U_n^{(2)}) = O[n^{-2}\sigma_n^2(f)\sigma_n^2(f_a^{(1)}) + n^{-3}M_n^2(f) M_n^2(f_a^{(1)})]$, where the result for $U_n^{(2)}$ follows from the fact that $\mathbb{E}(U_n^{(2)}) = O[n^{-1}\sigma_n(f)\sigma_n(f_a^{(1)})] = o[n^{-1/2}\sigma_n(f)]$.

When $f_a^{(1)} \equiv 1$, $\sigma_n^2(f_a^{(1)}) = O(\sqrt{\log n})$, $M_n(f_a^{(1)}) = O(\sqrt{\log n})$ which together with (3.1) provide the result. Indeed, since C_n is such that f_β satisfies Assumption (A5) we have $n^{-1}\sigma_n(f)(\log n)^{1/4} = o(n^{-1/2})$ and $n^{-3/2}M_n(f)\sqrt{\log n} = o(n^{-1/2})$. □

LEMMA A.4. *Assume that f and $f_a^{(1)}$ satisfy Assumptions* (A4) *for any integer a in $[1, m]$ and let $F(Z_i)$ be a continuous function. Let C_n be such that f and $f_a^{(1)}$ satisfy Assumptions* (A5) *for any integer a in $[1, m]$ and consider the following quantity*:

$$U_n = n^{-1}\sum_{i=1}^n W(Z_i)F(Z_i)[Y_i - \Phi_{\beta^0}(Z_i)]\left[\widehat{\Gamma}_f(Z_i)-\Gamma_f(Z_i)\right]\left[\widehat{h}(Z_i)-h(Z_i)\right].$$

Then $U_n = o_p[n^{-1/2}]$.

PROOF. We proceed as for the case of Lemma A.3 using the fact that $\mathbb{E}[Y_i - \Phi_{\beta^0}(Z_i)|Z_i] = 0$. □

Acknowledgments. Research for this paper was partially carried while on leave at Unité de Biométrie, INRA, 78352 Jouy-en-Josas Cedex, France.

I thank S. Huet and P. Massart for supervising during the preparation of my Ph. D thesis which includes this work. I also thank A.W. Van der Vaart for helpful discussions, an Associate Editor and the referee for helpful suggestions.

REFERENCES

[1] ANDERSON, T. W. (1984). Estimating linear statistical relationships. *Ann. Statist.* **12** 1–45.

[2] BICKEL, P. J., KLAASSEN, A. J. C., RITOV, Y. and WELLNER, J. A. (1993). *Efficient and Adaptative Estimation for Semiparametric Model.* Johns Hopkins Univ. Press.

[3] BICKEL, P. J. and RITOV, A. J. C. (1987). Efficient estimation in the errors-in-variables model. *Ann. Statist.* **15** 513–540.

[4] BIRGÉ, L. and MASSART, P. (1993). Rates of convergence for minimum contrast estimators. *Probab. Theory Related Fields* **97** 113–150.

[5] CARROLL, R. J., KÜCHENHOFF, H., LOMBARD, F. and STEFANSKI, L. A. (1996). Asymptotics for the SIMEX estimator in nonlinear measurement error models. *J. Amer. Statist. Assoc.* **91** 242–250.

[6] CARROLL, R. J. and HALL, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.* **83** 1184–1186.

[7] CARROLL, R. J. and STEFANSKI, L.A. (1990). Approximate quasilikelihood estimation in models with surrogate Predictors. *J. Amer. Statist. Assoc.* **85** 652–663.

[8] CHAN, L. K. and MAK, T. K. (1985). On the polynomial functional relationship. *J. Roy. Statist. Soc. Ser. B* **47** 510–518.

[9] CHENG, C. H. and VAN NESS, J. W. (1994). On estimating linear relationships when both variables are subject to errors. *J. Roy. Statist. Soc. Ser. B* **56** 167–183.

[10] COOK, J. R. and STEFANSKI, L. A. (1994). Simulation-extrapolation estimation in parametric measurement error models. *J. Amer. Statist. Assoc.* **89** 1314–1328.

[11] DACUNHA-CASTELLE, D. and DUFLO, M. (1986). *Probability and Statistics 2*. Springer, Berlin.

[12] DAVIS, K. B. (1975). Mean square error properties of density estimates. *Ann. Statist.* **3** 1025–1030.

[13] DAVIS, K. B. (1977). Mean integrated square error properties of density estimates. *Ann. Statist.* **5** 530–535.

[14] FAN, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19** 1257–1272.

[15] FAN, J. and MASRY, E. (1992). Multivariate regression estimation with errors-in-variables: asymptotic normality for mixing processes. *J. Multivariate Anal.* **43** 237–271.

[16] FAN, J., TRUONG, Y. K. and WANG, Y. (1991). Nonparametric function estimation involving errors-in-variables. In *Nonparametric Functional Estimation and Related Topics* 613-627. Kluwer, Dordrecht.

[17] FAN, J. and TRUONG, Y. K. (1993). Nonparametric regression with errors in variables. *Ann. Statist.* **21** 1900–1925.

[18] GLESER, L. J. (1981). Estimation in a multivariate “errors in variables” regression model: large sample results. *Ann. Statist.* **9** 24–44.

[19] GLESER, L. J. (1985). A note on G.R. Dolby’s unreplicated ultrastructural model. *Biometrika* **72** 117–124.

[20] GLESER, L. J. (1990). Improvements of the naive approach to estimation in nonlinear errors-in-variables regression models. Statistical analysis of measurement error models with applications. *Contemp. Math.* **112** 99–114

[21] GOLUBEV, G. K. and LEVIT, B. Y. (1996). Asymptotically efficient estimation for analytic distributions. *Math. Methods Statist.* **5** 357–368.

[22] HAUSMAN, J. A., NEWEY, W. K. ICHIMURA, K. and POWELL, J. L. (1991). Identification and estimation of polynomial errors-in-variables models. *J. Econometrics* **50** 273–295.

[23] 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.

[24] HSIAO, C. (1989) Consistent estimation for some nonlinear errors-in-variables models. *J. Econometrics* **41** 159–185.

[25] IBRAGIMOV, I. A. and HASMINSKII, R. Z. (1982). Estimation of distribution density belonging to a class of entire function. *Theory Probab. Appl.* **27** 551–562.

[26] IBRAGIMOV, I. A and HASMINSKII, R. Z. (1983). Estimation of distribution density. *J. Soviet Math.* **21** 40–57.

[27] KIEFER, J. and WOLFOWITZ, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many nuisance parameters. *Ann. Math. Statist.* **27** 887–906.

[28] KUKUSH, A. and ZWANZIG, S. (1996). On inconsistency of the least squares estimator in nonlinear functional errors-in-variables models. Unpublished manuscript.

[29] LEDOUX, M. (1996). On Talagrand's deviation inequalities for product measures. *ESAIM: Probab. Statist.* **1** 63–87.

[30] MASSART, P. and RIO, E. (1998). A uniform Marcinkiewicz-Zygmund strong law of large numbers for empirical processes. In *Asymptotic Methods in Probability and Statistics. A Volume in Honour of Miklós Csörgö* 199-211. North-Holland, Amsterdam.

[31] NIKOL'SKIĬ, S. M. (1975). *Approximations of Functions of Several Variables and Embedding Theorems.* Springer, Berlin.

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

[33] RUDIN, W. (1987). *Real and Complex Analysis* McGraw-Hill, New York.

[34] SEPANSKI, J. H. and CARROLL, R. J. (1993). Semiparametric quasilikelihood and variance function estimation in measurement error models. *J. Econometrics* **58** 223–256.

[35] TAUPIN, M. L. (1998). Semiparametric estimation in the nonlinear errors-in-variables model. Ph.D. thesis, Université Paris-Sud.

[36] VAN DER VAART, A. W. (1988). Estimating a real parameter in a class of semiparametric models. *Ann. Statist.* **16** 1450–1474.

[37] VAN DER VAART, A. W. (1996). Efficient maximum likelihood estimation in semiparametric mixture models. *Ann. Statist.* **24** 862–878.

[38] WOLTER, K. M. and FULLER, W. A. (1982). Estimation of nonlinear errors-in-variables models. *Ann. Statist.* **10** 539–548.

[39] ZHANG, C. H. (1990). Fourier methods for estimating mixing densities and distributions. *Ann. Statist.* **18** 806–831.

[40] ZWANZIG, S. (1990). On consistency in nonlinear functional relations. Unpublished manuscript.

LABORATOIRE DE PROBABILITÉS,
STATISTIQUE ET MODÉLISATION
UMR 8628
UNIVERSITÉ PARIS-SUD, BÂT. 425
91405 ORSAY CEDEX
FRANCE
E-MAIL: Marie-Luce.Taupin@math.u-psud.fr