NONPARAMETRIC ESTIMATION OF THE REGRESSION FUNCTION IN AN ERRORS-IN-VARIABLES MODEL

F. COMTE*,[1] AND M.-L. TAUPIN[2]

Abstract. In this paper, we consider the regression model with errors-in-variables: $Y_i = f(X_i) + \xi_i,\ Z_i = X_i + \sigma\varepsilon_i$, where ε and ξ are i.i.d. noises, independent of X, the density of $\sigma\varepsilon$, $(1/\sigma)f_\varepsilon(./\sigma)$, is known. Our aim is to estimate the regression function f from a sample (Y_i, Z_i), $i = 1, \cdots, n$, of observations. We propose to estimate f by the ratio of two penalized minimum contrast estimators of $\ell = fg$ and of g, where g denotes the density of the X_i's. We prove that the $\mathbb{L}_2$-risk of our estimator on a compact set is given by the sum of the $\mathbb{L}_2(\mathbb{R})$-risk of the estimator of $\ell = fg$ and of the $\mathbb{L}_2(\mathbb{R})$-risk of the estimator of g. We give the rate of convergence of such estimators for various smoothness classes for ℓ and g, when the errors ε are either ordinary smooth or super smooth. The resulting rate is optimal in a minimax sense when both ℓ and g belong to the same Sobolev class and when the errors are ordinary or super smooth. In the other cases, our estimator automatically adjusts.

October 2004

MSC 2000 Subject Classification. Primary 62G07, 62G08. Secondary 62G05, 62G20.
Keywords and phrases. Adaptive estimation. Errors-in-variables. Density deconvolution. $\mathbb{L}_2$ estimation. Minimax estimation. Nonparametric regression. Projection estimators.

1. Introduction

We consider in this paper the following errors-in-variables model

$$\begin{cases} Y_i = f(X_i) + \xi_i \\ Z_i = X_i + \sigma\varepsilon_i. \end{cases} \tag{1.1}$$

We assume that the errors $(\xi_1, \cdots, \xi_n)$ and $(\varepsilon_1, \cdots, \varepsilon_n)$ are two independent sequences of centered independent and identically distributed (i.i.d.), unobserved random variables. The explanatory random variables $X_1, \cdots, X_n$ are i.i.d. but unobserved with density denoted by g. The errors ε's have common density known density f_ε and the noise level σ is known . Our aim is to estimate the regression function f by using the observations (Y_i, Z_i) for $i = 1, \ldots, n$, on a compact set denoted by A.

Many papers on the subject study some parametric or semi-parametric frameworks but we consider here only previous known results in the general nonparametric case. In most previous such papers, the proposed estimators are some Nadaraya-Watson kernel type estimator, based on the classical idea that the regression function f at point x can be written as the conditional

[1] Université Paris V, MAP5, UMR CNRS 8145, 45 rue des Saints-Pères, 75 270 PARIS cedex 06, France. email: fabienne.comte@univ-paris5.fr.
[2] IUT de Paris V et Université d'Orsay, Laboratoire de Probabilités, Statistique et Modélisation, UMR 8628, Bâtiment 425,91405 Orsay Cedex, France.- email: Marie-Luce.Taupin@math.u-psud.fr.

expectation $f(x) = \mathbb{E}(Y|X = x)$, and so as the ratio

$$f(x) = \frac{\int y f_{X,Y}(x,y)dy}{g(x)} = \frac{(fg)(x)}{g(x)},$$

with $f_{X,Y}$ the joint density of (X, Y). All those estimators are constructed as the ratio of two deconvolution kernel type estimators that is

$$\hat{f}(x) = \frac{(h_n n)^{-1}\sum_{i=1}^n Y_i \widetilde{K_n}((x - Z_i)/h_n)}{(h_n n)^{-1}\sum_{i=1}^n \widetilde{K_n}[(x - Z_i)/h_n]},$$

with the deconvolution kernel $\widetilde{K_n}$ such that $\widetilde{K_n}^*(u) = K^*(u)/f^*_{\sigma\varepsilon}(u/h)$, where $r^*(t) = \int e^{itx} r(x)dx$ denotes the Fourier transform of r. Among the others, one can cite Fan *et al.* (1991), Fan and Masry (1992), Fan and Truong (1993), Masry (1993), Truong (1991), Ioannides and Alevizos (1997). In those papers, such estimators are studied from many point of view ($\mathbb{L}_p$-risks, asymptotic normality...), and under various assumptions (regularity, mixing...). One assumption usually done, is that the regularity of the regression function f and the regularity of the density g of the design are equal. In particular, when the regression function f and the density g admit kth derivatives, Fan and Truong (1993) give upper and lower bounds of the minimax risk for quadratic pointwise risk and for $\mathbb{L}_p$-risk on compact set for ordinary (polynomial decay) and super smooth (exponential decay) errors ε.

In a slightly different way, Koo and Lee (1998) propose a B-spline method for nonparametric regression function estimation when the errors are ordinary smooth. This method also relates to estimation of the regression function as a ratio of two estimators.

All those works show that, as in the deconvolution framework, the worst rates of convergence are obtained for the smoother error density f_ε. Moreover all those previous papers consider that the regression function belongs to a specified known smoothness class and assume the same order of regularity for both functions f and g.

We propose here to construct an estimator $\tilde{f}$ of the regression function f, without any prior information on the regularity of the unknown functions f and g. We propose to estimate f as a ratio of an adaptive estimator $\tilde{\ell}$ of $\ell = fg$ and of an adaptive estimator $\tilde{g}$ of g. Those estimators are based on projection methods, and choose automatically the relevant projection spaces, without any prior information on f or g.

We give upper bounds for the $\mathbb{L}_2$-risk on a compact set and show, in particular, that the rate of convergence of our estimator $\tilde{f}$ of f is given by the sum $\mathbb{E} \parallel \tilde{\ell} - \ell \parallel_2 + \mathbb{E} \parallel \tilde{g} - g \parallel_2$, that is by the slower rate between the rate for the adaptive estimation of g and the rate for the adaptive estimation of $\ell = fg$. Those rates of convergence are described by the smoothness properties of $\ell = fg$ and of g in term of the sets

$$\mathcal{S}_{\alpha,\nu,\rho}(\kappa_\alpha) = \{\psi \ : \ \text{such that} \int_{-\infty}^{+\infty} |\psi^*(x)|^2 (x^2+1)^\alpha \exp\{2\rho|x|^\nu\}dx \le \kappa_\alpha\},$$

for $\alpha, \nu, \rho, \kappa_\alpha$ some nonnegative real numbers, and also, as in deconvolution problems, by the rate of decrease of the Fourier transform of f_ε given by

$$\kappa_0(x^2+1)^{-\gamma/2}\exp\{-B|x|^\delta\} \le |f^*_\varepsilon(x)| \le \kappa'_0(x^2+1)^{-\gamma/2}\exp\{-B|x|^\delta\}, \text{ for all } x \text{ in } \mathbb{R}.$$

One consequence is that if $\ell = fg$ belongs to some set $\mathcal{S}_{\alpha_\ell,0,\rho_\ell}(\kappa_{\alpha_\ell})$ and g belongs to some $\mathcal{S}_{\alpha_g,0,\rho_g}(\kappa_{\alpha_g})$, that is if they belong to some Sobolev classes, with $\alpha_\ell = \alpha_g = k$ then the resulting rate is the same as the one obtained for regression function f and density g having kth derivatives, either for ordinary smooth errors ($\delta = 0$) or for super smooth errors $\delta > 0$. Therefore, our estimator $\tilde{f}$ achieves the optimal rate of convergence, established by Fan and Truong (1993). It seems reasonable to think that $\tilde{f}$ also achieves the optimal rate of convergence, in the case where $\alpha_g > \alpha_\ell$ and even in the case where $\nu_g > \nu_\ell > 0$, that is, in cases where the density g is smoother than f. In this last case the optimality may be up to some logarithmic loss, but such a loss only appears in cases where the rate is faster than any power of $\ln(n)$, as the price to pay for adaptation. The question of the optimality of our estimator when g is less regular than f remains open and raises some important question. Indeed, it is known that, when the X_i's are observed, that is in the classical regression framework, then the Nadaraya-Watson type estimator does not achieve the optimal rate of convergence since its rate also depends on the rate of convergence for the estimation of g, whereas other estimators like projection estimator or local polynomials estimators achieve optimal rates without prior regularity assumptions on the density of the design g. We refer for instance to Tsybakov (1986) or Fan and Gijbels (1996) for references on this subject. The question is thus : why does the rate of convergence of our estimator depend on the rate of convergence for the estimation of g? Does this dependence appear because the X_i's are not observable or because we estimate the conditional expectation by a ratio? This question remains open since classical methods like projection or local polynomials estimators do not seem to work in the context of errors-in-variables model.

Nevertheless, our procedure provides an adaptive estimator, in the sense that its construction does not require prior knowledge on f nor on g, which seems optimal in a large part of considered cases, and for various types of smoothness and various types of errors.

The paper is organized as follows. First, we describe the construction of the estimators, starting with projection spaces and contrast functions. Section 3 is devoted to the presentation of the non penalized estimators with some discussions about their optimality. The adaptive procedure and the results coming with are given in Section 4. All proofs and technical lemmas are gathered in Section 5.

2. Notation and assumptions

2.1. **Notations.** Subsequently we denote by u^* the Fourier transform of the function u defined as $u^*(t) = \int e^{itx}u(x)dx$, and by $\|u\|_1$, $\|u\|_2$, $\|u\|_\infty$, $\|u\|_{\infty,K}$, $< u, v >$ the quantities

$$\|u\|_1 = \int |u(x)|dx,\ \|u\|_2^2 = \int u^2(x)dx,\ \|u\|_\infty = \sup_{x\in\mathbb{R}} |u(x)|,$$

$$\|u\|_{\infty,K} = \sup_{x\in K} |u(x)| \quad \text{and} \quad < u, v >= \int u(x)\overline{v}(x)dx \text{ with } z\overline{z} = |z|^2.$$

2.2. **Assumptions.** Consider Model (1.1) under the following assumptions.

($\mathbf{A_1}$) The $X_i's$, the ξ_i's and the ε_i's are i.i.d. random variables.

($\mathbf{A_2}$) The sequences $(X_i)_{i\in\mathbb{N}}$, $(\xi_i)_{i\in\mathbb{N}}$ and $(\varepsilon_i)_{i\in\mathbb{N}}$ are independent.

($\mathbf{A_3}$) All variables have finite variance.

($\mathbf{A_4}$) f_ε^* belongs to $\mathbb{L}_2(\mathbb{R})$, for all $x \in \mathbb{R}, f_\varepsilon(-x) = f_\varepsilon(x)$ and $f_\varepsilon^*(x) \neq 0$.

($\mathbf{A_5}$) The function $\ell = fg$ belongs to the set

$$\mathcal{L} = \left\{\phi \text{ such that } \int \phi^2(x)dx < \infty, \text{ and } \int x^2\phi^2(x)dx < \kappa_{\mathcal{L}} < \infty\right\},$$

($\mathbf{A_6}$) Let G denote the support of g. The regression function f belongs to the set $\mathcal{F}_G = \{\phi \text{ such that } \sup_{x\in G} |\phi(x)| \leq \kappa_{\infty,G} < \infty\}$.

($\mathbf{A_7}$) The density g belongs to the set

$$\mathcal{G} = \{\phi, \text{ density, such that } \int \phi^2(x)dx < \infty, \text{ and } \int x^2\phi^2(x)dx < \kappa_{\mathcal{G}} < \infty\}.$$

($\mathbf{A_8}$) There exist g_0, g_1 positive constants such that for all $x \in A, g_0 \leq g(x) \leq g_1$.

Assumption ($\mathbf{A_4}$) which is quite usual in density deconvolution and also in nonparametric regression estimation in errors-in-variables models, ensures that the density g is identifiable.

Note that we do not assume that g is compactly supported but only that f is bounded on the support of g. It follows that if g is compactly supported then f has to be bounded on a compact set. But if g has $\mathbb{R}$ as support then the regression function has to be bounded on $\mathbb{R}$. Nevertheless, we estimate f, only on a compact set denoted by A. Moreover, Assumption ($\mathbf{A_8}$) implies that $A \subset G$ and therefore under ($\mathbf{A_6}$) and ($\mathbf{A_8}$), f is bounded on A. Assumptions ($\mathbf{A_6}$) and ($\mathbf{A_7}$) imply that ($\mathbf{A_5}$) holds, with $\kappa_{\mathcal{L}} = \kappa_{\infty,G}^2\kappa_{\mathcal{G}}$.

According to Assumption ($\mathbf{A_2}$) the (unknown density) h of the Z_i's is given by $h(z) = (1/\sigma)g * f_\varepsilon(./\sigma)(z)$ and its Fourier transform satisfies that $h^*(.) = g^*(.)f_\varepsilon^*(\sigma.)$. It is well known that, as in deconvolution problems, the rate of convergence for estimating f is strongly related to the decrease of $f_\varepsilon^*(x)$ when x goes to infinity, with the slowest rates of convergence obtained for the smoothest error density f_ε, that is for densities having a fast decreasing Fourier transform. This decrease is described as follows.

($\mathbf{A_9}$) $\kappa_0(x^2+1)^{-\gamma/2}\exp\{-B|x|^\delta\} \leq |f_\varepsilon^*(x)| \leq \kappa_0'(x^2+1)^{-\gamma/2}\exp\{-B|x|^\delta\}$, for all x in $\mathbb{R}$.

Let us comment Assumption ($\mathbf{A_9}$). First note that only its left-hand side is required for upper bounds whereas the right-hand side is needed when we consider lower bounds and optimality problems. This assumption implies the part $f_\varepsilon^*(x) \neq 0$ of ($\mathbf{A_4}$). Moreover ($\mathbf{A_4}$) and ($\mathbf{A_9}$) imply that γ has to be such that $\gamma > 1/2$ when $\delta = 0$. Furthermore, when $\delta = 0$ in Assumption ($\mathbf{A_9}$), it amounts to consider what is usually called "ordinary smooth" errors, and "super smooth" errors when $\delta > 0$. The standard examples are the following : Gaussian or Cauchy distributions are super smooth of order ($\gamma = 0$, $\delta = 2$) and ($\gamma = 0$, $\delta = 1$) respectively, and the double exponential distribution is ordinary smooth ($\delta = 0$) of order $\gamma = 2$. Classically, the slowest rate of convergence for estimating f and g are obtained for super-smooth error density. In particular, when f_ε is the Gaussian density the optimal rate of convergence obtained by Fan and Truong (1993) when f and g have the same Hölderian type regularity is of order power of $\ln(n)$. As in deconvolution problems, the rates of convergence can be improved by some additional regularity conditions on f and g. In that way we consider the set of functions

$$\mathcal{S}_{\alpha,\nu,\rho}(\kappa_\alpha) = \{\psi \ : \ \text{such that} \int_{-\infty}^{+\infty} |\psi^*(x)|^2(x^2+1)^\alpha \exp\{2\rho|x|^\nu\}dx \leq \kappa_\alpha\}, \tag{2.1}$$

for $\alpha, \nu, \rho, \kappa_\alpha$ some nonnegative real numbers. The smoothness class in (2.1) is classically considered in nonparametric estimation. When $\nu = 0$, this corresponds to Sobolev spaces of order α. The densities belonging to $\mathcal{S}_{\alpha,\nu,\rho}(\kappa_\alpha)$ with $\nu > 0, \rho > 0$ are infinitely many times differentiable. Moreover, such densities admit analytic continuation on a finite width strip when $\nu = 1$ and on the whole complex plane if $\nu = 2$.

By convention, we set $B = 0$ when $\delta = 0$ and we assume that $B > 0$ when $\delta > 0$. In the same way, if $\sigma = 0$, the X_i's are directly observed without noise and we set $B = \gamma = \delta = 0$ in this case.

3. Projection spaces and estimators

3.1. **Projection spaces.** Let us define $\varphi(x) = \sin(\pi x)/(\pi x)$, and $\varphi_{m,j}(x) = \sqrt{D_m}\varphi(D_m x - j)$, where $D_m = 2^m, D_m \leq n$, is the analogue of the dimension in usual finite dimensional problems. Here we take $D_m = 2^m$, but most of results still hold with $D_m \in \{1, \cdots, n\}$. It is well known (see Meyer (1990), p.22) that $\{\varphi_{m,j}\}_{j\in\mathbb{Z}}$ is an orthonormal basis of the space of square integrable function having Fourier transform with compact support included into $[-\pi D_m, \pi D_m]$. We denote by S_m such a space and by $(S_m)_{m\in\mathcal{M}_n}$, with $\mathcal{M}_n = \{1, \ldots, m_n\}$, this collection of linear spaces that is

$$S_m = \text{Vect}\{\varphi_{m,j},\ j \in \mathbb{Z}\} = \{u \in \mathbb{L}^2(\mathbb{R}),\ \text{with support}(u^*) \text{ included into } [-D_m\pi, D_m\pi]\}.$$

Denoting by ℓ_m and g_m the orthogonal projection of $\ell = fg$ and g on S_m, then ℓ_m and g_m are given by

$$\begin{aligned} \ell_m &= \sum_{j\in\mathbb{Z}} a_{m,j}(\ell)\varphi_{m,j} \text{ with } a_{m,j}(\ell) =< \varphi_{m,j}, \ell >, \\ \text{and } g_m &= \sum_{j\in\mathbb{Z}} a_{m,j}(g)\varphi_{m,j} \text{ with } a_{m,j}(g) =< \varphi_{m,j}, g > . \end{aligned}$$

Since those orthogonal projections ℓ_m and g_m involve infinite sums we shall consider some truncated versions. For k_n an integer depending on n, we consider the truncated spaces

$$S_m^{(n)} = \text{Vect}\{\varphi_{m,j},\ \ j \in \mathbb{Z},\ \ |j| \leq k_n\}.$$

Consequently, if $\ell_m^{(n)}$ and $g_m^{(n)}$ are the $\mathbb{L}_2$-orthogonal projection of ℓ and g on $S_m^{(n)}$, then

$$\ell_m^{(n)} = \sum_{|j|\leq k_n} a_{m,j}(\ell)\varphi_{m,j} \quad \text{and} \quad g_m^{(n)} = \sum_{|j|\leq k_n} a_{m,j}(g)\varphi_{m,j}$$

with still $a_{m,j}(\ell) =< \ell, \varphi_{m,j} >$ and $a_{m,j}(g) =< g, \varphi_{m,j} >$.

3.2. **First step: the estimator of $\ell = fg$.** Associate this collection of models $(S_m^{(n)})_{m\in\mathcal{M}_n}$ to the following contrast function, for t belonging to some $S_m^{(n)}$:

$$\gamma_n(t) = \|t\|^2 - \frac{2}{n}\sum_{i=1}^n (Y_i u_t^*(Z_i)), \quad u_t(x) = \frac{1}{2\pi}\frac{t^*}{f_\varepsilon^*}(-x). \tag{3.1}$$

By using that under Assumption $(\mathbf{A_4})$ f_ε^* is an even function combined with Parseval and inverse Fourier formulas we get that

$$\mathbb{E}(Y_1 u_t^*(Z_1)) \;=\; \mathbb{E}(f(X_1)u_t^*(Z_1)) = \langle u_t^* * f_\varepsilon, fg\rangle = \frac{1}{2\pi}\langle \frac{t^*}{f_\varepsilon^*}f_\varepsilon^*, (fg)^*\rangle = \frac{1}{2\pi}\langle t^*, (fg)^*\rangle = \langle t, \ell\rangle.$$

Therefore, we find that $\mathbb{E}(\gamma_n(t)) = \|t\|_2^2 - 2\langle \ell, t\rangle = \|t-\ell\|_2^2 - \|\ell\|_2^2$ which is minimal when $t = \ell$. This shows that $\gamma_n(t)$ suits well for the estimation of $\ell = fg$. Associated to those projection spaces and contrast, consider the collection of estimators $\hat\ell_m$ of ℓ built as

$$\hat\ell_m = \text{ Argmin }_{t\in S_m^{(n)}}\gamma_n(t). \tag{3.2}$$

Consequently we have

$$\hat\ell_m = \sum_{|j|\le k_n}\hat a_{m,j}(\ell)\varphi_{m,j} \;\text{ with }\; \hat a_{m,j}(\ell) = \frac{1}{n}\sum_{i=1}^n Y_i u^*_{\varphi_{m,j}}(Z_i), \tag{3.3}$$

with k_n some positive integer suitably chosen.

3.3. **Second step: the estimator of g.** The second stage of the estimation procedure is to find an estimator of g, and we simply use the one proposed in Comte and Taupin (2003). We define $\hat g_m$ the estimator of g on $S_m^{(n)}$ by

$$\hat g_m = \sum_{|j|\le k_n}\hat a_{m,j}(g)\varphi_{m,j} \;\text{ with }\; \hat a_{m,j}(g) = \frac{1}{n}\sum_{i=1}^n u^*_{\varphi_{m,j}}(Z_i), \tag{3.4}$$

which can also be viewed as the minimum contrast estimator on $S_m^{(n)}$, associated with

$$\gamma_{n,g}(t) = \|t\|_2^2 - \frac{2}{n}\sum_{i=1}^n u_t^*(Z_i).$$

3.4. **Last step: the estimator of f.** Under Assumption $(\mathbf{A_8})$, for some m_1 and m_2, we define the trimmed estimator

$$\hat f_{m_1,m_2} = \left(\frac{\hat\ell_{m_1}}{\hat g_{m_2}}\right)^{(a_n)}, \text{ where } \left(\frac{x}{y}\right)^{(d)} = \begin{cases} d \text{ sign }\left(\dfrac{x}{y}\right) & \text{if } |x| \ge d|y| \\ \dfrac{x}{y} \text{ else }, & \end{cases} \tag{3.5}$$

with $\hat\ell_{m_1}$ is associated to the projection estimator on a space $S_{m_1}^{(n)}$, $\hat g_{m_2}$ is obtained as a projection estimator on a space $S_{m_2}^{(n)}$ and a_n being suitably chosen. Clearly this construction is used in order to avoid the problems that may occur when $\hat g_{m_2}$ takes small values.

4. Risk bound of the non penalized estimator $\hat{f}_{m_\ell,m_g}$

4.1. **Risk bound of the non penalized estimator $\hat{\ell}_m$.**

Proposition 4.1. *Consider the estimator $\hat{\ell}_{D_m} = \hat{\ell}_m$ of ℓ defined by (3.2) where $\ell = fg$ with f and g as in model (1.1). Let Δ be the quantity*

$$\Delta(m) = \frac{D_m}{\pi} \int_0^{\pi D_m} |f_\varepsilon^*(D_m x\sigma)|^{-2} dx. \tag{4.1}$$

Then, under Assumptions ($\mathbf{A_1}$)–($\mathbf{A_5}$),

$$\mathbb{E}(\|\ell - \hat{\ell}_m\|_2^2) \leq \|\ell - \ell_m\|_2^2 + 2\frac{\Delta(m)}{n}\mathbb{E}(Y_1^2) + \frac{D_m^2}{k_n}(\kappa_{\mathcal{L}} + \|\ell\|_1). \tag{4.2}$$

First note that the bias term $\|\ell - \ell_m\|^2$ depends, as usual, on the smoothness of the function g. It has the expected order for classical smoothness classes since it is given by the distance between g and the classes of entire functions having Fourier transform compactly supported on $[-\pi D_m, \pi D_m]$ (see Ibragimov and Hasminskii (1983)).

Secondly, as in deconvolution problems, the variance term $\Delta(m)/n$ depends on the rate of decay of the Fourier transform of f_ε, with larger variance for smoother f_ε. Lastly, let us precise the conditions under which the third term is negligible. In a general setting,under Assumptions ($\mathbf{A_4}$) and ($\mathbf{A_9}$), γ has to be such that $\gamma > 1/2$. Since it is reasonable to consider dimensions D_m such that $\Gamma(m)/n$ is bounded, therefore we only consider $D_m \leq \sqrt{n}$. Hence a rough condition on k_n is $k_n \geq n^2$ in order to obtain $D_m^2/k_n \leq 1/n$ and the last term in (4.2) is negligible with respect to the others.

Those previous remarks are precised in the following corollary.

Corollary 4.1. *Consider the estimator $\hat{\ell}_{D_m} = \hat{\ell}_m$ of ℓ defined by (3.2) where $\ell = fg$ with f and g as in model (1.1). Let us denote by $\lambda_1 = \lambda_1(\gamma, \kappa_0, B, \sigma, \delta)$ and $\Gamma(m)$ the quantities defined by*

$$\Gamma(m) = D_m^{2\gamma+1-\delta} \exp(2B\sigma^\delta \pi^\delta D_m^\delta) \tag{4.3}$$

and

$$\lambda_1 = \frac{(\sigma^2\pi^2 + 1)^\gamma}{\pi^\delta \kappa_0^2 R(B,\sigma,\delta)} \quad \text{and } R(B,\sigma,\delta) = \begin{cases} 1 & \text{if } \delta = 0 \\ 2B\delta\sigma^\delta & \text{if } 0 < \delta \leq 1 \\ 2B\sigma^\delta & \text{if } \delta > 1. \end{cases} \tag{4.4}$$

Under Assumptions ($\mathbf{A_1}$)–($\mathbf{A_5}$) and ($\mathbf{A_9}$), if $D_m \leq \sqrt{n}$ and $k_n \geq n^2$, then

$$\mathbb{E}(\|\ell - \hat{\ell}_m\|_2^2) \leq \|\ell - \ell_m\|_2^2 + 2\lambda_1 \mathbb{E}(Y_1^2)\Gamma(m)/n + (\kappa_{\mathcal{L}} + \|\ell\|_1)/n. \tag{4.5}$$

Corollary 4.1 follows from (4.2) in Proposition 4.1 by applying Formula (10) in Comte and Taupin (2003).

The rate of convergence of $\hat{\ell}_m$ is obtained by selecting the space $S_m^{(n)}$ that makes $\|\ell - \ell_m\|_2^2 + 2\lambda_1\mathbb{E}(Y_1^2)n^{-1}\Gamma(m)$ as small as possible and by choosing k_n great enough (e.g. $k_n \geq n^2$ for $D_m \leq \sqrt{n}$).

Let us precise the order of the risk when ℓ belongs $\mathcal{S}_{\alpha_\ell,\nu_\ell,\rho_\ell}(\kappa_{\alpha_\ell})$ defined by (2.1). Since ℓ_m is the orthogonal projection of ℓ on S_m, $\|\ell - \ell_m\|_2^2 = (2\pi)^{-1}\int_{|x|\geq \pi D_m} |\ell^*|^2(x)dx$. As usual, it depends on the smoothness of the function ℓ and if we consider for a moment that ℓ belongs $\mathcal{S}_{\alpha_\ell,\nu_\ell,\rho_\ell}(\kappa_{\alpha_\ell})$ defined by (2.1), we find

$$\|\ell-\ell_m\|_2^2 \le [\kappa_{\alpha_\ell}/(2\pi)](D_m^2\pi^2+1)^{-\alpha_\ell}\exp\{-2\rho_\ell\pi^{\nu_\ell}D_m^{\nu_\ell}\}. \tag{4.6}$$

By taking $k_n \ge n^2$ and $D_m \le \sqrt{n}$, and by combining (4.5) with (4.6), we obtain

$$\mathbb{E}(\|\ell-\hat\ell_m\|_2^2) \le \frac{\kappa_{\alpha_\ell}}{2\pi}(D_m^2\pi^2+1)^{-\alpha_\ell}e^{-2\rho_\ell\pi^{\nu_\ell}D_m^{\nu_\ell}} + 2\lambda_1\mathbb{E}(Y_1^2)\frac{D_m^{(2\gamma+1-\delta)}e^{2B\sigma^\delta\pi^\delta D_m^\delta}}{n} + \frac{\kappa_{\mathcal{L}}+\|\ell\|_1}{n}.$$

The calculus of D_{m_ℓ} that minimizes $\mathbb{E}(\|\ell-\hat\ell_m\|_2^2)$ is summarized in Table 1.

ℓ \ f_ε	$\delta=0$	$\delta>0$
$\nu_\ell=0$	$\pi D_{m_\ell}=O(n^{1/(2\alpha_\ell+2\gamma+1)})$ rate $O(n^{-2\alpha_\ell/(2\alpha_\ell+2\gamma+1)})$	$\pi D_{m_\ell}=[\ln(n)/(2B\sigma^\delta+1)]^{1/\delta}$ rate $O((\ln(n))^{-2\alpha_\ell/\delta})$
$\nu_\ell>0$	$\pi D_{m_\ell}=\left(\frac{\ln(n)}{2\rho_\ell}\right)^{1/\nu_\ell}$ rate $O\left(\frac{\ln(n)^{(2\gamma+1)/\nu_\ell}}{n}\right)$	Non explicit optimal D_{m_ℓ}, D_{m_ℓ} solution of $nO(1)=D_{m_\ell}^{2\alpha_\ell+2\gamma+1-\nu_\ell}\exp\{2B\sigma^\delta\pi^\delta D_{m_\ell}^\delta+2\rho_\ell\pi^{\nu_\ell}D_{m_\ell}^{\nu_\ell}\}$.

Table 1. Choice of the dimension (D_{m_ℓ}) for the squared-bias/variance compromise and resulting rates for $\hat\ell_{m_\ell}$.

For the case $\nu_\ell>0$ and $\delta>0$, the choice $\pi D_{m_\ell}=[\ln(n)/(2B\sigma^\delta+1)]^{1/\delta}$ leads to a rate which is faster than any power of $\ln(n)$ and this means that the resulting rate has a fortiori this property. For instance if $\nu_\ell=\delta$, the rate is of order $[\ln(n)]^b n^{-\rho_\ell/(\rho_\ell+B\sigma^\delta)}$ with $b=[-2\alpha_\ell B\sigma^\delta+(2\gamma-\nu_\ell+1)\rho_\ell]/[\nu_\ell(B\sigma^\delta+\rho_\ell)]$. Lastly, all these rates depend on both the known parameters related to f_ε but also on the unknown ones related to the smoothness of ℓ.

Remark 4.1. Clearly we can find a relevant choice of k_n (independent of f and g) that makes the last supplementary term in the bound (4.2) unconditionally negligible with respect to the others. We already noticed that $k_n\ge n^2$ for $D_m\le\sqrt{n}$ could work in all cases. But we can use the information contained in Table 1 to give a weaker condition on k_n than the general one. Under $(\mathbf{A_5})$, we can ensure that $(\kappa_{\mathcal{L}}+\|\ell\|_1)D_m^2/k_n$ has the same order as the variance term under Assumptions $(\mathbf{A_9})$. When $\delta=0$, under Assumption $(\mathbf{A_4})$ γ has to be such that $\gamma>1/2$ and hence D_m^2/k_n is negligible with respect to $D_m^{2\gamma+1}/n$ if $k_n\ge n$. Therefore, $k_n\ge n$ suits when f_ε is ordinary smooth. When f_ε is super smooth, we can consider D_m's such that $\pi D_m\le\ln^{2/\delta}(n)$ since otherwise, the mean square integrated risk would not tend to zero. Therefore, $D_m^2/k_n\le\pi^{-2}\ln^{4/\delta}(n)/k_n\le 1/n$ if $k_n\ge\pi^{-2}n\ln^{4/\delta}(n)$. It follows that $k_n\ge n$ for $\delta=0$ and $k_n\ge\pi^{-2}n\ln^{4/\delta}(n)$ for $\delta>0$ ensures that the risk $\mathbb{E}(\|\ell-\hat\ell_m^{(n)}\|_2^2)$ has the order $[\kappa_1/(2\pi)](D_m^2\pi^2+1)^{-\alpha_\ell}\exp\{-2\rho_\ell\pi^{\nu_\ell}D_m^{\nu_\ell}\}+2\lambda_1 D_m^{(2\gamma+1-\delta)}\exp\left\{2B\sigma^\delta\pi^\delta D_m^\delta\right\}/n$.

4.2. **Risk bound of the non penalized estimator $\hat{g}_m$.** Here we just recall a result given in Comte and Taupin (2003), for the density deconvolution.

Proposition 4.2. *(Comte and Taupin (2003).) Consider the estimator $\hat{g}_m$ of g defined by (3.2) in model (1.1). If Assumptions $(\mathbf{A_1})$–$(\mathbf{A_4})$ and $(\mathbf{A_7})$ hold, then*

$$\mathbb{E}(\|g-\hat{g}_m\|_2^2) \leq \|g-g_m\|_2^2 + 2\frac{\Delta(m)}{n} + \frac{D_m^2}{k_n}(\kappa_{\mathcal{G}}+1). \tag{4.7}$$

It follows, that if $D_m \leq \sqrt{n}$ and $k_n \geq n^2$, if $(\mathbf{A_9})$ holds and if g belongs to some class $\mathcal{S}_{\alpha_g,\nu_g,\rho_g}(\kappa_{\alpha_g})$ defined by (2.1), then

$$\mathbb{E}\|g-\hat{g}_m^{(n)}\|^2 \leq [\kappa_{\alpha_g}/(2\pi)](D_m^2\pi^2+1)^{-\alpha_g}e^{-2\rho_g\pi^{\nu_g}D_m^{\nu_g}} + 2\lambda_1\Gamma(m)/n + \kappa_{\mathcal{G}}+1)/n. \tag{4.8}$$

For D_{m_g} chosen in the same way as D_{m_ℓ} in Table 1, the rate of convergence of $\hat{g}_{m_g}$ is the optimal rate of convergence in the minimax sense, as given in Fan (1991a) for $\nu_g = 0$, in Butucea (2004) for $\nu_g > 0$ and $\delta = 0$ and in Butucea and Tsybakov (2004) for $0 < \nu_g < \delta$ and $\alpha_g = 0$. Note that the remark 4.1 is still valid here.

4.3. **Risk bound of the non penalized estimator $\hat{f}_{m_\ell,m_g}$.**

By Definition (3.5), we have $\hat{f}_{m_\ell,m_g} = \left(\hat{\ell}_{m_\ell}/\hat{g}_{m_g}\right)^{(a_n)}$ where m_ℓ and m_g are chosen in order to minimize $\mathbb{E}(\|\ell-\hat{\ell}_m\|_2^2)$ and $\mathbb{E}(\|g-\hat{g}_m\|_2^2)$. The rate of convergence of $\hat{f}_{m_\ell,m_g}$ is given by the following proposition.

Proposition 4.3. *Assume that Assumptions $(\mathbf{A_1})$–$(\mathbf{A_4})$ and $(\mathbf{A_6})$–$(\mathbf{A_9})$ hold, that g belongs to some space $\mathcal{S}_{\alpha_g,\nu_g,\rho_g}(\kappa_{\alpha_g})$ defined by (2.1). Let $\hat{f}_{m_\ell,m_g}$ be defined by (3.5), with m_ℓ and m_g such that D_{m_ℓ} and D_{m_g} minimize the risks $\mathbb{E}(\|\ell-\hat{\ell}_m\|_2^2)$ and $\mathbb{E}(\|g-\hat{g}_m\|_2^2)$ respectively. If $\alpha_g > 1/2$, $a_n = n^k$ for $k > 0$, and $k_n \geq n^2$, then, for n great enough,*

$$\mathbb{E}\|(\hat{f}_{m_\ell,m_g}-f)\mathbf{1}_A\|_2^2 \leq K g_0^{-2}\left(1+g_1\kappa_{\infty,G}g_0^{-2}\right)\left[\mathbb{E}(\|\ell-\hat{\ell}_{m_\ell}\|_2^2)+\mathbb{E}(\|g-\hat{g}_{m_g}\|_2^2)\right] + o(n^{-1}). \tag{4.9}$$

Note that D_{m_g} and D_{m_ℓ} are automatically less than $\sqrt{n}$ (see Table 1 and remind that $\gamma > 1/2$ when $\delta = 0$). Note also that if $\alpha_g \leq 1/2$ then we only have a result of type $\|(f-\hat{f}_{m_\ell,m_g})\mathbf{1}_A\|_2^2 = O_p(\|\ell-\hat{\ell}_{m_\ell}\|_2^2+\|g-\hat{g}_{m_g}\|_2^2)$. Lastly note that the result holds when the constant $\kappa_{\infty,G}$ is replaced by $\| f \|_{\infty,A}$ if f is bounded on the compact set A.

Therefore, the performance of $\hat{f}_{m_\ell,m_g}$ is given by the worst performance between the one of $\hat{\ell}_{m_\ell}$ and the one of $\hat{g}_{m_g}$.

Let us give the resulting rates in some examples. Assume that the assumptions of Proposition 4.3 hold.

- If the ε_i's are ordinary smooth,
  - If $\nu_\ell = \nu_g = 0$ and $\pi D_{m_\ell} = O\left(n^{1/(2\alpha_\ell+2\gamma+1)}\right)$ and $\pi D_{m_g} = O\left(n^{1/(2\alpha_g+2\gamma+1)}\right)$, then
  $$\mathbb{E}(\|(f-\hat{f}_{m_\ell,m_g})\mathbf{1}_A\|_2^2) \leq O(n^{-2\alpha^*/(2\alpha^*+2\gamma+1)}) \quad \text{with} \quad \alpha^* = \inf(\alpha_\ell,\alpha_g).$$
  - If $\nu_\ell > 0$, $\nu_g > 0$, $\pi D_{m_\ell} = (\ln(n)/2\rho)^{1/\nu_\ell}$ and $\pi D_{m_g} = (\ln(n)/2\rho)^{1/\nu_g}$, then
  $$\mathbb{E}(\|(f-\hat{f}_{m_\ell,m_g})\mathbf{1}_A\|_2^2) \leq O(\frac{\ln(n)^{(2\gamma+1)/\nu^*}}{n}) \quad \text{with} \quad \nu^* = \inf(\nu_\ell,\nu_g).$$

- If the ε_i's are super smooth and πD_{m_ℓ} and πD_{m_g} are equal to $[\ln(n)/(2B\sigma^\delta+1)]^{1/\delta}$, then

$$\mathbb{E}(\|(f-\hat{f}_{m_\ell,m_g})\mathbf{I}_A\|_2^2) \leq O([\ln(n)]^{-2\alpha^*/\delta}) \quad \text{with} \quad \alpha^* = \inf(\alpha_\ell, \alpha_g).$$

Since $\ell = fg$ it is clear that the smoothness properties of ℓ are in fact related to smoothness properties of f and of g.

When ℓ belongs to $\mathcal{S}_{\alpha_\ell,0,\rho_\ell}(\kappa_{\alpha_\ell})$ and g belongs to $\mathcal{S}_{\alpha_g,0,\rho_g}(\kappa_{\alpha_g})$ with $\alpha_\ell = \alpha_g$, then the resulting rate is the optimal rate given in Fan and Truong (1993) for Hölderian regression functions and densities with the same regularity. It follows that our estimator seems then optimal in that case. It is easy to see that the estimator is also optimal if $\alpha_g \geq \alpha_\ell$, that is when the density g is smoother than the regression function f. But the optimality of the rate of $\hat{f}_{m_\ell,m_g}$ when $\alpha_\ell > \alpha_g$, that is when the regression function f is smoother than g, remains an open question. This is a known drawback of Nadaraya-Watson type estimators for regression functions, constructed as ratio of estimators. Nevertheless, in "classical" regression models, when the X_i's are observed, a lot of methods, like local polynomial estimators, projection estimators..., avoid the need of regularity conditions on g for the estimation of f. The point is that standard methods solving the regression problem do not seem work in the errors-in-variables model and it is an open problem to build an estimator which would generalize a method different of the Nadaraya-Watson technique.

The first important point is that we give the same upper bounds for the risk, when the errors are either ordinary smooth or supersmooth, for $\ell = fg$ and density g of Sobolev type ($\nu_\ell = \nu_g = 0$), as previous results by Fan and Truong (1993). But we also consider smoother regression function and density g (when $\nu_\ell > 0$ and $\nu_g > 0$). That is we give upper bounds for various type of regularity for f and various type of errors. In particular, we capture various type of errors, compared with the B-spline estimator described in Koo and Lee (1998), which only allows to consider ordinary smooth errors.

It is clear, from the above results that the choice of the dimensions D_{m_ℓ} and D_{m_g} that realize the best trade-off between the squared bias and the variance terms depends on the unknown regularity coefficients of the functions ℓ and g. Our goal is now to present some estimator that leads to an automatic choice of (nearly-) optimal dimensions, and thus, the second important point is that our procedure can be easily extended for the construction of an adaptive estimator, in the sense that the construction of this new estimator does not require any prior knowledge neither on f nor on g.

5. Adaptation

As for the construction of $\hat{f}_{m_\ell,m_g}$ in (3.5), we propose to estimate f by the ratio of an adaptive estimator of ℓ and an adaptive estimator of g. Let us start with the adaptive estimator of g.

5.1. **Adaptive estimation of g.** We consider the adaptive estimator of g is defined as in Comte and Taupin (2003), by:

$$\tilde{g} = \hat{g}_{\hat{m}_g} \text{ with } \hat{m}_g = \text{ Argmin }_{m\in\mathcal{M}_n} \left[\gamma_{n,g}(\hat{g}_m) + \text{ pen}_g(m)\right]. \tag{5.1}$$

where pen_g is a data driven penalty function. We recall the associated result.

Theorem 5.1. *(Comte and Taupin (2003)). Assume that Assumptions* $(\mathbf{A_1})$*,* $(\mathbf{A_2})$*,* $(\mathbf{A_4})$*,* $(\mathbf{A_7})$ *and* $(\mathbf{A_9})$ *are fulfilled and that* $k_n \geq n^2$*. Let* λ_1 *be given by (4.4) and* μ_1 *be defined by*

$$\mu_1 = \begin{cases} \lambda_1(\gamma, \kappa_0, B, \sigma, \delta) & \text{if } \delta > 1, \\ \lambda_1^{1/2}(\gamma, \kappa_0, B, \sigma, \delta)(1+\sigma^2\pi^2)^{\gamma/2}\kappa_0^{-1}(2\pi)^{-1/2} & \text{if } \delta \leq 1. \end{cases}$$

1) *For* $0 \leq \delta < 1/3$*, let* $\tilde{g} = \hat{g}_{\hat{m}}$ *be defined by (5.1) with*

$$\text{pen}_g(m) \geq 2x(1+\xi^2)\kappa\lambda_1 \frac{D_m^{2\gamma+1-\delta}\exp\{2B\sigma^\delta(\pi D_m)^\delta\}}{n}, \tag{5.2}$$

for some universal numerical constants κ*,* $\xi > 0$ *and* $x > 1$*.*
2) *If* $\delta = 1/3$*, let* $\tilde{g} = \hat{g}_{\hat{m}}$ *be defined by (5.1) with*

$$\text{pen}_g(m) \geq 2x(1+\xi^2)\kappa[\lambda_1 + B\sigma^\delta\pi^\delta\mu_1] \frac{D_m^{2\gamma+1-\delta}\exp\{2B\sigma^\delta(\pi D_m)^\delta\}}{n}, \tag{5.3}$$

for some universal numerical constants κ*,* $\xi > 0$ *and* $x > 1$*.*
3) *For* $\delta \geq 1/3$*, let* $\tilde{g} = \hat{g}_{\hat{m}}$ *be defined by (5.1) with*

$$\text{pen}_g(m) \geq 2x(1+\xi^2)\kappa[\lambda_1 + B\sigma^\delta\pi^\delta\mu_1] \frac{D_m^{2\gamma+((1/2+\delta/2)\wedge 1)}\exp\{2B\sigma^\delta(\pi D_m)^\delta\}}{n}, \tag{5.4}$$

for some universal numerical constants κ*,* $\xi > 0$ *and* $x > 1$*.*
If $\hat{m}_g \in \mathcal{M}_{n,g}$ *with*

$$\mathcal{M}_{n,g} = \{m \in \{1, \ldots, m_{n,g}\}, \text{ such that for all } m, D_m \leq \sqrt{n} \text{ and } \text{pen}_g(m) \leq C\},$$

then, in these three cases, $\tilde{g}$ *satisfies*

$$\mathbb{E}(\|g - \tilde{g}\|_2^2) \leq C_x \inf_{m \in \mathcal{M}_{n,g}} \left[\|g - g_m\|_2^2 + \text{pen}_g(m)\right] + c/n \tag{5.5}$$

where $C_x = \kappa_x^2 \vee 2\kappa_x$*,* $\kappa_x = (x+1)/(x-1)$ *and* c *is a constant depending on* f_ε *and* A_g*.*

The main consequence is that the rate of convergence of $\tilde{g}$ is the rate of convergence of $\hat{g}_{m_g}$ when $0 \leq \delta \leq 1/3$ or when $\delta > 1/3$ and $\nu_g = 0$ or $\nu_g < \delta$. And there is a logarithmic loss, as a price to pay for adaptation when $\nu_g \geq \delta > 1/3$. We refer to Comte and Taupin (2003) for further comments on optimality of $\tilde{g}$.

5.2. **Adaptive estimation of ℓ.** In that case, the model selection is performed in an automatic way, using the following penalization:

$$\tilde{\ell} = \hat{\ell}_{\hat{m}_\ell} \text{ with } \hat{m}_\ell = \text{ Argmin }_{m \in \mathcal{M}_{n,\ell}} \left[\gamma_n(\hat{\ell}_m) + \text{ pen}_\ell(m)\right]. \tag{5.6}$$

where pen_ℓ is a penalty function defined in the theorems and given by the theory, looking as the penalty given in Theorem 5.1.

Theorem 5.2. *Consider the estimator* $\tilde{\ell}$ *of* ℓ *defined by (5.6) where* $\ell = fg$ *and* f *and* g *as in model (1.1). Assume that Assumptions* $(\mathbf{A_1})$*–*$(\mathbf{A_6})$ *and* $(\mathbf{A_9})$ *are fulfilled, that* g *belongs to* $\mathbb{L}_2(\mathbb{R})$*, that the* ξ_i*'s admit moments of order* $p \geq 6$ *and that* $k_n \geq n^2$*. We recall that* $\lambda_1 = \lambda_1(\gamma, \kappa_0, B, \sigma, \delta)$ *is defined by (4.4) and we define* μ_2 *by*

$$\mu_2 = \begin{cases} \lambda_1(\gamma, \kappa_0, B, \sigma, \delta) & \text{if } \delta > 1, \\ \lambda_1^{1/2}(\gamma, \kappa_0, B, \sigma, \delta)(1+\sigma^2\pi^2)^{\gamma/2}\|f_\varepsilon\|_2(1+\mathbb{E}^{1/2}(Y_1^2))\kappa_0^{-1}(2\pi)^{-1/2} & \text{if } \delta \leq 1. \end{cases} \tag{5.7}$$

1) *For $0 \le \delta < 1/3$, let $\tilde{\ell} = \hat{\ell}_{\hat{m}}$ be defined by (5.6) with*

$$\text{pen}_\ell(m) = \kappa \lambda_1 \mathbb{E}(Y_1^2) \frac{D_m^{2\gamma+1-\delta} \exp\{2B\sigma^\delta(\pi D_m)^\delta\}}{n}, \tag{5.8}$$

for some universal numerical constant κ.
2) *If $\delta = 1/3$, let $\tilde{\ell} = \hat{\ell}_{\hat{m}}$ be defined by (5.6) with*

$$\text{pen}_\ell(m) = \kappa[\lambda_1 + B\sigma^\delta \pi^\delta \mu_2][1 + \mathbb{E}(Y_1^2)] \frac{D_m^{2\gamma+1-\delta} \exp\{2B\sigma^\delta(\pi D_m)^\delta\}}{n}, \tag{5.9}$$

where κ is some universal numerical constant.
3) *For $\delta > 1/3$, let $\tilde{\ell} = \hat{\ell}_{\hat{m}}$ be defined by (5.6) with*

$$\text{pen}_\ell(m) = \kappa[\lambda_1 + B\sigma^\delta \pi^\delta \mu_2][1 + \mathbb{E}(Y_1^2)] \frac{D_m^{2\gamma+((1/2+\delta/2)\wedge 1)} \exp\{2B\sigma^\delta(\pi D_m)^\delta\}}{n}, \tag{5.10}$$

where κ is some universal numerical constant.
If $\hat{m}_\ell \in \mathcal{M}_{n,\ell}$ with

$$\mathcal{M}_{n,\ell} = \{m \in \{1, \ldots, m_{n,\ell}\},\ \textit{such that for all } m, D_m \le \sqrt{n} \textit{ and } \text{pen}_\ell(m) \le C\},$$

then, in these three cases, $\tilde{\ell}$ satisfies

$$\mathbb{E}(\|\ell - \tilde{\ell}\|_2^2) \le K \inf_{m \in \mathcal{M}_{n,\ell}} \left[\|\ell - \ell_m\|_2^2 + \text{pen}_\ell(m)\right] + c/n \tag{5.11}$$

where K is a constant and c is a constant depending on f_ε, $\kappa_{\mathcal{L}}$, and $\|\ell\|_1$.

The rates, similar to the rates of $\tilde{g}$, are easy to deduce from (5.11) as soon as $\ell = fg$ belongs to some smoothness class, but the procedure can reach the rate of $\hat{\ell}_{m_\ell}$, without requiring the knowledge of any smoothness parameter. If $\text{pen}_\ell(m)$ has the same order as the variance order $\Gamma(m)/n$, then equation (5.11) guarantees an automatic trade-off between the squared bias term $\|\ell - \ell_m\|_2^2$ and the variance term, up to some multiplicative constant. If not there is some loss due to the adaptation. Let us be more precise.

The first two cases: $0 \le \delta \le 1/3$. In these cases, we consider ordinary smooth ε_i's or super-smooth ε_i's with $\delta \le 1/3$. Moreover, if ℓ satisfies (2.1), then the squared bias is bounded by $\|\ell - \ell_m\|_2^2 \le (\kappa_{\mathcal{L}}/2\pi) D_m^{-2\alpha_\ell} \exp\{-2\rho_\ell \pi^{\nu_\ell} D_m^{\nu_\ell}\}$, which combined with the value of $\text{pen}_\ell(m)$, of order $D_m^{2\gamma+1-\delta} \exp(2B\sigma^\delta(\pi D_m)^\delta)/n$, gives that the estimator $\tilde{g}$ automatically reaches best rate achievable by the estimator $\hat{\ell}_{m_\ell}$, given in Table 1, just as if the smoothness parameters α_ℓ, ν_ℓ and ρ_ℓ were known, whereas they are not.

The last case : $\delta > 1/3$. In that case, the penalty function $\text{pen}_\ell(m)$ has not exactly the order of the variance, but a loss of order $D_m^{(3\delta/2-1/2)\wedge\delta}$ occurs, that is of order $D_m^{(3\delta-1)/2}$ if $1/3 < \delta \le 1$ and of order D_m^δ if $\delta > 1$. In this context, the rate remains optimal if the bias $\|\ell - \ell_m\|_2^2$ is the dominating term in the trade-off between $\|\ell - \ell_m\|_2^2$ and $\text{pen}_\ell(m)$. More precisely, when $\nu_\ell = 0$ and $\delta > 0$, the optimal rate of order $(\ln(n))^{-2\alpha_\ell/\delta}$ is given by the bias term, and the loss in the penalty function does not change the rate achieved by the adaptive estimator $\tilde{\ell}$, which remains the best achievable rate for one single estimator $\hat{\ell}_m$.

When $0 < \nu_\ell < \delta$, the rate is given by the bias term and thus this loss does not affect the rate of convergence of $\tilde{\ell}$ either.

Let us now focus our discussion on the case where $\text{pen}_\ell(m)$ can be the dominating term in the trade-off between $\|\ell - \ell_m\|_2^2$ and $\text{pen}_\ell(m)$, that is when $\nu_\ell \ge \delta > 1/3$. In that case, there is a loss of order $D_m^{(3\delta/2-1/2)\wedge\delta}$ in the penalty function, compared to the variance term. The main point is that it happens in cases where the order of the optimal D_m is less than $(\ln n)^{1/\delta}$ and consequently the loss in the rate is at most of order $\ln n$, when the rate is faster than logarithmic: therefore the loss appears only in cases where it can be seen as negligible.

In particular, there is not price to pay for the adaptation if the ξ_i's are Gaussian and the ε_i's are ordinary smooth. Indeed, in that case, the rate of convergence of the penalized estimator $\tilde{\ell}$, without any knowledge on ℓ or g, is the same as the rate given by the non penalized estimator $\hat{\ell}_{m_\ell}$, requiring the knowledge of smoothness parameters. But, if both the ξ_i's and the ε_i's are Gaussian, then $\delta = 2$ and there may be a logarithmic negligible loss in the rate of $\tilde{\ell}$ compared to the rate of $\hat{\ell}_{m_\ell}$, if $\nu_\ell \geq 2$.

In Theorem 5.2, the penalty involves the unknown quantity $\mathbb{E}(Y_1^2)$. Consequently it has to be estimated, in order to have a penalty term only depending on the observations and on parameters related to f_ε. We define as an obvious estimator of $m_2(Y) = \mathbb{E}(Y_1^2)$ the quantity

$$\hat{m}_2(Y) = \frac{1}{n}\sum_{i=1}^{n} Y_i^2. \tag{5.12}$$

Then we can easily extend the result of Theorem 5.2 as follows.

Corollary 5.1. *If the ξ_i's admit moments of order 8 and under the assumptions of Theorem 5.2, with $\mathbb{E}(Y_1^2)$ in the penalties (5.8), (5.9) or (5.10) replaced by $\hat{m}_2(Y)$, then the inequality (5.11) keeps holding with* $\mathrm{pen}_\ell(m)$ *replaced by its expectation* $\mathbb{E}(\mathrm{pen}_\ell(m))$.

The only price to pay for the substitution is the increase in the moment condition on the ξ_i's that must be now finite until order 8 instead of 6 in Theorem 5.2.

5.3. **Adaptive estimation of f.** Under Assumption $(\mathbf{A_8})$, as in (3.5), we define

$$\tilde{f} = \left(\frac{\tilde{\ell}}{\tilde{g}}\right)^{(a_n)}. \tag{5.13}$$

Its rate of convergence is thus given by the following theorem.

Theorem 5.3. *Let $\tilde{f}$ be defined by (5.13) with $\tilde{\ell}$ defined in Theorem 5.2 and $\tilde{g}$ defined in Theorem 5.1 under Assumptions $(\mathbf{A_1})$–$(\mathbf{A_4})$ and $(\mathbf{A_6})$–$(\mathbf{A_9})$ with $\hat{m}_g \in \mathcal{M}_{n,g}$ with $D_{m_{n,g}} \leq (n/\ln(n))^{1/(2\gamma+2)}$ and $\hat{m}_\ell \in \mathcal{M}_{n,\ell}$. Assume that g belongs to some space $\mathcal{S}_{\alpha_g,\nu_g,\rho_g}(\kappa_{\alpha_g})$ defined by (2.1) with $\alpha_g > 1/2$. Assume that the ξ_i's admit moments of order $p \geq 8$. If $k_n \geq n^2$, $a_n = n^k$ for $k > 0$, then, for n large enough,*

$$\begin{aligned}\mathbb{E}(\|(f-\tilde{f})\mathbf{1}_A\|_2^2) \;\leq\; & 8Kg_0^{-2}\inf_{m\in\mathcal{M}_{n,\ell}}\left[\|\ell-\ell_m\|_2^2 + \mathrm{pen}_\ell(m)\right] + c/n \\ & +4K'g_0^{-2}(2g_1^2+1)\kappa_{\infty,G}^2 \inf_{m\in\mathcal{M}_{n,g}}\left[\|g-g_m\|_2^2 + \mathrm{pen}_g(m)\right]\end{aligned} \tag{5.14}$$

where K and K' are constants depending on f_ε, and c is a constant depending on f_ε, f and g.

As in Theorem 3.5, if $\alpha_g \leq 1/2$ then it may happen that $D_{\tilde{m}_g} \geq n^{1/(2\gamma+2)}$, and in this case we only have a result in probability: $\|(f-\tilde{f})\mathbf{1}_A\|_2^2 = O_p(\|\ell-\tilde{\ell}\|_2^2 + \|g-\tilde{g}\|_2^2)$. And again as in Theorem 3.5, the result holds when the constant $\kappa_{\infty,G}$ is replaced by $\| f \|_{\infty,A}$ if f is bounded on the compact set A. Also note that the remark 4.1 is still valid for all adaptive estimators.

Comments about the resulting rates for estimating f. The main one is that the rate of convergence of $\tilde{f}$ is given by the worst rate of convergence between the rate of $\tilde{\ell}$ and $\tilde{g}$. Obviously all the comments about $\hat{f}_{m_\ell,m_g}$, related to this fact keep holding here.

Consequently, when $0 \leq \delta < 1/3$ or when $\nu_\ell \leq \delta$ and $\nu_g \leq \delta$, then $\tilde{f}$ achieves the rate of convergence of $\hat{f}_{m_\ell,m_g}$, given by the worst rate of convergence between $\| \hat{\ell}_{m_\ell} - \ell \|_2^2 + \| \hat{g}_{m_g} - g \|_2^2$. And when $\nu_g > \delta > 1/3$ or $\nu_\ell > \delta > 1/3$, there is a logarithmic loss in the rate of convergence of $\tilde{f}$ compared to the rate of convergence of $\hat{f}_{m_\ell,m_g}$.

Let us go further in the analysis by noting that the regularity of ℓ is by definition the regularity of fg. It follows that the rate of convergence of $\tilde{\ell}$ in fact depends on smoothness properties of f and on g through ℓ. As a consequence, if ℓ and g belong respectively to $\mathcal{S}_{\alpha_\ell,\nu_\ell,\rho_f}(\kappa_{\alpha_\ell})$ and $\mathcal{S}_{\alpha_g,\nu_g,\rho_g}(\kappa_{\alpha_g})$, then the rate of convergence of $\tilde{f}$ is the rate of $\hat{f}_{m_\ell,m_g}$ when $0 \leq \delta \leq 1/3$. According to Fan and Truong (1993), this rate seems the optimal one when $\alpha_\ell \leq \alpha_g$ and $\nu_\ell = \nu_g = 0$. In the other cases, the optimality question remains open. Even if the regression function is smoother than g and $0 \leq \delta \leq 1/3$, the rate of convergence of $\tilde{f}$ has the order of the rate of convergence of $\hat{f}_{m_\ell,m_g}$, but we don't know if the rate of $\hat{f}_{m_\ell,m_g}$ is the optimal one (see comments following Theorem 3.5).

Furthermore when $\delta > 1/3$, a loss appears between the rate of convergence of $\tilde{f}$ and the rate of convergence of $\hat{f}_{m_\ell,m_g}$. This loss only appears, when $\nu_\ell > \delta$ or $\nu_g > \delta$ (see the comments after Theorem 5.2) and these are cases where it is negligible with respect to the rate. This logarithmic loss appears as a price to pay for ignoring all about f and g.

Note that this loss appears when $\delta > 1/3$ and $\nu_\ell > \delta$ or $\nu_g > \delta$, that is when the dominating term is the variance term, that is when the noise influence dominates the smoothness properties of ℓ and g, that is of f and g.

Remark 5.1. We clearly see that the resulting rates for all estimators depend on the noise level σ. The first point is to note that if $\sigma = 0$, then by convention $B = 0 = \delta = 0, \lambda = 1$, and $Z = X$ is observed. It follows that the quantity $\Gamma(m)/n$ is of order D_m/n that is the expected order for the variance term in "usual regression", when the explanatory variables are observed. This continue to hold for the penalties pen_ℓ and pen_g which become of order D_m/n when $\sigma = 0$, that is of the expected orders for penalties, for density estimation and nonparametric regression estimation, when there is one model per dimension, as in our case.

The second point is to note that if σ is small, then the procedure automatically selects a dimension D_m closed to the dimension that would be selected in "usual" density estimation and nonparametric regression estimation.

6. Proofs

6.1. **Proof of Proposition 4.1.** By applying Definition (3.2), for any m belonging to $\mathcal{M}_n$, $\hat{\ell}_m$ satisfies $\gamma_n(\hat{\ell}_m) - \gamma_n(\ell_m^{(n)}) \leq 0$. Denoting by $\nu_n(t)$ the centered empirical process,

$$\nu_n(t) = \frac{1}{n} \sum_{i=1}^{n} \left(Y_i u_t^*(Z_i) - \langle t, \ell \rangle \right), \tag{6.1}$$

and by using that $t \mapsto u_t^*$ is linear we get the following decomposition

$$\gamma_n(t) - \gamma_n(s) = \|t-\ell\|_2^2 - \|s-\ell\|_2^2 - 2\nu_n(t-s) \tag{6.2}$$

and therefore, since by Pythagoras Theorem, $\|\ell-\ell_m^{(n)}\|_2^2 = \|\ell-\ell_m\|_2^2 + \|\ell-\ell_m^{(n)}\|_2^2$, we infer that $\|\ell-\hat{\ell}_m\|_2^2 \le \|\ell-\ell_m\|_2^2 + \|\ell_m - \ell_m^{(n)}\|_2^2 + 2\nu_n(\hat{\ell}_m - \ell_m^{(n)})$. Using that $\hat{a}_{m,j}(\ell) - a_{m,j}(\ell) = \nu_n(\varphi_{m,j})$, we get

$$\nu_n(\hat{\ell}_m - \ell_m^{(n)}) = \sum_{|j|\le k_n} (\hat{a}_{m,j}(\ell) - a_{m,j}(\ell))\nu_n(\varphi_{m,j}) = \sum_{|j|\le k_n} [\nu_n(\varphi_{m,j})]^2, \tag{6.3}$$

and consequently

$$\mathbb{E}\|\ell-\hat{\ell}_m\|_2^2 \le \|\ell-\ell_m\|_2^2 + \|\ell_m-\ell_m^{(n)}\|_2^2 + 2\sum_{j\in\mathbb{Z}} \mathrm{Var}[\nu_n(\varphi_{m,j})]. \tag{6.4}$$

Now, since the (Y_i, Z_i)'s are independent, $\mathrm{Var}[\nu_n(\varphi_{m,j})] = n^{-1}\mathrm{Var}\left[Y_1 u^*_{\varphi_{m,j}}(Z_1)\right]$, and consequently,

$$\sum_{j\in\mathbb{Z}} \mathrm{Var}[\nu_n(\varphi_{m,j})] \le n^{-1} \, \| \sum_{j\in\mathbb{Z}} |u^*_{\varphi_{m,j}}|^2 \, \|_\infty \, \mathbb{E}(Y_1^2). \tag{6.5}$$

Arguing as in Comte and Taupin (2003) we find, by using Parseval's formula that

$$\| \sum_{j\in\mathbb{Z}} |u^*_{\varphi_{m,j}}|^2 \, \|_\infty \le \Delta(m), \tag{6.6}$$

where Δ is defined by (4.1). Finally, by gathering (6.4), (6.5) and by using the definition of $\Gamma(m)$ given in (4.3) we get

$$\mathbb{E}(\|\ell-\hat{\ell}_m\|_2^2) \le \|\ell-\ell_m\|_2^2 + \|\ell_m-\ell_m^{(n)}\|_2^2 + 2\mathbb{E}(Y_1^2)\Delta(m)/n. \tag{6.7}$$

For the study of the residual term $\|\ell_m - \ell_m^{(n)}\|_2^2$, we simply write

$$\|\ell_m - \ell_m^{(n)}\|_2^2 = \sum_{|j|\ge k_n} a_{m,j}^2(\ell) \le (\sup_j j a_{m,j}(\ell))^2 \sum_{|j|\ge k_n} j^{-2}.$$

Now by definition

$$\begin{aligned} j a_{m,j} &= j\sqrt{D_m}\int \varphi(D_m x - j)\ell(x)dx \\ \le&\ D_m^{3/2}\int |x||\varphi(D_m x-j)||\ell(x)|dx + \sqrt{D_m}\int |D_m x - j||\varphi(D_m x - j)||\ell(x)|dx \\ \le&\ D_m^{3/2}\left(\int |\varphi(D_m x-j)|^2 dx\right)^{1/2}\kappa_{\mathcal{L}}^{1/2} + \sqrt{D_m}\sup_x |x\varphi(x)|\|\ell\|_1. \end{aligned}$$

This implies finally that $j a_{m,j} \le D_m\|\varphi\|_2\kappa_{\mathcal{L}}^{1/2} + \sqrt{D_m}\|\ell\|_1/\pi$, and therefore

$$\|\ell_m - \ell_m^{(n)}\|_2^2 \le \kappa(\kappa_{\mathcal{L}} + \|\ell\|_1^2)D_m^2/k_n. \tag{6.8}$$

Under $(\mathbf{A_5})$, $\kappa_{\mathcal{L}}$ is finite and under $(\mathbf{A_3})$, $\mathbb{E}(f^2(X_1))$ is finite, which implies $\|\ell\|_1 < \infty$. The result follows by combining (6.7) and (6.8). $\square$

6.2. **Proof of Proposition 4.3.** Let E_n be the event $E_n = \{\| g - \hat{g}_{m_g} \|_{\infty,A} \leq g_0/2\}$. Since $g(x) \geq g_0$ for any x in A, then, on E_n, for any x in A, $\hat{g}_{m_g}(x) \geq g_0/2$. It follows that

$$\mathbb{E}(\|(f - \hat{f}_{m_\ell,m_g})\mathbf{I}_A \mathbf{I}_{E_n}\|_2^2) \leq \frac{8}{g_0^2}\mathbb{E}\|\hat{\ell}_{m_\ell} - \ell\|_2^2 + \frac{8\|\ell\|_{\infty,A}^2}{g_0^4}\mathbb{E}\|\hat{g}_{m_g} - g\|_2^2, \tag{6.9}$$

where $\|\ell\|_{\infty,A} \leq g_1\|f\|_{\infty,A} \leq g_1\kappa_{\infty,G}$, and using that $\|\hat{f}_{m_\ell,m_g}\|_{\infty,A} \leq a_n$

$$\mathbb{E}\|(f - \hat{f}_{m_\ell,m_g})\mathbf{I}_A \mathbf{I}_{E_n^c}\|_2^2 \leq 2(a_n^2 + \|f\|_{\infty,A}^2)\lambda(A)\mathbb{P}(E_n^c), \tag{6.10}$$

where $\lambda(A) = \int_A dx$. It follows that for $m_\ell = m_\ell(n)$, $m_g = m_g(n)$, if $a_n\mathbb{P}(E_n^c) = o(n^{-1})$, then (4.9) is proved. We now come to the study of $\mathbb{P}(E_n^c)$ by writing that

$$\mathbb{P}(E_n^c) \;=\; \mathbb{P}\left(\|g - \hat{g}_{m_g}\|_\infty > g_0/2\right) = \mathbb{P}\left(\|g - g_{m_g}^{(n)} + g_{m_g}^{(n)} - \hat{g}_{m_g}\|_\infty > g_0/2\right).$$

By applying Lemma 7.1, and arguing as for (6.8) (with g replacing ℓ), we get that

$$\begin{aligned}\|g - g_{m_g}^{(n)}\|_\infty &\leq \|g_{m_g} - g_{m_g}^{(n)}\|_\infty + \|g - g_{m_g}\|_\infty \\ &\leq \sqrt{\kappa(\kappa_{\mathcal{G}}+1)\frac{D_{m_g}^3}{k_n}} + \frac{\sqrt{\pi D_{m_g}}}{2\pi}((\pi D_{m_g})^2+1)^{-\alpha_g/2}\exp(-\rho_g|\pi D_{m_g}|^{\nu_g})A_g^{1/2},\end{aligned}$$

if g belongs to some space $\mathcal{S}_{\alpha_g,\nu_g,\rho_g}(\kappa_{\alpha_g})$ defined by (2.1) with $\alpha_g > 1/2$. Consequently, $\|g - g_{m_g}^{(n)}\|_\infty$ tends to zero as soon as $\alpha_g > 1/2$ and since $k_n \geq n^2$ and $D_{m_g} \leq \sqrt{n}$. It follows that for n large enough, $\|g - g_{m_g}^{(n)}\|_\infty \leq g_0/4$. This combined with the fact that $\|g - \hat{g}_{m_g}\|_\infty \leq \|g - g_{m_g}^{(n)}\|_\infty + \|g_{m_g}^{(n)} - \hat{g}_{m_g}\|_\infty$ we infer that

$$\mathbb{P}(E_n^c) \leq \mathbb{P}\left[\|g_{m_g}^{(n)} - \hat{g}_{m_g}\|_\infty > g_0/4\right]. \tag{6.11}$$

By applying again Lemma 7.1, we get that $\mathbb{P}(E_n^c) \leq \mathbb{P}(\|g_{m_g}^{(n)} - \hat{g}_{m_g}\|_2 > g_0/(4\sqrt{D_{m_g}}))$. Now, according to (3.4), and by arguing as in (6.3), $\|g_{m_g}^{(n)} - \hat{g}_{m_g}\|_2^2 = \sum_{|j|\leq k_n}\nu_{n,g}^2(\varphi_{m_g,j})$ where $\nu_{n,g}(t) = n^{-1}\sum_{i=1}^n(u_t^*(Z_i) - <t,g>)$. If $B_m(0,1) = \{t \in S_m^{(n)} \;/\; \|t\|_2 = 1\}$, then by combining Cauchy-Schwarz inequality and the fact that $\nu_{n,g}$ is linear, we get that

$$\sum_{|j|\leq k_n}\nu_{n,g}^2(\varphi_{m_g,j}) = \sup_{t\in B_{m_g}(0,1)}\nu_{n,g}^2(t). \tag{6.12}$$

Consequently, $\mathbb{P}(E_n^c) \leq \mathbb{P}\left[\sup_{t\in B_{m_g}(0,1)}|\nu_{n,g}(t)| \geq g_0/(4D_{m_g}^{1/2})\right]$. We apply the version of Talagrand's (1996) inequality given in Birgé and Massart (1998) with f replaced by $r = f - \mathbb{E}f(X_1)$ and M_1 by $2M_1$ to get

$$\mathbb{P}\left(\sup_{r\in\mathcal{R}}|\nu_{n,g}(r)| \geq (1+\eta)H + \lambda\right) \leq 3\exp\left[-K_1 n\left(\frac{\lambda^2}{v} \wedge \frac{(\eta\wedge 1)\lambda}{M_1}\right)\right]. \tag{6.13}$$

Usual density arguments show that this result can be applied to the class of functions $\mathcal{R} = B_{m_g}(0,1)$. If we take $\eta = 1$, and if $2H + \lambda < g_0/(4\sqrt{D_{m_g}})$, that is, for instance if we take

$\lambda = g_0/(8\sqrt{D_{m_g}})$ and if we ensure $2H < g_0/(8\sqrt{D_{m_g}})$, we get

$$\mathbb{P}\left(\sup_{t\in B_{m_g}(0,1)} |\nu_{n,g}(t)| \geq g_0/(4D_{m_g}^{1/2})\right) \leq \mathbb{P}\left(\sup_{t\in B_{m_g}(0,1)} |\nu_{n,g}(t)| \geq 2H+\lambda\right) \leq 3\exp\left[-K_1 n\left(\frac{\lambda^2}{v}\wedge\frac{\lambda}{M_1}\right)\right],$$

that is

$$\mathbb{P}\left(\sup_{t\in B_{m_g}(0,1)} |\nu_{n,g}(t)| \geq g_0/(4\sqrt{D_{m_g}})\right) \leq 3\exp\left[-K_1' n\left(\frac{1}{D_{m_g}v}\wedge\frac{1}{M_1\sqrt{D_{m_g}}}\right)\right].$$

According to (6.5), (6.6) and (6.12), with λ_1 given in (4.3), and μ_1 recalled in Theorem 5.1,

$$H^2 = \lambda_1\Gamma(m_g)/n,\ M_1 = \sqrt{nH^2},\ \text{and } v = \mu_1 D_{m_g}^{2\gamma+(1/2-\delta/2)\wedge(1-\delta)}\exp\{2B\sigma^\delta(\pi D_{m_g})^\delta\}.$$

Now we use that $\exp(a\wedge b)\leq \exp(a)+\exp(b)$ and study

$$\exp[-K_1'n/(M_1\sqrt{D_{m_g}})] \text{ and } \exp[-K_1'n/(D_{m_g}v)]. \tag{6.14}$$

First, $M_1=\sqrt{nH^2}$ and $D_{m_g}\leq\sqrt{n}$ imply that

$$n/(M_1\sqrt{D_{m_g}}) = \sqrt{n}/(H\sqrt{D_{m_g}}) = n/\sqrt{D_{m_g}\Gamma(m_g))} \leq K''n^{1/4}.$$

It follows that

$$\exp[-K_1'n/(M_1\sqrt{D_{m_g}})] \leq \exp[-K\text{"}n^{1/4}] = o(n^{-1}).$$

Second, under Assumption $(\mathbf{A_9})$, when $\delta=0$, then $v=\mu_1 D_{m_g}^{2\gamma+1/2}$, $n/(D_{m_g}v) = O(n^{1-\frac{2\gamma+3/2}{2\alpha+2\gamma+1}})$ and consequently

$$\exp[-K_1'n/(D_{m_g}v)] \leq \exp[-K''n^{1-\frac{2\gamma+3/2}{2\alpha+2\gamma+1}}].$$

When $\delta>0$, we only use that $v\leq D_{m_g}^{2\gamma+1-\delta}\exp[2B\sigma^\delta\pi^\delta D_{m_g}^\delta]$ which implies that

$$\exp\left[-\frac{K_1'n}{D_{m_g}v}\right] \leq \exp\left[-K^{(3)}\frac{n}{D_{m_g}^{2\gamma+2-\delta}\exp\{2B\sigma^\delta(\pi D_{m_g})^\delta\}}\right].$$

Now use that, according to Table 1 and the remark following it, $D_{m_g}\leq \left(\ln n/(2B\sigma^\delta+1)\right)^{1/\delta}$ to get that

$$\exp\left[-K^{(3)}\frac{n}{D_{m_g}^{2\gamma+2-\delta}\exp\{2B\sigma^\delta(\pi D_{m_g})^\delta\}}\right] \ \leq\ \exp\left[-K^{(3)}n^{1/(2B\sigma^\delta+1)}(\ln n)^{-\frac{2\gamma+2-\delta}{2\alpha+2\gamma+1}}\right].$$

Lastly, if g belongs to some space $\mathcal{S}_{\alpha_g,\nu_g,\rho_g}(\kappa_{\alpha_g})$ defined by (2.1) with $\alpha_g>1/2$, $\sqrt{D_{m_g}}H$ tends to zero when n grows to infinity, and therefore is less than $g_0/8$ for n great enough. Consequently, if $\delta>0$ then

$$a_n\mathbb{P}(E_n^c) \leq a_n\kappa\left[\exp(-K''n^{1/4}) + \exp(K^{(3)}n^{1/(2B\sigma^\delta+1)}\ln n^{-\frac{2\gamma+2-\delta}{2\alpha+2\gamma+1}})\right] \tag{6.15}$$

and if $\delta=0$

$$a_n\mathbb{P}(E_n^c) \leq a_n\kappa\left[\exp(-K''n^{1/4}) + \exp(-K''n^{\frac{2\alpha-1/2}{2\alpha+2\gamma+1}})\right]. \tag{6.16}$$

We conclude that if $a_n = n^k$ for $k > 0$, then $a_n\mathbb{P}(E_n^c) = o(n^{-1})$ if $\delta > 0$ and if $\delta = 0$ and $\alpha_g > 1/2$. The result follows by combining this with the inequalities (6.9) and (6.10).

□

6.3. **Proof of Theorem 5.2.** For the study of $\tilde{\ell}$, the main difficulty compared to the study of $\tilde{g}$ lies in avoiding any assumption of boundedness on the noise ξ_i. By definition, $\tilde{\ell}$ satisfies that for all $m \in \mathcal{M}_{n,\ell}$, $\gamma_n(\tilde{\ell}) + \mathrm{pen}_\ell(\hat{m}) \leq \gamma_n(\ell_m^{(n)}) + \mathrm{pen}_\ell(m)$. Therefore, by applying (6.2) we get that

$$\| \tilde{\ell} - \ell \|_2^2 \leq \| \ell - \ell_m^{(n)} \|_2^2 + 2\nu_n(\tilde{\ell} - \ell_m^{(n)}) + \mathrm{pen}_\ell(m) - \mathrm{pen}_\ell(\hat{m}). \tag{6.17}$$

Next, we use that if $t = t_1 + t_2$ with t_1 in S_m and t_2 in $S_{m'}$, then t is such that t^* has its support in $[-\pi D_{m\vee m'}, \pi D_{m\vee m'}]$ and therefore t belongs to $S_{m\vee m'}$. Denote by $B_{m,m'}(0,1)$ the set

$$B_{m,m'}(0,1) = \{t \in S_{m\vee m'}^{(n)} \ / \ \|t\|_2 = 1\}.$$

It follows that

$$|\nu_n(\tilde{\ell} - \ell_m^{(n)})| \leq \|\tilde{\ell} - \ell_m^{(n)}\|_2 \sup_{t\in B_{m,\hat{m}}(0,1)} |\nu_n(t)|,$$

where $\nu_n(t)$ is defined by (6.1). Consequently, by using that $2ab \leq x^{-1}a^2 + xb^2$

$$\|\tilde{\ell} - \ell\|_2^2 \ \leq \ \|\ell_m^{(n)} - \ell\|_2^2 + \frac{1}{x}\|\tilde{\ell} - \ell_m^{(n)}\|_2^2 + x \sup_{t\in B_{m,\hat{m}}(0,1)} \nu_n^2(t) + \mathrm{pen}_\ell(m) - \mathrm{pen}_\ell(\hat{m})$$

and therefore, writing that $\|\tilde{\ell} - \ell_m^{(n)}\|_2^2 \leq (1 + y^{-1})\|\tilde{\ell} - \ell\|_2^2 + (1 + y)\|\ell - \ell_m^{(n)}\|_2^2$, with $y = (x+1)/(x-1)$ for $x > 1$, we infer that

$$\|\tilde{\ell} - \ell\|_2^2 \leq \left(\frac{x+1}{x-1}\right)^2 \|\ell - \ell_m^{(n)}\|_2^2 + \frac{x(x+1)}{x-1} \sup_{t\in B_{m,\hat{m}}(0,1)} \nu_n^2(t) + \frac{x+1}{x-1}(\mathrm{pen}_\ell(m) - \mathrm{pen}_\ell(\hat{m})).$$

Choose some positive function $p_\ell(m, m')$ such that $xp_\ell(m, m') \leq \mathrm{pen}_\ell(m) + \mathrm{pen}_\ell(m')$. Then, by denoting by $\kappa_x = (x + 1)/(x - 1)$,

$$\begin{aligned}\|\tilde{\ell} - \ell\|_2^2 \ &\leq \ \kappa_x^2\|\ell - \ell_m^{(n)}\|_2^2 + x\kappa_x[\sup_{t\in B_{m,\hat{m}}(0,1)} |\nu_n|^2(t) - p(m, \hat{m})]_+ \\ &\quad + \kappa_x \left(xp_\ell(m, \hat{m}) + \mathrm{pen}_\ell(m) - \mathrm{pen}_\ell(\hat{m})\right)\end{aligned} \tag{6.18}$$

that is

$$\|\tilde{\ell} - \ell\|_2^2 \leq \kappa_x^2\|\ell - \ell_m^{(n)}\|_2^2 + 2\kappa_x\mathrm{pen}_\ell(m) + x\kappa_x W_n(\hat{m}), \tag{6.19}$$

where

$$W_n(m') = [\sup_{t\in B_{m,m'}(0,1)} |\nu_n(t)|^2 - p_\ell(m, m')]_+. \tag{6.20}$$

The main point of the proof lies in studying $W_n(m')$, more precisely in finding $p_\ell(m, m')$ such that

$$\mathbb{E}(W_n(\hat{m})) \leq \sum_{m'\in\mathcal{M}_{n,\ell}} \mathbb{E}(W_n(m'))) \leq C/n, \tag{6.21}$$

where C is a constant. In this case, combining (6.19) and (6.21) we infer that, for all m in $\mathcal{M}_{n,\ell}$,

$$\mathbb{E}\|\ell - \tilde{\ell}\|_2^2 \leq \kappa_x^2 \|\ell - \ell_m^{(n)}\|_2^2 + 2\kappa_x \mathrm{pen}_\ell(m) + x\kappa_x C/n,$$

which can also be written

$$\mathbb{E}\|\ell - \tilde{\ell}\|_2^2 \leq C_x \inf_{m \in \mathcal{M}_{n,\ell}} \left[\|\ell - \ell_m\|_2^2 + \mathrm{pen}_\ell(m)\right] + C_x C'/n, \tag{6.22}$$

where $C_x = \kappa_x^2 \vee (2\kappa_x)$ suits, by using (6.8) when $k_n \geq n^2$, $D_{m_{n,\ell}} \leq \sqrt{n}$. It remains thus to find $p_\ell(m, m')$ such that (6.21) holds.

The study of $W_n(m')$ is done in two steps, coming from the decomposition of $\nu_n(t)$

$$\nu_n(t) = \nu_{n,1}(t) + \nu_{n,2}(t) \text{ with } \begin{cases} \nu_{n,1}(t) = \dfrac{1}{n}\displaystyle\sum_{i=1}^n (f(X_i)u_t^*(Z_i) - \langle t, \ell\rangle) \\ \nu_{n,2}(t) = \dfrac{1}{n}\displaystyle\sum_{i=1}^n \xi_i u_t^*(Z_i) \end{cases}$$

Both terms would be treated in the same way if the variables ξ_i were assumed to be bounded. It follows from the above decomposition that $W_n(m') \leq 2W_{n,1}(m') + 2W_{n,2}(m')$ where for $i = 1, 2$,

$$W_{n,i}(m') = \left[\sup_{t \in B_{m,m'}(0,1)} |\nu_{n,i}(t)|^2 - p_i(m, m')\right]_+, \tag{6.23}$$

and

$$p_\ell(m, m') = 2p_1(m, m') + 2p_2(m, m'). \tag{6.24}$$

• Study of $W_{n,1}$.
Since under $(\mathbf{A_6})$, f is bounded on the support of g, we apply a standard Talagrand's (1996) inequality (see Lemma 6.1 below that can be a fortiori applied to identically distributed variables):

Lemma 6.1. *Let $U_1, \ldots, U_n$ be independent random variables and $\nu_n(r) = (1/n)\sum_{i=1}^n [r(U_i) - \mathbb{E}(r(U_i))]$ for r belonging to a countable class $\mathcal{R}$ of uniformly bounded measurable functions. Then for $\epsilon > 0$*

$$\mathbb{E}\left[\sup_{r \in \mathcal{R}} |\nu_n(r)|^2 - 2(1+2\epsilon)H^2\right]_+ \leq \frac{6}{K_1}\left(\frac{v}{n} e^{-K_1 \epsilon \frac{nH^2}{v}} + \frac{8M_1^2}{K_1 n^2 C^2(\epsilon)} e^{-\frac{K_1 C(\epsilon)\sqrt{\epsilon}}{\sqrt{2}}\frac{nH}{M_1}}\right), \tag{6.25}$$

with $C(\epsilon) = \sqrt{1+\epsilon} - 1$, K_1 is a universal constant, and where

$$\sup_{r \in \mathcal{R}} \|r\|_\infty \leq M_1, \quad \mathbb{E}\left(\sup_{r \in \mathcal{R}} |\nu_n(r)|\right) \leq H, \quad \sup_{r \in \mathcal{R}} \frac{1}{n}\sum_{i=1}^n \mathrm{Var}(r(U_i)) \leq v.$$

The inequality (6.25) is a straightforward consequence of Talagrand's (1996) inequality given in Ledoux (1996) (or Birgé and Massart (1998)) and recalled in (6.13), with f replaced by $r = f - \mathbb{E}f(X_1)$ and M_1 by $2M_1$, and by taking $\eta = (\sqrt{1+\epsilon} - 1) \wedge 1 = C(\epsilon) \leq 1$. Therefore

$$\mathbb{E}\left[\sup_{t \in B_{m,m'}(0,1)} |\nu_{n,1}(t)|^2 - 2(1+2\epsilon_1)\mathbb{H}_1^2\right]_+ \leq \kappa_1\left(\frac{v_1}{n} e^{-K_1 \epsilon_1 \frac{n\mathbb{H}_1^2}{v_1}} + \frac{M_1^2}{n^2} e^{-K_2\sqrt{\epsilon_1}C(\epsilon_1)\frac{n\mathbb{H}_1}{M_1}}\right), \tag{6.26}$$

where $K_2 = K_1/\sqrt{2}$ and $\mathbb{H}_1$, v_1 and M_1 are defined by $\mathbb{E}(\sup_{t\in B_{m,m'}(0,1)} |\nu_{n,1}(t)|^2) \leq \mathbb{H}_1^2$,

$$\sup_{t\in B_{m,m'}(0,1)} \mathrm{Var}(f(X_1)u_t^*(Z_1)) \leq v_1, \text{ and } \sup_{t\in B_{m,m'}(0,1)} \|f(X_1)u_t^*(Z_1)\|_\infty \leq M_1.$$

According to (6.6), we propose to take

$$M_1 = M_1(m,m') = \kappa_{\infty,G}\sqrt{\lambda_1\Gamma(m^*)} \tag{6.27}$$

where $\Gamma(m)$ is defined by (4.3) and $m^* = m \vee m'$. For v_1, write

$$\sup_{t\in B_{m,m'}(0,1)} \mathrm{Var}(f(X_1)u_t^*(Z_1)) \;\; \leq \;\; (\sum_{j,k\in\mathbb{Z}} |P_{j,k}(m^*)|^2)^{1/2}$$

where $P_{j,k}(m) = \mathbb{E}\left[f^2(X_1)u^*_{\varphi_{m,j}}(Z_1)u^*_{\varphi_{m,k}}(-Z_1)\right]$. Arguing as in Comte and Taupin (2003), let us define $\Delta_2(m,\theta)$ by

$$\begin{aligned}\Delta_2(m,\theta) &= D_m^2 \int\int \left|\frac{\varphi^*(x)\varphi^*(y)}{f_\varepsilon^*(D_m x)f_\varepsilon^*(D_m y)}\theta^*(D_m(x-y))\right|^2 dxdy \\ &\leq \lambda_2^2(\|\theta\|_2)\Gamma_2^2(m^*),\end{aligned} \tag{6.28}$$

with

$$\Gamma_2(m^*) = D_{m^*}^{2\gamma+(1/2-\delta/2)\wedge(1-\delta)} \exp\{2B\sigma^\delta(\pi D_{m^*})^\delta\} \tag{6.29}$$

and

(6.30)

$$\lambda_2(\|\theta\|_2) = \lambda_2(\gamma,\kappa_0,B,\sigma,\delta,\|\theta\|_2) = \begin{cases} \lambda_1(\gamma,\kappa_0,B,\sigma,\delta) & \text{if } \delta > 1, \\ \kappa_0^{-1}(2\pi)^{-1/2}\lambda_1^{1/2}(\gamma,\kappa_0,B,\sigma,\delta)(1+\sigma^2\pi^2)^{\gamma/2}\|\theta\|_2 & \text{if } \delta \leq 1. \end{cases}$$

Here,

$$\begin{aligned}P_{j,k}(m) &= \iint f^2(x)u^*_{\varphi_{m,j}}(x+y)u^*_{\varphi_{m,k}}(-(x+y))g(x)f_\varepsilon(y)dxdy \\ &= D_m \iint f^2(x) \iint e^{-i(x+y)uD_m}\frac{\varphi^*(u)e^{iju}}{f_\varepsilon^*(D_m u)}e^{i(x+y)vD_m}\frac{\varphi^*(v)e^{ikv}}{f_\varepsilon^*(D_m v)}dudvg(x)f_\varepsilon(y)dxdy \\ &= D_m \iint \frac{e^{iju+ikv}\varphi^*(u)\varphi^*(v)}{f_\varepsilon^*(D_m u)f_\varepsilon^*(D_m v)}\left(\iint e^{-i(x+y)(u-v)D_m}f^2(x)g(x)f_\varepsilon(y)dxdy\right)dudv \\ &= D_m \iint \frac{e^{iju+ikv}\varphi^*(u)\varphi^*(v)}{f_\varepsilon^*(D_m u)f_\varepsilon^*(D_m v)}[(f^2g)*f_\varepsilon]^*((u-v)D_m)dudv.\end{aligned}$$

Therefore by Parseval's formula

$$\begin{aligned}\sum_{j,k}|P_{j,k}(m)|^2 &= D_m^2\int\int\left|\frac{\varphi^*(u)\varphi^*(v)}{f_\varepsilon^*(D_m u)f_\varepsilon^*(D_m v)}[(f^2g)*f_\varepsilon]^*((u-v)D_m)\right|^2 dudv \\ &= \Delta_2(m,(f^2g)*f_\varepsilon).\end{aligned}$$

Since $\|(f^2g)*f_\varepsilon\|_2 \leq \|f^2g\|_2\|f_\varepsilon\|_2 = \mathbb{E}^{1/2}(f^2(X_1))\|f_\varepsilon\|_2$, and $\lambda_2(\|f^2g\|_2\|f_\varepsilon\|_2) \leq \mu_2$ by using the definition of μ_2 given in (5.7), we propose to take

$$v_1 = v_1(m,m') = \mu_2\Gamma_2(m^*). \tag{6.31}$$

Lastly, arguing as for (6.12), we have $\mathbb{E}[\sup_{t\in B_{m,m'}(0,1)}|\nu_{n,1}(t)|^2] \le \mathbb{E}(f^2(X_1))\lambda_1\Gamma(m^*)/n$ and thus we propose to take

$$\mathbb{H}_1^2 = \mathbb{H}_1^2(m,m') = \mathbb{E}(f^2(X_1))\lambda_1\Gamma(m^*)/n. \tag{6.32}$$

It follows from (6.26), (6.27), (6.31) and (6.32) that if

$$p_1(m,m') = 2(1+2\epsilon_1)\mathbb{H}_1^2 = 2(1+2\epsilon_1)\mathbb{E}(f^2(X_1))\lambda_1\Gamma(m^*)/n$$

then

$$\mathbb{E}(W_{n,1}(m')) \le \mathbb{E}\left[\sup_{t\in B_{m,m'}(0,1)}|\nu_{n,1}(t)|^2 - 2(1+2\epsilon_1)\mathbb{H}_1^2\right]_+ \le A_1(m^*) + B_1(m^*) \tag{6.33}$$

with

$$A_1(m) = K_3\frac{\mu_2\Gamma_2(m)}{n}\exp\left(-K_1\epsilon_1\mathbb{E}(f^2(X_1))\frac{\lambda_1\Gamma(m)}{\mu_2\Gamma_2(m)}\right) \tag{6.34}$$

$$\text{and } B_1(m) = K_3\frac{\kappa_{\infty,G}^2\lambda_1\Gamma(m)}{n^2}\exp\left\{-K_2\sqrt{\epsilon_1}C(\epsilon_1)\frac{\sqrt{\mathbb{E}(f^2(X_1))}}{\kappa_{\infty,G}}\sqrt{n}\right\}. \tag{6.35}$$

Since for any $m\in\mathcal{M}_{n,\ell}$, $\Gamma(m)\le n$ and $|\mathcal{M}_{n,\ell}|\le n$, it follows that there exist some constants K_4 and c such that

$$\sum_{m\in\mathcal{M}_{n,\ell}} B_1(m^*) \le K_3\|f\|_{\infty,G}^2\lambda_1\exp[-K_4\sqrt{\mathbb{E}(f^2(X_1))}\sqrt{n}/\kappa_{\infty,G}] \le c/n.$$

Let us now come to the study of $A_1(m^*)$.

1) Case $0\le\delta<1/3$. In that case, $\delta\le(1/2-\delta/2)_+$ and the choice $\epsilon_1 = 1/2$ ensures the convergence of $\sum_{m'\in\mathcal{M}_{n,\ell}} A_1(m^*)$. Indeed, if we denote by $\psi = 2\gamma + (1/2-\delta/2)\wedge(1-\delta)$, $\omega = (1/2-\delta/2)_+$, $K' = \kappa_2\lambda_1/\mu_2$, then for $a,b\ge 1$, we infer that

$$\begin{aligned}(a\vee b)^\psi e^{2B\sigma^\delta\pi^\delta(a\vee b)^\delta}e^{-K'\xi^2(a\vee b)^\omega} &\le (a^\psi e^{2B\sigma^\delta\pi^\delta a^\delta} + b^\psi e^{2B\sigma^\delta\pi^\delta b^\delta})e^{-(K'\xi^2/2)(a^\omega+b^\omega)}\\ &\le a^\psi e^{2B\sigma^\delta\pi^\delta a^\delta}e^{-(K'\xi^2/2)a^\omega}e^{-(K'\xi^2/2)b^\omega} + b^\psi e^{2B\sigma^\delta\pi^\delta b^\delta}e^{-(K'\xi^2/2)b^\omega}.\end{aligned} \tag{6.36}$$

Since the function $a\mapsto a^\psi e^{2B\sigma^\delta\pi^\delta a^\delta}e^{-(K'\xi^2/2)a^\omega}$ is bounded on $\mathbb{R}^+$ by a constant, depending on γ, δ and K' only, and since $Ak^\delta - Bk^\omega \le -(B/2)k^\omega$ for any $k\ge 1$, it follows that $\sum_{m'\in\mathcal{M}_{n,\ell}} A_1(m^*) \le C/n$.

2) Case $\delta = 1/3$. In that case, $\delta = (1/2-\delta/2)_+$, and $\omega=\delta$. We choose $\epsilon_1 = \epsilon_1(m,m')$ such that $2B\sigma^\delta\pi^\delta D_{m^*}^\delta - K'\mathbb{E}(f^2(X_1))\epsilon_1 D_{m^*}^\delta = -2B\sigma^\delta\pi^\delta D_{m^*}^\delta$ that is, since $K' = K_1\lambda_1/\mu_2$, $\epsilon_1 = \epsilon_1(m,m') = (4B\sigma^\delta\pi^\delta\mu_2)/(K_1\lambda_1\mathbb{E}(f^2(X_1)))$.

3) Case $\delta>1/3$. In that case, $\delta > (1/2-\delta/2)_+$. Bearing in mind the inequality (6.36) we choose $\epsilon_1 = \epsilon_1(m,m')$ such that $2B\sigma^\delta\pi^\delta D_{m^*}^\delta - K'\mathbb{E}(f^2(X_1))\epsilon_1 D_{m^*}^\omega = -2B\sigma^\delta\pi^\delta D_{m^*}^\delta$ that is, since $K' = K_1\lambda_1/\mu_2$, $\epsilon_1 = \epsilon_1(m,m') = (4B\sigma^\delta\pi^\delta\mu_2)/(K_1\lambda_1\mathbb{E}(f^2(X_1)))D_{m^*}^{\delta-\omega}$.

These choices ensure that $\sum_{m'\in\mathcal{M}_{n,\ell}} A_1(m^*) \le C/n$.

• Study of $W_{n,2}$.

Let us define

$$\mathbb{H}_\xi^2(m,m') = \left(\frac{1}{n}\sum_{i=1}^n \xi_i^2\right)\frac{\lambda_1\Gamma(m^*)}{n}, \tag{6.37}$$

and write

$$\begin{aligned}\left(\frac{1}{n}\sum_{i=1}^n \xi_i^2\right)\frac{\lambda_1\Gamma(m)}{n} &= \left(\frac{1}{n}\sum_{i=1}^n \xi_i^2-\sigma_\xi^2\right)\frac{\lambda_1\Gamma(m)}{n}+\sigma_\xi^2\frac{\lambda_1\Gamma(m)}{n}\\ &\leq \left(\frac{1}{n}\sum_{i=1}^n \xi_i^2-\sigma_\xi^2\right)\mathbb{I}_{\{n^{-1}|\sum_{i=1}^n(\xi_i^2-\sigma_\xi^2)|>\sigma_\xi^2/2\}}\frac{\lambda_1\Gamma(m)}{n}+\frac{3\sigma_\xi^2}{2}\frac{\lambda_1\Gamma(m)}{n}.\end{aligned}$$

It follows that $\mathbb{H}_\xi^2(m,m')\leq \mathbb{H}_{\xi,1}(m,m')+\mathbb{H}_{\xi,2}(m,m')$ where

$$\mathbb{H}_{\xi,1}(m,m')=\left(\frac{1}{n}\sum_{i=1}^n \xi_i^2-\sigma_\xi^2\right)\mathbb{I}_{\{n^{-1}|\sum_{i=1}^n\xi_i^2-\sigma_\xi^2|>\sigma_\xi^2/2\}}\frac{\lambda_1\Gamma(m^*)}{n},$$

$$\text{and}\ \ \mathbb{H}_{\xi,2}(m,m')=\frac{3\sigma_\xi^2}{2}\frac{\lambda_1\Gamma(m^*)}{n}.$$

Then using that $\nu_{n,2}(t)=n^{-1}\sum_{i=1}^n\xi_i(u_t^*(Z_i)-\langle t,g\rangle)+\langle t,g\rangle n^{-1}\sum_{i=1}^n\xi_i$, we can write consecutively

$$\begin{aligned}&\mathbb{E}\left[\sup_{t\in B_{m,m'}(0,1)}|\nu_{n,2}(t)|^2-p_2(m,m')\right]_+\\ \leq\ &\mathbb{E}\left[2\sup_{t\in B_{m,m'}(0,1)}\left(\frac{1}{n}\sum_{i=1}^n\xi_i(u_t^*(Z_i)-\langle t,g\rangle\rangle\right)^2-4(1+2\epsilon_2)\mathbb{H}_\xi^2(m,m')\right]_+\\ &+2\|g\|_2^2\mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^n\xi_i\right)^2\right]+\mathbb{E}[4(1+2\epsilon_2)\mathbb{H}_\xi^2(m,m')-p_2(m,m')]_+,\end{aligned}$$

that is

$$\begin{aligned}(6.38)\quad &\mathbb{E}\left[\sup_{t\in B_{m,m'}(0,1)}|\nu_{n,2}(t)|^2-p_2(m,m')\right]_+\\ &\leq 2\mathbb{E}\left[\sup_{t\in B_{m,m'}(0,1)}\left(\frac{1}{n}\sum_{i=1}^n\xi_i(u_t^*(Z_i)-\langle t,g\rangle)\right)^2-2(1+2\epsilon_2)\mathbb{H}_\xi^2(m,m')\right]_+\\ &+2\frac{\|g\|_2^2\sigma_\xi^2}{n}+4(1+2\epsilon_2)\mathbb{E}|\mathbb{H}_{\xi,1}(m,m')|+\mathbb{E}[4(1+2\epsilon_2)\mathbb{H}_{\xi,2}(m,m')-p_2(m,m')]_+.\end{aligned}$$

Since we only consider dimensions D_m such that, $\Gamma(m)/n$ is bounded by some constant κ, we get that for some $p\geq 2$,

$$\mathbb{E}|\mathbb{H}_{\xi,1}(m,m')|\ \leq\ \kappa\lambda_1\mathbb{E}\left[|\frac{1}{n}\sum_{i=1}^n\xi_i^2-\sigma_\xi^2|\mathbb{I}_{\{n^{-1}|\sum_{i=1}^n(\xi_i^2-\sigma_\xi^2)|>\sigma_\xi^2/2\}}\right]\leq\frac{\kappa\lambda_1 2^{p-1}}{\sigma_\xi^{2(p-1)}}\mathbb{E}\left[|n^{-1}\sum_{i=1}^n\xi_i^2-\sigma_\xi^2|^p\right]$$

By using Rosenthal's inequality (7.2) recalled in Appendix, we find

$$\mathbb{E}|n^{-1}\sum_{i=1}^{n}\xi_i^2-\sigma_\xi^2|^p \le C'(p)\left(\frac{\sigma_{\xi,2p}^{2p}}{n^{p-1}}+\frac{\sigma_{\xi,4}^{2p}}{n^{p/2}}\right),$$

where $\sigma_{\xi,p}^p := \mathbb{E}(|\xi|^p)$, $\sigma_{\xi,2}^2=\sigma_\xi^2$. Now, Assumption $(\mathbf{A_4})$ implies that $\gamma>1/2$, therefore $|\mathcal{M}_n|\le\sqrt{n}$ and consequently, by choosing $p=3$ this leads to

$$\sum_{m'\in\mathcal{M}_n}\mathbb{E}|\mathbb{H}_{\xi,1}(m,m')|\le C(\sigma_{\xi,6},\sigma_\xi)/n.$$

Moreover, the last term of the inequality (6.38) vanishes as soon as

$$p_2(m,m')=4(1+2\epsilon_2)\mathbb{H}_{\xi,2}(m,m')=6(1+2\epsilon_2)\lambda_1\sigma_\xi^2\Gamma(m^*)/n. \tag{6.39}$$

Therefore, for this choice of $p_2(m,m')$, the inequality (6.38) becomes

$$\begin{aligned}\mathbb{E}&\left[\left(\sup_{t\in B_{m,\hat m}(0,1)}|\nu_{n,2}(t)|\right)^2-p_2(m,\hat m)\right]_+\\ &\le 2\sum_{m'\in\mathcal{M}_{n,\ell}}\mathbb{E}\left[\sup_{t\in B_{m,m'}(0,1)}\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i(u_t^*(Z_i)-\langle t,g\rangle)\right)^2-2(1+2\epsilon_2)\mathbb{H}_\xi^2(m,m')\right]_+\\ &\qquad\qquad+2\frac{\|g\|_2^2\sigma_\xi^2}{n}+4(1+2\epsilon_2)\frac{C}{n}.\end{aligned}$$

Then we apply the following Lemma to reach the same kind of result as (6.26) for $W_{n,1}$.

Lemma 6.2. *Under the assumptions of Theorem 5.2, if the ξ_i's admit moments of order $p\ge 6$ and if $m^*=m\vee m'$, then the following inequality holds, for some given $\epsilon_2>0$:*

$$\begin{aligned}&\sum_{m'\in\mathcal{M}_{n,\ell}}\mathbb{E}\left[\sup_{t\in B_{m,m'}(0,1)}\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i(u_t^*(Z_i)-\langle t,g\rangle\right)^2-2(1+2\epsilon_2)\mathbb{H}_\xi^2(m,m')\right]_+\\ &\le K_1\left\{\sum_{m'\in\mathcal{M}_{n,\ell}}\left[\frac{\sigma_\xi^2\mu_2\Gamma_2(m^*)}{n}\exp\left(-K_1\epsilon_2\frac{\lambda_1\Gamma(m^*)}{\mu_2\Gamma_2(m^*)}\right)\right]+\left(1+\frac{\ln^4(n)}{\sqrt{n}}\right)\frac{1}{n}\right\},\end{aligned}\tag{6.40}$$

where μ_2 and $\Gamma_2(m)$ are defined by (5.7) and (6.29) and K_1 is a constant depending on the moments of ξ. The constant μ_2 can be replaced by $\lambda_2(\|h\|_2)$ where λ_2 is defined by (6.30).

By analogy with (6.34) we denote by

$$A_2(m^*)=\frac{K_1\sigma_\xi^2}{n}\mu_2\Gamma_2(m^*)\exp\left(-K_1\epsilon_2\frac{\lambda_1\Gamma(m^*)}{\mu_2\Gamma_2(m^*)}\right)=\frac{K_1\sigma_\xi^2\mu_2\Gamma_2(m^*)}{n}\exp\left(-\kappa_2\epsilon_2\frac{\lambda_1}{\mu_2}D_{m^*}^{(1/2-\delta/2)_+}\right).\tag{6.41}$$

With $p_2(m,m')$ given by (6.39), by gathering (6.26) and (6.40), we find, for $W_{n,2}$ defined by (6.23),

$$\mathbb{E}(W_{n,2}(\hat m))\le K\sum_{m'\in\mathcal{M}_n}A_2(m^*)+\frac{C}{n}\left(1+\frac{\ln(n)^6}{n}\right)+\frac{K'}{n}.$$

The sum $\sum_{m'\in\mathcal{M}_n} A_2(m^*)$ is bounded in the same way as the sum $\sum_{m'\in\mathcal{M}_n} A_1(m^*)$ with $\epsilon_2 = \epsilon_1 = 1/2$ if $0 \le \delta < 1/3$ and $\epsilon_1(m,m')$ replaced by $\epsilon_2 = \epsilon_2(m,m') = \mathbb{E}(f^2(X_1))\epsilon_1(m,m')$, when $\delta \ge 1/3$ that is $\epsilon_2(m,m') = (4B\sigma^\delta\pi^\delta\mu_2)/(K_1\lambda_1)D_{m^*}^{\delta-\omega}$. These choices ensure that $\sum_{m'\in\mathcal{M}_{n,\ell}} A_2(m^*) \le C/n$. The result follows by taking as announced in (6.24)

$$\begin{aligned} p_\ell(m,m') &= 2p_1(m,m') + 2p_2(m,m') \\ &= 4[(1+2\epsilon_1(m,m'))\mathbb{E}(f^2(X_1)) + 3(1+2\epsilon_2(m,m'))\sigma_\xi^2]\lambda_1\Gamma(m^*)/n, \end{aligned}$$

that is, if $0 \le \delta < 1/3$,

$$p_\ell(m,m') = 24\mathbb{E}(Y_1^2)\frac{\lambda_1\Gamma(m^*)}{n}, \tag{6.42}$$

and if $\delta \ge 1/3$,

$$p_\ell(m,m') = 4\left[3\mathbb{E}(Y_1^2) + \frac{32B\sigma^\delta\pi^\delta\mu_2}{k_1\lambda_1}D_{m^*}^{\delta-\omega}\right]\frac{\lambda_1\Gamma(m^*)}{n}. \tag{6.43}$$

Consequently if $0 \le \delta < 1/3$, we take

$$\mathrm{pen}_\ell(m) = \kappa\mathbb{E}(Y_1^2)\lambda_1\frac{\Gamma(m)}{n},$$

and if $\delta \ge 1/3$ we take

$$\mathrm{pen}_\ell(m) = \kappa\left[\mathbb{E}(Y_1^2) + \frac{B\sigma^\delta\pi^\delta\mu_2}{k_1\lambda_1}D_m^{\delta-\omega}\right]\frac{\lambda_1\Gamma(m)}{n},$$

for some numerical constants κ. Note that for $\delta = 1/3$, $\delta - \omega = 0$ and the second penalty has the same order as the first one with a modified multiplicative constant. $\square$

6.4. **Proof of Lemma 6.2, by using a conditioning argument.** We work conditionally to the ξ_i's and $\mathbb{E}_\xi$ and $\mathbb{P}_\xi$ denote the conditional expectations and probability for fixed $\xi_1, \dots, \xi_n$.

We apply Lemma 6.1, conditionally to the ξ_i's to the random variables $(\xi_1, Z_1), \dots, (\xi_n, Z_n)$ which are independent but non identically distributed since the ξ_i's are fixed constants, and with $f_t(\xi_i, Z_i) = \xi_i u_t^*(Z_i)$. Let $Q_{j,k} = \mathbb{E}[u^*_{\varphi_{m,j}}(Z_1)u^*_{\varphi_{m,k}}(-Z_1)]$. Straightforward calculations give that

$$\mathbb{E}_\xi^2\left(\sup_{t\in B_{m,m'}(0,1)}\frac{1}{n}\sum_{i=1}^n \xi_i(u_t^*(Z_i) - \langle t, g\rangle)\right) \le \mathbb{H}_\xi^2(m,m')$$

with $\mathbb{H}_\xi(m,m')$ defined in (6.37). Again, arguing as in Comte and Taupin (2003), $\sum_{j,k}|Q_{j,k}|^2 \le \Delta_2(m,h) \le \lambda_2(\|h\|_2)\Gamma_2(m, \|f_\varepsilon\|_2)$ with $\|h\|_2 \le \|f_\varepsilon\|_2$, where $\Delta_2(m,h)$ is defined by (6.28), λ_2 by (6.30), $\Gamma_2(m)$ by (6.29), μ_2 by (5.7). We now write that

$$\sup_{t\in B_{m,m'}(0,1)}\left(\frac{1}{n}\sum_{i=1}^n \mathrm{Var}_\xi(\xi_i u_t^*(Z_i))\right) \;\le\; \left(\frac{1}{n}\sum_{i=1}^n \xi_i^2\right)\mu_2\Gamma_2(m^*, \|f_\varepsilon\|_2)$$

and thus we take

$$v_\xi(m,m') = \left(\frac{1}{n}\sum_{i=1}^n \xi_i^2\right)\mu_2\Gamma_2(m^*, \|f_\varepsilon\|_2).$$

Lastly,

$$\sup_{t\in B_{m,m'}(0,1)} \|f_t\|_\infty \le 2\max_{1\le i\le n}|\xi_i|\sqrt{\Delta(m^*)} \le 2\max_{1\le i\le n}|\xi_i|\sqrt{\lambda_1\Gamma(m^*)}$$

and we define $M_{1,\xi}(m,m') = 2\max_{1\le i\le n}|\xi_i|\sqrt{\lambda_1\Gamma(m^*)}$. By applying Lemma 6.1, we get for some constants κ_1, κ_2, κ_3

$$\begin{aligned}\mathbb{E}_\xi\left[\sup_{t\in B_{m,m'}(0,1)} \nu_{n,1}^2(t) - 2(1+2\epsilon)\mathbb{H}_\xi^2\right]_+ &\le \kappa_1\left[\frac{\mu_2\Gamma_2(m^*)}{n^2}(\sum_{i=1}^n \xi_i^2)\exp\left\{-\kappa_2\epsilon\frac{\lambda_1\Gamma(m^*)}{\mu_2\Gamma_2(m^*)}\right\}\right.\\ &\left.\quad+\frac{\lambda_1\Gamma(m^*)}{n^2}(\max_{1\le i\le n}\xi_i^2)\exp\left\{-\kappa_3\sqrt{\epsilon}C(\epsilon)\frac{\sqrt{\sum_{i=1}^n\xi_i^2}}{\max_i|\xi_i|}\right\}\right]\end{aligned}$$

To relax the conditioning, it suffices to integrate with respect to the law of the ξ_i's the above expression. The first term in the bound simply becomes:

$$\frac{\sigma_\xi^2\mu_2\Gamma_2(m^*)}{n}\exp\left(-\kappa_2\epsilon\lambda_1\Gamma(m^*)/(\mu_2\Gamma_2(m^*))\right)$$

and has the same order as in the case of bounded variables. The second term is bounded by

$$\frac{\lambda_1\Gamma(m^*)}{n^2}\mathbb{E}\left[(\max|\xi_i|^2)\exp\left(-\kappa_3\sqrt{\epsilon}C(\epsilon)\frac{\sqrt{\sum_{i=1}^n\xi_i^2}}{\max_{1\le i\le n}|\xi_i|}\right)\right]. \tag{6.44}$$

Since we only consider dimensions D_m such that the penalty term is bounded, we have $\Gamma(m)/n\le K$ and the sum of the above terms for $m\in\mathcal{M}_{n,\ell}$ and $|\mathcal{M}_{n,\ell}|\le n$ is less than

$$\lambda_1\mathbb{E}\left[\left(\max_{1\le i\le n}\xi_i^2\right)\exp\left(-\kappa_3\sqrt{\epsilon}C(\epsilon)\frac{\sqrt{\sum_{i=1}^n\xi_i^2}}{\max_{1\le i\le n}|\xi_i|}\right)\right].$$

We need to study when such a term is less than c/n for some constant c. We bound $\max_i|\xi_i|$ by b on the set $\{\max_i|\xi_i|\le b\}$ and the exponential by 1 on the set $\{\max_i|\xi_i|\ge b\}$ and by denoting $\mu=\kappa_3\sqrt{\epsilon}C(\epsilon)$, this yields

$$\begin{aligned}&\mathbb{E}\left[\max_{1\le i\le n}\xi_i^2\exp\left(-\mu\sqrt{\frac{\sum_{i=1}^n\xi_i^2}{\max_{1\le i\le n}\xi_i^2}}\right)\right]\\ \le\ & b^2\mathbb{E}\left(\exp(-\mu\frac{\sqrt{\sum_{i=1}^n\xi_i^2}}{b})\right)+\mathbb{E}\left(\max_{1\le i\le n}\xi_i^2\mathbb{I}_{\{\max_{1\le i\le n}|\xi_i|\ge b\}}\right)\\ \le\ & b^2\left[\mathbb{E}\left(\exp(-\mu\sqrt{n\sigma_\xi^2/(2b^2)}\right)+\mathbb{P}\left(|\frac{1}{n}\sum_{i=1}^n\xi_i^2-\sigma_\xi^2|\ge\sigma_\xi^2/2\right)\right]+b^{-r}\mathbb{E}(\max_{1\le i\le n}|\xi_i|^{r+2})\\ \le\ & b^2e^{-\mu\sqrt{n}\sigma_\xi/(\sqrt{2}b)}+b^22^p\sigma_\xi^{-2p}\mathbb{E}\left(\left|\frac{1}{n}\sum_{i=1}^n\xi_i^2-\sigma_\xi^2\right|^p\right)+b^{-r}\mathbb{E}(\max_{1\le i\le n}|\xi_i|^{r+2}).\end{aligned}$$

Again by applying Rosenthal's inequality (7.2) in Appendix, we get that

$$\begin{aligned}
&\mathbb{E}\left[\max_{1\le i\le n}\xi_i^2\exp\left(-\mu\sqrt{\frac{\sum_{i=1}^n\xi_i^2}{\max_{1\le i\le n}\xi_i^2}}\right)\right]\\
\le\ & b^2e^{-\mu\sqrt{n}\sigma_\xi/(\sqrt{2}b)}+b^2\frac{2^p}{\sigma_\xi^{2p}}\frac{C(p)}{n^p}\left[n\mathbb{E}(|\xi_1^2-\sigma_\xi^2|^p)+(n\mathbb{E}(\xi_1^4))^{p/2}\right]+n\frac{\mathbb{E}(|\xi_1|^{r+2})}{b^r}\\
\le\ & b^2e^{-\mu\sqrt{n}\sigma_\xi/(\sqrt{2}b)}+C'(p)b^2\sigma_{\xi,2p}^{2p}\frac{2^p}{\sigma_\xi^{2p}}\left[n^{1-p}+n^{-p/2}\right]+n\frac{\sigma_{\xi,r+2}^{r+2}}{b^r}.
\end{aligned}$$

Since the ξ_i's admit moments of order 6, we take $p=3$, $r=4$ and $b=\sigma_\xi\sqrt{\epsilon}C(\epsilon)\kappa_3\sqrt{n}/[2\sqrt{2}(\ln(n)-\ln\ln n)]$ and we find that for any $n\ge 3$,

$$\mathbb{E}\left\{\left(\max_{1\le i\le n}\xi_i^2\right)\exp\left(-\kappa_3\sqrt{\epsilon}C(\epsilon)\sqrt{\sum_{i=1}^n\xi_i^2/\max_{1\le i\le n}\xi_i^2}\right)\right\}\le\frac{C_1}{\sqrt{n}}+C_2\left(\frac{\ln^4(n)}{\sqrt{n}}\right)\frac{1}{\sqrt{n}},$$

with C_1 and C_2 constants depending on the moments of ξ. Then the sum over $\mathcal{M}_{n,\ell}$ with cardinality less than $\sqrt{n}$ of the terms in (6.44) is bounded by $C(1+\ln(n)^4/\sqrt{n})/n$ for some constant C, by using again that $\Gamma(m^*)/n$ is bounded.

6.5. **Proof of Corollary 5.1.** Consider the set $\Omega_b=\left\{\left|\frac{\hat m_2(Y)}{\mathbb{E}(Y_1^2)}-1\right|<b\right\}$, $0<b<1$. On Ω_b, we have $\hat m_2(Y)<(b+1)\mathbb{E}(Y_1^2)$ and $\mathbb{E}(Y_1^2)<\hat m_2(Y)/(1-b)$. It follows from (6.18), that

$$\|\tilde\ell-\ell_m\|_2^2\le\kappa_x\|\ell-\ell_m^{(n)}\|_2^2+\kappa_x(\widehat{\mathrm{pen}}_\ell(m)-\widehat{\mathrm{pen}}_\ell(m')+xp_\ell(m,\hat m))+x\kappa_x\sum_{m\in\mathcal{M}_{n,\ell}}W_n(m'),$$

where, according to (6.42), (6.43) and (6.43), $p_\ell(m,m')=12\lambda_1\mathbb{E}(Y_1^2)D_{m^*}^{2\gamma+1-\delta}\Gamma(m^*)/n$, if $0\le\delta<1/3$ and $p_\ell(m,m')=2\left[3\mathbb{E}(Y_1^2)+32B\sigma^\delta\pi^\delta\mu_2D_{m^*}^{\delta-\omega}\right]\lambda_1\Gamma(m^*)/(k_1\lambda_1 n)$ if $\delta\ge 1/3$. Therefore, for $0\le\delta<1/3$, we choose

$$\widehat{\mathrm{pen}}_\ell(m)=\frac{x\kappa}{(1-b)}\lambda_1\hat m_2(Y)\frac{\Gamma(m)}{n},$$

and for $\delta\ge 1/3$, we choose

$$\widehat{\mathrm{pen}}_\ell(m)=\frac{x\kappa\lambda_1\Gamma(m)}{n}\left[\frac{\hat m_2(Y)}{1-b}+\frac{B\sigma^\delta\pi^\delta\mu_2}{K_1\lambda_1}D_m^{\delta-\omega}\right].$$

For those choices, if $0\le\delta<1/3$, on Ω_b, we have

$$\widehat{\mathrm{pen}}_\ell(m)-\widehat{\mathrm{pen}}_\ell(\hat m)-p_\ell(m,\hat m)\le x\kappa\lambda_1(\hat m_2(Y)/(1-b)+\mathbb{E}(Y_1^2))\Gamma(m)/n,$$

and if $\delta\ge 1/3$, on Ω_b

$$\widehat{\mathrm{pen}}_\ell(m)-\widehat{\mathrm{pen}}_\ell(\hat m)-p_\ell(m,\hat m)\le x\kappa\lambda_1\frac{\Gamma(m)}{n}\left[(\hat m_2(Y)/(1-b)+\mathbb{E}(Y_1^2))+\frac{B\sigma^\delta\pi^\delta\mu_2}{K_1\lambda_1}D_m^{\delta-\omega}\right].$$

Therefore, since (6.21) still holds, we obtain that if $0\le\delta<1/3$, then

$$\mathbb{E}(\|\tilde\ell-\ell\|_2^2\mathbf{I}_{\Omega_b})\le\kappa_x^2\|\ell-\ell_m\|_2^2+2\kappa_x\kappa\lambda_1(\frac{1}{1-b}+1)\mathbb{E}(Y_1^2)\frac{\Gamma(m)}{n}+\frac{x\kappa_xc}{n},$$

and finally if $\delta \geq 1/3$,

$$\mathbb{E}(\|\tilde{\ell}-\ell\|_2^2 \mathbf{1}_{\Omega_b}) \leq \kappa_x^2\|\ell-\ell_m\|_2^2 + 2\kappa_x\kappa\left[(\frac{1}{1-b}+1)\lambda_1\mathbb{E}(Y_1^2) + B\sigma^\delta\pi^\delta\mu_2 D_m^{\delta-\omega}\right]\frac{\Gamma(m)}{n} + \frac{x\kappa_x c}{n}.$$

That is

$$\mathbb{E}(\|\tilde{\ell}-\ell\|_2^2 \mathbf{1}_{\Omega_b}) \leq \kappa_x^2\|\ell-\ell_m\|_2^2 + 2\kappa_x\kappa \mathrm{pen}_\ell(m) + x\kappa_x c/n \tag{6.45}$$

where $\mathrm{pen}_\ell(m)$ is defined in Theorem 5.2. It remains to bound $\mathbb{E}(\|\tilde{\ell}-\ell\|_2^2 \mathbf{1}_{\Omega_b^c})$. On the one hand, we write that

$$\mathbb{P}(\Omega_b^c) = \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n (Y_i^2 - \mathbb{E}(Y_i^2))\right| > b\mathbb{E}(Y_1^2)\right) \leq \frac{1}{b^p\mathbb{E}^p(Y_1^2)}\mathbb{E}\left(\left|\frac{1}{n}\sum_{i=1}^n (Y_i^2 - \mathbb{E}(Y_i^2))\right|^p\right).$$

Then we can apply Rosenthal's inequality (7.2) as recalled in Appendix and we find

$$\begin{aligned}\mathbb{P}(\Omega_b^c) &\leq \frac{C(p)}{(nb\mathbb{E}(Y_1^2))^p}\left\{\sum_{i=1}^n \mathbb{E}|Y_i^2-\mathbb{E}(Y_i^2)|^p + \left(\sum_{i=1}^n \mathbb{E}[(Y_i^2-\mathbb{E}(Y_i^2))^2]\right)^{p/2}\right\}\\ &\leq \frac{C(p)}{(nb\mathbb{E}(Y_1^2))^p}\left\{n2^p\mathbb{E}(|Y_1|^{2p}) + n^{p/2}[\mathbb{E}(Y_1^4)]^{p/2}\right\}.\end{aligned}$$

Since the Y_i's admit moments of order more than 8, we can choose $p=4$ which leads to

$$\mathbb{P}(\Omega_b^c) \leq \frac{C(4)(2^4+1)(\mathbb{E}(Y_1^8))}{b^4\mathbb{E}^4(Y_1^2)}\frac{1}{n^2}. \tag{6.46}$$

On the other hand, we write that for any m in $\mathcal{M}_n$, $\|\hat{\ell}_m\|_2^2 = \sum_{|j|\leq k_n}|\hat{a}_{m,j}(\ell)|^2$ with

$$|\hat{a}_{m,j}(\ell)|^2 = \frac{1}{n^2}\left|\sum_{i=1}^n Y_i u^*_{\varphi_{m,j}}(Z_i)\right|^2 \leq \frac{1}{n^2}\left(\sum_{i=1}^n Y_i^2\right)\left(\sum_{i=1}^n |u^*_{\varphi_{m,j}}(Z_i)|^2\right).$$

It follows that

$$\sum_{|j|\leq k_n}|\hat{a}_{m,j}(\ell)|^2 \leq \frac{1}{n^2}\left(\sum_{i=1}^n Y_i^2\right)\left(\sum_{i=1}^n\sum_{|j|\leq k_n}|u^*_{\varphi_{m,j}}(Z_i)|^2\right) \leq \left(\frac{1}{n}\sum_{i=1}^n Y_i^2\right)\Delta(m).$$

Since we only consider dimension D_m such that $\Delta(m)/n \leq K$ where K is some constant, we find that for all m in $\mathcal{M}_n = \{1,\ldots,m_{n,\ell}\}$, $\|\hat{\ell}_m\|_2^2 \leq K\sum_{i=1}^n Y_i^2$ and therefore $\|\tilde{\ell}\|_2^2 \leq K\sum_{i=1}^n Y_i^2$ and

$$\begin{aligned}\mathbb{E}(\|\tilde{\ell}\|_2^2\mathbf{1}_{\Omega_b^c}) &\leq K\mathbb{E}\left(\sum_{i=1}^n Y_i^2\mathbf{1}_{\Omega_b^c}\right) \leq Kn\mathbb{E}\left(|n^{-1}\sum_{i=1}^n Y_i^2 - \mathbb{E}(Y_1^2)|\mathbf{1}_{\Omega_b^c}\right) + Kn\mathbb{E}(Y_1^2)\mathbb{P}(\mathbf{1}_{\Omega_b^c})\\ &\leq \frac{Kn\mathbb{E}|n^{-1}\sum_{i=1}^n Y_i^2 - \mathbb{E}(Y_i^2)|^p}{b^{p-1}\mathbb{E}^{p-1}(Y_1^2)} + \frac{KC'\mathbb{E}(Y_1^8)}{b^4\mathbb{E}^3(Y_1^2)}\frac{1}{n}\\ &\leq \frac{KC(p)2^p\mathbb{E}(|Y_1|^{2p})}{n^{p-2}b^{p-1}\mathbb{E}^{p-1}(Y_1^2)} + \frac{KC(p)\mathbb{E}^{p/2}(Y_1^4)}{n^{p/2-1}b^{p-1}\mathbb{E}^{p-1}(Y_1^2)} + \frac{KC'\mathbb{E}(Y_1^8)}{b^4\mathbb{E}^3(Y_1^2)}\frac{1}{n}.\end{aligned}$$

By taking $p = 4$ we get that

$$\mathbb{E}(\|\tilde\ell\|_2^2 \mathbf{I}_{\Omega_b^c}) \leq \frac{KC(4)2^4\mathbb{E}(|Y_1|^8)}{n^2b^3\mathbb{E}^3(Y_1^2)} + \frac{KC(4)\mathbb{E}^2(Y_1^4)}{nb^3\mathbb{E}^3(Y_1^2)} + \frac{KC'\mathbb{E}(Y_1^8)}{b^4\mathbb{E}^3(Y_1^2)}\frac{1}{n}. \tag{6.47}$$

By combining (6.46) and (6.47), we find that

$$\mathbb{E}(\|\tilde\ell - \ell\|_2^2 \mathbf{I}_{\Omega_b^c}) \leq 2\mathbb{E}(\|\tilde\ell\|_2^2\mathbf{I}_{\Omega_b^c}) + 2\|\ell\|_2\mathbb{P}(\Omega_b^c) \leq c'/n, \tag{6.48}$$

where c' is a constant depending on $\|\ell\|_2^2$ and on the moments of Y_1. The result follows by combining the inequalities (6.45) and (6.48). □

6.6. **Proof of Theorem 5.3.** Let $\tilde E_n$ be the event $\tilde E_n = \{\| g - \tilde g \|_{\infty,A} \leq g_0/2\}$. Since $g(x) \geq g_0$ for any x in A, then, on $\tilde E_n$, $\tilde g(x) \geq g_0/2$ also for any x in A. It follows that

$$\mathbb{E}\|(f - \tilde f)\mathbf{I}_A\mathbf{I}_{\tilde E_n}\|_2^2 \leq \frac{8}{g_0^2}\mathbb{E}\|\tilde\ell - \ell\|_2^2 + \frac{8\|\ell\|_{\infty,A}^2}{g_0^4}\mathbb{E}\|\tilde g - g\|_2^2, \tag{6.49}$$

where $\|\ell\|_{\infty,A} \leq g_1\kappa_{\infty,G}$. Using that $\|\tilde f\|_{\infty,A} \leq a_n$

$$\mathbb{E}\left[\|(f - \tilde f)\mathbf{I}_A\mathbf{I}_{\tilde E_n^c}\|_2^2\right] \leq 2(a_n^2 + \|f\|_{\infty,A}^2)\lambda(A)\mathbb{P}(\tilde E_n^c), \tag{6.50}$$

where $\lambda(A) = \int_A dx$. It follows that for $\hat m_\ell = \hat m_\ell(n)$, $\hat m_g = \hat m_g(n)$, if $a_n\mathbb{P}(\tilde E_n^c) = o(n^{-1})$, then (5.14) is proved by applying Theorem 5.1 and Theorem 5.2. We now come to the study of $\mathbb{P}(\tilde E_n^c)$ by writing that $\mathbb{P}(\tilde E_n^c) = \mathbb{P}(\|g - \tilde g\|_\infty > g_0/2) = \mathbb{P}\left(\|g - g_{\hat m_g}^{(n)} + g_{\hat m_g}^{(n)} - \tilde g\|_\infty > g_0/2\right)$. By applying Lemma 7.1, and by arguing as for (6.8) (see also Comte and Taupin (2003)), we get that

$$\begin{aligned}\|g - g_{\hat m_g}^{(n)}\|_\infty &\leq \|g - g_{\hat m_g}\|_\infty + \|g_{\hat m_g} - g_{\hat m_g}^{(n)}\|_\infty \\ &\leq \sqrt{\kappa(\kappa_{\mathcal{G}}+1)\frac{D_{\hat m_g}^3}{k_n}} + \frac{\sqrt{\pi D_{\hat m_g}}}{2\pi}((\pi D_{\hat m_g})^2+1)^{-\alpha_g/2}\exp(-\rho_g|\pi D_{\hat m_g}|^{\nu_g})A_g^{1/2}.\end{aligned}$$

Consequently, $\|g - g_{\hat m_g}^{(n)}\|_\infty$ tends to zero as soon as g belongs to some space $\mathcal{S}_{\alpha_g,\nu_g,\rho_g}(\kappa_{\alpha_g})$ defined by (2.1) with $\alpha_g > 1/2$, since $k_n \geq n^2$ and $D_{\hat m_g} \leq \sqrt{n}$. It follows that for n large enough, $\|g - g_{\hat m_g}^{(n)}\|_\infty \leq g_0/4$. It follows that

$$\mathbb{P}(\tilde E_n^c) \leq \mathbb{P}\left[\|g_{\hat m_g}^{(n)} - \tilde g\|_\infty > g_0/4\right]. \tag{6.51}$$

By applying again Lemma 7.1, $g_{\hat m_g}^{(n)} - \tilde g$ belongs to $S_{\hat m_g}$, and consequently

$$\mathbb{P}(\tilde E_n^c) \leq \mathbb{P}\left(\|g_{\hat m_g}^{(n)} - \tilde g\|_2 > g_0/(4\sqrt{D_{\hat m_g}})\right). \tag{6.52}$$

In this context, (6.12) becomes

$$\|g_{m_g}^{(n)} - \tilde g_{\hat m_g}\|_2^2 = \sum_{|j|\leq k_n}(\hat a_{\hat m_g,j} - a_{\hat m_g,j})^2 = \sum_{|j|\leq k_n}\nu_{n,g}^2(\varphi_{\hat m_g,j}) = \sup_{t\in B_{\hat m_g}(0,1)}\nu_{n,g}^2(t). \tag{6.53}$$

Consequently,

$$\begin{aligned}\mathbb{P}(\tilde{E}_n^c) &\leq \mathbb{P}\left[\sup_{t\in B_{\hat{m}_g}(0,1)} |\nu_{n,g}(t)| \geq g_0/(4\sqrt{D_{\hat{m}_g}})\right] \leq \sup_{m\in\mathcal{M}_n} \mathbb{P}\left[\sup_{t\in B_{m_g}(0,1)} |\nu_{n,g}(t)| \geq g_0/(4\sqrt{D_m})\right]\\ &\leq \sum_{m\in\mathcal{M}_n} \mathbb{P}\left[\sup_{t\in B_{m_g}(0,1)} |\nu_{n,g}(t)| \geq g_0/(4\sqrt{D_m})\right].\end{aligned}$$

We apply again the version of Talagrand's (1996) inequality given in Birgé and Massart (1998), recalled in (6.13) to get that if we take $\lambda = g_0/(8\sqrt{D_m})$ and if we ensure $2H < g_0/(8\sqrt{D_m})$, then $\mathbb{P}[\sup_{t\in B_m(0,1)} |\nu_{n,g}(t)| \geq g_0/(4\sqrt{D_m})] \leq 3\exp\left[-K_1'n\left((D_m v)^{-1} \wedge (M_1\sqrt{D_m})^{-1}\right)\right]$. As in the proof of Proposition 3.5,

$$\mathbb{P}(\tilde{E}_n^c) \leq K \sum_{m\in\mathcal{M}_n} \left[\exp[-K_1'n/(M_1\sqrt{D_m})] + \exp[-K_1'n/(D_m v)]\right]. \tag{6.54}$$

Since we only consider D_m such that $D_m \leq \sqrt{n}$,

$$a_n|\mathcal{M}_n| \exp[-K_1'n/(M_1\sqrt{D_m})] \leq a_n|\mathcal{M}_n|\exp(-K\text{"}n^{1/4}) = o(n^{-1}).$$

Again as in the proof of Proposition 3.5, we only consider D_m such that $\Gamma(m)/n$ tends to zero. Consequently, when $\delta > 0$ then $D_m \leq \left(\ln n/(2B\sigma^\delta + 1)\right)^{1/\delta}$ which combined with the fact that $v \leq D_m^{2\gamma+1-\delta}\exp(2B\sigma^\delta\pi^\delta D_m^\delta)$ gives that $a_n|\mathcal{M}_n|\exp\left(-K_1'n/(D_m v)\right) = o(1/n)$.

When $\delta = 0$, then $v = \mu_1 D_m^{2\gamma+1/2}$ and consequently, as $D_m \leq (n/\ln(n))^{1/(2\gamma+1)} \leq n^{1/(2\gamma+1)}$,

$$\exp(-K_1'n/(D_m v)) \leq \exp(-K\text{"}n/(D_m^{2\gamma+3/2})) \leq \exp(-K\text{"}n^{1/(4(\gamma+1))}).$$

Analogously, $\sqrt{D_m}H \leq 1/\sqrt{\ln(n)}$ in the worst case corresponding to $\delta = 0$, for $D_m \leq (n/\ln(n))^{1/(2\gamma+2)}$, tends to zero and therefore is bounded by $g_0/8$ for n great enough. We conclude that if we only consider D_m such that $D_m \leq n^{1/(2\gamma+2)}$ then $a_n\mathbb{P}(\tilde{E}_n^c) = o(1/n)$, and the result follows by applying the inequalities (6.49) and (6.50). □

7. Appendix

Lemma 7.1. *Assume that g belongs to $\mathcal{S}_{\alpha_g,\nu_g,\rho_g}(\kappa_{\alpha_g})$ defined by (2.1) with $\alpha_g > 1/2$. Then*

$$\|g - g_m\|_\infty \leq \frac{\sqrt{\pi D_m}}{2\pi}((\pi D_m)^2 + 1)^{-\alpha_g/2}\exp(-\rho_g|\pi D_m|^{\nu_g})A_g^{1/2}. \tag{7.1}$$

Moreover for any $t \in S_m$ we have that $\|t\|_\infty \leq \sqrt{D_m}\|t\|_2$.

Proof of Lemma 7.1. Let β such that $1/2 < \beta < \alpha_g$.

$$\begin{aligned}
\|g - g_m\|_\infty &\leq \frac{1}{2\pi}\int_{|x|\geq \pi D_m} |g^*(x)|dx \\
&\leq \frac{1}{2\pi}((\pi D_m)^2+1)^{-(\alpha_g-\beta)/2} e^{-\rho_g|\pi D_m|^{\nu_g}} \int_{|x|\geq \pi D_m} |g^*(x)|(x^2+1)^{(\alpha_g-\beta)/2} e^{\rho_g |x|^{\nu_g}} dx \\
&\leq \frac{1}{2\pi}((\pi D_m)^2+1)^{-(\alpha_g-\beta)/2} \exp(-\rho_g|\pi D_m|^{\nu_g})\kappa_{\alpha_g}^{1/2} \sqrt{\int_{|x|\geq \pi D_m} (x^2+1)^{-\beta} dx} \\
&\leq \frac{1}{2\pi}((\pi D_m)^2+1)^{-(\alpha_g-\beta)/2} \exp(-\rho_g|\pi D_m|^{\nu_g})\kappa_{\alpha_g}^{1/2} (\pi D_m)^{1/2-\beta} \\
&\leq \frac{\sqrt{\pi D_m}}{2\pi}((\pi D_m)^2+1)^{-\alpha_g/2} \exp(-\rho_g|\pi D_m|^{\nu_g})\kappa_{\alpha_g}^{1/2},
\end{aligned}$$

and (7.1) follows. Let $t \in S_m$, $t(x) = \sum_{j\in\mathbb{Z}} \langle t, \varphi_{m,j}\rangle \varphi_{m,j}(x)$ and

$$|t(x)|^2 \leq \sum_{j\in\mathbb{Z}} |\langle t, \varphi_{m,j}\rangle|^2 \sum_{j\in\mathbb{Z}} |(\varphi_{m,j}^*)^*(-x)|^2/(2\pi)^2.$$

The first term is simply $\|t\|_2^2$. For the second, $\int e^{itx}\varphi_{m,j}^*(x)dx = D_m^{-1/2}\int e^{itx} e^{ijx/D_m}\varphi^*(x/D_m)dx = D_m^{1/2}\int e^{iju} e^{itD_m u}\varphi^*(u)du$ which implies the result by applying Parseval's Formula since $\sum_{j\in\mathbb{Z}} |(\varphi_{m,j}^*)(-x)|^2/(2\pi)^2 = D_m \int \varphi^*(u)^2 du/(2\pi) = D_m$. □

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 $p \geq 1$,

$$\mathbb{E}[|\sum_{i=1}^n U_i|^p] \leq C(p)[\sum_{i=1}^n \mathbb{E}[|U_i|^p] + (\sum_{i=1}^n \mathbb{E}[U_i^2])^{p/2}]. \tag{7.2}$$

References

[1] Birgé, L. and Massart, P. (1998) Minimum contrast estimators on sieves : exponential bounds and rates of convergence. *Bernoulli* **4** 329-375.

[2] Butucea, C. (2004) Deconvolution of supersmooth densities with smooth noise. *The Canadian Journal of Statistics* **32**(2). 181–192

[3] Butucea, C. and Tsybakov, A.B. (2004) Sharp optimality and some effects of dominating bias in density deconvolution, Preprint LPMA-898, http://www.proba.jussieu.fr/mathdoc/preprints/index.html#2004.

[4] Carroll, R. J. (1999) Nonparametric regression in the presence of measurement error. *Biometrika* **86** (3) 541-554

[5] Comte, F. and Taupin, M.-L. (2003). Penalized contrast estimator for density deconvolution. Working paper 2003-2, http://www.math-info.univ-paris5.fr/map5/.

[6] Fan, J. (1991a) On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19**, 1257-1272.

[7] Fan, J. (1991b). Global behaviour of deconvolution kernel estimates. *Statist. Sinica* **1 (2)** 541-551.

[8] Fan, J. and Gijbels, I. (1996) Local polynomial modelling and its applications. Monographs on Statistics and Applied Probability. 66. London: Chapman & Hall.

[9] Fan, J. and Masry, M. (1992) Multivariate regression estimation with errors-in-variables: asymptotic normality for mixing processes. *J. Multivariate Anal.* **43**, 237-272.

[10] Fan, J. and Truong, Y.K. (1993) Nonparametric regression with errors-in-variables. *Ann. Statist.* **21**, 4, 1900-1925.

[11] Fan, J., Truong, Y. K. and Wang, Y. (1991) Nonparametric function estimation involving errors-in-variables. *Nonparametric Functional Estimation and Related Topics* 613-627.
[12] Ioannides, D. A. and Alevizos, P. D. (1997) Nonparametric regression with errors in variables and applications. *Statist. Probab. Lett.* **32**, 35-43.
[13] Koo, J.-Y. and Lee K.-W. (1998) B-spline estimation of regression functions with errors in variable. *Statist. Probab. Lett.* **40**, 57-66.
[14] Ledoux, M. (1996) On Talagrand's deviation inequalities for product measures. *ESAIM Probab. Statist.*, 63-87.
[15] Masry, E. (1993). Multivariate regression estimation with errors-in-variables for stationary processes. *Nonparametric statistics.* **3**, 13-36.
[16] Petrov, V.V. (1995). Limit theorems of probability theory. Sequences of independent random variables. Oxford Studies in Probability 4. Oxford Science Publications. The Clarendon Press, Oxford University Press, New York.
[17] Rosenthal, H. P. (1970) On the subspaces of L^p, $(p > 2)$, spanned by sequences of independent random variables. *Israel J. Math.* **8**, 273–303.
[18] Talagrand, M. (1996) New concentration inequalities in product spaces. *Invent. Math.* **126**, 505-563.
[19] Tsybakov, A. (1986) Robust reconstruction of functions by the local-approximation method. *Problems of Information Transmission* **22**, 133-146.