Statistica Sinica **17**(2007), 1065-1090

ADAPTIVE ESTIMATION IN A NONPARAMETRIC REGRESSION MODEL WITH ERRORS-IN-VARIABLES

F. Comte and M.-L. Taupin

Université Paris Descartes and IUT de Paris 5, *Université d'Orsay*

Abstract: We consider a regression model with errors-in-variables. Let (Y_i, Z_i), $i = 1, \ldots, n$ be n i.i.d. copies of (Y, Z) satisfying $Y = f(X) + \xi$, $Z = X + \sigma\varepsilon$, involving independent and unobserved random variables X, ξ, ε. The density of ε and the constant noise level σ are known while the densities of X and ξ are unknown. Using the observations (Y_i, Z_i), $i = 1, \cdots, n$, we propose an estimator $\tilde{f}$ of the regression function f which is defined as the ratio of two adaptive estimators $-$ an estimator of $\ell = fg$ divided by an estimator of g, the density of X. Both estimators are obtained by minimization of penalized contrast functions. We prove that the MISE of $\tilde{f}$ on a compact set is bounded by the sum of the two MISEs of the estimators of ℓ and g. Rates of convergence are given when ℓ and g belong to various smoothness classes and when the error ε is either ordinary smooth or super smooth. The rate of $\tilde{f}$ is optimal in a minimax sense in all cases where lower bounds are available.

Key words and phrases: Adaptive estimation, density deconvolution, errors-in-variables, minimax estimation, nonparametric regression, projection estimators.

1. Introduction

Nonparametric estimation in a regression model when variables are observed with measurement errors has been the subject of several recent contributions. The model is the following. Let $(Y_i, Z_i)_{1\le i\le n}$ be a n-sample of independent and identically distributed (i.i.d.) two-dimensional random variables such that

$$Y_i = f(X_i) + \xi_i, \qquad \mathbb{E}(\xi_i) = 0, \tag{1.1}$$

$$Z_i = X_i + \sigma\varepsilon_i, \tag{1.2}$$

where $X_i, \xi_i, \varepsilon_i$ are independent. The problem is to estimate the unknown regression function f from the observations $(Y_i, Z_i)_{i=1,\cdots,n}$. The random variables $(X_i)_{i=1,\cdots,n}$ are unobserved and have common unknown density g. In (1.1), the ξ_i's are supposed to be centered with unknown density. In (1.2), the measurement errors $(\sigma\varepsilon_i)_{i=1,\cdots,n}$ are decomposed into random variables $(\varepsilon_i)_{i=1,\cdots,n}$ with known density f_ε, multiplied by a noise level σ which is also known.

When estimating f in this framework, nonparametric rates of convergence depend not only on the smoothness of f and g, but also on the smoothness of the error density f_ε. As in deconvolution problems, the worst rates correspond to the smoothest error densities. It is classical to distinguish between two kinds of smoothness. A density is called ordinary smooth if its Fourier Transform has a polynomial decay; it is super smooth if its Fourier Transform has an exponential decay.

Nonparametric estimation in Model (1.1)−(1.2) has been initiated by Fan, Truong and Wang (1991) and Fan and Truong (1993), who propose a Nadaraya-Watson type estimator obtained as the ratio of two deconvolution kernel estimators. This problem has been further studied by Fan and Masry (1992), Masry (1993), Truong (1991), and Ioannides and Alevizos (1997), among others. These authors investigate different approaches ($\mathbb{L}_p$-risks, asymptotic normality, ...) under various assumptions on the model (e.g., mixing assumptions) or on the regularity of functions. In particular, when both the regression function f and the density g admit kth-order derivatives, Fan and Truong (1993) study the minimax risk. When ε is either ordinary or super smooth, they give upper and lower bounds for the pointwise quadratic risk and for the $\mathbb{L}_p$-risk restricted to compact sets.

Using a different approach, Koo and Lee (1998) propose an estimation method based on B-splines. Their results are only valid for ordinary smooth errors.

In all the papers quoted above, the regression function f and the density g belong to the same known smoothness class, which is clearly unrealistic.

In this paper, we propose an estimation procedure for f that does not require any prior knowledge on the smoothness of f and g. Our estimation procedure is based on the classical idea that the regression function f at point x can be written as the ratio

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

with $f_{X,Y}$ the joint density of (X, Y). Hence f is estimated by the ratio $\tilde{f} = \tilde{\ell}/\tilde{g}$, where $\tilde{\ell}$ is an adaptive estimator of $\ell = fg$ and $\tilde{g}$ an adaptive estimator of g. Both estimators are obtained by minimization of penalized contrast functions. The contrasts are determined by projection methods and the penalizations give an automatic choice of the relevant projection spaces.

We give upper bounds for the Mean Integrated Squared Error (MISE) of $\tilde{\ell}$ and $\tilde{g}$ when errors are either ordinary or super smooth. We show that the MISE of $\tilde{f}$ on a compact set A is bounded by the sum of the MISEs of $\tilde{\ell}$ and $\tilde{g}$. Consequently, the rate of $\tilde{f}$ is given by the slowest rate of the two adaptive estimators $\tilde{\ell}$ and $\tilde{g}$. The estimator $\tilde{f}$ automatically reaches the minimax rate in all standard cases where lower bounds are available. Other cases are discussed in detail.

Estimators are described in Section 2. In Section 3, we give the upper bounds for MISEs and discuss the minimax properties of estimators. Proofs and technical lemmas are to be found in Section 4.

2. Description of the estimators

For complex-valued functions u and v in $\mathbb{L}_2(\mathbb{R}) \cap \mathbb{L}_1(\mathbb{R})$, u^* denotes the Fourier transform of u with $u^*(x) = \int e^{itx} u(t)dt$, $u * v$ the convolution product, with $u*v(x) = \int u(y)v(x-y)dy$ and $< u, v >= \int u(x)\overline{v}(x)dx$ with $\overline{z}$ the conjugate of a complex number z.

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

We assume that f_ε and f_ε^* belong to $\mathbb{L}_2(\mathbb{R})$, that $f_\varepsilon^*(x) \neq 0$ for all $x \in \mathbb{R}$, and assume that all random variables have a second order moment.

2.1. Projection spaces

Let $\varphi(x) = \sin(\pi x)/(\pi x)$. For $m \in \mathbb{N}$ and $j \in \mathbb{Z}$, set $\varphi_{m,j}(x) = \sqrt{m}\varphi(mx-j)$. The functions $\{\varphi_{m,j}\}_{j\in\mathbb{Z}}$ constitute an orthonormal system in $\mathbb{L}^2(\mathbb{R})$ (see *e.g.* Meyer (1990), p.22). For $m = 2^k$, it is known as the Shannon basis. Though we choose integer values for m here, a thinner grid would also be possible. Define $S_m = \overline{\text{span}}\{\varphi_{m,j},\ j \in \mathbb{Z}\}$, $m \in \mathbb{N}$. The space S_m is exactly the subspace of $\mathbb{L}_2(\mathbb{R})$ of functions having a Fourier transform with compact support contained in $[-\pi m, \pi m]$.

The orthogonal projections of g and ℓ on S_m are given, respectively, by $g_m = \sum_{j\in\mathbb{Z}} a_{m,j}(g)\varphi_{m,j}$ and $\ell_m = \sum_{j\in\mathbb{Z}} a_{m,j}(\ell)\varphi_{m,j}$, where $a_{m,j}(g) =< \varphi_{m,j}, g >$ and $a_{m,j}(\ell) =< \varphi_{m,j}, \ell >$. To obtain representations having a finite number of "coordinates", take $S_m^{(n)} = \text{span}\{\varphi_{m,j},\ |j| \leq k_n\}$, with integers k_n to be specified later. The family $\{\varphi_{m,j}\}_{|j|\leq k_n}$ is an orthonormal basis of $S_m^{(n)}$ and the orthogonal projections of g and ℓ on $S_m^{(n)}$ are given by $g_m^{(n)} = \sum_{|j|\leq k_n} a_{m,j}(g)\varphi_{m,j}$ and $\ell_m^{(n)} = \sum_{|j|\leq k_n} a_{m,j}(\ell)\varphi_{m,j}$, respectively.

2.2. Constructing minimum contrast estimators for ℓ and g

For an arbitrary fixed integer m, an estimator of ℓ belonging to $S_m^{(n)}$ is defined by

$$\hat{\ell}_m = \arg\min_{t\in S_m^{(n)}} \gamma_{n,\ell}(t), \tag{2.1}$$

where, for $t \in S_m^{(n)}$,

$$\gamma_{n,\ell}(t) = \|t\|^2 - 2n^{-1}\sum_{i=1}^{n}(Y_i u_t^*(Z_i)) \text{ with } u_t(x) = (2\pi)^{-1}\frac{t^*(-x)}{f_\varepsilon^*(-x)}. \tag{2.2}$$

It is easy to see that $\hat{\ell}_m = \sum_{|j|\le k_n} \hat{a}_{m,j}(\ell)\varphi_{m,j}$ with $\hat{a}_{m,j}(\ell) = n^{-1}\sum_{i=1}^n Y_i u_{\varphi_{m,j}}^*(Z_i)$. By using the Parseval and inverse Fourier formulae, we get that

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

Since $\mathbb{E}(\gamma_{n,\ell}(t)) = \|t\|_2^2 - 2\langle \ell,\, t\rangle = \|t-\ell\|_2^2 - \|\ell\|_2^2$ is minimal when $t = \ell$, we conclude that $\gamma_{n,\ell}(t)$ suits well for the estimation of $\ell = fg$.

As in Comte, Rozenholc and Taupin (2006), the estimator of g in $S_m^{(n)}$ is defined by

$$\hat{g}_m = \sum_{|j|\le k_n} \hat{a}_{m,j}(g)\varphi_{m,j} \quad \text{with} \quad \hat{a}_{m,j}(g) = n^{-1}\sum_{i=1}^{n} u_{\varphi_{m,j}}^*(Z_i).$$

In other words,

$$\hat{g}_m = \arg\min_{t\in S_m^{(n)}} \gamma_{n,g}(t) \tag{2.3}$$

where, for $t \in S_m^{(n)}$, $\gamma_{n,g}(t) = \|t\|_2^2 - 2n^{-1}\sum_{i=1}^n u_t^*(Z_i)$ and u_t is defined in (2.2).

2.3. Minimum penalized contrast estimators for ℓ and g

In order to construct the minimum penalized contrast estimators, we must define the penalty functions. They are related to the behavior of f_ε^*. We assume that, for all x in $\mathbb{R}$,

$$\kappa_0(x^2+1)^{-\frac{\alpha}{2}}\exp\{-\beta|x|^\rho\} \le |f_\varepsilon^*(x)| \le \kappa_0'(x^2+1)^{-\frac{\alpha}{2}}\exp\{-\beta|x|^\rho\}. \tag{$\mathbf{A_1}$}$$

We only need the left-hand side of ($\mathbf{A_1}$) to define penalties and obtain the upper bounds. The right-hand side of ($\mathbf{A_1}$) is needed when we consider lower bounds and minimax properties. Since f_ε^* must belong to $\mathbb{L}_2(\mathbb{R})$, we require that $\alpha > 1/2$ if $\rho = 0$. The errors are usually called "ordinary smooth" when $\rho = 0$, and "super smooth" when $\rho > 0$. Standard examples are the following: Gaussian or Cauchy distributions are super smooth of order ($\alpha = 0$, $\rho = 2$) and ($\alpha = 0$, $\rho = 1$) respectively. The double exponential distribution is ordinary smooth ($\rho = 0$) of order $\alpha = 2$.

By convention, we set $\beta = 0$ when $\rho = 0$ and assume that $\beta > 0$ when $\rho > 0$. If $\sigma = 0$, i.e. the X_i's are observed without noise, we set $\beta = \alpha = \rho = 0$.

The minimum penalized estimator of ℓ and g are defined as $\tilde{\ell} = \hat{\ell}_{\hat{m}_\ell}$ and $\tilde{g} = \hat{g}_{\hat{m}_g}$ where $\hat{m}_\ell$ and $\hat{m}_g$ are chosen in a purely data-driven way. The main point of the estimation procedure lies in the choice of $m = \hat{m}_h$ for the estimators $\hat{h}_m$ from Section 2.2 in order to mimic the oracle parameter

$$\breve{m}_h = \arg\min_m \mathbb{E} \parallel \hat{h}_m - h \parallel_2^2, \tag{2.4}$$

where h stands for ℓ or g.

More precisely, $\tilde{\ell}$ is defined by

$$\tilde{\ell} = \hat{\ell}_{\hat{m}_\ell} \text{ with } \hat{m}_\ell = \arg\min_{m \in \mathcal{M}_{n,\ell}} [\gamma_{n,\ell}(\hat{\ell}_m) + \mathrm{pen}_\ell(m)], \tag{2.5}$$

and $\tilde{g}$ is defined, as in Comte et al. (2006), by

$$\tilde{g} = \hat{g}_{\hat{m}_g} \text{ with } \hat{m}_g = \arg\min_{m \in \mathcal{M}_{n,g}} [\gamma_{n,g}(\hat{g}_m) + \mathrm{pen}_g(m)], \tag{2.6}$$

with $\mathcal{M}_{n,\ell} = \{1, \cdots, m_{n,\ell}\}$ and $\mathcal{M}_{n,g} = \{1, \cdots, m_{n,g}\}$, $m_{n,\ell}$ and $m_{n,g}$ being specified later. The penalties are data driven and given by

$$\mathrm{pen}_\ell(m) = \frac{\kappa'(\lambda_1 + \mu_2)[1 + \hat{m}_2(Y)]\tilde{\Gamma}(m)}{n}, \quad \mathrm{pen}_g(m) = \frac{\kappa(\lambda_1 + \mu_1)\tilde{\Gamma}(m)}{n}, \text{ with} \tag{2.7}$$

$$\hat{m}_2(Y) = \frac{1}{n}\sum_{i=1}^n Y_i^2 \text{ and } \tilde{\Gamma}(m) = (\pi m)^{2\alpha + \max(1-\rho, \min(\frac{(1+\rho)}{2}, 1))} \exp\{2\beta\sigma^\rho(\pi m)^\rho\}. \tag{2.8}$$

The constants λ_1, μ_1 and μ_2 only depend on f_ε and σ, which are known. They are defined below (see (3.5), (3.13) and (3.14)).

The quantities κ and κ' are universal constants. In practice, they are calibrated by intensive simulation studies. We refer to Comte et al. (2006, 2005b) for further details on penalty calibration and implementation of analogous estimators in density deconvolution.

Remark 2.1. The penalty functions defined by (2.7) and (2.8) have the same order. More precisely, both penalties are of order $m^{2\alpha+1-\rho}\exp(2\beta\sigma^\rho(\pi m)^\rho)/n$ if $0 \le \rho \le 1/3$, of order $m^{2\alpha+(1+\rho)/2}\exp(2\beta\sigma^\rho(\pi m)^\rho)/n$ if $1/3 \le \rho \le 1$ and of order $m^{2\alpha+1}\exp(2\beta\sigma^\rho(\pi m)^\rho)/n$ if $\rho \ge 1$.

2.4. Estimation of f itself

For $r \in \mathbb{R}$ and $d > 0$, we write $r^{(d)} = \mathrm{sign}(r)\min(|r|, d)$. The estimator $\tilde{f}$ of f is defined as

$$\tilde{f} = \left(\frac{\tilde{\ell}}{\tilde{g}}\right)^{(a_n)}, \tag{2.9}$$

with a_n suitably chosen. We have to use trimming to avoid problems due to small values of $\tilde{g}$.

3. Rates of convergence and adaptivity

3.1. Assumptions

We consider Model (1.1) under $(\mathbf{A_1})$ and the following additional assumptions.

$$\ell \in \mathbb{L}_2(\mathbb{R}) \text{ and } \ell \in \mathcal{L} = \Big\{\phi \text{ such that } \int x^2\phi^2(x)dx \leq \kappa_{\mathcal{L}} < \infty\Big\}, \qquad (\mathbf{A_2})$$

$$f \in \mathcal{F}_G = \{\phi \text{ such that } \sup_{x\in G} |\phi(x)| \leq \kappa_{\infty,G} < \infty\},$$
$$\text{where } G \text{ is the support of } g. \qquad (\mathbf{A_3})$$

$$g \in \mathbb{L}_2(\mathbb{R}) \text{ and } g \in \mathcal{G} = \{\phi, \text{ density such that } \int x^2\phi^2(x)dx < \kappa_{\mathcal{G}} < \infty\}. \quad (\mathbf{A_4})$$

There exist positive constants g_0, g_1 such that for all $x \in A$,
$$g_0 \leq g(x) \leq g_1. \qquad (\mathbf{A_5})$$

Assumption $(\mathbf{A_3})$ states that f is bounded on the support of g. If g is compactly supported, f has to be bounded on a compact set. Otherwise f has to be bounded on $\mathbb{R}$. We estimate f only on a compact set denoted by A. Hence, Assumption $(\mathbf{A_5})$ implies that $A \subset G$. Therefore, under $(\mathbf{A_3})$ and $(\mathbf{A_5})$, f is bounded on A. Assumptions $(\mathbf{A_3})$ and $(\mathbf{A_4})$ imply that $(\mathbf{A_2})$ holds with $\kappa_{\mathcal{L}} = \kappa_{\infty,G}^2\kappa_{\mathcal{G}}$.

Classically, the slowest rates of convergence for estimating f and g are obtained for super smooth error densities. In particular, when f_ε is Gaussian and f and g have the same Hölderian type regularity, the minimax rates of convergence are negative powers of $\ln(n)$ (see Fan (1991) and Fan and Truong (1993)). Nevertheless, we prove below that rates are improved if ℓ and g have stronger smoothness properties, described by the smoothness classes

$$\mathcal{S}_{a,r,B}(C_1) = \Big\{\psi \in \mathbb{L}_2(\mathbb{R}) : \text{such that} \int_{-\infty}^{+\infty} |\psi^*(x)|^2(x^2+1)^a \exp\{2B|x|^r\}dx \leq C_1\Big\}, \tag{3.1}$$

for a, r, B, C_1 nonnegative real numbers. Such smoothness classes have already been considered in density deconvolution (see Pensky and Vidakovic (1999) and Comte et al. (2006)). To our knowledge, it is the first time they have been considered in regression with errors-in-variables. When $r = 0$, (3.1) corresponds to a Sobolev ball. With $r > 0$, $B > 0$, functions in (3.1) are infinitely differentiable, they admit analytic continuation on a finite width strip when $r = 1$ and on the whole complex plane if $r = 2$.

3.2. Risk bounds for the minimum contrast estimators

We start by presenting general bound for the risks.

Proposition 3.1. *Consider the estimators* $\hat{\ell}_m$ *and* $\hat{g}_m$ *of* ℓ *and* g *defined by* (2.1) *and* (2.3). *Let* $\Delta(m) = \pi^{-1}\int_0^{\pi m} |f_\varepsilon^*(x\sigma)|^{-2}dx$. *Then, under* $(\mathbf{A_2})$ *and* $(\mathbf{A_4})$,

$$\mathbb{E}(\|\hat{\ell}_m - \ell\|_2^2) \leq \|\ell - \ell_m\|_2^2 + \frac{2\mathbb{E}(Y_1^2)\Delta(m)}{n} + \frac{(\kappa_{\mathcal{L}} + \| \ell \|_1^2)(\pi m)^2}{k_n}, \tag{3.2}$$

$$\mathbb{E}(\|\hat{g}_m - g\|_2^2) \leq \|g - g_m\|_2^2 + \frac{2\Delta(m)}{n} + \frac{(\kappa_{\mathcal{G}} + 1)(\pi m)^2}{k_n}. \tag{3.3}$$

The variance term $\Delta(m)/n$ depends on the rate of decay of the Fourier transform f_ε^*. Under $(\mathbf{A_1})$, it is bounded as follows

$$\Delta(m) \leq \lambda_1 \Gamma(m) \text{ with } \Gamma(m) = (\pi m)^{2\alpha+1-\rho}\exp(2\beta\sigma^\rho(\pi m)^\rho), \tag{3.4}$$

$$\lambda_1(\alpha, \kappa_0, \beta, \sigma, \rho) = \lambda_1 = \frac{(\sigma^2+1)^\alpha}{(\pi^\rho \kappa_0^2 R(\beta, \sigma, \rho))} \tag{3.5}$$

and $R(\beta,\sigma,\rho) = \mathbb{I}_{\rho=0} + 2\beta\rho\sigma^\rho \mathbb{I}_{0<\rho\leq 1} + 2\beta\sigma^\rho \mathbb{I}_{\rho>1}$.

To ensure that $\Gamma(m)/n$ is bounded, we only consider models such that $m \leq m_n$, with

$$\pi m_n \leq \begin{cases} n^{\frac{1}{(2\alpha+1)}} & \text{if } \rho = 0 \\ \Big[\dfrac{\ln(n)}{2\beta\sigma^\rho} + \dfrac{2\alpha+1-\rho}{2\rho\beta\sigma^\rho}\ln\Big(\dfrac{\ln(n)}{2\beta\sigma^\rho}\Big)\Big]^{\frac{1}{\rho}} & \text{if } \rho > 0. \end{cases} \tag{3.6}$$

Let us come to the bias terms which depend, as usual, on the smoothness properties of ℓ and g. Since ℓ_m is the orthogonal projection of ℓ on S_m, when ℓ belongs to $\mathcal{S}_{a_\ell,r_\ell,B_\ell}(\kappa_{a_\ell})$ (see (3.1)),

$$\|\ell - \ell_m\|_2^2 = (2\pi)^{-1}\int_{|x|\geq \pi m} |\ell^*|^2(x)dx \leq [\frac{\kappa_{a_\ell}}{(2\pi)}](m^2\pi^2+1)^{-a_\ell}\exp\{-2B_\ell \pi^{r_\ell} m^{r_\ell}\}. \tag{3.7}$$

The same holds for $\| g - g_m \|_2^2$ when g belongs to $\mathcal{S}_{a_g,r_g,B_g}(\kappa_{a_g})$ with (a_ℓ, B_ℓ, r_ℓ) replaced by (a_g, B_g, r_g).

Corollary 3.1. *Assume* $(\mathbf{A_1})$, $(\mathbf{A_2})$ *and* $(\mathbf{A_4})$, *and let* $\Gamma(m)$ *and* λ_1 *be defined by* (3.4) *and* (3.5). *Assume that* ℓ *belongs to* $\mathcal{S}_{a_\ell,r_\ell,B_\ell}(\kappa_{a_\ell})$ *and that* g *belongs to* $\mathcal{S}_{a_g,r_g,B_g}(\kappa_{a_g})$ *(see* (3.1)*). Then for* $k_n \geq n$,

$$\mathbb{E}(\|\ell - \hat{\ell}_m\|_2^2) \leq \kappa_{a_\ell}(2\pi)^{-1}(m^2\pi^2+1)^{-a_\ell}e^{-2B_\ell\pi^{r_\ell}m^{r_\ell}} + \frac{2\lambda_1\mathbb{E}(Y_1^2)\Gamma(m)}{n} + \frac{(\kappa_{\mathcal{L}}+\parallel \ell \parallel_1^2)(\pi m)^2}{n},$$

$$\mathbb{E}(\|g - \hat{g}_m\|_2^2) \leq \kappa_{a_g}(2\pi)^{-1}(m^2\pi^2+1)^{-a_g}e^{-2B_g\pi^{r_g}m^{r_g}} + \frac{2\lambda_1\Gamma(m)}{n} + \frac{(\kappa_{\mathcal{G}}+1)(\pi m)^2}{n}.$$

Remark 3.1. We point out that the $\{\varphi_{m,j}\}$ are $\mathbb{R}$-supported (and not compactly supported). Hence, we obtain estimations of ℓ and g on the whole line and not only on a compact set as is the case for the usual projection estimators. This is the advantage of the basis. A drawback is that we have to choose k_n, but is not difficult. Under $(\mathbf{A_2})$ and $(\mathbf{A_4})$, the choice $k_n \geq n$ ensures that terms involving k_n are negligible with respect to the variance terms. The choice of a large k_n will not change the accuracy of our estimator. From a practical point of view, it will reduce the speed of the algorithm.

Table 1 gives the rates for $\hat{\ell}_{\breve{m}_\ell}$. The results are also valid for $\hat{g}_{\breve{m}_g}$. The latter estimator has the minimax rate of convergence in all cases where lower bounds are known. See Fan (1991) for $r_g = 0$, Butucea (2004) for $r_g > 0$, $\rho = 0$, and Butucea and Tsybakov (2005) for $0 < r_g < \rho$, $a_g = 0$. We refer to Comte et al. (2006) for further details concerning density deconvolution.

Table 1. Best choices of $\breve{m}_\ell$ minimizing $\mathbb{E}(\|\ell - \hat{\ell}_m\|_2^2)$ and resulting rates for $\hat{\ell}_{\breve{m}_\ell}$.

		f_ε	
		$\rho = 0$ ordinary smooth	$\rho > 0$ super smooth
ℓ	$r_\ell = 0$ Sobolev(a_ℓ)	$\breve{m}_\ell = O(n^{1/(2\alpha+2a_\ell+1)})$ rate $= O(n^{-2a_\ell/(2\alpha+2a_\ell+1)})$	$\pi\breve{m}_\ell = [\ln(n)/(2\beta\sigma^\rho+1)]^{1/\rho}$ rate $= O((\ln(n))^{-2a_\ell/\rho})$
	$r_\ell > 0$ $\mathcal{C}^\infty$	$\pi\breve{m}_\ell = [\ln(n)/2B_\ell]^{1/r_\ell}$ rate $= O\Big(\frac{\ln(n)^{(2\alpha+1)/r_\ell}}{n}\Big)$	$\pi\breve{m}_\ell$ implicit solution of $(\pi\breve{m}_\ell)^{2\alpha+2a_\ell+1-r_\ell}e^{2\beta\sigma^\rho(\pi\breve{m}_\ell)^\rho+2B(\pi\breve{m}_\ell)^{r_\ell}} = O(n)$ rate : see comments below

When $r_\ell > 0$, $\rho > 0$, the optimal parameter $\breve{m}_\ell$ is not explicitly given. It is obtained as the solution of

$$(\pi\breve{m}_\ell)^{2\alpha+2a_\ell+1-r_\ell}\exp\{2\beta\sigma^\rho(\pi\breve{m}_\ell)^\rho+2B_\ell(\pi\breve{m}_\ell)^{r_\ell}\} = O(n). \tag{3.8}$$

Consequently, the rate of $\hat{\ell}_{\breve{m}_\ell}$ is not explicit and depends on the ratio r_ℓ/ρ. If r_ℓ/ρ or ρ/r_ℓ belongs to $]k/(k+1);(k+1)/(k+2)]$ with k an integer, the rate of convergence can be expressed as a function of k. For instance, if $r_\ell = \rho$, the rate is of order $[\ln(n)]^b n^{-B_\ell/(B_\ell+\beta\sigma^\rho)}$ with $b = [-2a_\ell\beta\sigma^\rho+(2\alpha-r_\ell+1)B_\ell]/[r_\ell(\beta\sigma^\rho+B_\ell)]$. It is of order $\ln(n)^{-2a_\ell/\rho}\exp[-2B_\ell(\ln(n)/(2\beta\sigma^\rho))^{r_\ell/\rho}]$ for $0 < r_\ell/\rho \le 1/2$, and of order $\ln(n)^{(2\alpha+1-\rho)/r_\ell}\exp[2\beta\sigma^\rho(\ln(n)/(2B_\ell))^{\rho/r_\ell}]/n$ for $0 < \rho/r_\ell \le 1/2$.

The case $\rho > 0$ is important since it contains Gaussian densities. When $\rho > 0$, $r_\ell = 0$ (Sobolev balls), rates are logarithmic. Now, as can be seen from the discussion above, faster rates can be obtained with ρ and r_ℓ are positive.

Proposition 3.2. *Assume* $(\mathbf{A_1})$–$(\mathbf{A_5})$ *and that g belongs to a space* $\mathcal{S}_{a_g,r_g,B_g}(\kappa_{a_g})$ *with $a_g > 1/2$ if $r_g = 0$ (see* (3.1)*). Let* $\hat{f}_{\breve{m}_\ell,\breve{m}_g} = \hat{\ell}_{\breve{m}_\ell}/\hat{g}_{\breve{m}_g}$*, with $\breve{m}_\ell$ and $\breve{m}_g$ that realize the best trade-off in Corollary* 3.1*. If $a_n = n^k$ with $k > 0$ and $k_n \ge n^{3/2}$, then for n large enough,*

$$\mathbb{E}\|(\hat{f}_{\breve{m}_\ell,\breve{m}_g} - f)\mathbb{I}_A\|_2^2 \le C_0[\mathbb{E}(\|\ell - \hat{\ell}_{\breve{m}_\ell}\|_2^2) + \mathbb{E}(\|g - \hat{g}_{\breve{m}_g}\|_2^2)] + o(n^{-1}), \qquad (3.9)$$

where $C_0 = Kg_0^{-2}(1 + g_1 g_0^{-2}\kappa_{\infty,G})$,

Let us make some comments. If $a_g \le 1/2$, we only have $\|(f - \hat{f}_{\breve{m}_\ell,\breve{m}_g})\mathbb{I}_A\|_2^2 = O_p(\|\ell - \hat{\ell}_{\breve{m}_\ell}\|_2^2 + \|g - \hat{g}_{\breve{m}_g}\|_2^2)$. If f is bounded on A, Proposition 3.2 still holds with $\kappa_{\infty,G}$ replaced by $\| f \|_{\infty,A}$.

The rate of $\hat{f}_{\breve{m}_\ell,\breve{m}_g}$ is given by the slowest term on the right-hand side of (3.9). Let us illustrate this result through examples.

- Suppose that the ε_i's are ordinary smooth.
  If $r_\ell = r_g = 0$, then $\breve{m}_\ell = O(n^{1/(2a_\ell+2\alpha+1)})$, $\breve{m}_g = O(n^{1/(2a_g+2\alpha+1)})$, and

$$\mathbb{E}(\|(f - \hat{f}_{\breve{m}_\ell,\breve{m}_g})\mathbb{I}_A\|_2^2) \le O(n^{-\frac{2a^*}{(2a^*+2\alpha+1)}}) \text{ with } a^* = \inf(a_\ell, a_g).$$

  If $r_\ell > 0$, $r_g > 0$, then $\pi\breve{m}_\ell = (\ln(n)/2B)^{1/r_\ell}$, $\pi\breve{m}_g = (\ln(n)/2B)^{1/r_g}$, and

$$\mathbb{E}(\|(f - \hat{f}_{\breve{m}_\ell,\breve{m}_g})\mathbb{I}_A\|_2^2) \le O\Big(\frac{\ln(n)^{\frac{(2\alpha+1)}{r^*}}}{n}\Big) \text{ with } r^* = \inf(r_\ell, r_g).$$

- Suppose that the ε_i's are super smooth.
  If $r_\ell = r_g = 0$, then $\pi\breve{m}_\ell = \pi\breve{m}_g = [\ln(n)/(2\beta\sigma^\rho + 1)]^{1/\rho}$, and

$$\mathbb{E}(\|(f - \hat{f}_{\breve{m}_\ell,\breve{m}_g})\mathbb{I}_A\|_2^2) \le O([\ln(n)]^{-\frac{2a^*}{\rho}}) \text{ with } a^* = \inf(a_\ell, a_g).$$

Since $\ell = fg$, the smoothness properties of ℓ are related to those of f and g. When ℓ belongs to $\mathcal{S}_{a_\ell,0,B_\ell}(\kappa_{a_\ell})$ and g belongs to $\mathcal{S}_{a_g,0,B_g}(\kappa_{a_g})$ with $a_\ell = a_g$, the resulting rate is the minimax rate given in Fan and Truong (1993) when both f

and g are Hölder of the same order. In that case, $\hat{f}_{\breve{m}_\ell,\breve{m}_g}$ is optimal in a minimax sense. When g is smoother than f, it is reasonable to believe that $\hat{f}_{\breve{m}_\ell,\breve{m}_g}$ is minimax. The case f smoother than g seems different. The optimality of $\hat{f}_{\breve{m}_\ell,\breve{m}_g}$ is not clear. In regression models without errors, when the X_i's are observed, there are estimation procedures that do not require any comparison of smoothness parameters of f and g (e.g., procedures based on local polynomial estimators). However, these methods do not seem to work on models with errors-in-variables.

The choices of $\breve{m}_\ell$ and $\breve{m}_g$ are optimal when they realize the best trade-off between the squared bias and the variance terms (see Corollary 3.1). These optimal values depend on the unknown smoothness parameters of ℓ and g. In the next section, we study penalized estimators which are constructed without smoothness knowledge. We provide upper bounds for their risks.

3.3. Risk bounds of the minimum penalized contrast estimators

In adaptive estimation of ℓ and g, one would expect to obtain bounds such as

$$\mathbb{E}\parallel \tilde{\ell}-\ell \parallel^2 \leq \inf_{m\in\mathcal{M}_{n,\ell}}\Big[\parallel \ell-\ell_m \parallel^2+\frac{2\lambda_1\mathbb{E}(Y_1^2)\Gamma(m)}{n}+\frac{(\pi m)^2(M_{\mathcal{L}}+\parallel \ell \parallel_1^2)}{n}\Big], \tag{3.10}$$

$$\mathbb{E}\parallel \tilde{g}-g \parallel^2 \leq \inf_{m\in\mathcal{M}_{n,g}}\Big[\parallel g-g_m \parallel^2+\frac{2\lambda_1\Gamma(m)}{n}+\frac{(\pi m)^2(M_{\mathcal{G}}+1)}{n}\Big]. \tag{3.11}$$

The following theorem describes the cases where the oracle inequalities (3.10) and (3.11) are reached.

Theorem 3.1. *Assume* ($\mathbf{A_1}$), ($\mathbf{A_2}$) *and* ($\mathbf{A_4}$). *Consider the collection of estimators $\hat{\ell}_m$ and $\hat{g}_m$ defined by* (2.1) *and* (2.3) *with $k_n > n$, $m_{n,\ell} \leq m_n$, $m_{n,g} \leq m_n$, and m_n satisfying*

$$\pi m_n \leq \Big[\frac{\ln(n)}{2\beta\sigma^\rho}+\frac{2\alpha+\min[(\frac{1}{2}+\frac{\rho}{2}),1]}{2\rho\beta\sigma^\rho}\ln\Big(\frac{\ln(n)}{2\beta\sigma^\rho}\Big)\Big]^{\frac{1}{\rho}}, \tag{3.12}$$

if $\rho > 1/3$, and (3.6) *if $\rho \leq 1/3$. Let*

$$\mu_1 = \begin{cases} 0 & \text{if } \rho < \frac{1}{3}, \\ \beta(\sigma\pi)^\rho\lambda_1^{\frac{1}{2}}(\alpha,\kappa_0,\beta,\sigma,\rho)(1+\sigma^2)^{\frac{\alpha}{2}}\kappa_0^{-1}(2\pi)^{-\frac{1}{2}} & \text{if } \frac{1}{3}\leq\rho\leq 1, \\ \beta(\sigma\pi)^\rho\lambda_1(\alpha,\kappa_0,\beta,\sigma,\rho) & \text{if } \rho > 1, \end{cases} \tag{3.13}$$

$$\mu_2 = \mu_1 1\!\!1_{\{0\leq\rho<\frac{1}{3}\}\cup\{\rho>1\}} + \mu_1\|f_\varepsilon\|_2 1\!\!1_{\{\frac{1}{3}\leq\rho\leq 1\}}, \tag{3.14}$$

and let pen_ℓ *and* pen_g *be given by* (2.7)−(2.8) *with* μ_1 *and* μ_2 *defined in* (3.13)−(3.14).

(1) Adaptive estimation of g. (Comte et al. (2006)).
For $\tilde{g} = \hat{g}_{\hat{m}_g}$ *defined in* (2.6),

$$\mathbb{E}(\|g - \tilde{g}\|_2^2) \leq K \inf_{m \in \mathcal{M}_{n,g}} \Big[\|g - g_m\|_2^2 + \frac{(\pi m)^2(\kappa_{\mathcal{G}} + 1)}{n} + \mathrm{pen}_g(m)\Big] + \frac{c}{n},$$

where K *is a constant and* c *another constant depending on* f_ε *and* A_g.

(2) Adaptive estimation of ℓ. *Under Assumption* $(\mathbf{A_3})$, *if* $\mathbb{E}|\xi_1|^8 < \infty$, $\tilde{\ell} = \hat{\ell}_{\hat{m}_\ell}$ *defined in* (2.5) *satisfies*

$$\mathbb{E}(\|\ell - \tilde{\ell}\|_2^2) \leq K' \inf_{m \in \mathcal{M}_{n,\ell}} \Big[\|\ell - \ell_m\|_2^2 + \frac{(\pi m)^2(\kappa_{\mathcal{L}} + \|\ \ell\ \|_1^2)}{n} + \mathbb{E}(\mathrm{pen}_\ell(m))\Big] + \frac{c'}{n},$$

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

Remarks. (1) According to Remark 2.1, the penalty functions $\mathrm{pen}_g(m)$ and $\mathrm{pen}_\ell(m)$ are of order $\Gamma(m)/n$ if $0 \leq \rho \leq 1/3$, and of order $m^{\min[(3\rho/2-1/2),\rho]}\Gamma(m)/n$, if $\rho > 1/3$. Hence, a loss of order $m^{\min[(3\rho/2-1/2),\rho]}$ may occur if $\rho > 1/3$.

(2) When $0 \leq \rho \leq 1/3$ or ($\rho > 1/3$, $r_g = 0$) or ($\rho > 1/3$, $r_g < \rho$), the rate of convergence of $\tilde{g}$ is that of $\hat{g}_{\breve{m}_g}$. It is minimax in all cases where lower bounds are known. When $r_g \geq \rho > 1/3$, there is a logarithmic loss due to adaptation (see Comte et al. (2006)).

(3) The rates of $\tilde{\ell}$ are easily deduced from Theorem 3.1. If $\mathrm{pen}_\ell(m)$ has the variance order $\Gamma(m)/n$, Theorem 3.1 guarantees an automatic trade-off between $\|\ell - \ell_m\|_2^2$ and the variance term up to some multiplicative constant.

In particular when $0 \leq \rho \leq 1/3$, the ε_i's are ordinary or super smooth. Whenever ℓ belongs to $\mathcal{S}_{a_\ell,r_\ell,B_\ell}(\kappa_{a_\ell})$ defined by (3.1), if we combine (3.7) and the order $\mathrm{pen}_\ell(m) = O(\Gamma(m)/n)$ (see (3.4)) we get that $\hat{m}_\ell$ mimics $\breve{m}_\ell$ (in (2.4)). The estimator $\tilde{\ell}$ automatically reaches the rate of $\hat{\ell}_{\breve{m}_\ell}$ given in Table 1.

If $\rho > 1/3$, $\mathrm{pen}_\ell(m)$ is slightly bigger than $\Gamma(m)/n$. If the bias $\|\ell - \ell_m\|_2^2$ is the main term, the rate of $\tilde{\ell}$ is still that of $\hat{\ell}_{\breve{m}_\ell}$. This holds in particular when ($r_\ell = 0$, $\rho > 0$) or $0 < r_\ell < \rho$.

In the case where $\mathrm{pen}_\ell(m)$ dominates $\|\ell - \ell_m\|_2^2$, i.e. $r_\ell \geq \rho > 1/3$, there is a loss of order at most $\ln n$: the rate of $\tilde{f}$ is equal to that of $\hat{f}_{\breve{m}_\ell}$ multiplied by $\ln n$. This has little importance since the main order term is faster than logarithmic (see comments on Table 1.).

Theorem 3.2. ***Adaptive estimation of*** f. *Under* $(\mathbf{A_1})$−$(\mathbf{A_5})$, *assume that* g *belongs to* $\mathcal{S}_{a_g,r_g,B_g}(\kappa_{a_g})$ *defined in* (3.1), *with* $a_g > 1/2$ *if* $r_g = 0$. *Assume moreover that* $\mathbb{E}|\xi_1|^8 < \infty$. *Let* $\tilde{f}$ *be defined by* (2.9), *where* $\tilde{\ell}$ *and* $\tilde{g}$ *are as in Theorem* 3.1 *and, in addition,* $m_{n,g} \leq (n/\ln(n))^{1/(2\alpha+2)}$ *if* $\rho = 0$.

If $k_n \geq n^{3/2}$, $a_n = n^k$ with $k > 0$, then for n large enough,

$$\mathbb{E}(\|(f-\tilde{f})\mathbb{1}_A\|_2^2) \leq C_0 \inf_{m\in\mathcal{M}_{n,\ell}} \Big[\|\ell-\ell_m\|_2^2 + \frac{(\pi m)^2(\kappa_{\mathcal{L}} + \parallel \ell \parallel_1^2)}{n} + \mathbb{E}(\text{pen}_\ell(m))\Big]$$
$$+C_1 \inf_{m\in\mathcal{M}_{n,g}} \Big[\|g-g_m\|_2^2 + \frac{(\pi m)^2(\kappa_{\mathcal{G}}+1)}{n} + \text{pen}_g(m)\Big] + \frac{c}{n}, \quad (3.15)$$

where $C_0 = 8Kg_0^{-2}$, $C_1 = 4K'g_0^{-2}(2g_1^2+1)\kappa_{\infty,G}^2$, K and K' are constants depending on f_ε, and c is a constant depending on f_ε, f and g.

As in Proposition 3.2, if $a_g \leq 1/2$ we only have $\|(f-\tilde{f})\mathbb{1}_A\|_2^2 = O_p(\|\ell - \tilde{\ell}\|_2^2 + \|g-\tilde{g}\|_2^2)$. If f is bounded on the compact set A, Theorem 3.2 holds with $\kappa_{\infty,G}$ replaced by $\parallel f \parallel_{\infty,A}$.

Let us make some additional comments. As for $\hat{f}_{\breve{m}_\ell,\breve{m}_g}$, the rate of $\tilde{f}$ is given by the largest MISE of $\tilde{\ell}$ and $\tilde{g}$. If ℓ and g belong to $\mathcal{S}_{a_\ell,r_\ell,B_\ell}(\kappa_{a_\ell})$ and $\mathcal{S}_{a_g,r_g,B_g}(\kappa_{a_g})$, respectively, with $0 \leq \rho \leq 1/3$ or $\rho > 1/3$ and $r_\ell, r_g \leq \rho$, $\tilde{f}$ achieves the rate of convergence of $\hat{f}_{\breve{m}_\ell,\breve{m}_g}$. We have already given some indications of the minimax properties of $\hat{f}_{\breve{m}_\ell,\breve{m}_g}$ (see comments after Theorem 3.2).

If ℓ and g belong to $\mathcal{S}_{a_\ell,r_\ell,B_f}(\kappa_{a_\ell})$ and $\mathcal{S}_{a_g,r_g,B_g}(\kappa_{a_g})$, respectively, with $r_g > \rho > 1/3$ or $r_\ell > \rho > 1/3$, there is a loss of order at most $\ln n$ (see comments after Theorem 3.1).

The rates for all estimators depend on the noise level σ. If $\sigma = 0$, $Z = X$ is observed. By convention $\beta = \alpha = \rho = 0$, hence $\lambda_1 = 1$. In that case, $\Gamma(m)/n$ has the order m/n, exactly as in density and nonparametric regression without errors. Analogously, pen_ℓ and pen_g have the order m/n, used for adaptive density estimation and nonparametric regression without errors (at least when there is, as here, one model per dimension).

For small σ, the procedure automatically selects a value of m close to the one that would be selected without errors in variables.

4. Proofs

4.1. Proof of Proposition 3.1

It follows from Definition (2.1) that, for any m belonging to $\mathcal{M}_n$, $\hat{\ell}_m$ satisfies $\gamma_{n,\ell}(\hat{\ell}_m) - \gamma_{n,\ell}(\ell_m^{(n)}) \leq 0$. Denote by $\nu_n(t)$ the centered empirical process,

$$\nu_n(t) = \frac{1}{n}\sum_{i=1}^n \Big(Y_i u_t^*(Z_i) - \langle t, \ell\rangle\Big). \quad (4.1)$$

Since $t \mapsto u_t^*$ is linear, we get the following decomposition

$$\gamma_{n,\ell}(t) - \gamma_{n,\ell}(s) = \|t-\ell\|_2^2 - \|s-\ell\|_2^2 - 2\nu_n(t-s). \quad (4.2)$$

Since $\|\ell - \ell_m^{(n)}\|_2^2 = \|\ell - \ell_m\|_2^2 + \|\ell_m - \ell_m^{(n)}\|_2^2$, we get $\|\ell - \hat{\ell}_m\|_2^2 \leq \|\ell - \ell_m\|_2^2 + \|\ell_m - \ell_m^{(n)}\|_2^2 + 2\nu_n(\hat{\ell}_m - \ell_m^{(n)})$. Using $\hat{a}_{m,j}(\ell) - a_{m,j}(\ell) = \nu_n(\varphi_{m,j})$, we obtain

$$\nu_n(\hat{\ell}_m - \ell_m^{(n)}) = \sum_{|j|\leq k_n} (\hat{a}_{m,j}(\ell) - a_{m,j}(\ell))\nu_n(\varphi_{m,j}) = \sum_{|j|\leq k_n} [\nu_n(\varphi_{m,j})]^2. \tag{4.3}$$

Consequently,

$$\mathbb{E}\|\ell - \hat{\ell}_m\|_2^2 \leq \|\ell - \ell_m\|_2^2 + \|\ell_m - \ell_m^{(n)}\|_2^2 + 2\sum_{j\in\mathbb{Z}} \mathrm{Var}[\nu_n(\varphi_{m,j})]. \tag{4.4}$$

Since the (Y_i, Z_i)'s are independent, $\mathrm{Var}[\nu_n(\varphi_{m,j})] = n^{-1}\mathrm{Var}[Y_1 u^*_{\varphi_{m,j}}(Z_1)]$. By Parseval's formula (see Comte et al. (2006)), we get that

$$\sum_{j\in\mathbb{Z}} \mathrm{Var}[\nu_n(\varphi_{m,j})] \leq n^{-1} \, \| \sum_{j\in\mathbb{Z}} |u^*_{\varphi_{m,j}}|^2 \, \|_\infty \, \mathbb{E}(Y_1^2) \leq \frac{\mathbb{E}(Y_1^2)\Delta(m)}{n}, \tag{4.5}$$

where $\Delta(m)$ is defined in Proposition 3.1.

Let us study the residual term $\|\ell_m - \ell_m^{(n)}\|_2^2$. We have

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

Now, by definition,

$$\begin{aligned} j a_{m,j}(\ell) &= j\sqrt{m}\int \varphi(mx - j)\ell(x)dx \\ &\leq m^{\frac{3}{2}} \int |x||\varphi(mx-j)||\ell(x)|dx + \sqrt{m}\int |mx - j||\varphi(mx-j)||\ell(x)|dx \\ &\leq m^{\frac{3}{2}} \Big(\int |\varphi(mx-j)|^2 dx\Big)^{\frac{1}{2}} \kappa_{\mathcal{L}}^{\frac{1}{2}} + \sqrt{m}\sup_x |x\varphi(x)|\|\ell\|_1. \end{aligned}$$

Thus $j a_{m,j} \leq m\|\varphi\|_2 \kappa_{\mathcal{L}}^{1/2} + \sqrt{m}\|\ell\|_1/\pi$ and $\|\ell_m - \ell_m^{(n)}\|_2^2 \leq (\kappa_{\mathcal{L}} + \|\ell\|_1^2)(\pi m)^2/k_n$.

4.2. Proof of Proposition 3.2

The proof of Proposition 3.2 is similar to the proof of Theorem 3.2 and is omitted. We refer to Comte and Taupin (2004) for further details.

4.3. Proof of Theorem 3.1

Point (1) is proved in Comte and Taupin (2004). We prove Point (2) with $\mathbb{E}(Y^2)$ in the penalty, this only requires $\mathbb{E}|\xi_1|^6 < \infty$. The complete proof with $\hat{m}_2(Y)$ in the penalty is obtained as an application of Rosenthal's inequality (see Rosenthal (1970)) and requires the stronger condition $\mathbb{E}|\xi_1|^8 < \infty$ (see Comte and Taupin (2004) for a complete proof).

For the study of $\tilde{\ell}$, the main difficulty compared with the study of $\tilde{g}$ lies in the fact that the ξ_i's are not necessarily bounded. By definition, $\tilde{\ell}$ satisfies, for all $m \in \mathcal{M}_{n,\ell}$, $\gamma_{n,\ell}(\tilde{\ell}) + \mathrm{pen}_\ell(\hat{m}_\ell) \le \gamma_{n,\ell}(\ell_m^{(n)}) + \mathrm{pen}_\ell(m)$. Therefore, (4.2) yields

$$\| \tilde{\ell} - \ell \|_2^2 \le \| \ell - \ell_m^{(n)} \|_2^2 + 2\nu_n(\tilde{\ell} - \ell_m^{(n)}) + \mathrm{pen}_\ell(m) - \mathrm{pen}_\ell(\hat{m}_\ell). \tag{4.6}$$

Next, we use that, if $t = t_1 + t_2$ with t_1 in $S_m^{(n)}$ and t_2 in $S_{m'}^{(n)}$, then t is such that t^* has its support included in $[-\pi\max(m,m'), \pi\max(m,m')]$. Therefore t belongs to $S_{\max(m,m')}^{(n)}$. Let $B_{m,m'}(0,1) = \{t \in S_{\max(m,m')}^{(n)} / \|t\|_2 = 1\}$. For $\nu_n(t)$ defined by (4.1), we get $|\nu_n(\tilde{\ell} - \ell_m^{(n)})| \le \|\tilde{\ell} - \ell_m^{(n)}\|_2 \sup_{t\in B_{m,\hat{m}}(0,1)} |\nu_n(t)|$. Consequently, using that $2ab \le x^{-1}a^2 + xb^2$, we have

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

Therefore, we can write $\|\tilde{\ell} - \ell_m^{(n)}\|_2^2 \le (1+y^{-1})\|\tilde{\ell} - \ell\|_2^2 + (1+y)\|\ell - \ell_m^{(n)}\|_2^2$, with $y = (x+1)/(x-1)$ for $x > 1$. Thus,

$$\begin{aligned}\|\tilde{\ell} - \ell\|_2^2 \le {}& \Big(\frac{x+1}{x-1}\Big)^2 \|\ell - \ell_m^{(n)}\|_2^2 + \frac{x(x+1)}{x-1} \sup_{t\in B_{m,\hat{m}_\ell}(0,1)} \nu_n^2(t) \\ &+ \frac{x+1}{x-1}\Big(\mathrm{pen}_\ell(m) - \mathrm{pen}_\ell(\hat{m}_\ell)\Big).\end{aligned}$$

Choose a positive function $p_\ell(m,m')$ such that $xp_\ell(m,m') \le \mathrm{pen}_\ell(m) + \mathrm{pen}_\ell(m')$. Setting $\kappa_x = (x+1)/(x-1)$, we obtain

$$\begin{aligned}\|\tilde{\ell} - \ell\|_2^2 \le {}& \kappa_x^2\|\ell - \ell_m^{(n)}\|_2^2 + x\kappa_x[\sup_{t\in B_{m,\hat{m}}(0,1)} |\nu_n|^2(t) - p(m,\hat{m}_\ell)]_+ \\ &+ \kappa_x(xp_\ell(m,\hat{m}_\ell) + \mathrm{pen}_\ell(m) - \mathrm{pen}_\ell(\hat{m})).\end{aligned} \tag{4.7}$$

If we set

$$W_n(m') = \Big[\sup_{t\in B_{m,m'}(0,1)} |\nu_n(t)|^2 - p_\ell(m,m')\Big]_+, \tag{4.8}$$

(4.7) can be written as

$$\|\tilde{\ell} - \ell\|_2^2 \le \kappa_x^2\|\ell - \ell_m^{(n)}\|_2^2 + 2\kappa_x \mathrm{pen}_\ell(m) + x\kappa_x W_n(\hat{m}_\ell). \tag{4.9}$$

The key point of the proof lies in the study of $W_n(m')$. More precisely, we search for $p_\ell(m,m')$ such that, for C a constant,

$$\mathbb{E}(W_n(\hat{m}_\ell)) \le \sum_{m'\in\mathcal{M}_{n,\ell}} \mathbb{E}(W_n(m'))) \le \frac{C}{n}. \tag{4.10}$$

Now, combining (4.9) and (4.10) we have, for all m in $\mathcal{M}_{n,\ell}$,

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

This is also

$$\mathbb{E}\|\ell-\tilde{\ell}\|_2^2 \le C_x \inf_{m\in\mathcal{M}_{n,\ell}}\Big[\|\ell-\ell_m\|_2^2+\frac{(\kappa_{\mathcal{L}}+\parallel \ell \parallel_1^2)(\pi m)^2}{k_n}+\mathrm{pen}_\ell(m)\Big]+\frac{C_xC'}{n}, \quad (4.11)$$

where $C_x = \max(\kappa_x^2, 2\kappa_x)$ suits. Hence Theorem 3.1 will hold if (4.10) is proved.

It remains thus to find $p_\ell(m,m')$ such that (4.10) holds.

The process $W_n(m')$ is studied using the decomposition of $\nu_n(t) = \nu_{n,1}(t) + \nu_{n,2}(t)$ with

$$\nu_{n,1}(t) = \frac{1}{n}\sum_{i=1}^n (f(X_i)u_t^*(Z_i) - \langle t,\ell\rangle) \text{ and } \nu_{n,2}(t) = \frac{1}{n}\sum_{i=1}^n \xi_i u_t^*(Z_i). \quad (4.12)$$

So $W_n(m') \le 2W_{n,1}(m') + 2W_{n,2}(m')$ where for $i=1,2$,

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

$$p_\ell(m,m') = 2p_1(m,m') + 2p_2(m,m'). \quad (4.13)$$

• **Study of $\mathbf{W_{n,1}}(m')$.**

Under $(\mathbf{A_3})$, f is bounded on the support of g. So we apply a standard Talagrand (1996) inequality recalled in Lemma 4.1 below.

Lemma 4.1. *Let $U_1,\dots,U_n$ 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 $\epsilon>0$,*

$$\mathbb{E}\Big[\sup_{r\in\mathcal{R}}|\nu_n(r)|^2 - 2(1+2\epsilon)H^2\Big]_+ \le \frac{6}{K_1}\Big(\frac{v}{n}e^{-K_1\epsilon\frac{nH^2}{v}} + \frac{8M_1^2}{K_1 n^2 C^2(\epsilon)}e^{-\frac{K_1C(\epsilon)\sqrt{\epsilon}}{\sqrt{2}}\frac{nH}{M_1}}\Big), \quad (4.14)$$

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

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

Inequality (4.14) is a straightforward consequence of Talagrand (1996) inequality given in Ledoux (1996) (or Birgé and Massart (1998)). The application of (4.14) to $\nu_{n,1}(t)$ gives

$$\mathbb{E}\Big[\sup_{t\in B_{m,m'}(0,1)}|\nu_{n,1}(t)|^2 - 2(1+2\epsilon_1)\mathbb{H}_1^2\Big]_+ \le \kappa_1\Big(\frac{v_1}{n}e^{-K_1\epsilon_1\frac{n\mathbb{H}_1^2}{v_1}} + \frac{M_1^2}{n^2}e^{-K_2\sqrt{\epsilon_1}C(\epsilon_1)\frac{n\mathbb{H}_1}{M_1}}\Big), \quad (4.15)$$

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

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

According to (3.4) and (4.5), we can take

$$M_1 = M_1(m,m') = \kappa_{\infty,G}\sqrt{\lambda_1\Gamma(m^*)}, \text{ where } m^* = \max(m,m'). \tag{4.16}$$

To compute v_1, we set $P_{j,k}(m) = \mathbb{E}[f^2(X_1)u^*_{\varphi_{m,j}}(Z_1)u^*_{\varphi_{m,k}}(-Z_1)]$, and write

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

Following Comte et al. (2006), take

$$\Delta_2(m,\Psi) = m^2\iint \Big|\frac{\varphi^*(x)\varphi^*(y)}{f_\varepsilon^*(mx)f_\varepsilon^*(my)}\Psi^*(m(x-y))\Big|^2 dxdy \le \lambda_2^2(\|\Psi\|_2)\Gamma_2^2(m), \tag{4.17}$$

with

$$\Gamma_2(m) = (\pi m)^{2\alpha+\min[(\frac{1}{2}-\frac{\rho}{2}),(1-\rho)]}\exp\{2\beta\sigma^\rho(\pi m)^\rho\}, \tag{4.18}$$

$$\lambda_2(\|\Psi\|_2) = \begin{cases} \lambda_1(\alpha,\kappa_0,\beta,\sigma,\rho) & \text{if } \rho > 1,\\ \kappa_0^{-1}(2\pi)^{-\frac{1}{2}}\lambda_1^{\frac{1}{2}}(\alpha,\kappa_0,\beta,\sigma,\rho)(1+\sigma^2)^{\frac{\alpha}{2}}\|\Psi\|_2 & \text{if } \rho \le 1.\end{cases} \tag{4.19}$$

Now, write $P_{j,k}(m) = \iint f^2(x)u^*_{\varphi_{m,j}}(x+y)u^*_{\varphi_{m,k}}(-(x+y))g(x)f_\varepsilon(y)dxdy$ so that

$$\begin{aligned} P_{j,k}(m) &= m\iint f^2(x)\iint e^{-i(x+y)um}\frac{\varphi^*(u)e^{iju}}{f_\varepsilon^*(mu)}e^{i(x+y)vm}\frac{\varphi^*(v)e^{ikv}}{f_\varepsilon^*(mv)}dudvg(x)f_\varepsilon(y)dxdy\\ &= m\iint \frac{e^{iju+ikv}\varphi^*(u)\varphi^*(v)}{f_\varepsilon^*(mu)f_\varepsilon^*(mv)}\Big(\iint e^{-i(x+y)(u-v)m}f^2(x)g(x)f_\varepsilon(y)dxdy\Big)dudv\\ &= m\iint \frac{e^{iju+ikv}\varphi^*(u)\varphi^*(v)}{f_\varepsilon^*(mu)f_\varepsilon^*(mv)}[(f^2g)*f_\varepsilon]^*((u-v)m)dudv. \end{aligned}$$

By Parseval's formula, we get that $\sum_{j,k}|P_{j,k}(m)|^2$ equals

$$m^2\iint \Big|\frac{\varphi^*(u)\varphi^*(v)}{f_\varepsilon^*(mu)f_\varepsilon^*(mv)}[(f^2g)*f_\varepsilon]^*((u-v)m)\Big|^2 dudv = \Delta_2(m,(f^2g)*f_\varepsilon).$$

Since $\|(f^2g)*f_\varepsilon\|_2 \le \|f^2g\|_2\|f_\varepsilon\|_2 = \mathbb{E}^{1/2}(f^2(X_1))\|f_\varepsilon\|_2$, and $\lambda_2(\|f^2g\|_2\|f_\varepsilon\|_2) \le \mu_2$, with μ_2 defined in (3.13), we can take

$$v_1 = v_1(m,m') = \mu_2\Gamma_2(m^*). \tag{4.20}$$

Lastly, we have $\mathbb{E}[\sup_{t\in B_{m,m'}(0,1)} |\nu_{n,1}(t)|^2] \le \mathbb{E}(f^2(X_1))\lambda_1\Gamma(m^*)/n$. Thus,

$$\mathbb{H}_1^2 = \mathbb{H}_1^2(m,m') = \frac{\mathbb{E}(f^2(X_1))\lambda_1\Gamma(m^*)}{n}. \tag{4.21}$$

From (4.15), (4.16), (4.20) and (4.21) we choose $p_1(m,m') = 2(1+2\epsilon_1)\mathbb{H}_1^2 = 2(1+2\epsilon_1)\mathbb{E}(f^2(X_1))\lambda_1\Gamma(m^*)/n$. Then,

$$\begin{aligned}\mathbb{E}(W_{n,1}(m')) &\le \mathbb{E}\Big[\sup_{t\in B_{m,m'}(0,1)} |\nu_{n,1}(t)|^2 - 2(1+2\epsilon_1)\mathbb{H}_1^2\Big]_+ \\ &\le A_1(m^*) + B_1(m^*), \end{aligned}\tag{4.22}$$

with

$$A_1(m) = K_3\frac{\mu_2\Gamma_2(m)}{n}\exp\Big(-K_1\epsilon_1\mathbb{E}(f^2(X_1))\frac{\lambda_1\Gamma(m)}{\mu_2\Gamma_2(m)}\Big), \tag{4.23}$$

$$B_1(m) = K_3\frac{\kappa_{\infty,G}^2\lambda_1\Gamma(m)}{n^2}\exp\Big\{-K_2\sqrt{\epsilon_1}C(\epsilon_1)\frac{\sqrt{\mathbb{E}(f^2(X_1))}}{\kappa_{\infty,G}}\sqrt{n}\Big\}. \tag{4.24}$$

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

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

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

(1) Case $0\le\rho<1/3$. Here $\rho\le(1/2-\rho/2)_+$ and the choice $\epsilon_1=1/2$ ensures the convergence of $\sum_{m'\in\mathcal{M}_{n,\ell}} A_1(m^*)$. Indeed, set $\psi = 2\alpha+\min[(1/2-\rho/2),(1-\rho)]$, $\omega=(1/2-\rho/2)_+$, and $K'=\kappa_2\lambda_1/\mu_2$. For $a,b\ge1$,

$$\begin{aligned}&\max(a,b)^{\psi}e^{2\beta\sigma^\rho\pi^\rho\max(a,b)^\rho}e^{-K'\xi^2\max(a,b)^\omega}\\ &\quad\le (a^\psi e^{2\beta\sigma^\rho\pi^\rho a^\rho}+b^\psi e^{2\beta\sigma^\rho\pi^\rho b^\rho})e^{-(\frac{K'\xi^2}{2})(a^\omega+b^\omega)}\end{aligned}$$

is bounded by

$$a^\psi e^{2\beta\sigma^\rho\pi^\rho a^\rho}e^{-(\frac{K'\xi^2}{2})a^\omega}e^{-(\frac{K'\xi^2}{2})b^\omega}+b^\psi e^{2\beta\sigma^\rho\pi^\rho b^\rho}e^{-(\frac{K'\xi^2}{2})b^\omega}. \tag{4.25}$$

Since the function $a\mapsto a^\psi e^{2\beta\sigma^\rho\pi^\rho a^\rho}e^{-(K'\xi^2/2)a^\omega}$ is bounded on $\mathbb{R}^+$ by a constant only depending on α, ρ and K', and since $Ak^\rho-\beta k^\omega\le-(\beta/2)k^\omega$ for any $k\ge1$, it follows that $\sum_{m'\in\mathcal{M}_{n,\ell}} A_1(m^*)\le C/n$.

(2) Case $\rho=1/3$. Here $\rho=(1/2-\rho/2)_+$, and $\omega=\rho$. We choose $\epsilon_1=\epsilon_1(m,m')$ such that $2\beta\sigma^\rho\pi^\rho m^{*\rho}-K'\mathbb{E}(f^2(X_1))\epsilon_1 m^{*\rho}=-2\beta\sigma^\rho\pi^\rho m^{*\rho}$. Since $K'=K_1\lambda_1/\mu_2$, $\epsilon_1=\epsilon_1(m,m')=(4\beta\sigma^\rho\pi^\rho\mu_2)/(K_1\lambda_1\mathbb{E}(f^2(X_1)))$.

(3) Case $\rho > 1/3$. Here $\rho > (1/2 - \rho/2)_+$. According to (4.25), we choose $\epsilon_1 = \epsilon_1(m, m')$ such that $2\beta\sigma^\rho\pi^\rho m^{*\rho} - K'\mathbb{E}(f^2(X_1))\epsilon_1 m^{*\omega} = -2\beta\sigma^\rho\pi^\rho m^{*\rho}$. Since $K' = K_1\lambda_1/\mu_2$, $\epsilon_1 = \epsilon_1(m, m') = (4\beta\sigma^\rho\pi^\rho\mu_2)/(K_1\lambda_1\mathbb{E}(f^2(X_1)))m^{*\rho-\omega}$.

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

• **Study of $\mathbf{W_{n,2}}(m')$.**
Let

$$\mathbb{H}_\xi^2(m, m') = \frac{\left(n^{-1}\sum_{i=1}^n \xi_i^2\right)\lambda_1\Gamma(m^*)}{n}, \tag{4.26}$$

with $(n^{-1}\sum_{i=1}^n \xi_i^2)\lambda_1\Gamma(m)/n = (n^{-1}\sum_{i=1}^n \xi_i^2 - \sigma_\xi^2)\lambda_1\Gamma(m)/n + \sigma_\xi^2\lambda_1\Gamma(m)/n$. The latter term is less than

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

Consequently $\mathbb{H}_\xi^2(m, m') \le \mathbb{H}_{\xi,1}(m, m') + \mathbb{H}_{\xi,2}(m, m')$ where

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

and $\mathbb{H}_{\xi,2}(m, m') = 3\sigma_\xi^2\lambda_1\Gamma(m^*)/(2n)$. By (4.12), $\mathbb{E}[\sup_{t\in B_{m,m'}(0,1)}|\nu_{n,2}(t)|^2 - p_2(m, m')]_+$ is bounded by

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

Hence, we obtain

$$\begin{aligned}
&\mathbb{E}[\sup_{t\in B_{m,m'}(0,1)}|\nu_{n,2}(t)|^2 - p_2(m, m')]_+ \\
&\quad \le 2\mathbb{E}\Big[\sup_{t\in B_{m,m'}(0,1)}\Big(n^{-1}\sum_{i=1}^n \xi_i(u_t^*(Z_i) - \langle t, g\rangle)\Big)^2 - 2(1+2\epsilon_2)\mathbb{H}_\xi^2(m, m')\Big]_+ \\
&\qquad +\frac{2\|g\|_2^2\sigma_\xi^2}{n} + 4(1+2\epsilon_2)\mathbb{E}|\mathbb{H}_{\xi,1}(m, m')| \\
&\qquad +\mathbb{E}[4(1+2\epsilon_2)\mathbb{H}_{\xi,2}(m, m') - p_2(m, m')]_+.
\end{aligned} \tag{4.27}$$

Since we only consider m such that $\Gamma(m)/n$ is bounded by some constant κ, we get that for some $p \geq 2$, $\mathbb{E}|\mathbb{H}_{\xi,1}(m,m')|$ is bounded by

$$\kappa\lambda_1 \mathbb{E}\Big[|\frac{1}{n}\sum_{i=1}^n \xi_i^2 - \sigma_\xi^2| \mathbb{I}_{\{n^{-1}|\sum_{i=1}^n(\xi_i^2-\sigma_\xi^2)|\geq \sigma_\xi^2/2\}}\Big] \leq \frac{\kappa\lambda_1 2^{p-1}\mathbb{E}\Big[|n^{-1}\sum_{i=1}^n \xi_i^2 - \sigma_\xi^2|^p\Big]}{\sigma_\xi^{2(p-1)}}.$$

According to Rosenthal's inequality (see Rosenthal (1970)), we find that, for $\sigma_{\xi,p}^p := \mathbb{E}(|\xi|^p)$, $\sigma_{\xi,2}^2 = \sigma_\xi^2$,

$$\mathbb{E}\Big|n^{-1}\sum_{i=1}^n \xi_i^2 - \sigma_\xi^2\Big|^p \leq C'(p)\Big(\sigma_{\xi,2p}^{2p} n^{1-p} + \sigma_{\xi,4}^{2p} n^{-\frac{p}{2}}\Big).$$

Note that, $\alpha > 1/2$ since f_ε^* in $\mathbb{L}_2(\mathbb{R})$. Therefore $|\mathcal{M}_{n,\ell}| \leq \sqrt{n}$. The choice $p = 3$ leads to $\sum_{m'\in\mathcal{M}_{n,\ell}} \mathbb{E}|\mathbb{H}_{\xi,1}(m,m')| \leq C(\sigma_{\xi,6},\sigma_\xi)/n$. The last term in (4.27) vanishes as soon as

$$p_2(m,m') = 4(1+2\epsilon_2)\mathbb{H}_{\xi,2}(m,m') = \frac{6(1+2\epsilon_2)\lambda_1\sigma_\xi^2\Gamma(m^*)}{n}. \tag{4.28}$$

With this choice for $p_2(m,m')$, (4.27) leads to

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{t\in B_{m,\hat{m}_\ell}(0,1)} |\nu_{n,2}(t)|^2 - p_2(m,\hat{m}_\ell)\Big]_+ \\
&\quad\leq 2\sum_{m'\in\mathcal{M}_{n,\ell}} \mathbb{E}\Big[\sup_{t\in B_{m,m'}(0,1)} \Big(n^{-1}\sum_{i=1}^n \xi_i(u_t^*(Z_i) - \langle t,g\rangle)\Big)^2 \\
&\qquad -2(1+2\epsilon_2)\mathbb{H}_\xi^2(m,m')\Big]_+ + \frac{2\|g\|_2^2\sigma_\xi^2}{n} + \frac{4C(1+2\epsilon_2)}{n}.
\end{aligned} \tag{4.29}$$

To deal with the right-hand side of (4.29), we apply the following lemma.

Lemma 4.2. *Under the assumptions of Theorem* 3.1, *if* $\mathbb{E}|\xi_1|^6 < \infty$, *then for some given* $\epsilon_2 > 0$*:*

$$\begin{aligned}
&\sum_{m'\in\mathcal{M}_{n,\ell}} \mathbb{E}\Big[\sup_{t\in B_{m,m'}(0,1)} \Big(\frac{1}{n}\sum_{i=1}^n \xi_i(u_t^*(Z_i) - \langle t,g\rangle)\Big)^2 - 2(1+2\epsilon_2)\mathbb{H}_\xi^2(m,m')\Big]_+ \\
&\quad\leq K_1\Big\{\sum_{m'\in\mathcal{M}_{n,\ell}} \Big[\frac{\sigma_\xi^2\mu_2\Gamma_2(m^*)}{n}\exp\Big(-K_1\epsilon_2\frac{\lambda_1\Gamma(m^*)}{\mu_2\Gamma_2(m^*)}\Big)\Big] + \Big(1+\frac{\ln^4(n)}{\sqrt{n}}\Big)\frac{1}{n}\Big\},
\end{aligned} \tag{4.30}$$

where μ_2 *and* $\Gamma_2(m)$ *are defined by* (3.13) *and* (4.18), *and* K_1 *is a constant depending on the moments of* ξ. *The constant* μ_2 *can be replaced by* $\lambda_2(\|f_Z\|_2)$ *where* λ_2 *is defined by* (4.19) *and* f_Z *denotes the density of* Z.

In analogy with (4.23), set

$$A_2(m^*) = \frac{K_1\sigma_\xi^2\mu_2\Gamma_2(m^*)}{n}\exp\Big(-\kappa_2\epsilon_2\frac{\lambda_1}{\mu_2}m^{*(\frac{1}{2}-\frac{\rho}{2})_+}\Big). \tag{4.31}$$

With (4.28), (4.15) and (4.30), we find, for $W_{n,2}$ (see (4.13)),

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

Now $\sum_{m'\in\mathcal{M}_{n,\ell}}A_2(m^*)$ is bounded as is the analogous sum $\sum_{m'\in\mathcal{M}_{n,\ell}}A_1(m^*)$ with $\epsilon_2=\epsilon_1=1/2$ if $0\le\rho<1/3$, and $\epsilon_1(m,m')$ replaced by $\epsilon_2=\epsilon_2(m,m')=\mathbb{E}(f^2(X_1))\epsilon_1(m,m')$, when $\rho\ge 1/3$, that is $\epsilon_2(m,m')=(4\beta\sigma^\rho\pi^\rho\mu_2)/(K_1\lambda_1)m^{*\rho-\omega}$. These choices ensure that $\sum_{m'\in\mathcal{M}_{n,\ell}}A_2(m^*)$ is less than C/n. As announced in (4.13), we take

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

More precisely, if $0\le\rho<1/3$,

$$p_\ell(m,m') = \frac{24\mathbb{E}(Y_1^2)\lambda_1\Gamma(m^*)}{n}. \tag{4.32}$$

If $\rho\ge 1/3$,

$$p_\ell(m,m') = 4\Big[3\mathbb{E}(Y_1^2)+\frac{32\beta\sigma^\rho\pi^\rho\mu_2 m^{*\rho-\omega}}{k_1\lambda_1}\Big]\frac{\lambda_1\Gamma(m^*)}{n}. \tag{4.33}$$

Consequently if $0\le\rho<1/3$, we take $\mathrm{pen}_\ell(m)=\kappa\mathbb{E}(Y_1^2)\lambda_1\Gamma(m)/n$. If $\rho\ge1/3$ we take $\mathrm{pen}_\ell(m)=\kappa[\mathbb{E}(Y_1^2)+\beta\sigma^\rho\pi^\rho\mu_2 m^{\rho-\omega}/k_1\lambda_1]\lambda_1\Gamma(m)/n$, where κ is a universal constant. Note that for $\rho=1/3$, $\rho-\omega=0$ and both penalties have same order.

4.4. Proof of Lemma 4.2

We work conditionally on $\sigma(\xi_i,\ i=1,\cdots,n)$, and we denote by $\mathbb{E}_\xi$ and $\mathbb{P}_\xi$ the conditional expectation and probability given $\xi_1,\ldots,\xi_n$.

Conditioning on $\sigma(\xi_i,\ i=1,\cdots,n)$, we apply Lemma 4.1 to the random variables $f_t(\xi_i,Z_i)=\xi_i u_t^*(Z_i)$, which are independent but non-identically distributed. Let $Q_{j,k}=\mathbb{E}[u^*_{\varphi_{m,j}}(Z_1)u^*_{\varphi_{m,k}}(-Z_1)]$. Straightforward calculations give that, for $\mathbb{H}_\xi(m,m')$ defined in (4.26), we have

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

As in Comte et al. (2006), we write that $\sum_{j,k} |Q_{j,k}|^2 \leq \Delta_2(m, f_Z) \leq \lambda_2(\|f_Z\|_2)$ $\Gamma_2(m, \|f_\varepsilon\|_2)$ with $\|f_Z\|_2 \leq \|f_\varepsilon\|_2$, where $\Delta_2(m, f_Z)$ is defined by (4.17), λ_2 by (4.19), $\Gamma_2(m)$ by (4.18), μ_2 by (3.13). We now write that

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

and thus we take $v_\xi(m, m') = (n^{-1}\sum_{i=1}^n \xi_i^2)\mu_2\Gamma_2(m^*, \|f_\varepsilon\|_2)$. Lastly, since

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

we take $M_{1,\xi}(m, m') = 2\max_{1\leq i\leq n}|\xi_i|\sqrt{\lambda_1\Gamma(m^*)}$. Applying Lemma 4.1, we get for some constants κ_1, κ_2, κ_3,

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

To conclude we integrate the above expression with respect to the law of the ξ_i's. The first term $\sigma_\xi^2\mu_2\Gamma_2(m^*)\exp[-\kappa_2\epsilon\lambda_1\Gamma(m^*)/(\mu_2\Gamma_2(m^*)])/n$ has the same order as in the study of $W_{n,1}$. The second term is bounded by

$$\frac{\lambda_1\Gamma(m^*)}{n^2}\mathbb{E}\Big[(\max|\xi_i|^2)\exp\Big(-\kappa_3\sqrt{\epsilon}C(\epsilon)\frac{\sqrt{\sum_{i=1}^n \xi_i^2}}{\max_{1\leq i\leq n}|\xi_i|}\Big)\Big]. \tag{4.34}$$

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

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

We need to prove that this is less than c/n for some constant c. We bound $\max_i|\xi_i|$ by b on the set $\{\max_i|\xi_i| \leq b\}$ and the exponential by 1 on the set $\{\max_i|\xi_i| \geq b\}$. Setting $\mu_\epsilon = \kappa_3\sqrt{\epsilon}C(\epsilon)$, we get

$$\mathbb{E}\Big[\max_{1\leq i\leq n}\xi_i^2\exp\Big(-\mu_\epsilon\sqrt{\frac{\sum_{i=1}^n \xi_i^2}{\max_{1\leq i\leq n}\xi_i^2}}\Big)\Big]$$

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

Again by Rosenthal's inequality, we obtain

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

This is bounded by

$$
b^2e^{-\frac{\mu_\epsilon\sqrt{n}\sigma_\xi}{(\sqrt{2}b)}}+C'(p)b^2\sigma_{\xi,2p}^{2p}2^p\sigma_\xi^{-2p}[n^{1-p}+n^{-\frac{p}{2}}]+n\sigma_{\xi,r+2}^{r+2}b^{-r}.
$$

Since $\mathbb{E}|\xi_1|^6<\infty$, we take $p=3$, $r=4$, $b=\sigma_\xi\sqrt{\epsilon}C(\epsilon)\kappa_3\sqrt{n}/[2\sqrt{2}(\ln(n)-\ln\ln n)]$. For any $n\ge 3$, and for C_1 and C_2 constants depending on the moments of ξ, we find that

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

We sum (4.34) over $\mathcal{M}_{n,\ell}$ with cardinality less than $\sqrt{n}$. The result is bounded by $C(1+\ln(n)^4/\sqrt{n})/n$ for some constant C, since $\Gamma(m^*)/n$ is bounded.

4.5. Proof of Theorem 3.2

Consider the event $\tilde{E}_n=\{\|g-\tilde{g}\|_{\infty,A}\le g_0/2\}$. Since $g(x)\ge g_0$ for all x in A, $\tilde{g}(x)\ge g_0/2$ for all x in A on $\tilde{E}_n$ as well. It follows that

$$
\mathbb{E}\|(f-\tilde{f})\mathbb{I}_A\mathbb{I}_{\tilde{E}_n}\|_2^2\le 8g_0^{-2}\mathbb{E}\|\tilde{\ell}-\ell\|_2^2+8\|\ell\|_{\infty,A}^2g_0^{-4}\mathbb{E}\|\tilde{g}-g\|_2^2, \tag{4.35}
$$

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

$$
\mathbb{E}[\|(f-\tilde{f})\mathbb{I}_A\mathbb{I}_{\tilde{E}_n^c}\|_2^2]\le 2(a_n^2+\|f\|_{\infty,A}^2)\lambda(A)\mathbb{P}(\tilde{E}_n^c), \tag{4.36}
$$

where $\lambda(A) = \int_A dx$. So, for $\hat{m}_\ell = \hat{m}_\ell(n)$ and $\hat{m}_g = \hat{m}_g(n)$, if $a_n \mathbb{P}(\tilde{E}_n^c) = o(n^{-1})$, (3.15) is proved by applying Theorem 3.1. We now come to the study of $\mathbb{P}(\tilde{E}_n^c)$. We write $\mathbb{P}(\tilde{E}_n^c) = \mathbb{P}(\|g - \tilde{g}\|_{\infty,A} > g_0/2) = \mathbb{P}(\|g - g_{\hat{m}_g}^{(n)} + g_{\hat{m}_g}^{(n)} - \tilde{g}\|_{\infty,A} > g_0/2)$ and use the following lemma.

Lemma 4.3. *Let g belongs to $\mathcal{S}_{a_g,\nu_g,B_g}(\kappa_{a_g})$, defined by* (3.1) *with $a_g > 1/2$. Then for $t \in S_m$, $\|t\|_\infty \leq \sqrt{m}\|t\|_2$ and $\|g - g_m\|_\infty \leq (2\pi)^{-1}\sqrt{\pi m}((\pi m)^2 + 1)^{-a_g/2} \exp(-B_g|\pi m|^{r_g}) A_g^{1/2}$.*

Hence, we get that (see the study of $\| \ell_m - \ell_m^{(n)} \|_2^2$),

$$\begin{aligned}
&\|g - g_{\hat{m}_g}^{(n)}\|_{\infty,A}\\
&\quad \leq \|g - g_{\hat{m}_g}\|_\infty + \|g_{\hat{m}_g} - g_{\hat{m}_g}^{(n)}\|_\infty\\
&\quad \leq \frac{\sqrt{\kappa(\kappa_{\mathcal{G}}+1)}\hat{m}_g^{\frac{3}{2}}}{\sqrt{k_n}} + (2\pi)^{-1}\sqrt{\pi\hat{m}_g}((\pi\hat{m}_g)^2 + 1)^{-\frac{a_g}{2}} \exp(-B_g|\pi\hat{m}_g|^{\nu_g}) A_g^{\frac{1}{2}}.
\end{aligned}$$

For g in $\mathcal{S}_{a_g,\nu_g,B_g}(\kappa_{a_g})$ with $a_g > 1/2$ if $r_g = 0$, since $k_n \geq n^{3/2}$ and $\hat{m}_g = o(\sqrt{n})$ for $\alpha > 1/2$, $\|g - g_{\hat{m}_g}^{(n)}\|_\infty$ tends to zero. It follows that, for n large enough, $\|g - g_{\hat{m}_g}^{(n)}\|_{\infty,A} \leq g_0/4$ and consequently $\mathbb{P}(\tilde{E}_n^c) \leq \mathbb{P}[\|g_{\hat{m}_g}^{(n)} - \tilde{g}\|_\infty > g_0/4]$. Again, by Lemma 4.3, since $g_{\hat{m}_g}^{(n)} - \tilde{g}$ belongs to $S_{\hat{m}_g}$, we get that

$$\mathbb{P}(\tilde{E}_n^c) \leq \mathbb{P}\Big[\|g_{\hat{m}_g}^{(n)} - \tilde{g}\|_2 > \frac{g_0}{(4\sqrt{\hat{m}_g})}\Big]. \tag{4.37}$$

Hence,

$$\begin{aligned}
\|g_{\hat{m}_g}^{(n)} - \tilde{g}_{\hat{m}_g}\|_2^2 &= \sum_{|j|\leq k_n} (\hat{a}_{\hat{m}_g,j} - a_{\hat{m}_g,j})^2 = \sum_{|j|\leq k_n} \nu_{n,g}^2(\varphi_{\hat{m}_g,j})\\
&= \sup_{t\in B_{\hat{m}_g}(0,1)} \nu_{n,g}^2(t).
\end{aligned} \tag{4.38}$$

Consequently,

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

We use Lemma 4.1 with $M_1 = \sqrt{nH^2}$, $H \geq \mathbb{E}(\sup_{t\in B_{m,m'}(0,1)} |\nu_{n,g}(t|))$ and $v \geq \sup_{t\in B_{m,m'}(0,1)} \mathrm{Var}(u_t^*(Z_1))$. If we take $\lambda = g_0/(8\sqrt{m})$ and ensure that $2H <$

$g_0/(8\sqrt{m})$, then $\mathbb{P}[\sup_{t\in B_m(0,1)}|\nu_{n,g}(t)|\geq g_0/(4\sqrt{m})]\leq 3\exp[-K_1'n(\min[(mv)^{-1}, (M_1\sqrt{m})^{-1}])]$. This yields

$$\mathbb{P}(\tilde{E}_n^c)\leq K\sum_{m\in\mathcal{M}_{n,\ell}}\Big\{\exp\Big[-\frac{K_1'n}{(M_1\sqrt{m})}\Big]+\exp\Big[-\frac{K_1'n}{(mv)}\Big]\Big\}. \tag{4.39}$$

Since we only consider $m\leq\sqrt{n}$,

$$a_n|\mathcal{M}_{n,\ell}|\exp\Big[-\frac{K_1'n}{(M_1\sqrt{m})}\Big]\leq a_n|\mathcal{M}_{n,\ell}|\exp(-K"n^{\frac{1}{4}})=o(n^{-1}).$$

We also consider m such that $\Gamma(m)/n$ tends to zero. Consequently, when $\rho>0$, $\pi m\leq(\ln n/(2\beta\sigma^\rho+1))^{1/\rho}$. We combine this with $v\leq(\pi m)^{2\alpha+1-\rho}\exp(2\beta\sigma^\rho\pi^\rho m^\rho)$ and obtain the bound $a_n|\mathcal{M}_{n,\ell}|\exp(-K_1'n/(mv))=o(1/n)$.

When $\rho=0$, $v=\mu_1(\pi m)^{2\alpha+1/2}$. Since $\pi m\leq(n/\ln(n))^{1/(2\alpha+1)}\leq n^{1/(2\alpha+1)}$, we get that

$$\exp\Big(-\frac{K_1'n}{(mv)}\Big)\leq\exp\Big(-\frac{K"n}{(m^{2\alpha+\frac{3}{2}})}\Big)\leq\exp\Big(-K"n^{\frac{1}{(4(\alpha+1))}}\Big)=o(n^{-1}).$$

To control $\sqrt{m}H$, the worst case is $\rho=0$. In that case, for $\pi m\leq(n/\ln(n))^{1/(2\alpha+2)}$, we get that $\sqrt{m}H\leq 1/\sqrt{\ln(n)}$, which tends to zero. Therefore $\sqrt{m}H\leq 1/\sqrt{\ln(n)}$ is bounded by $g_0/8$ for n large enough. We conclude that $a_n\mathbb{P}(\tilde{E}_n^c)=o(1/n)$. The result follows by the inequalities (4.35) and (4.36).

Proof of Lemma 4.3. For $t\in S_m$, $t(x)=\sum_{j\in\mathbb{Z}}\langle t,\varphi_{m,j}\rangle\varphi_{m,j}(x)$ and $|t(x)|^2\leq\sum_{j\in\mathbb{Z}}|\langle t,\varphi_{m,j}\rangle|^2\sum_{j\in\mathbb{Z}}|(\varphi_{m,j}^*)^*(-x)|^2/(2\pi)^2$. By Parseval's Formula,

$$\sum_{j\in\mathbb{Z}}|\langle t,\varphi_{m,j}\rangle|^2\sum_{j\in\mathbb{Z}}\frac{|(\varphi_{m,j}^*)^*(-x)|^2}{(2\pi)^2}=\|t\|_2^2 m\int\frac{\varphi^*(u)^2du}{(2\pi)}=m\|t\|_2^2.$$

Let b be such that $1/2<b<a_g$. Since $\|g-g_m\|_\infty\leq(2\pi)^{-1}\int_{|x|\geq\pi m}|g^*(x)|dx$, we get that

$$\begin{aligned}&\|g-g_m\|_\infty\\&\quad\leq(2\pi)^{-1}((\pi m)^2+1)^{-\frac{(a_g-b)}{2}}e^{-B_g|\pi m|^{r_g}}\int_{|x|\geq\pi m}|g^*(x)|(x^2+1)^{\frac{(a_g-b)}{2}}e^{B_g|x|^{r_g}}dx\end{aligned}$$

and so

$$\begin{aligned}&\frac{1}{2\pi}((\pi m)^2+1)^{-\frac{(a_g-b)}{2}}\exp(-B_g|\pi m|^{r_g})\kappa_{a_g}^{\frac{1}{2}}\sqrt{\int_{|x|\geq\pi m}(x^2+1)^{-b}dx}\\&\quad\leq(2\pi)^{-1}((\pi m)^2+1)^{-\frac{(a_g-b)}{2}}\exp(-B_g|\pi m|^{r_g})\kappa_{a_g}^{\frac{1}{2}}(\pi m)^{\frac{1}{2}-b}\end{aligned}$$

$$\leq (2\pi)^{-1}\sqrt{\pi m}((\pi m)^2 + 1)^{-\frac{a_g}{2}} \exp(-B_g|\pi m|^{r_g})\kappa_{a_g}^{\frac{1}{2}}.$$

Acknowledgements

We thank the Co-Editor and the Associate Editor for very helpful suggestions and constructive remarks. We thank also Valentine Genon-Catalot for numerous fruitful discussions on the subject.

References

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

Butucea, C. (2004). Deconvolution of supersmooth densities with smooth noise. *Canadian J. Statist.* **32**, 181-192

Butucea, C. and Tsybakov, A.B. (2005). Sharp optimality for density deconvolution with dominating bias. To appear in *Theory Probab. Appl.*, http://fr.arxiv.org/abs/math.ST/0409471.

Comte, F. and Taupin, M.-L. (2004). Nonparametric estimation of the regression function in an errors-in-variables model. *Preprint MAP5* 2004-20, http://fr.arxiv.org/abs/math.ST/0511111.

Comte, F., Rozenholc, Y. and Taupin, M.-L. (2006). Penalized contrast estimator for density deconvolution. *Canadian J. Statist.* **34**, 431-452.

Comte, F., Rozenholc, Y. and Taupin, M.-L. (2005). Finite sample penalization in adaptive deconvolution. To appear in *J. Statist. Comput. Simulation*, available online.

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

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

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

Fan, J., Truong, Y. K. and Wang, Y. (1991). Nonparametric function estimation involving errors-in-variables. *Nonparametric Functional Estimation and Related Topics*, 613-627.

Ioannides, D. A. and Alevizos, P. D. (1997). Nonparametric regression with errors in variables and applications. *Statist. Probab. Lett.* **32**, 35-43.

Koo, J.-Y. and Lee K.-W. (1998). B-spline estimation of regression functions with errors in variable. *Statist. Probab. Lett.* **40**, 57-66.

Ledoux, M. (1996). On Talagrand's deviation inequalities for product measures. *ESAIM Probab. Statist.*, 63-87.

Masry, E. (1993). Multivariate regression estimation with errors-in-variables for stationary processes. *J. Nonparametr. Statist.* **3**, 13-36.

Meyer, E. (1990). Ondelettes et opérateurs. I., Ondelettes, *Actualités Mathématiques*. (Wavelets and operators. I., Wavelets, Current Mathematical Topics), Hermann, Paris.

Pensky, M. and Vidakovic, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Statist.* **27**, 2033-2053.

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.

Talagrand, M. (1996). New concentration inequalities in product spaces. *Invent. Math.* **126**, 505-563.

Truong, Y. K. (1991). Nonparametric curve estimation with time series errors. *J. Statist Plann Inference* **28**, 167-183.

Tsybakov, A. B. (1986). Robust reconstruction of functions by the local-approximation method. *Probl. Inf. Transm.* **22**, 133-146.

Université Paris Descartes, MAP5, UMR CNRS 8145, 45 rue des Saints-Pères, 75 270 PARIS cedex 06, France.

E-mail: fabienne.comte@univ-paris5.fr.

IUT de Paris V et Université d'Orsay, Laboratoire de Probabilités, Statistique et Modélisation, UMR 8628, Bâtiment 425,91405 Orsay Cedex, France.

E-mail: Marie-Luce.Taupin@math.u-psud.fr

(Received October 2004; revised January 2006)