Adaptive estimation of linear functionals in the convolution model and applications.

Cristina Butucea, Fabienne Comte

▶ To cite this version:

Adaptive estimation of linear functionals in the convolution model and applications.

C. Butucea, F. Comte$^{(*)}$

February 2007

Abstract

We consider the model $Z_i = X_i + \varepsilon_i$ for i.i.d. X_i's and ε_i's and independent sequences $(X_i)_{i\in\mathbb{N}}$ and $(\varepsilon_i)_{i\in\mathbb{N}}$. The density of ε is assumed to be known whereas the one of X_1 denoted by g is unknown. Our aim is to study the estimation of linear functionals of g, $\langle\psi, g\rangle$ for a known function ψ. We propose a general estimator of $\langle\psi, g\rangle$ and study the rate of convergence of its quadratic risk in function of the smoothness of g, f_ε and ψ. Different dependency contexts are also considered. An adaptive estimator is then proposed, following a method studied by Laurent *et al.* [23] in another context. The quadratic risk of this estimator is studied. The results are applied to adaptive pointwise deconvolution, in which context losses in the adaptive rates are shown to be optimal in the minimax sense. They are also applied to pointwise Laplace transform estimation in the standard context and in the context of the stochastic volatility model. Estimation in the context of ARCH-type models lastly illustrates the method.

Authors. C. Butucea, Université Paris X, Modal'X and LPMA, UMR CNRS 7599, 175 rue du Chevaleret, 75013 Paris, FRANCE.
F. Comte, *corresponding author*, MAP5, UMR CNRS 8145, Université Paris 5, 45 rue des Saints-Pères, 75270 Paris cedex 06, FRANCE.

1 Introduction

We consider the convolution model

$$Z_i = X_i + \varepsilon_i. \tag{1}$$

The sequences $(X_i)_{i\in\mathbb{N}}$ and $(\varepsilon_i)_{i\in\mathbb{N}}$ are independent. The X_i's are i.i.d with unknown density g, the ε_i's are i.i.d. with known density f_ε, whose smoothness is described by the following

assumption.

Suppose there exist nonnegative numbers κ_0, κ_0', β, α, and ρ such that f_ε^* satisfies

$$\kappa_0(x^2+1)^{-\beta/2}\exp\{-\alpha|x|^\rho\} \leq |f_\varepsilon^*(x)| \leq \kappa_0'(x^2+1)^{-\beta/2}\exp\{-\alpha|x|^\rho\}, \tag{2}$$

with $\beta > 1$ when $\rho = 0$. Since f_ε is known, the constants $\alpha, \rho, \kappa_0, \kappa_0'$ and β defined in (2) are known. When $\rho = 0$ in (2), the errors are called "ordinary smooth" errors. When $\alpha > 0$ and $\rho > 0$, they are called "super smooth". The standard examples for super smooth densities are Gaussian or Cauchy distributions (super smooth of order $\beta = 0, \rho = 2$ and $\beta = 0, \rho = 1$ respectively). An example of ordinary smooth density is the Laplace distribution ($\rho = 0 = \alpha$ and $\beta = 2$).

In this context, many papers studied the so-called "deconvolution problem". In other words, many strategies have been developed in order to estimate the distribution g of the unobserved X_i's, when assuming that g belongs to some smoothness class defined by:

$$\mathcal{S}(b,a,r,L) = \left\{f \text{ such that } \int_{-\infty}^{+\infty} |f^*(x)|^2 (x^2+1)^b \exp\{2a|x|^r\}dx \leq 2\pi L\right\} \tag{3}$$

for b, a, r, L some unknown non-negative numbers, such that $b > 1/2$ when $r = 0$.

Kernel estimators were first widely studied (see Carroll and Hall [10], Stefanski and Carroll [29], Fan [17]), in the case of Sobolev balls (case $r = 0$ in (3)). Classically in this context, the slowest rates of convergence for estimating g are obtained for super smooth error densities. Then adaptive strategies have been examined, using wavelets (see Pensky and Vidakovic [27]), or model selection methods (see Comte *et al.* [15]). These works, together with those of Butucea [5], Butucea and Tsybakov [7] and Lacour [22], studied cases $r > 0, a > 0$ in (3) involving thus infinitely many times differentiable functions and lead to improved but non standard rates whose optimality in the minimax sense was detailed in Fan [17], Butucea [5], Butucea and Tsybakov [7].

In this paper, we are interested in the problem of estimating $\theta(g) = \langle \psi, g \rangle = \mathbb{E}(\psi(X_1))$ in model (1), where ψ is a known integrable function.

For the sake of clarity, we first define the three types of estimators and associated rates discussed in this paper: minimax, adaptive minimax and adaptive. Let $\Lambda = [\underline{b}, \overline{b}] \times [\underline{a}, \overline{a}] \times [\underline{r}, \overline{r}] \times [\underline{L}, \overline{L}] \subset [0, \infty) \times [0, \infty) \times (0, 2] \times (0, \infty)$ be a set of parameters $\lambda = (b, a, r, L)$.

Definition 1.1 *A sequence $\varphi_{n,\lambda}$ which tends to 0 with n is a minimax rate of convergence over the class of density functions $\mathcal{S}(\lambda)$ if there exists an estimator θ_n^* of θ and a constant $C > 0$ such that*

$$\sup_{g \in \mathcal{S}(\lambda)} \varphi_{n,\lambda}^{-2}\mathbb{E}_g[|\theta_n^* - \theta(g)|^2] \leq C, \text{ for } n \text{ large enough,}$$

and if for some $c > 0$ we have

$$\inf_{\theta_n} \sup_{g \in \mathcal{S}(\lambda)} \varphi_{n,\lambda}^{-2}\mathbb{E}_g[|\theta_n - \theta(g)|^2] \geq c, \text{ for } n \text{ large enough,}$$

where the infimum is taken over all estimators θ_n of θ.

Definition 1.2 *An estimator $\hat{\theta}_n$ is adaptive minimax over the family of classes $\bigcup_{\lambda \in \Lambda} \mathcal{S}(\lambda)$ if there exists some constant $C > 0$ such that*

$$\sup_{\lambda \in \Lambda} \sup_{g \in \mathcal{S}(\lambda)} \varphi_{n,\lambda}^{-2}\mathbb{E}_g[|\hat{\theta}_n - \theta(g)|^2] \leq C, \text{ for } n \text{ large enough,}$$

2

where $\varphi_{n,\lambda}$ is the minimax rate of convergence of the pointwise risk (i.e. for fixed values of b, a, r and L).

It is not always possible to attain the minimax rate uniformly over a set of parameters Λ. Most often there is a loss in the rate due to adaptation.

Definition 1.3 *We say that an estimator $\hat{\theta}_n^*$ is adaptive if it attains a rate of convergence $\psi_{n,\lambda}$ uniformly in λ over Λ, i.e. there exists a constant $C > 0$ such that*

$$\sup_{\lambda \in \Lambda} \sup_{g \in \mathcal{S}(\lambda)} \psi_{n,\lambda}^{-2} \mathbb{E}_g[|\hat{\theta}_n^* - \theta(g)|^2] \leq C, \text{ for } n \text{ large enough.}$$

and if the loss of rate with respect to the minimax rate is optimal, i.e. it satisfies the following lower bounds

$$\inf_{\theta_n} \sup_{\lambda \in \Lambda} \sup_{g \in \mathcal{S}(\lambda)} \psi_{n,\lambda}^{-2} \mathbb{E}_g[|\theta_n - \theta(g)|^2] \geq c,$$

for n large enough, where the infimum is taken over all possible estimators θ_n.

Comte *et al.* [15] developed model selection techniques to provide an adaptive estimator of g. Using the same collection of spaces S_m, we can build an estimator of $\theta(g) = \langle \psi, g \rangle$ on a given S_m, for which we can exhibit various rates of the mean square error. Then, in the spirit of Laurent *et al.* [23], we build an adaptive procedure for automatic selection of the space S_m in a collection $(S_m)_{m \in \mathcal{M}_n}$. The difficulty here lies in finding an adequate penalization of an empirical error of the estimator. We adapt Laurent *et al.* [23]'s methodology by defining the selection spaces in the frequency domain. Moreover, our setting is not gaussian.

To compute the rates, we have to take into account the regularity parameters of the function ψ, which is thus assumed to satisfy, $\forall x \in \mathbb{R}$,

$$|\psi^*(x)|^2 \leq C_\psi (x^2 + 1)^{-B} \exp(-2A|x|^R). \tag{4}$$

We also extend the result to different dependency contexts, in view of particular hidden markov models or ARCH-type models.

Adaptive estimation of linear functionals has been widely studied in the context of the white noise model and regression (direct observation), see e.g. Lepski [24], Tsybakov [30], Cai and Low [9, 8], Artiles and Levit [3] Laurent *et al.* [23] and in the context of density models with direct observations Lepski and Levit [25], Butucea [4], Artiles [2]. For the model of Gaussian sequences Golubev and Levit [20] and Golubev [19] considered adaptive estimation of linear functionals in both direct and inverse setup. Note also that in some particular inverse problems the pointwise adaptive estimation was solved by Klemelä and Tsybakov [21] for the Riesz transform, by Cavalier [11] for tomography problem. To our knowledge, we present the first work on model selection based adaptation for density estimation in the convolution model (1).

Our findings are interesting, in term of rates and loss due to adaptation. We provide a new adaptation procedure and when applying our general results to pointwise estimation of g, we recover as a particular case, the upper bound rates obtained by Fan [17], Butucea [5] and Butucea and Tsybakov [7], directly in a general context. Moreover, we prove the optimality in the minimax sense of the loss due to adaptation for Sobolev smooth densities and supersmooth densities in presence of ordinary smooth noise and for supersmooth densities in presence of supersmooth noise with $r \geq \rho$ and $0 < \rho \leq 1$ (in the case $r < \rho$ no loss occurs, while the case

3

$r \geq \rho$ and $1 < \rho < 2$ is still open). As a by-product we also prove in the last case that the rates of our estimator (which requires knowledge of a, r) are optimal in the minimax sense, which was not yet known in the literature.

We also apply the procedure to pointwise Laplace transform estimation, when X_1 is a positive random variable, even when the Laplace transform of the noise is infinite. Lastly, we illustrate our method with an application to the discrete stochastic volatility model, where derivatives of the Laplace transform of the volatility can be estimated with good rates. All the upper bounds given in the applications correspond to particular values for the parameters B, A, R of the function ψ in (4). Lastly, we show how the methods can be applied in the context of general ARCH-type models.

The plan of the paper is the following. Section 2 defines the estimators and studies their rates with squared loss function, and the adaptive procedure is detailed in Section 3. Both independent and β-mixing contexts are studied. In Section 4, several applications of our general results are detailed. Section 4.1 is devoted to the application of the results to adaptive pointwise deconvolution, upper bounds are deduced from Section 3 and the associated lower bounds are proven when a loss occurs. Section 4.2 presents application to Laplace transform estimation, in the standard context and 4.3 to the context of the stochastic volatility model. Lastly, Section 4.4 explains how the procedure applies to ARCH-type processes. Some proofs are gathered in Section 5.

2 Study of strategies for estimation

Recall that we want to estimate $\theta(g) = \langle \psi, g \rangle = \mathbb{E}(\psi(X_1))$ where X_1 follows model (1) and is unobserved. Only the Z_i's, for $i = 1, \ldots, n$ are available.

We assume in all the following that:

$$f_\varepsilon \text{ belongs to } \mathbb{L}_2(\mathbb{R}) \text{ and is such that } \forall x \in \mathbb{R}, \ f_\varepsilon^*(x) \neq 0. \tag{5}$$

Note that the square integrability of f_ε requires that $\beta > 1/2$ when $\rho = 0$ in (2).

In the sequel, we denote by $\star$ the convolution product of functions $(u \star v(x) = \int u(t)v(t-x)dt)$ and by u^* the Fourier Transform of u: $u^*(x) = \int e^{itx} u(t) dt$.

2.1 Two strategies

Two ideas can be investigated.

The first one is to write $\langle \psi, g \rangle = (1/2\pi)\langle \psi^*, g^* \rangle$. As the density f_Z of Z_1 satisfies $f_Z = g \star f_\varepsilon$, we have $f_Z^* = g^* f_\varepsilon^*$. In other words, under (5), $\langle \psi, g \rangle = (1/2\pi)\langle \psi^*, f_Z^*/f_\varepsilon^* \rangle$. Replacing $f_Z^*(t)$ by its empirical version $(1/n)\sum_{k=1}^n e^{itZ_k}$, this leads to the estimator

$$\check{\theta} = \frac{1}{2\pi n} \sum_{k=1}^n \int e^{itZ_k} \frac{\psi^*(t)}{f_\varepsilon^*(t)} dt. \tag{6}$$

This estimator is built directly and seems attractive. Unfortunately, the term f_ε^* in the denominator should make in many cases the integral divergent (think of a Gaussian noise ε for instance). Thus, for the estimator to be well defined, it is wise to take as an estimator of $\theta(g)$,

$$\check{\theta}_m = \frac{1}{2\pi n} \sum_{k=1}^n \int_{|t| \leq \pi m} e^{itZ_k} \frac{\psi^*(t)}{f_\varepsilon^*(t)} dt. \tag{7}$$

4

The second strategy is less direct but natural as well: we can use some estimator of g, $\hat{g}_m$ and set

$$\hat{\theta}_m = \theta(\hat{g}_m) = \langle \psi, \hat{g}_m \rangle. \tag{8}$$

It happens that if $\hat{g}_m$ is the projection estimator defined in Comte et al. [15], then $\hat{\theta}_m = \check{\theta}_m$. To see this, we need to recall the definition of $\hat{g}_m$.

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 [26], p.22). Let us 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 $[-\pi m, \pi m]$. Here Condition (5) allows to define the following contrast function: for t in S_m, let

$$\gamma_n(t) = \frac{1}{n}\sum_{i=1}^{n}\left[\|t\|^2 - 2u_t^*(Z_i)\right], \quad \text{with} \quad u_t(x) = \frac{1}{2\pi}\frac{t^*(-x)}{f_\varepsilon^*(x)}. \tag{9}$$

Then, for an arbitrary fixed integer m, an estimator of g belonging to S_m is defined by

$$\hat{g}_m = \arg\min_{t \in S_m} \gamma_n(t). \tag{10}$$

By using Parseval and inverse Fourier formulae we obtain that $\mathbb{E}[u_t^*(Z_i)] = \langle t, g \rangle$, so that $\mathbb{E}(\gamma_n(t)) = \|t - g\|^2 - \|g\|^2$ is minimal when $t = g$. This explains why $\gamma_n(t)$ is well-suited for the estimation of g. Note that the orthogonal projection of g on S_m is $g_m = \sum_{j \in \mathbb{Z}} a_{m,j}(g)\varphi_{m,j}$ where $a_{m,j}(g) = < \varphi_{m,j}, g >$ and that

$$\hat{g}_m = \sum_{j \in \mathbb{Z}} \hat{a}_{m,j}\varphi_{m,j} \ \text{with} \ \hat{a}_{m,j} = \frac{1}{n}\sum_{i=1}^{n}u_{\varphi_{m,j}}^*(Z_i), \quad \text{and} \quad \mathbb{E}(\hat{a}_{m,j}) = a_{m,j}.$$

It is then easy to see that

$$
\begin{aligned}
\hat{\theta}_m &= \langle \hat{g}_m, \psi \rangle = \sum_{j \in \mathbb{Z}} a_{m,j}\langle \varphi_{m,j}, \psi \rangle = \sum_{j \in \mathbb{Z}}\frac{1}{n}\sum_{k=1}^{n}\int e^{iuZ_k}\frac{\varphi_{m,j}^*(u)}{f_\varepsilon^*(u)}du\frac{1}{2\pi}\langle \psi^*, \varphi_{m,j}^* \rangle \\
&= \frac{1}{2\pi n}\sum_{k=1}^{n}\int e^{iuZ_k}\frac{\sum_{j \in \mathbb{Z}}\langle \psi^*, \varphi_{m,j}^* \rangle \varphi_{m,j}^*(u)}{f_\varepsilon^*(u)}du \\
&= \check{\theta}_m,
\end{aligned}
$$

because $\sum_{j \in \mathbb{Z}}\langle \psi^*, \varphi_{m,j}^* \rangle \varphi_{m,j}^*(u) = \psi^*(u)\mathbb{I}_{|u| \leq \pi m}$. We have proved that:

Proposition 2.1 *Let $\hat{\theta}_m = \theta(\hat{g}_m)$ be defined by (8) with $\hat{g}_m$ defined by (9)-(10) and $\check{\theta}_m$ defined by (7), then $\hat{\theta}_m = \check{\theta}_m$.*

We can note that in practice $\hat{g}_m$ involves an infinite sum which should be truncated. The study of the impact of this truncation is in Comte *et al* [15]. For this reason, (7) may be a better way to write the estimator.

2.2 Risk bounds and rates for independent variables

The direct strategy has the advantage that $m = +\infty$ can be chosen. If this is possible, then the estimate is unbiased and its rate can reach the parametric rate. Indeed the variance is

$$
\begin{aligned}
\mathrm{Var}(\check{\theta}) &= \mathrm{Var}(\check{\theta}_\infty) = \frac{1}{4\pi^2 n}\mathrm{Var}\left(\int e^{iuZ_1}\frac{\psi^*(u)}{f_\varepsilon^*(u)}du\right) \\
&= \frac{1}{4\pi^2 n}\iint (f_Z^*(u-v) - f_Z^*(u)f_Z^*(-v))\frac{\psi^*(u)\psi^*(-v)}{f_\varepsilon^*(u)f_\varepsilon^*(-v)}dudv \\
&\leq \frac{1}{4\pi^2 n}\int \frac{|\psi^*(u)|^2}{|f_\varepsilon^*(u)|^2}du \int |f_Z^*(x)|dx.
\end{aligned}
$$

Another bound for the variance shall prove useful in the sequel:

$$
\mathrm{Var}(\check{\theta}) \leq \frac{1}{4\pi^2 n}\mathbb{E}\left(\left|\int e^{iuZ_1}\frac{\psi^*(u)}{f_\varepsilon^*(u)}du\right|^2\right) \leq \frac{1}{4\pi^2 n}\left(\int \frac{|\psi^*(u)|}{|f_\varepsilon^*(u)|}du\right)^2.
$$

Finally,

$$
\mathrm{Var}(\check{\theta}) \leq \frac{1}{4\pi^2 n}\min\left\{\int |f_Z^*|\int \frac{|\psi^*(u)|^2}{|f_\varepsilon^*(u)|^2}du, \left(\int \frac{|\psi^*(u)|}{|f_\varepsilon^*(u)|}du\right)^2\right\}.
$$

Thus if all integrals are finite, the estimator has a quadratic risk $\mathbb{E}(\theta - \check{\theta})^2$ of order $1/n$. As $\int |f_Z^*(x)|dx \leq \int |f_\varepsilon^*(x)|dx < \infty$ by (2), we have the following result:

Proposition 2.2 *Assume that f_ε and ψ are such that $\int |f_\varepsilon^*(x)|dx < +\infty$ and*

$$
\int |\psi^*(x)/f_\varepsilon^*(x)|^2 dx < +\infty \ \text{ or } \ \int |\psi^*(x)/f_\varepsilon^*(x)|dx < \infty. \tag{11}
$$

Then $\check{\theta}$ given by (6) is well defined. It is an unbiased estimator of $\theta(g) = \langle \psi, g\rangle$, and $\mathbb{E}[(\check{\theta} - \theta(g))^2] \leq C/n$.

Remark 2.1 Condition (11) is fulfilled if ψ^* decreases faster than f_ε^* near infinity, which corresponds to the intuitive idea that ψ is a smoother function than f_ε. For example, this happens if ψ is supersmooth when ε is ordinary smooth.

In the general case, a bound for the squared bias can be found, using that $\mathbb{E}(\theta - \hat{\theta}_m)^2 = b^2(\hat{\theta}_m) + \mathrm{Var}(\hat{\theta}_m)$ with $b(\hat{\theta}_m) = \theta - \mathbb{E}(\hat{\theta}_m)$. As $\mathbb{E}(\hat{\theta}_m) = (1/(2\pi))\int_{|t|\leq \pi m} g^*(t)\psi^*(t)dt$, we obtain

$$
b(\hat{\theta}_m) = \frac{1}{2\pi}\left(\int g^*(t)\psi^*(t)dt - \int_{|t|\leq \pi m} g^*(t)\psi^*(t)dt\right) = \frac{1}{2\pi}\int_{|t|\geq \pi m} g^*(t)\psi^*(t)dt.
$$

Therefore, the squared-bias variance decomposition is here

$$
\mathbb{E}(\theta - \hat{\theta}_m)^2 \leq b^2(\hat{\theta}_m) + \frac{1}{4\pi^2 n}\min\left\{\int_{|u|\leq \pi m}\frac{|\psi^*(u)|^2}{|f_\varepsilon^*(u)|^2}du\int |f_Z^*|, \left(\int_{u\leq \pi m}\frac{|\psi^*(u)|}{|f_\varepsilon^*(u)|}du\right)^2\right\}.
$$

Thus we can study the rates that can be deduced from the previous upper bounds, in function of the smoothness parameters of the three involved functions: g, ψ, f_ε.

6

Proposition 2.3 *Assume that $C_\varepsilon = \int |f_\varepsilon^*(x)|dx < +\infty$, and let $\hat\theta_m$ be defined by (8) or (7). Then*

$$\mathbb{E}(\theta - \hat\theta_m)^2 \leq \left(\frac{1}{2\pi}\int_{|t|\geq\pi m}|g^*(t)\psi^*(t)|dt\right)^2 + \frac{1}{4\pi^2 n}\min\left\{C_\varepsilon\int_{-\pi m}^{\pi m}\frac{|\psi^*|^2}{|f_\varepsilon^*|^2}, \left(\int_{-\pi m}^{\pi m}\frac{|\psi^*|}{|f_\varepsilon^*|}\right)^2\right\}.$$

Let us assume thus that ψ satisfies (4), that g belongs to $\mathcal{S}(b,a,r,L)$ as defined by (3) and that f_ε^* fulfills (2). Then

$$
\begin{aligned}
b^2(\hat\theta_m) &\leq \left|\int_{|x|\geq\pi m}g^*(x)\psi^*(x)dx\right|^2\\[2mm]
&\leq \left|\int_{|x|\geq\pi m}|g^*(x)|(1+x^2)^{b/2}\exp(a|x|^r)(|\psi^*(x)|(1+x^2)^{-b/2}\exp(-a|x|^r))dx\right|^2\\[2mm]
&\leq \int_{|x|\geq\pi m}|g^*(x)|^2(1+x^2)^b\exp(2a|x|^r)dx\int_{|x|\geq\pi m}|\psi^*(x)|^2(1+x^2)^{-b}\exp(-2a|x|^r)dx\\[2mm]
&\leq LC\int_{|x|\geq\pi m}(1+x^2)^{-b-B}\exp(-2a|x|^r - 2A|x|^R)dx\\[2mm]
&\leq C_1 m^{-2b-2B-\max(r,R)+1}\exp(-2a(\pi m)^r - 2A(\pi m)^R).
\end{aligned}
$$

On the other hand, for the variance, we find:

$$
\mathrm{Var}(\hat\theta_m) \leq
\begin{cases}
\dfrac{C'}{n}\ \text{(case (I))}, & \begin{cases}\text{if }(\rho = R = 0, \beta < B - 1/2)\\ \text{or }(\rho = R > 0, \alpha = A, \beta < B - 1/2)\\ \text{or }(\rho = R, \alpha < A)\\ \text{or }(\rho < R)\end{cases}\\[8mm]
\dfrac{C'\ln(m)}{n}\ \text{(case (II))}, & \begin{cases}\text{if }(\rho = R > 0, \alpha = A, \beta = B - 1/2)\\ \text{or }(\rho = R = 0, \beta = B - 1/2)\end{cases}\\[6mm]
\dfrac{C'}{n}m^{2\beta-2B+1},\ \text{(case (III))} & \begin{cases}\text{if }(\rho = R > 0, \alpha = A, \beta > B - 1/2)\\ \text{or }(\rho = R = 0, \beta > B - 1/2)\end{cases}\\[6mm]
\dfrac{C'}{n}m^{2\beta-2B+1-\rho+(1-\rho)_+}e^{2\alpha(\pi m)^\rho - 2A(\pi m)^R} & \text{if }(\rho > R)\text{ or }(\rho = R > 0, \alpha > A).\\ \text{(case (IV))}
\end{cases}
$$

The term $1 - \rho + (1 - \rho)_+$, where $x_+ = \max\{x, 0\}$, comes from comparisons of the two possible variance orders as, e.g. for $R = 0$:

$$\int_{|u|\leq\pi m}\frac{|\psi^*(u)|^2}{|f_\varepsilon^*(u)|^2}du \leq C_2 m^{2\beta-2B+1-\rho}\exp(2\alpha(\pi m)^\rho)$$

and

$$\int_{|u|\leq\pi m}\frac{|\psi^*(u)|}{|f_\varepsilon^*(u)|}du \leq C_3 m^{\beta-B+1-\rho}\exp(\alpha(\pi m)^\rho).$$

Table 1: Upper bounds for the minimax rates of convergence

Parameters			Rate
$\rho < R$			n^{-1}
$\rho = R \quad \alpha < A$			n^{-1}
$(\rho = R = 0)$ or $(\rho = R > 0, \alpha = A)$	$\beta \leq B - 1/2$		n^{-1}
	$\beta = B - 1/2$	$\max\{r, R\} = 0$	$(\ln \ln n) n^{-1}$
		$\max\{r, R\} > 0$	$(\ln n) n^{-1}$
	$\beta > B - 1/2$	$\max\{r, R\} > 0$	$(\ln n)^{(2\beta - 2B + 1)/\{r \vee R\}} n^{-1}$
		$\max\{r, R\} = 0$	$n^{-(b+B-1/2)/(b+B)}, \ (b > 1/2)$
$(\rho = R > 0, \alpha > A)$			v_n
$\rho > R$	$\max\{r, R\} > 0$		v_n
	$\max\{r, R\} = 0$		$\ln(n)^{-(2(b+B)-1)/\rho}, \ (b > 1/2)$

More generally, several cases can arise, detailed here and summarized in Table 1. Note that in case $(\rho = R > 0, \ \alpha > A, \ \min\{r, R\} > 0)$ or $(\rho = R = 0, \ \max r, R > 0$ the rate is given by

$$
\begin{aligned}
v_n \quad = \quad &\arg\min_m \Big\{ C_B m^{-2b - 2B + 1 - r \vee R} e^{-2a(\pi m)^r - 2A(\pi m)^R} \\
&+ m^{2\beta - 2B + 1 - \rho + (1-\rho)_+} e^{2\alpha(\pi m)^\rho - 2A(\pi m)^R} \frac{1}{n} \Big\}.
\end{aligned}
$$

These rates are faster than $(\ln(n))^{-\lambda_1}$ and slower than $n^{-\lambda_2}$ for any $\lambda_1, \lambda_2 > 0$.

Remark 2.2 Here, the smoothness of ψ seems to have no influence on the optimal choice for m. Nevertheless, the dependence on the unknown parameters related to g of the different optimal choices of m enhances the interest of an automatic selection of m.

Note that $\psi(x) = x$ or $\psi(x) = x^p$ are not integrable on $\mathbb{R}$ so that the moments of X_1 can not be estimated in that way. But if ε_1 admits moments of the same order, since they are known, they can be used together with empirical moments of the Z_i's to obtain estimated moments of X_1.

2.3 Extension to mixing contexts

In view of applications, it is natural to study the robustness of the results with respect to dependency in the variables, and in particular to β-mixing properties.

To be more precise, two dependency contexts are considered. First, we can assume:

(D1) In Model (1), the sequences (X_i) and (ε_i) are independent and the ε_i's are i.i.d. The sequence (X_i) is strongly stationary and β-mixing, with β-mixing coefficients denoted by $(\beta_k)_k$.

otherwise we assume:

(D2) In Model (1), the ε_i's are i.i.d and for any given i, X_i and ε_i are independent (but the sequences (X_i) and (ε_i) are not independent). The sequence $(Z_i, X_i)_{i \in \mathbb{Z}}$ is strongly stationary and β-mixing, with β-mixing coefficients denoted by $(\beta_k)_k$.

Context (D1) encompasses the case of particular Hidden Markov Models, when the noise is additive and (X_i) is a β-mixing Markov process. As many Markov chain models or other standard models can be proved to have such mixing properties (see Doukhan [16] for a large set of examples and study of their mixing properties), this means that our results can be applied to many classical models. In that case, we can prove the following result:

Proposition 2.4 *Consider the model (1) under* (D1) *with moreover* $\sum_{k \geq 0} \beta_k < +\infty$. *Assume that* $C_\varepsilon = \int |f_\varepsilon^*(x)| dx < +\infty$. *Let* $\hat{\theta}_m$ *be defined by (8) or (7). Then*

$$\mathbb{E}(\theta - \hat{\theta}_m)^2 \leq \left(\frac{1}{2\pi} \int_{|t| \geq \pi m} |g^*(t)\psi^*(t)| dt \right)^2 + \frac{C_\varepsilon}{4\pi^2 n} \min \left\{ \int_{-\pi m}^{\pi m} \frac{|\psi^*|^2}{|f_\varepsilon^*|^2}, \left(\int_{-\pi m}^{\pi m} \frac{|\psi^*|}{|f_\varepsilon^*|} \right)^2 \right\}$$
$$+ \frac{2(\int_{|t| \leq \pi m} |\psi^*|(t) dt)^2 \sum_{k \geq 0} \beta_k}{n}. \tag{12}$$

In particular, if $K_\psi := \int |\psi^*(t)| dt < +\infty$, *then*

$$\mathbb{E}(\theta - \hat{\theta}_m)^2 \leq \left(\frac{1}{\pi} \int_{\pi m}^{+\infty} |g^*\psi^*| \right)^2 + \frac{C_\varepsilon}{4\pi^2 n} \min \left\{ \int_{-\pi m}^{\pi m} \frac{|\psi^*|^2}{|f_\varepsilon^*|^2}, \left(\int_{-\pi m}^{\pi m} \frac{|\psi^*|}{|f_\varepsilon^*|} \right)^2 \right\} + \frac{K}{n}, \tag{13}$$

where $K = 2K_\psi^2 \sum_k \beta_k$.

Note that, in any case, we have in (12),

$$\int_{|t| \leq \pi m} |\psi^*(t)| dt \leq \min \left\{ 2\pi \|f_\varepsilon\|^2 \int_{-\pi m}^{\pi m} \frac{|\psi^*|^2}{|f_\varepsilon^*|^2}, \left(\int_{-\pi m}^{\pi m} \frac{|\psi^*|}{|f_\varepsilon^*|} \right)^2 \right\},$$

so that the last term is always less or equal than the variance term. It follows that the rates, in the context of mixing X_k's described by assumption (D1), remain the same as in the independent setting.

We explain in Section 4.4, how context (D2) is linked with ARCH models. Let us for now only state the result:

Proposition 2.5 *Consider the model (1) under* (D2) *with moreover* $\sum_{k \geq 0} \beta_k < +\infty$. *Assume that* $C_\varepsilon = \int |f_\varepsilon^*(x)| dx < +\infty$. *Let* $\hat{\theta}_m$ *be defined by (8) or (7). Then*

$$\mathbb{E}(\theta - \hat{\theta}_m)^2 \leq \left(\frac{1}{2\pi} \int_{|t| \geq \pi m} |g^*(t)\psi^*(t)| dt \right)^2 + \frac{C_\varepsilon}{4\pi^2 n} \min \left\{ \int_{-\pi m}^{\pi m} \frac{|\psi^*|^2}{|f_\varepsilon^*|^2}, \left(\int_{-\pi m}^{\pi m} \frac{|\psi^*|}{|f_\varepsilon^*|} \right)^2 \right\}$$
$$+ \frac{2 \sum_{k \geq 0} \beta_k}{n} \left(\int_{-\pi m}^{\pi m} |\psi^*| \right) \left(\int_{-\pi m}^{\pi m} \frac{|\psi^*|}{|f_\varepsilon^*|} \right). \tag{14}$$

9

In particular, if f_ε satisfies (2) and if ψ satisfies (4) and

$$|\psi^*(x)| \geq C'_\psi (x^2+1)^{-B} \exp(-2A|x|^R), \tag{15}$$

with $\beta > \max(B,1)$ or $(A > 0, \rho > 0)$, then

$$\mathbb{E}(\theta - \hat{\theta}_m)^2 \leq \left(\frac{1}{\pi} \int_{\pi m}^{+\infty} |g^*\psi^*|\right)^2 + \frac{K}{4\pi^2 n} \min\left\{\int_{-\pi m}^{\pi m} \frac{|\psi^*|^2}{|f_\varepsilon^*|^2}, \left(\int_{-\pi m}^{\pi m} \frac{|\psi^*|}{|f_\varepsilon^*|}\right)^2\right\}, \tag{16}$$

where K is a constant.

Note that condition (2) contains two inequalities analogous to (15) added to (4): they ensure that the orders are exact and not only upper bounds. They are required to compare the order of the additional mixing term to them.

It appears from Inequality (14) that

$$\left(\int_{|t| \leq \pi m} |\psi^*|(t)dt\right)\left(\int_{|t| \leq \pi m} |\psi^*/f_\varepsilon^*|(t)dt\right) \leq \left(\int_{|t| \leq \pi m} |\psi^*/f_\varepsilon^*|(t)dt\right)^2$$

but the comparison with $\int_{|t| \leq \pi m} |\psi^*/f_\varepsilon^*|^2(t)dt$ requires a case study which explains the validity conditions $\beta > \max(B,1)$ or $(A > 0, \rho > 0)$ given for (16). It follows from (16) that the rates given in Table 1 are preserved whenever the ε_i's are supersmooth.

3 Adaptive estimation

3.1 The adaptation problem for linear functionals

The problem of adaptation with linear functionals can be understood by comparison with what happens for global estimation of g for instance. In this context, we can see that $\|g - \hat{g}_m\|^2 = \|g - g_m\|^2 + \|g_m - \hat{g}_m\|^2$ with Pythagoras theorem. Then to mimic and perform the squared-bias / variance compromise, both terms must be approximated. But as g_m is the orthogonal projection of g on S_m, $\|g - g_m\|^2 = \|g\|^2 - \|g_m\|^2$. Then the squared bias can be reduced to $-\|g_m\|^2$, the other term being a constant. A natural estimation of $-\|g_m\|^2$ is $\gamma_n(\hat{g}_m) = -\|\hat{g}_m\|^2$. This explains why model selection in that case is performed by setting $\hat{m} = \arg\min_m\{\gamma_n(\hat{g}_m) + \text{pen}(m)\}$ where the penalty generally has roughly the order of the variance ($\mathbb{E}(\|\hat{g}_m - g_m\|^2)$).

For linear functionals, let us describe the heuristics. As now only a standard square in involved, $(\theta(g) - \theta(g_m))^2 = \theta^2(g) - 2\theta(g)\theta(g_m) + \theta^2(g_m)$, no simplification occurs in the cross product. Therefore, the best approximation of the bias is obtained by replacing it by $(\theta(g_j) - \theta(g_m))^2$ for $j \geq m$, j great enough, and then by $(\theta(\hat{g}_j) - \theta(\hat{g}_m))^2 = (\hat{\theta}_j - \hat{\theta}_m)^2$. This approximation in turn implies a bias which must be corrected. This explains why the theoretical criterion is

$$Crit(m) = \sup_{j \geq m}(\theta(g_j) - \theta(g_m))^2 + \text{pen}(m),$$

where $\text{pen}(m)$ has the order of the variance, and its empirical version is

$$\widehat{Crit}(m) = \sup_{j \geq m, j \in \mathcal{M}}[(\theta(\hat{g}_m) - \theta(\hat{g}_j))^2 - H(j,m)] + \text{pen}(m),$$

10

where $H(j, m)$ is an additional bias correction. We can then define

$$\hat{m} = \inf \left\{ m \in \mathcal{M}, \widehat{Crit}(m) \leq \inf_{j \in \mathcal{M}} \widehat{Crit}(j) + \frac{1}{n} \right\} \tag{17}$$

as the model selection procedure. It remains to find $\mathrm{pen}(\cdot)$ and $H(j, m)$ that make the procedure work and give good rates for $\hat{\theta}_{\hat{m}}$.

3.2   Model selection

First, note that model selection has an interest only in the case $\int |\psi^*/f_\varepsilon^*| = +\infty$ and $\int |\psi^*/f_\varepsilon^*|^2 = +\infty$ since otherwise the variance is of order $1/n$ and the rate is parametric. As ψ and f_ε are assumed to be known, these conditions can be explicitly checked.

Let $C_\varepsilon = \int |f_\varepsilon^*(x)| dx$. Let x_m, be some positive weights to be chosen, and let a > 0, we define:

$$\mathrm{pen}(m) = 4(1 + \frac{1}{\mathrm{a}})(x_m \sigma_m^2 + x_m^2 c_m^2) \tag{18}$$

where $\sigma_m^2 = \sigma_{0,m}^2$, $c_m = c_{0,m}$, with $\sigma_{j,m}^2$ and $c_{j,m}$ defined by

$$\sigma_{j,m}^2 = \frac{1}{2\pi n} \min \left\{ C_\varepsilon \int_{\pi(j \wedge m) \leq |x| \leq \pi(j \vee m)} \left| \frac{\psi^*(x)}{f_\varepsilon^*(x)} \right|^2 dx, \left(\int_{\pi(j \wedge m) \leq |x| \leq \pi(j \vee m)} \frac{|\psi^*(x)|}{|f_\varepsilon^*(x)|} dx \right)^2 \right\}$$

$$\text{and } c_{j,m} = \frac{1}{2\pi n} \int_{\pi(j \wedge m) \leq |x| \leq \pi(j \vee m)} \left| \frac{\psi^*(x)}{f_\varepsilon^*(x)} \right| dx.$$

Let also

$$H(j, m) = 4(1 + \frac{1}{\mathrm{a}})(x_j \sigma_{j,m}^2 + x_j^2 c_{j,m}^2). \tag{19}$$

We can prove the following Theorem:

Theorem 3.1 *Consider model (1) for $(X_i)_{1 \leq i \leq n}$ and $(\varepsilon_i)_{1 \leq i \leq n}$ independent sequences of i.i.d. random variables and assume that f_ε satisfies (5). Let $\hat{\theta}_{\hat{m}}$ be defined by (7) or (8) and (17)-(18)-(19) when $\int |\psi^*/f_\varepsilon^*| = +\infty$ and $\int |\psi^*/f_\varepsilon^*|^2 = +\infty$. Then there exists some positive constant $C(\mathrm{a})$ depending on a only, such that*

$$\mathbb{E}[(\hat{\theta}_{\hat{m}} - \theta)^2] \leq C(\mathrm{a}) \inf_{m \in \mathcal{M}} \left\{ \left(\int_{|x| \geq \pi m} |\psi^*(x)g^*(x)| dx \right)^2 + \mathrm{pen}(m) \right\} + C(\mathrm{a}) \sum_{m \in \mathcal{M}} e^{-x_m} \omega_m^2 + \frac{1}{n},$$

where $\omega_m^2 = \sigma_m^2 \vee c_m + 2(\sigma_m^2 \vee c_m)^2$.

Theorem 3.1 states that $\hat{\theta}_{\hat{m}}$ leads to an automatic tradeoff between the squared bias term $(\int_{|x| \geq \pi m} |\psi^*(x)g^*(x)| dx)^2$ and $\mathrm{pen}(m)$ if the residual $\sum_m e^{-x_m} \omega_m^2$ is negligible, that is $O(1/n)$. In other words, x_m is not free but chosen so that $\sum_m e^{-x_m} \omega_m^2 = O(1/n)$. In turn, as the main term in $\mathrm{pen}(m)$ is clearly $x_m \sigma_m^2$ and σ_m^2 is the variance of $\hat{\theta}_m$, x_m represents a loss in the variance (not necessarily in the rate).

Now, let us discuss the possible choices for the x_m's in order to see what loss occurs, if any, when using the adaptive procedure. The cases are discussed with respect to cases (II), (III) and (IV) of the variance which contain only known parameters, under the assumption that f_ε fulfills (2), ψ fulfills (4) and g belongs to the set defined by (3).

- Case (II). We take $x_m = 2\ln(m)$ and the rate become of order $(\ln\ln(n))^2/n$ instead of $\ln\ln(n)/n$ or of order $\ln^2(n)/n$ instead of $\ln(n)/n$.

- Case (III). We take $x_m = (2\beta - 2B + 3)\ln(n)$, and the rate becomes of order $\ln\ln(n)\ln^\delta(n)/n$ instead of $\ln^\delta(n)/n$ and of order $(n/\ln(n))^{-[(b+B)-1/2]/(b+\beta)}$ instead of $n^{-[(b+B)-1/2]/(b+\beta)}$.

- Case (IV). We take $x_m = 4\alpha(\pi m)^\rho$, and there is no loss in case with logarithmic rate and a loss of logarithmic order in case where the rate is such that powers of logarithms are negligible with respect to it.

Remark 3.1 Comte *et al.* [15] provide a model selection procedure for selecting an optimal m with respect to the $\mathbb{L}_2$ risk for the estimator $\hat{g}_m$; let $\tilde{g}$ be the resulting estimator. Then $\langle\psi, \tilde{g}\rangle$ is an estimator of $\theta(g)$ on a randomly selected space among the S_m's. The inequality

$$\mathbb{E}\left[(\langle\psi, \tilde{g}\rangle - \langle\psi, g\rangle)^2\right] \leq \|\psi\|^2 \mathbb{E}(\|\tilde{g} - g\|^2)$$

explains why the rate of this estimator does not benefit of the improvement brought by the known regularity of ψ and is therefore not optimal.

Moreover, if we want to extend the adaptive result to the mixing case, we can use the Bernstein inequality given in Doukhan [16] or in Butucea and Neumann [6], provided that the mixing is geometrical. We can prove the following Corollary of Theorem 3.1:

Corollary 3.1 *Consider model (1) under* (D1) *or under* (D2) *with f_ε satisfying (2) and ψ satisfying (4) and (15) with $\beta > \max(B, 1)$ or $A, \rho > 0$, and assume in both case that $\beta_k \leq e^{-ck}$ for any $k \in \mathbb{N}$. Then if f_ε satisfies (5), if $\int |\psi^*(t)|dt < +\infty$ and $\int |\psi^*/f_\varepsilon^*| = +\infty$, $\int |\psi^*/f_\varepsilon^*|^2 = +\infty$ then the result of Theorem 3.1 for $\hat{\theta}_{\hat{m}}$ defined in the same way, holds with c_m, $c_{j,m}$ replaced by $2c_m \ln(n)/c$, $2c_{j,m}\ln(n)/c$ and σ_m^2, $\sigma_{j,m}^2$ multiplied by 2.*

Clearly, the constant c appearing in the c_m's, $c_{m,j}$'s is unknown, but these terms have in general negligible orders when compared to the σ_m^2's, $\sigma_{j,m}^2$'s.

4 Applications

4.1 Pointwise estimation

For pointwise estimation of g, we can take $\psi(x) = \mathbf{I}_{\{x_0\}}(x)$ for any given x_0, which implies $\psi^*(t) = e^{itx_0}$, $|\psi^*(t)| = 1$. Therefore, the rates of convergence are the same as usual in pointwise deconvolution, as recalled in Table 2.

When $r > 0, \rho > 0$ the value of $\breve{m}$ is not explicitly given. It is obtained as the solution of the equation

$$\breve{m}^{2b+2\beta+(1-\rho)_+} \exp\{2\alpha(\pi\breve{m})^\rho + 2a(\pi\breve{m})^r\} = O(n). \tag{20}$$

Consequently, the rate of $\hat{g}_{\check{m}}$ is not easy to give explicitly and depends on the ratio r/ρ. If r/ρ or ρ/r belongs to $]k/(k+1); (k+1)/(k+2)]$ with k integer, the rate of convergence can be expressed as a function of k. For explicit formulae for the rates, see Lacour [22].

These rates are known to be optimal in the minimax sense as indicated in Table 2. The case $r = 0$ is done in Fan [17], the case $r > 0, \rho = 0$ in Butucea [5]. The rate in the case $r > 0$, $\rho > 0$, $\beta = 0$ is proven optimal in the minimax sense in Butucea and Tsybakov [7] for $r \leq \rho$ and by using their construction we get by following the same proof near optimality (within a log factor) in the case $r > \rho$.

For adaptive pointwise estimation, using $|\psi^*(x)| = 1$ again, we have $c_m \leq \sigma_m^2$ and $x_m^2 c_m^2 \leq C x_m \sigma_m^2$ for all the choices of x_m that will be found. Clearly, if f_ε is ordinary smooth, the choice $x_m = (2\beta+3)\ln(m)$ suits and if f_ε is supersmooth, we can choose $x_m = 4\alpha(\pi m)^\rho$. These choices coincide with the general case detailed above for $b = 0$. Then we have $\sum_{m\in\mathcal{M}} e^{-x_m}\omega_m^2 \leq C/n$. This implies that

$$\mathbb{E}[(\hat{\theta}_{\hat{m}} - \theta)^2] \leq C \inf_{m\in\mathcal{M}} \left(\left(\int_{\pi m}^{+\infty} |g^*| \right)^2 + \frac{x_m}{n} \min \left\{ \int_{-\pi m}^{\pi m} |f_\varepsilon^*|^{-2}, \left(\int_{-\pi m}^{\pi m} |f_\varepsilon^*|^{-1} \right)^2 \right\} \right) + \frac{C'}{n}.$$

Table 2: Choice of $\check{m}$ for pointwise deconvolution and corresponding rates under Assumptions (2) and (3). Adaptive rates for comparison. B_m is abbreviated for $m^{-2b+1-r}\exp(-2a(\pi m)^r)$ and V_m for $m^{2\beta+1-\rho+(1-\rho)_+}\exp(2\alpha(\pi m)^\rho)/n$.

| | f_ε | |
| | $\rho = 0$ | $\rho > 0$ |
	ordinary smooth	supersmooth
$r = 0$ Sob.(b)	$\pi\check{m} = n^{1/(2b+2\beta)}$ $\varphi_n^2 = O(n^{-(2b-1)/(2b+2\beta)})$ *minimax rate*	$\pi\check{m} = [\ln(n)/(2\alpha+1)]^{1/\rho}$ $\varphi_n^2 = O((\ln(n))^{-(2b-1)/\rho})$ *minimax rate*
	$\psi_n^2 = O((n/\ln(n))^{-(2b-1)/(2b+2\beta)})$ *adaptive rate*	$\psi_n^2 = O((\ln(n))^{-(2b-1)/\rho})$ *adaptive minimax rate (no loss)*
$r > 0$ C^∞	$\pi\check{m} = [\ln(n)/2b]^{1/r}$ $\varphi_n^2 = O\left(\dfrac{\ln(n)^{(2\beta+1)/r}}{n}\right)$ *minimax rate*	$\check{m}$ solution of (20) $= \ln(n) - (\ln\ln(n))^2$ $\varphi_n^2 = O(B_{\check{m}})$: *minimax rate if $r < \rho$ and $b = 0$* $\varphi_n^2 = O(V_{\check{m}})$: *minimax rate if $r \geq \rho$, $\rho \leq 1$ and $b = 0$*
	$\psi_n^2 = O\left(\dfrac{\ln\ln(n)\ln(n)^{(2\beta+1)/r}}{n}\right)$ *adaptive rate*	$\psi_n^2 = O\left(\check{m}^{\rho I(r\geq\rho)}\varphi_n^2\right)$ *adaptive minimax rate if $r < \rho$ and $b = 0$* *adaptive rate if $r \geq \rho$, $\rho \leq 1$ and $b = 0$*

The rates correspond to $B = A = R = 0$ in (4) and are the following (see also Table 2):

- Case $\rho = r = 0$, f_ε and g are ordinary smooth, $x_m = (2\beta+3)\ln(m) \leq (2\beta+3)\ln(n)$, choose m of order $(n/\ln(n))^{1/(2b+2\beta)}$, the rate of the adaptive estimator is $(n/\ln(n))^{-(2b-1)/(2b+2\beta)}$.

13

- Case $\rho = 0, r > 0, a > 0$, f_ε is ordinary smooth and g is super-smooth, $x_m = (2\beta+3)\ln(m)$, the optimal m is of order $(\ln(n)/2a)^{1/r}$ and the rate of the adaptive estimator is of order $\ln(\ln(n))[\ln(n)]^{(2\beta+1)/r}/n$, so that the loss is of order $\ln(\ln(n))$.

- Case $\rho > 0, \alpha > 0$ and $r = 0$, i.e. f_ε is super-smooth and g is ordinary smooth. Then x_m is of order m^ρ, the optimal πm is $(\ln(n)/(2\alpha+1))^{1/\rho}$ and the rate of the adaptive estimator is of order $[\ln(n)]^{-(2b-1)/\rho}$, i.e. there is no loss due to adaptation.

- Case $\rho > 0, \alpha > 0$ and $r > 0, a > 0$, i.e. both f_ε and g are super-smooth. Here x_m is of order m^ρ, there is no loss if $r < \rho$, a loss of order $[\ln(n)]^{\rho/r}$ for $r > \rho$ for a rate faster than any power of logarithm. If $r = \rho$ the loss is logarithmic and the rate polynomial.

Now, we want to prove that the losses which occur are optimal in the minimax sense.

The previously defined estimator $\hat{\theta}_{\check{m}}$ with $\check{m}$ defined in Table 2 is adaptive minimax in the cases: $(r = 0$ and $\rho > 0)$ and $(r > 0, \rho > 0$ and $r < \rho)$.

As we already noticed, estimators $\hat{\theta}_{\hat{m}}$ which are free of parameters may attain a slower rate of convergence ψ_n, i.e. it may happen that $\varphi_n = o(\psi_n)$. Therefore, we check that the loss, when it occurs, is unavoidable.

Theorem 4.1 *The rates ψ_n defined in Table 2 are adaptive rates and whenever a loss with respect to the minimax rate appears (compare in Table 2 φ_n^2 and ψ_n^2) it is optimal in the sense of Definition 1.3, under the additional hypothesis that the noise density is 3-times continuously differentiable and*

$$\text{for polynomial noise } |f_\varepsilon'(u)| \leq C\frac{1}{|u|^{\beta+1}}, \ \ as \ |u| \to \infty \tag{21}$$

$$\text{for exponential noise } |f_\varepsilon'(u)| \leq C|u|^{\rho-1}\exp(-\alpha|u|^\rho), \ \ as \ |u| \to \infty. \tag{22}$$

Moreover, when $r > 0$, $r \geq \rho$ and $0 < \rho \leq 1$ the rate φ_n^2 is minimax rate of estimation.

Remark 4.1 Note that the adaptive property of $\hat{\theta}_{\hat{m}}$ in the case $r \geq \rho$ is proved only for $\rho \leq 1$, which is a technical restriction. Nevertheless, it is worth noticing that, still under the restriction that $\rho \leq 1$, we obtain as a by-product in Theorem 4.1 the minimaxity of the rate for $r \geq \rho$. This is a new result since the latest result on the subject was proving minimaxity in the case $r < \rho$ only (see Butucea and Tsybakov [7]).

Proof of Theorem 4.1. We describe first the general procedure for proving the theorem and postpone details of constructions and proofs to Section 5.5. As the adaptation loss is different according to whether $r = 0$ or $r \neq 0$, $\rho = 0$ or $\rho \neq 0$, explicit constructions are needed for each of the following setups:

1. $r = 0, \rho = 0$;

2. $b = 0, r > 0, \rho = 0$;

3. $b = 0, r > 0, 0 < \rho \leq 1$ and $r \geq \rho$.

Classically, we take $b = 0$ without loss of generality.

Typically, we construct two probability densities $g_0 \in \mathcal{S}(\overline{\lambda})$ and $g_{1,n} \in \mathcal{S}(\underline{\lambda})$ where $\overline{\lambda}, \underline{\lambda} \in \Lambda$. Moreover

$$g_{1,n}(x) = g_0(x) + G(x - x_0, m), \text{ for } m = m_n \to \infty \text{ with } n \text{ and } \int G(\cdot, m) = 0, \forall m.$$

14

Note that the likelihoods of the model become $f_0^Z = g_0 \star f_\varepsilon$ under g_0 and

$$f_{1,n}^Z(x) = [g_{1,n} \star f_\varepsilon](x) = f_0^Z(x) + [G(\cdot, m) \star f_\varepsilon](x - x_0)$$

under $g_{1,n}$. Then

$$\inf_{\theta_n} \sup_{\lambda \in \Lambda} \sup_{g \in \mathcal{S}(\lambda)} \psi_{n,\lambda}^{-2} \mathbb{E}_g[|\theta_n - \theta(g)|^2]$$

$$\geq \inf_{\theta_n} \max \left\{ \psi_{n,\overline{\lambda}}^{-2} \mathbb{E}_{g_0}[|\theta_n - \theta(g_0)|^2], \ \psi_{n,\underline{\lambda}}^{-2} \mathbb{E}_{g_{1,n}}[|\theta_n - \theta(g_{1,n})|^2] \right\}$$

$$\geq \inf_{T_n} \max \left\{ q_n^2 \mathbb{E}_{g_0}[T_n^2], \mathbb{E}_{g_{1,n}}[|T_n - G(0,m)/\psi_{n,\underline{\lambda}}|^2] \right\},$$

where $q_n = \psi_{n,\underline{\lambda}}/\psi_{n,\overline{\lambda}} \to \infty$ when $n \to \infty$, with a proper choice of $\underline{\lambda}$, $\overline{\lambda}$ and $T_n = (\theta_n - \theta(g_0))/\psi_{n,\underline{\lambda}}$.

¿From now on we denote $\mathbb{P}_0 = \mathbb{P}_{g_0}$, $\mathbb{E}_0 = \mathbb{E}_{g_0}$ and $\mathbb{P}_1 = \mathbb{P}_{g_{1,n}}$, $\mathbb{E}_1 = \mathbb{E}_{g_{1,n}}$. Following Theorem 6 in Tsybakov [30] we can deduce that, if $|G(0,m)/\psi_{n,\underline{\lambda}}| \geq c > 0$ and if for some fixed $0 < \epsilon < 1$ and $\tau > 0$

$$\mathbb{P}_1 \left(\frac{d\mathbb{P}_0}{d\mathbb{P}_1} \geq \tau \right) \geq 1 - \epsilon \tag{23}$$

then

$$\inf_{T_n} \max \left\{ q_n^2 \mathbb{E}_0[T_n^2], \mathbb{E}_1[|T_n - G(0,m)/\psi_{n,\underline{\lambda}}|^2] \right\} \geq \frac{\tau q_n^2 \epsilon^2 c^4 (1-\epsilon)^2}{\tau q_n^2 \epsilon^2 c^2 + (1-\epsilon)^2 c^2}. \tag{24}$$

If we can choose $\tau = \tau_n$ such that $\tau_n q_n^2 \to \infty$ with n, then the bound from below in (24) tends to $c^2(1-\epsilon)^2$ so it will be larger than $c^2(1-\epsilon)^4 > 0$ for n large enough.

Note also that this Lemma may provide the exact asymptotic constant in case $c \to 1$ and $\mathbb{P}_1(d\mathbb{P}_0/d\mathbb{P}_1 \geq \tau_n) \to 1$ as $n \to \infty$.

In order to deal with (23), we proceed as follows:

$$\mathbb{P}_1 \left(\frac{d\mathbb{P}_0}{d\mathbb{P}_1} \geq \tau \right) = \mathbb{P}_1 \left(\prod_{i=1}^n \frac{g_0 \star f_\varepsilon}{g_{1,n} \star f_\varepsilon}(Y_i) \geq \tau \right)$$

$$= \mathbb{P}_1 \left(\sum_{i=1}^n \ln \left(1 - \frac{G(\cdot - x_0) \star f_\varepsilon}{g_{1,n} \star f_\varepsilon}(Y_i) \right) \geq \ln(\tau) \right)$$

$$= \mathbb{P}_1 \left(\frac{\sum_{i=1}^n Z_{i,n} - n \mathbb{E}_1(Z_{1,n})}{(n \mathrm{Var}_1(Z_{1,n}))^{1/2}} \geq \frac{\ln(\tau) - n \mathbb{E}_1(Z_{1,n})}{(n \mathrm{Var}_1(Z_{1,n}))^{1/2}} \right),$$

where $Z_{i,n} = \ln(1 - [G(\cdot - x_0) \star f_\varepsilon](Y_i)/g_{1,n} \star f_\varepsilon(Y_i))$ form a triangular array of independent variables. Denote

$$U_n := \frac{\sum_{i=1}^n Z_{i,n} - n E_1(Z_{1,n})}{(n Var_1(Z_{1,n}))^{1/2}}.$$

We shall prove, for each setup, Lyapounov's central limit theorem for U_n. Moreover, we give an upper bound $\mathbb{E}_1(Z_{1,n}) \geq -c_e \kappa_n$ and a lower bound for $\mathrm{Var}_1(Z_{1,n}) \leq c_v \kappa_n$, where κ_n is such that

$$\chi^2(g_0 \star f_\varepsilon, g_{1,n} \star f_\varepsilon) := \int \frac{(g_{1,n} \star f_\varepsilon - g_0 \star f_\varepsilon)^2}{g_{1,n} \star f_\varepsilon} \leq \kappa_n$$

15

as $n \to \infty$. Choose then $\tau_n \to 0$ such that

$$u_n := \frac{\ln(\tau_n) + c_e n \kappa_n}{(c_v n \kappa_n)^{1/2}} \to -\infty$$

with n, giving that $\mathbb{P}_1(U_n \geq u_n) \geq 1 - \epsilon$, for some $0 < \epsilon < 1$ and large enough n and thus concluding the proof of the Theorem. $\square$

4.2 Pointwise Laplace Transform estimation

Let us denote for any positive real number λ the Laplace Transform of a function g by

$$Lg(\lambda) = \int_{\mathbb{R}} e^{-\lambda x} g(x) dx.$$

In other words, $Lg(\lambda) = \mathbb{E}(e^{-\lambda X_1}) = \langle \psi_\lambda, g \rangle$ with $\psi_\lambda(x) = e^{-\lambda x}$, for any $\lambda > 0$.

If X_1 is a nonnegative random variable, then its density g is a $\mathbb{R}^+$-supported density which admits a finite Laplace Transform. In that case, we can write $Lg(\lambda) = \langle g, \psi_\lambda \rangle$ with $\psi_\lambda(x) = e^{-\lambda x} \mathbb{I}_{\{x>0\}}$, and

$$\psi_\lambda^*(x) = \int_0^{+\infty} e^{ixu} e^{-\lambda u} du = \frac{1}{\lambda - ix}, \quad |\psi_\lambda^*(x)|^2 = \frac{1}{\lambda^2 + x^2}. \tag{25}$$

On the other hand, the noise ε is not necessarily positive random variable. If ε_1 also admits a Laplace Transform, then so does Z_1 and the Laplace Transform of X_1 can be estimated by using the empirical version of the relation $Lf_Z(\lambda) = Lg(\lambda) Lf_\varepsilon(\lambda)$. Thus, by setting

$$\widehat{Lg}(\lambda) = \left(\frac{1}{n} \sum_{k=1}^{n} e^{-\lambda Z_k} \right) / Lf_\varepsilon(\lambda),$$

we get an unbiased estimate of Lg with quadratic risk of order $1/n$.

Now, if ε *does not admit a Laplace Transform* (e.g. for $f_\varepsilon(x) = 1/[\pi(1 + x^2)]$, $\mathbb{E}(e^{-\lambda \varepsilon_1}) = +\infty$), the method developed in this paper still allows a pointwise estimation of Lg. We can define

$$\widehat{Lg}_m(\lambda) = \frac{1}{2\pi n} \sum_{k=1}^{n} \int_{|t| \leq \pi m} e^{it Z_k} \frac{\psi_\lambda^*(t)}{f_\varepsilon^*(t)} dt, \tag{26}$$

with ψ_λ^* given by (25). Then we know that $\widehat{Lg}_m(\lambda)$ is a consistent estimator of $Lg(\lambda)$, provided that m is well chosen:

Proposition 4.1 *Let X be a positive random variable, with a Laplace transform denoted by Lg. For all $\lambda > 0$, the estimate of Lg defined by (26) is such that*

$$\mathbb{E}\left[\widehat{Lg}_m(\lambda) - Lg(\lambda) \right]^2 \leq \frac{\left(\int_{|t| \geq \pi m} |g^*(t)| dt \right)^2}{4\pi^4 m^2}$$

$$+ \frac{1}{4\pi^2 n} \min \left\{ \int |f_\varepsilon^*| \int_{-\pi m}^{\pi m} \frac{dt}{(\lambda^2 + t^2)|f_\varepsilon^*(t)|^2}, \left(\int_{-\pi m}^{\pi m} \frac{dt}{\sqrt{\lambda^2 + t^2} |f_\varepsilon^*(t)|} \right)^2 \right\}.$$

16

Moreover, the adaptive procedure works for automatic selection of m. The rates are easily computed by changing β into $\beta - 1$ (for $b > 1$) and b into $b + 1$ in Table 2 or by setting $B = 1$ in Table 1. We have

Proposition 4.2 *Let X be a nonnegative random variable with Laplace transform denoted by Lg and estimated by $\widehat{Lg}_m$ given by (26). Let $\hat{m}$ be defined by (17), pen(m) by (18) and $H(j,m)$ by (19), with $|\psi^*(x)| = |\psi_\lambda^*(x)| = 1/\sqrt{\lambda^2 + x^2}$. Then, for all $\lambda > 0$,*

$$\mathbb{E}\left[\widehat{Lg}_{\hat{m}}(\lambda) - Lg(\lambda)\right]^2 \leq C(\mathrm{a}) \inf_{m \in \mathcal{M}_n} \left(\frac{\left(\int_{|t| \geq \pi m} |g^*(t)|dt\right)^2}{4\pi^4 m^2} + \mathrm{pen}(m) \right) + \frac{C(\mathrm{a})}{n}.$$

In the same way, we can estimate the symmetrized version of the Laplace Transform namely $SLg(\lambda) = \mathbb{E}(e^{-\lambda|X_1|})$. In that case, $\psi_\lambda(x) = e^{-\lambda|x|}$ and $\psi^*(x) = \frac{2\lambda}{\lambda^2+x^2}$. The rates are obtained by changing β into $\beta - 2$ (for $b > 2$) and b into $b + 2$ in Tables 2 or by setting $B = 2$ in Table 1.

4.3 Stochastic volatility model

Let us consider the discrete time stochastic volatility model:

$$U_i = \sqrt{V_i}\eta_i, \; i = 1, \ldots, n, \tag{27}$$

where η_i is an i.i.d. centered noise process while V_i is a volatility process of interest. Moreover, (V_i) and (η_i) are independent and (V_i) is a stationary β-mixing process with β-mixing coefficients denoted by (β_k). When this model is obtained as the discretization of a set of continuous time stochastic differential equations, V_i is in fact an integrated volatility process, it is geometrically β-mixing, and η_i follows a $\mathcal{N}(0,1)$ distribution, see Comte and Genon-Catalot [14].

Now, Model (27) is considered in this form by van Es *et al.* [31] among others, under the additional assumption $\eta_i \sim \mathcal{N}(0,1)$. Setting

$$Z_i = \ln(U_i^2), \;\; X_i = \ln(V_i) \;\; \text{and} \; \varepsilon_i = \ln(\eta_i^2)$$

allows to recover model (1). Then, we note that if $\eta_1 \sim \mathcal{N}(0,1)$, then

$$f_\varepsilon^*(x) = \frac{2^{ix}}{\sqrt{\pi}}\Gamma(1 + ix), \;\; \text{and} \; |f_\varepsilon^*(x)| \sim_{|x| \to +\infty} \sqrt{2/e}e^{-\pi|x|/2}, \tag{28}$$

by using that $\Gamma(z) \sim_{|z| \to +\infty} \sqrt{2\pi}z^{z-1/2}e^{-z}$, see Abramowitz and Stegun [1].

Applying the results of Section 4.1 in the mixing context (D1) (see Proposition 2.4 and Corollary 3.1), we deduce that, if V is geometrically β-mixing, we have a pointwise estimator of g,

$$\hat{g}_m(x) = \frac{1}{2\pi n} \int_{|t| \leq \pi m} \frac{e^{it(x+Z_k)}}{f_\varepsilon^*(t)}dt$$

for which we can propose an automatic selection of m which reaches the adaptive or adaptive minimax rate. The resulting rate is of order a negative power of $\ln(n)$ if g is in a Sobolev space but it is much faster if g is supersmooth (a case which is easy to meet, see Comte and

17

Genon-Catalot [14]). Therefore, we recover as a particular case, and substantially improve the result of van Es *et al.* [31], who propose a non adaptive kernel estimator of g, assuming that g is known to be twice continuously differentiable.

Now, extensions of the class of discrete time stochastic volatility models have been studied (see Genon-Catalot and Kessler [18], or Chaleyat-Maurel and Genon-Catalot [12]) and in particular, it is natural to consider more general types of distributions for η. For instance, we suppose now that η^2 follows a Gamma distribution, i.e. that $f_{\eta_1^2}(x) = (e^{-x}x^{p-1}/\Gamma(p))\mathbf{I}_{x>0}$. In that case, we find

$$f_\varepsilon^*(x) = \frac{\Gamma(ix+p)}{\Gamma(p)}, \text{ and } |f_\varepsilon^*(x)| \sim_{|x|\to+\infty} \frac{\sqrt{2\pi}e^{-p}}{\Gamma(p)}|x|^{p-1/2}e^{-\pi|x|/2}, \tag{29}$$

that is ε is super-smooth with $\beta = p - 1/2$, $\alpha = \pi/2$ and $\rho = 1$. The Gaussian case corresponds to $p = 1/2$. In this context, let π denotes the density of V_1, and consider that we are interested in estimating its Laplace transform. In fact, our general method provides an estimator of $h(\lambda) = -(L\pi)'(\lambda) = \mathbb{E}(V_1 e^{-\lambda V_1})$, the opposite of the derivative of the Laplace Transform of π. In other words, we can estimate $h(\lambda) = \langle \psi_\lambda, g \rangle = \mathbb{E}(V_1 e^{-\lambda V_1}) = \mathbb{E}(e^{X_1 - \lambda e^{X_1}})$. Actually we have, for $\lambda > 0$,

$$h(\lambda) = \langle \psi_\lambda, g \rangle, \text{ with } \psi_\lambda(x) = e^{x - \lambda e^x},$$

and

$$\psi_\lambda^*(x) = \lambda^{-1-ix}\Gamma(1+ix) \sim_{|x|\to+\infty} \frac{\sqrt{2\pi}}{e\lambda}\sqrt{|x|}e^{-\pi|x|/2}, \tag{30}$$

(i.e. $B = 1/2$, $A = \pi/2$ and $R = 1$). Let us define

$$\hat{h}_m(\lambda) = \frac{1}{2\pi n}\sum_{k=1}^n \int_{|t|\leq \pi m} e^{itZ_k}\frac{\psi_\lambda^*(t)}{f_\varepsilon^*(t)}dt \tag{31}$$

with f_ε^* and ψ_λ^* given by (29) and (30). Then, taking into account the orders of f_ε^* and ψ_λ^*, we obtain, by applying Inequality (13) of Proposition 2.4 and if $p \neq 3/2$:

$$\mathbb{E}[(\hat{h}_m(\lambda) - h(\lambda))^2] \leq Kme^{-\pi^2 m} + \frac{K'm^{(3-2p)\vee 0}}{n} + \frac{K''\sum_{k\geq 0}\beta_k}{n},$$

where K, K' and K'' are positive constants, $K'' = 2(\int |\psi^*|)^2$. If $p = 3/2$ the variance term has order $\ln(m)/n$. Then, as (D1) is satisfied in our model, we get

Proposition 4.3 *Consider model (27) with (D1), (29) and (30). Assume that $(X_k) = (\ln(V_k))$ is β-mixing with $\sum_k \beta_k < +\infty$, then $\hat{h}_m$ defined by (31) satisfies, for $\lambda > 0$,*

$$\mathbb{E}[(\hat{h}_m(\lambda) - h(\lambda))^2] \leq Kme^{-\pi^2 m} + \frac{K'(m^{(3-2p)\vee 0}\mathbf{I}_{p\neq 3/2} + \ln(m)\mathbf{I}_{p=3/2})}{n} + \frac{K''\sum_k \beta_k}{n},$$

where K, K' and K'' are positive constants.

In other words, applying the orders detailed in Table 1 to the model (27), we obtain a rate of order $[\ln(n)]^{(3-2p)\vee 1}/n$ (i.e. always less than $\ln^3(n)/n$), whatever the smoothness of g.

No adaptation is required if $p > 3/2$. If $p \leq 3/2$, the risk of the adaptive estimator is obtained by applying Corollary 3.1 and by choosing $x_m = 4\ln(m)$:

18

Proposition 4.4 *Consider the stochastic volatility model (27) with (D1), (29) and (30). Assume that (X_i) is geometrically β-mixing and consider $\hat{h}_m$ defined by (31), with $\hat{m}$ defined by (17). For any $\lambda > 0$, and $p \leq 3/2$*

$$
\mathbb{E}[(\hat{h}_{\hat{m}}(\lambda) - h(\lambda))^2] \leq K \inf_{m \in \mathcal{M}} \left[\left(\int_{|u| \geq \pi m} |g^*(u)\psi_\lambda^*(u)| du \right)^2 \right.
$$
$$
\left. + \frac{(m^{3-2p}\mathbb{I}_{p<3/2} + \ln(m)\mathbb{I}_{p=3/2}) \ln(m)}{n} \right] + K' \frac{\ln(n)}{n}.
$$

This corresponds to the case where a loss of order $\ln(\ln(n))$ occurs with respect to the non-adaptive rate.

Remark 4.2 The Gaussian case, for $p = 1/2$ is not especially studied here because another strategy is available then. Indeed for $\eta \sim \mathcal{N}(0, 1)$,

$$
\mathbb{E}(e^{i\sqrt{2\lambda}U_1}) = \mathbb{E}\left[\mathbb{E}\left(e^{i\sqrt{2\lambda V_1}\eta_1} | V_1 \right) \right] = \mathbb{E}\left(e^{-\lambda V_1} \right).
$$

Therefore the Laplace transform of π, $L\pi(\lambda)$ can be directly estimated by an empirical mean of the $\exp(i\sqrt{2\lambda}U_k)$'s, which is an unbiased estimator reaching the parametric rate $1/n$. The rate would be the same for estimating h, as by differentiating,

$$
h(\lambda) = \mathbb{E}(V_1 e^{-\lambda V_1}) = (-i/\sqrt{2\lambda})\mathbb{E}(U_1 e^{i\sqrt{2\lambda}U_1}).
$$

The method above reaches for $p = 1/2$, the rate $\ln^w(n) \ln(\ln(n))/n$, where $1 \leq w \leq 2$. Therefore, it is not optimal for any p.

4.4 ARCH models

General ARCH models can be formulated as follows. Let (η_i) be an i.i.d. noise sequence.

$$
Y_i = \sigma_i \eta_i \text{ with } \sigma_i = F(\eta_{i-1}, \eta_{i-2}, \dots), \tag{32}
$$

for some measurable functions F, or

$$
Y_i = \sigma_i \eta_i \text{ with } \sigma_i = F(\sigma_{i-1}, \eta_{i-1}) \text{ and } \sigma_0 \text{ independent of } (\eta_i)_{i \geq 0}. \tag{33}
$$

Many examples can be found in the literature, and conditions can be given under which the process $(Y_i, \sigma_i)_{i \in \mathbb{Z}}$ is geometrically β-mixing, we refer to Comte *et al.* [13] for a review of the examples and to the references therein. Clearly then, $Z_i = \ln(Y_i^2)$, $X_i = \ln(\sigma_i^2)$ and $\varepsilon_i = \ln(\eta_i^2)$ follow Model (1) and satisfy conditions given by (D2).

Therefore, taking $\psi(t) = \mathbb{I}_{\{x_0\}}(t)$ for any x_0, as in Section 4.1, allows to provide a pointwise density estimator and to recover the results obtained by the kernel estimator of van Es *et al.* [31]. Our results are more general since van Es *et al.* [31] only consider Gaussian noise η_t (implying super-smooth ε_i's, see Section 4.3), and do not study adaptation (which is not useful in their particular case).

Other functionals $\langle \psi, g \rangle$ may be estimated with our procedure.

5 Proofs

5.1 Proof of Theorem 3.1.

We insert here general weights $x_{j,m}$ such that

$$H(j,m) = 4(1 + \frac{1}{a})(x_{j,m}\sigma_{j,m}^2 + x_{j,m}^2 c_{j,m}^2).$$

We define

$$\Gamma(m) = [\theta(g_m) - \theta(g)]^2 + \sigma_m^2 + \sup_{j \leq m} x_{j,m}\sigma_{j,m}^2.$$

and

$$m_{opt} = \inf\left\{m \in \mathcal{M}, / \, Crit(m) \leq \inf_{l \in \mathcal{M}} Crit(l) + \frac{1}{n}\right\}.$$

Then we prove the following Theorem:

Theorem 5.1 *There exists some positive constant $C(a)$ depending on a only, such that*

$$\mathbb{E}[(\hat{\theta}_{\hat{m}} - \theta)^2] \leq C(a)(Crit(m_{opt}) + \Gamma(m_{opt}))$$

$$+ C(a)\left(\sum_{m \in \mathcal{M}} e^{-x_m}\omega_m^2 + \sum_{j \geq m_{opt}} e^{-x_{j,m_{opt}}}\omega_{j,m}^2 + \frac{1}{n}\right),$$

where $\omega_m^2 = \sigma_m^2 \vee c_m + 2(\sigma_m^2 \vee c_m)^2$ and $\omega_{j,m}^2 = \sigma_{j,m_{opt}}^2 \vee c_{j,m} + 2(\sigma_{j,m_{opt}}^2 \vee c_{j,m})^2$.

First, note that Theorem 5.1 implies the result. Indeed we observe that for $j \geq m$, $\sigma_{m,j}^2 \leq \sigma_j^2$ and $c_{m,j} \leq c_j$, so that choosing $x_{m,j} = x_j$ implies that

$$\sum_{j \geq m_{opt}} e^{-x_{j,m_{opt}}}\omega_{j,m}^2 \leq \sum_{m \in \mathcal{M}} e^{-x_m}\omega_m^2.$$

Moreover $Crit(m) \leq \left(\int_{|x| \geq \pi m} |\psi^*(x)g^*(x)|dx\right)^2 + pen(m)$ and $\Gamma(m) \leq \left(\int_{|x| \geq \pi m} |\psi^*(x)g^*(x)|dx\right)^2 + 2pen(m)$. This implies Theorem 3.1.

Now we establish the following Lemma.

Lemma 5.1 *For all $m \in \{1, \ldots, m_n\} := \mathcal{M}$, for all $x > 0$,*

$$\mathbb{P}\left(\widehat{Crit}(m) > (1+a)Crit(m) + 4(1+\frac{1}{a})(x+x^2)\right) \leq \sum_{j \geq m, j \in \mathcal{M}} e^{-x_{j,m}}e^{-x/(\sigma_{j,m}^2 \vee c_{j,m})}.$$

Proof of Lemma 5.1. We use Bernstein inequality which, for i.i.d. Y_k's such that $var(Y_1) \leq v^2$, $\|Y_1\|_\infty \leq 1/a$, gives, for $S_n = \sum_{k=1}^n Y_k$

$$\mathbb{P}\left(\frac{S_n - \mathbb{E}(S_n)}{n} \geq \sqrt{\frac{2uv^2}{n}} + \frac{u}{an}\right) \leq \exp(-u).$$

We take, for $j \geq m$,

$$Y_k = Y_k(j, m) = \frac{1}{2\pi} \int_{\pi m \leq |t| \leq \pi j} e^{itZ_k} \frac{\psi^*(t)}{f_\varepsilon^*(t)} dt. \tag{34}$$

Then $S_n/n = \hat{\theta}_j - \hat{\theta}_m$ and $\mathbb{E}(S_n/n) = \mathbb{E}(\hat{\theta}_j - \hat{\theta}_m) = \theta(g_j) - \theta(g_m)$. Moreover, we obtain that $v^2/n \leq \sigma_{j,m}^2$ and $1/(\mathrm{a}n) = c_{j,m}$. It follows that

$$\mathbb{P}\left\{ [(\hat{\theta}_j - \hat{\theta}_m) - (\theta(g_j) - \theta(g_m))]^2 \geq \left(\sigma_{j,m}\sqrt{2u} + c_{j,m}u \right)^2 \right\} \leq 2e^{-u}.$$

Now, we use that $(A + B)^2 \leq 2(A^2 + B^2)$ and that $(x + y)^2 \leq (1 + 1/\mathrm{a})x^2 + (1 + \mathrm{a})y^2$ gives, by setting $u = y$ and $v = x + y$, that $(v - u)^2 \geq (1/(1 + 1/\mathrm{a}))v^2 - (1 + \mathrm{a})/(1 + 1/\mathrm{a})u^2$. We obtain

$$\mathbb{P}\{(\hat{\theta}_j - \hat{\theta}_m)^2 \geq (1 + \mathrm{a})(\theta_j - \theta_m)^2 + 2(1 + 1/\mathrm{a})(2\sigma_{j,m}^2 u + c_{j,m}^2 u^2)\} \leq 2e^{-u}.$$

Now we set $u = x_{j,m} + x/(\sigma_{j,m}^2 \vee c_{j,m})$ and we find

$$\mathbb{P}\{(\hat{\theta}_j - \hat{\theta}_m)^2 - H(j, m) \geq (1 + \mathrm{a})(\theta_j - \theta_m)^2 + 4(1 + \frac{1}{\mathrm{a}})(x + x^2)\} \leq 2e^{-x_{j,m}} e^{-x/(\sigma_{j,m}^2 \vee c_{j,m})}.$$

To conclude we write

$$\mathbb{P}\left(\widehat{Crit}(m) < (1 + \mathrm{a})Crit(m) + 4(1 + \frac{1}{\mathrm{a}})(x + x^2) \right)$$

$$\leq \mathbb{P}\left\{ \exists j \geq m, j \in \mathcal{M}, (\hat{\theta}_j - \hat{\theta}_m)^2 - H(j, m) \geq (1 + \mathrm{a})(\theta_j - \theta_m)^2 + 4(1 + \frac{1}{\mathrm{a}})(x + x^2) \right\}$$

$$\leq 2 \sum_{j \geq m, j \in \mathcal{M}} e^{-x_{j,m}} e^{-x/(\sigma_{j,m}^2 \vee c_{j,m})}.$$

This ends the proof of Lemma 5.1. $\square$

Now we follow the steps of the proof of Laurent *et al.* [23].

- We first consider the case where $\hat{m} \leq m_{opt}$. The proof is exactly the same and we obtain

$$\mathbb{P}\left(\frac{1}{2}(\hat{\theta}_{\hat{m}} - \theta(g))^2 > (1 + \mathrm{a})Crit(m_{opt}) + 4\left(1 + \frac{1}{\mathrm{a}}\right)(x + x^2) + \sup_{j \leq m_{opt}} H(m_{opt}, j) \right.$$

$$\left. + (\theta(\hat{g}_{m_{opt}}) - \theta(g))^2 + \frac{1}{n} \cap \{\hat{m} \leq m_{opt}\} \right) \leq \sum_{j \geq m_{opt}} e^{-x_{j,m_{opt}}} e^{-x/(\sigma^2 j, m_{opt} \vee c_{j,m_{opt}})}. \tag{35}$$

- Now we consider the case $\hat{m} > m_{opt}$. We apply Bernstein Inequality to

$$\tilde{Y}_k = \tilde{Y}_k(m) = \frac{1}{2\pi} \int_{|t| \leq \pi m} e^{itZ_k} \frac{\psi^*(t)}{f_\varepsilon^*(t)} dt,$$

in the same way as in Lemma 1. We obtain, for all $m \in \mathcal{M}$,

$$\mathbb{P}\left((\theta(\hat{g}_m) - \theta(g))^2 \geq (1 + \mathrm{a})(\theta(g_m) - \theta(g))^2 + 4(1 + \frac{1}{\mathrm{a}})(x + x^2) + \mathrm{pen}(m) \right) \leq 2e^{-x_m} e^{-x/(\sigma_m^2 \vee c_m)}.$$

This implies that

$$\mathbb{P}\left((\theta(\hat{g}_{\hat{m}}) - \theta(g))^2 \ge (1+\mathrm{a})(\theta(g_{\hat{m}}) - \theta(g))^2 + 4(1+\frac{1}{\mathrm{a}})(x+x^2) + \mathrm{pen}(\hat{m}) \right)$$
$$\le \sum_{m\in\mathcal{M}} 2e^{-x_m} e^{-x/(\sigma_m^2 \vee c_m)}.$$

As $\sup_{j\ge m}[(\hat{\theta}_m - \hat{\theta}_j)^2 - H(j,m)] \ge (\hat{\theta}_m - \hat{\theta}_m)^2 - H(m,m) = 0$, we have $\widehat{Crit}(m) \ge \mathrm{pen}(m)$. Using the inequalities, $\mathrm{pen}(m) \le \widehat{Crit}(\hat{m}) \le \widehat{Crit}(m_{opt}) + 1/n$, we obtain

$$\mathbb{P}\left((\theta(\hat{g}_{\hat{m}}) - \theta(g))^2 \ge (1+\mathrm{a})(\theta(g_{\hat{m}}) - \theta(g))^2 + 4(1+\frac{1}{\mathrm{a}})(x+x^2) + \widehat{Crit}(m_{opt}) + \frac{1}{n} \right)$$
$$\le \sum_{m\in\mathcal{M}} 2e^{-x_m} e^{-x/(\sigma_m^2 \vee c_m)}.$$

If $\hat{m} > m_{opt}$, then $(\hat{\theta}_m - \theta(g))^2 \le \sup_{j\ge m_{opt}}(\theta_j - \theta(g))^2$ and we apply Lemma 1 with $m = m_{opt}$. This yields

$$\mathbb{P}\left((\theta(\hat{g}_{\hat{m}}) - \theta(g))^2 \ge (1+\mathrm{a})(\sup_{j\ge m_{opt}} (\theta(g_j) - \theta(g))^2 + 4(1+\frac{1}{\mathrm{a}})(x+x^2) \right.$$
$$\left. +(1+\mathrm{a})Crit(m_{opt}) + \frac{1}{n} \cap \{\hat{m} > m_{opt}\} \right)$$
$$\le \sum_{m\in\mathcal{M}} 2e^{-x_m} e^{-x/(\sigma_m^2 \vee c_m)} + \sum_{j\ge m_{opt}} 2e^{-x_{j,m_{opt}}} e^{-x/(\sigma_{j,m_{opt}}^2 \vee c_{j,m_{opt}})}. \tag{36}$$

Let

$$C_{m_{opt}} = 3(1+\mathrm{a})Crit(m_{opt}) + 2\sup_{j\le m_{opt}} H(m_{opt}, j) + (1+\mathrm{a})\sup_{j\ge m_{opt}} (\theta_j - \theta(g))^2 + \frac{3}{n}$$

and $X = (\hat{\theta}_{\hat{m}} - \theta(g))^2$, $Y = 2(\hat{\theta}_{m_{opt}} - \theta(g))^2$. It follows from (35) and (36) that, for all $x > 0$,

$$\mathbb{P}\left(X - Y > C_{m_{opt}} + 24(1+\frac{1}{\mathrm{a}})(x\vee x^2) \right) \le \sum_{m\in\mathcal{M}} 2e^{-x_m} e^{-x/(\sigma_m^2 \vee c_m)}$$
$$+ \sum_{j\ge m_{opt}} 2e^{-x_{j,m_{opt}}} e^{-x/(\sigma_{j,m_{opt}}^2 \vee c_{j,m_{opt}})}.$$

We write that $\mathbb{E}(X) = \mathbb{E}(X\mathbf{I}_{X\ge Y+C_{m_{opt}}}) + \mathbb{E}(X\mathbf{I}_{X\le Y+C_{m_{opt}}}) \le \mathbb{E}[(X - Y - C_{m_{opt}})_+] + \mathbb{E}(Y + C_{m_{opt}})$. Then, setting $C_{\mathrm{a}} = 24(1+1/\mathrm{a})$ and $Z = X - Y - C_{m_{opt}}$

$$\mathbb{E}[Z_+] = \int_0^{+\infty} \mathbb{P}(Z > t)dt = C_{\mathrm{a}}\left(\int_0^1 \mathbb{P}(Z > C_{\mathrm{a}}u)du + \int_1^\infty \mathbb{P}(Z > C_{\mathrm{a}}u)du \right)$$
$$= C_{\mathrm{a}}\left(\int_0^1 \mathbb{P}(Z > C_{\mathrm{a}}(u\vee u^2))du + 2\int_1^\infty \mathbb{P}(Z > C_{\mathrm{a}}v^2)vdv \right)$$
$$= C_{\mathrm{a}}\left(\int_0^1 \mathbb{P}(Z > C_{\mathrm{a}}(u\vee u^2))du + 2\int_1^\infty \mathbb{P}(Z > C_{\mathrm{a}}(v\vee v^2))vdv \right)$$

22

$$\mathbb{E}[(X - Y - C_{m_{opt}})_+] \leq C_{\mathrm{a}} \sum_{m \in \mathcal{M}} 2e^{-x_m}(\sigma_m^2 \vee c_m + 2(\sigma_m^2 \vee c_m)^2)$$

$$+ C_{\mathrm{a}} \sum_{j \geq m_{opt}} 2e^{-x_{j,m_{opt}}}(\sigma_{j,m_{opt}}^2 \vee c_{j,m_{opt}} + 2(\sigma_{j,m_{opt}}^2 \vee c_{j,m_{opt}})^2)$$

$$= C_{\mathrm{a}} \left(\sum_{m \in \mathcal{M}} 2e^{-x_m} \omega_m^2 + \sum_{j \geq m_{opt}} 2e^{-x_{j,m_{opt}}} \omega_{j,m_{opt}}^2 \right).$$

The end of the proof is the same as in Laurent et al. [23]. $\square$

5.2 Proof of Proposition 2.4.

The decomposition of the risk is the same and the bound for the bias also. Only the variance has to be re-examined. The basic idea is that, for $k \neq \ell$,

$$\mathrm{cov}(e^{itZ_k}, e^{isZ_\ell}) = f_\varepsilon^*(t) f_\varepsilon^*(-s) \mathrm{cov}(e^{itX_k}, e^{isX_\ell})$$

by conditioning on (X_k, X_ℓ). The additional trick is the standard covariance inequality for β-mixing variables (see e.g. Doukhan [16]) which implies that

$$|\mathrm{cov}(e^{itX_k}, e^{isX_\ell})| \leq \beta_{|k-\ell|}.$$

$$
\begin{aligned}
\mathrm{Var}(\breve{\theta}_m) &= \frac{1}{4\pi^2 n^2} \mathrm{Var}\left(\sum_{k=1}^n \int_{|t| \leq \pi m} e^{itZ_k} \frac{\psi^*(t)}{f_\varepsilon^*(t)} dt \right) \\
&= \frac{1}{4\pi^2 n^2} \sum_{k=1}^n \sum_{\ell=1}^n \iint_{|t| \leq \pi m, |s| \leq \pi m} \mathrm{cov}(e^{itZ_k}, e^{isZ_\ell}) \frac{\psi^*(t)}{f_\varepsilon^*(t)} \frac{\psi^*(-s)}{f_\varepsilon^*(-s)} ds dt \\
&= \frac{1}{4\pi^2 n^2} \sum_{k,\ell=1, k \neq \ell}^n \iint_{|t| \leq \pi m, |s| \leq \pi m} \mathrm{cov}(e^{itX_k}, e^{isX_\ell}) \psi^*(t) \psi^*(-s) ds dt \\
&\quad + \frac{1}{4\pi^2 n^2} \sum_{k=1}^n \iint_{|t| \leq \pi m, |s| \leq \pi m} \mathrm{cov}(e^{itZ_k}, e^{isZ_k}) \frac{\psi^*(t)}{f_\varepsilon^*(t)} \frac{\psi^*(-s)}{f_\varepsilon^*(-s)} ds dt. \quad (37)
\end{aligned}
$$

The last term is the standard variance term of the independent case. The first one is bounded in modulus by

$$\frac{2}{4\pi^2 n^2} \sum_{1 \leq k < \ell \leq n}^n \iint_{\substack{|t| \leq \pi m, \\ |s| \leq \pi m}} |\mathrm{cov}(e^{itX_1}, e^{isX_{\ell-k}})| |\psi^*(t)\psi^*(-s)| ds dt$$

$$\leq \frac{1}{2\pi n} \sum_{k=1}^n \beta_k \left(\int_{|t| \leq \pi m} |\psi^*(t)| dt \right)^2.$$

This gives the result. $\square$

5.3 Proof of Proposition 2.5.

Under (D2), we only obtain tha for $k < \ell$,

$$\mathrm{cov}(e^{itZ_k}, e^{isZ_\ell}) = f_\varepsilon^*(-s)\mathrm{cov}(e^{itZ_k}, e^{isX_\ell})$$

by conditioning on (X_ℓ). The covariance inequality for β-mixing variables (see e.g. Doukhan [16]) still applies (but to the variables (X_k, Z_k) and (X_ℓ, Z_ℓ) and implies that

$$|\mathrm{cov}(e^{itZ_k}, e^{isX_\ell})| \le \beta_{|k-\ell|}.$$

Then (37) remains true but leads for the bound of the modulus of the last term, to:

$$\frac{2}{4\pi^2 n^2} \sum_{1 \le k < \ell \le n}^{n} \iint_{\substack{|t| \le \pi m, \\ |s| \le \pi m}} |\mathrm{cov}(e^{itZ_1}, e^{isX_{\ell-k}})| |\frac{\psi^*(t)}{f_\varepsilon^*(t)}\psi^*(-s)| ds dt$$

$$\le \frac{1}{2\pi n} \sum_{k=1}^{n} \beta_k \left(\int_{|t| \le \pi m} |\psi^*(t)| dt \right) \left(\int_{|t| \le \pi m} \left| \frac{\psi^*(t)}{f_\varepsilon^*(t)} \right| dt \right).$$

This gives Inequality (14).

For the proof of (16), the result follows from the fact that the new mixing term is always negligible with respect to the independent variance term if ε is super-smooth (case $A, \rho > 0$). If ε is ordinary smooth, then we only have to study when $m^{(-B+1)_+ + \beta - B + 1}$ is less than $m^{2\beta - 2B + 1}$, which occurs if $\beta > \max(B, 1)$. $\square$

5.4 Proof of Corollary 3.1

The main difference with respect to the proof of Theorem 5.1 lies in the Bernstein inequality which must be written in the mixing context. For geometrically mixing variables (and $q = q_n = 2\ln(n)/c$ if $\beta_k \le e^{-ck}$), we get from Theorem 4 p.36 in Doukhan [16], that

$$\mathbb{P}\left(\frac{S_n - \mathbb{E}(S_n)}{n} \ge \sqrt{\frac{2u\tilde{v}^2}{n}} + \frac{2\ln(n)u}{can} \right) \le e^{-u} + \frac{2}{n^2},$$

with $\|Y_1\|_\infty \le 1/a$ and

$$\frac{1}{q}\mathrm{Var}\left(\sum_{k=1}^{q} Y_k \right) \le \tilde{v}^2.$$

In all cases, $|\mathcal{M}| \le n$, so that summing up the residuals of order $1/n^2$ will give negligible terms of order $1/n$. Next, the variables are still given by (34) and we can see from the bound that the upper bound being multiplied by $2\ln(n)/c$ in the Bernstein Inequality, all $c_{j,m}$'s and c_m's are the same as previously multiplied by $2\ln(n)/c$, this gives $\tilde{c}_{j,m} = 2c\ln(n)c_{j,m}$ and $\tilde{c}_m = 2c\ln(n)c_m$. Lastly, it follows from the above computation of $\mathrm{Var}(\breve{\theta}_m)$ that the new variance terms denoted by $\tilde{\sigma}_{j,m}^2$, $\tilde{\sigma}_m^2$ can be bounded under (D1) by

$$\tilde{\sigma}_{j,m}^2 \le \sigma_{j,m}^2 + \frac{1}{\pi n} \sum_{k \ge 1} \beta_k \left(\int_{\pi(m \wedge j) \le |t| \le \pi(m \vee j)} |\psi^*(t)| dt \right)^2,$$

and analogously for $\tilde{\sigma}_m^2$. It follows from our set of assumptions that $\tilde{\sigma}_{j,m}^2 \le \sigma_{j,m}^2 + c/n \le 2\sigma_{j,m}^2$ and $\tilde{\sigma}_m^2 \le 2\sigma_m^2$. The case (D2) is analogous under the more restrictive assumptions given. The result of Corollary 3.1 follows. $\square$

24

5.5 Proof of Theorem 4.1

1) Case $r = 0$, $\rho = 0$ and $\Lambda = [\underline{b}, \overline{b}] \times [\underline{L}, \overline{L}] \subset (1/2, \infty) \times (0, \infty)$.

Let us choose g_0 in the class $\mathcal{S}(\overline{b}, \overline{L}/2)$ such that $g_0 > 0$ and $g_0(x) \geq c|x|^{-2}$ as $|x| \to \infty$. We choose next the function G such that $G(x, m) = m^{-\underline{b}+1/2} G(mx)$ and with G^* at least 3-times continuously differentiable having the property

$$\frac{I(1/2 \leq |u| \leq 3/4)}{c(1 + u^{2\underline{b}})} \leq G^*(u) \leq \frac{I(1/4 \leq |u| \leq 1)}{c(1 + u^{2\underline{b}})}.$$

Here, $m = (c_0 \ln(n)/n)^{-1/(2\underline{b}+2\beta)}$. Note that $G^*(0) = \int G = 0$. Firstly, $g_{1,n}$ is a positive function with integral equal to 1 and belongs to $\mathcal{S}(\underline{b}, \overline{L})$. Indeed, for each fixed x we have $G(x, m) \to 0$ when $n \to \infty$ and as G^* is 3 times continuously differentiable that means $|G(x, m)| \leq O(|x|^{-3}) = o(g_0(x))$ as $|x| \to \infty$, giving that $g_{1,n} \geq 0$ for n large enough. Moreover,

$$\left(\int |g_{1,n}^*(u)|^2 |u|^{2\underline{b}} du \right)^{1/2}$$

$$\leq \left(\int |g_0^*(u)|^2 |u|^{2\underline{b}} du \right)^{1/2} + m^{-\underline{b}-1/2} \left(\int_{1/4 \leq |u|/m \leq 1} |G^*(u/m)|^2 |u|^{2\underline{b}} du \right)^{1/2}$$

$$\leq \sqrt{2\pi \overline{L}/2} + \frac{C}{c} \left(\int_{1/4}^1 \frac{|u|^{2\underline{b}}}{(1 + u^{2\underline{b}})^2} du \right)^{1/2} \leq (2\pi \overline{L})^{1/2},$$

for $c > 0$ large enough. Secondly,

$$\left| \frac{G(0, m)}{\psi_{n,\underline{b}}} \right| = (\psi_{n,\underline{b}})^{-1} m^{-\underline{b}+1/2} \frac{1}{2\pi} \int G^*(u) du \geq \frac{c_0^{-\underline{b}+1/2}}{2\pi} \int_{1/2}^{3/4} du \geq c_1 \cdot c_0^{-\underline{b}+1/2} > 0.$$

We shall prove that (23) holds with

$$\tau = n^{-\frac{2\beta+1}{2\underline{b}+2\beta}}$$

and together with the fact that

$$\tau q_n^2 = \tau \frac{\psi_{n,\underline{b}}^2}{\psi_{n,\overline{b}}^2} = \tau \left(\frac{\ln(n)}{n} \right)^{-\frac{(2\beta+1)(\overline{b}-\underline{b})}{(2\underline{b}+2\beta)(2\overline{b}+2\beta)}}$$

$$= (\ln(n))^{-\frac{(2\beta+1)(\overline{b}-\underline{b})}{(2\underline{b}+2\beta)(2\overline{b}+2\beta)}} n^{\frac{2\beta+1}{2\underline{b}+2\beta} \frac{\underline{b}+\beta}{\overline{b}+\beta}}$$

tends to infinity, with n, the proof of (24) and hence of the Theorem is finished.

We can prove that for each x_0

$$\sup_x \frac{|[G(m(\cdot - x_0)) \star f_\varepsilon](x)|}{f_0^Z(x)} = o(1), \text{ as } n \to \infty, \tag{38}$$

therefore $f_{1,n}^Z(x) = f_0^Z(x)(1 + o(1))$, where $o(1) \to 0$, $n \to \infty$ uniformly in x. As we chose $g > 0$ then $f_0^Z > 0$ and together with the previous statement it means that for any $M > 0$ we can find a constant $c_2 > 0$ such that $f_{1,n}^Z \geq 1/c_2$ on $[-M, M]$. Moreover, for some $M > 0$ large enough, see Butucea and Tsybakov [7],

$$f_0^Z(x) = g_0 \star f_\varepsilon(x) \geq \frac{C_2}{x^2}, \text{ as } |x| \geq M.$$

25

Therefore, for large enough $M > 0$, $f_{1,n}^Z(x) \geq 1/(c_3|x|^2)$, for some constant $c_3 > 0$ and for $|x| \geq M$. Finally, we deal with

$$\chi^2(f_0^Z, f_{1,n}^Z) = m^{-2\underline{b}+1} \int \frac{[G(m(\cdot - x_0)) \star f_\varepsilon]^2(x)}{f_{1,n}^Z(x)} dx$$

$$\leq m^{-2\underline{b}+1} \left(c_2 \int_{|x| \leq M} [G(m(\cdot - x_0)) \star f_\varepsilon]^2(x) dx + c_3 \int_{|x| > M} |x|^2 [G(m(\cdot - x_0)) \star f_\varepsilon]^2(x) dx \right),$$

say T_1 and T_2, for some fixed, large $M > 0$. Then

$$T_1 \leq m^{-2\underline{b}-1} \frac{c_2}{2\pi} \int \left| G^*\left(\frac{u}{m}\right) f_\varepsilon^*(u) \right|^2 du$$

$$\leq c_4 m^{-2\underline{b}-1} \int_{m/4}^m \frac{1}{|u|^{2\beta}} du \leq c_5 m^{-2\underline{b}-2\beta} \leq c_6 \frac{c_0 \ln(n)}{n}. \quad (39)$$

For T_2 we follow the similar proof in Butucea and Tsybakov [7] and use condition (21) to get

$$T_2 \leq m^{-2\underline{b}+1} \frac{c_3}{2\pi} \int \left| \frac{\partial}{\partial u} \left(\frac{1}{m} G^*\left(\frac{u}{m}\right) f_\varepsilon^*(u) \right) \right|^2 du$$

$$\leq c_6 m^{-2\underline{b}-1} m^{-2\beta} = o(T_1), \ n \to \infty. \quad (40)$$

Therefore, from (39) and (40) we have $\chi^2(f_0^Z, f_{1,n}^Z) \leq \kappa_n$, with $\kappa_n = c_\chi c_0 \ln(n)/n$. We use the fact that $-u(1 + u) \leq \ln(1 - u) \leq -u$ for all $u \in [0, 1/2]$ and that (38) implies that $|u| = |[G(m(\cdot - x_0)) \star f_\varepsilon](x)|/f_{1,n}^Z(x) \leq 1/2$ for n large enough to get

$$\mathbb{E}_1 Z_{1,n} = \int \ln \left(1 - \frac{[G(\cdot, m) \star f_\varepsilon](x - x_0)}{f_{1,n}^Z(x)} \right) f_{1,n}^Z(x) dx$$

$$\geq -\int [G(\cdot, m) \star f_\varepsilon](x - x_0) dx - \int \frac{[G(\cdot, m) \star f_\varepsilon]^2(x - x_0)}{f_{1,n}^Z(x)} dx$$

$$\geq -\chi^2(f_0^Z, f_{1,n}^Z) \geq -\kappa_n,$$

for n large enough. Indeed, note that $\int G(\cdot, m) = 0$ and therefore

$$\int [G(\cdot, m) \star f_\varepsilon](x - x_0) dx = 0.$$

Moreover,

$$V_1(Z_{1,n}) \leq \mathbb{E}_1(Z_{1,n}^2) = \int \ln^2 \left(1 - \frac{[G(\cdot, m) \star f_\varepsilon](x - x_0)}{f_{1,n}^Z(x)} \right) f_{1,n}^Z(x) dx$$

$$\leq \int \frac{[G(\cdot, m) \star f_\varepsilon]^2(x - x_0)}{f_{1,n}^Z(x)^2} \left(1 + \frac{[G(\cdot, m) \star f_\varepsilon]^2(x - x_0)}{f_{1,n}^Z(x)} \right)^2 f_{1,n}^Z(x) dx$$

$$\leq c_v \chi^2(f_0^Z, f_{1,n}^Z) \leq c_v \kappa_n,$$

as by (38): $\sup_x |f_0^Z(x)/f_{1,n}^Z(x)|$ is bounded from above by some constant depending only on g_0 and f_ε. By similar calculations, we also check that

$$
\begin{aligned}
V_1(Z_{1,n}) &\geq \frac{1}{2}\mathbb{E}_1(Z_{1,n}^2) = \frac{1}{2}\int \ln^2\left(1 - \frac{[G(\cdot,m)\star f_\varepsilon](x-x_0)}{f_{1,n}^Z(x)}\right) f_{1,n}^Z(x)dx \\
&\geq \frac{1}{2}\int \frac{[G(\cdot,m)\star f_\varepsilon]^2(x-x_0)}{f_{1,n}^Z(x)}dx \geq \frac{1}{2\|f_{1,n}^Z\|_\infty}\int [G(\cdot,m)\star f_\varepsilon]^2(x-x_0)dx \\
&\geq c_v'\kappa_n,
\end{aligned}
$$

and that

$$
\begin{aligned}
\sum_{i=1}^n \mathbb{E}_1\left|\frac{Z_{i,n}-\mathbb{E}_1(Z_{i,n})}{\sqrt{n\cdot V_1(Z_{1,n})}}\right|^4 &\leq \frac{n\mathbb{E}_1|Z_{1,n}|^4}{(c_v')^2 n^2\kappa_n^2} \leq \frac{n\int[G(\cdot,m)\star f_\varepsilon]^4(x-x_0)dx(1+o(1))}{(c_v')^2\ln^2(n)} \\
&\leq \frac{nc\int|G^*(u,m)f_\varepsilon^*(u)|^2du(\int|G^*(u,m)f_\varepsilon^*(u)|du)^2}{(c_v')^2\ln^2(n)} \\
&\leq c\frac{\ln(n)\cdot m^{-2\underline{b}-2\beta+1}}{\ln^2(n)} = o(1),
\end{aligned}
$$

as $n\to\infty$ and since $\underline{b} > 1/2$. Next we apply Lyapounov's central limit theorem for triangular arrays, see Petrov [28], to get $\mathbb{P}_1(U_n \geq u_n) \geq 1-\epsilon$, as

$$
0 \geq u_n = \frac{\ln(\tau)+\kappa_n}{\sqrt{c_v\kappa_n}} = \frac{-\frac{2\beta+1}{2\underline{b}+2\beta}+c_\chi c_0}{\sqrt{c_v c_\chi c_0}}\sqrt{\ln(n)} \to -\infty,
$$

with n.

2) Case $\alpha,\ r > 0$ and $\rho = 0$. Without loss of generality we consider $b = 0$.

In this case, take some $a \in [\underline{a},\overline{a}]$ and g_0 belonging to $\mathcal{S}(a,\overline{r},\overline{L}/2)$ such that $g_0 > 0$ and $g_0(x) \geq c|x|^{-2}$ as $|x|\to\infty$. Let us consider a function G as for the case 1 such that G^* is 3-times continuously differentiable having the property

$$
\frac{I(\pi/2 \leq |u| \leq 3\pi/4)}{c(1+u^4)} \leq G^*(u) \leq \frac{I(\pi/4 \leq |u| \leq \pi)}{c(1+u^4)}.
$$

Next,

$$
g_{1,n}(x) = g_0(x) + \sqrt{c_0\frac{\ln\ln n}{n}}m^{\beta+1/2}G(m(x-x_0)),
$$

where m is such that

$$
c_0\frac{\ln\ln n}{n}m^{2\beta+\underline{r}-1}\exp\left(2\underline{a}(\pi m)^{\underline{r}}\right) \leq 2\pi\overline{L}/2. \tag{41}
$$

Note that this gives a first order approximation of $m = (\log n/(2\underline{a}))^{1/\underline{r}}$. Then, similarly to the case 1, $g_{1,n}$ is a proper density function as soon as n is large enough and for some $M > 0$ we have $f_{1,n}^Z(x) = g_{1,n}*f_\varepsilon(x) \geq C|x|^{-2}$ for all $|x| \geq M$.

By using (41), we get that $g_{1,n}$ belongs to $S(\underline{a},\underline{r},\overline{L})$ for any $a \geq \underline{a}$.

Next,

$$
\frac{|g_{1,n}(x_0)-g_0(x_0)|}{\psi_{n,\underline{a},\underline{r}}} = c_0|G(0)| > 0
$$

27

and we get, by the same procedure as for the case 1,

$$\chi^2(f_0^Z, f_{1,n}^Z) = c_0 \frac{\ln \ln n}{n} m^{2\beta+1} \int \frac{[G(m(\cdot - x_0)) \star f_\varepsilon]^2(x)}{f_{1,n}^Z(x)} dx$$

$$\leq c_0 \frac{\ln \ln n}{n} m^{2\beta+1} c_1 \int [G(m(\cdot - x_0)) \star f_\varepsilon]^2(x) dx (1 + o(1))$$

$$\leq c_0 c_\chi \frac{\ln \ln n}{n} =: \kappa_n.$$

Let us choose c_0 small such that $c_0 c_\chi < (\bar{r} - \underline{r})(2\beta + 1)/(\bar{r}\underline{r})$ and let ξ and τ be defined by

$$c_0 c_\chi < \xi < \frac{\bar{r} - \underline{r}}{\bar{r}\underline{r}}(2\beta + 1) \text{ and } \tau = \ln(n)^{-\xi}.$$

On the one hand, this implies $\tau q_n^2 \to \infty$ with n. On the other hand, after checking again that Lyapounov's central limit theorem holds in this case we get

$$\mathbb{P}_1 \left(\frac{d\mathbb{P}_0}{d\mathbb{P}_1} \geq \tau \right) \geq \mathbb{P}_1(U_n \geq u_n) \geq 1 - \epsilon,$$

as $u_n = (-\ln(\tau) + n\kappa_n)(c_v n\kappa_n)^{-1/2} = (-\xi + c_0 c_\chi)(c_v c_0 c_\chi)^{-1/2}\sqrt{\ln \ln(n)} \to -\infty$.

3) Case $r > 0$, $0 < \rho \leq 1$ and $r \in [\underline{r}, \bar{r}]$ such that $\underline{r} \geq \rho$. Without loss of generality we consider $b = 0$.

As in the second case, take some $a \in [\underline{a}, \bar{a}]$ and g_0 belonging to $\mathcal{S}(a, \bar{r}, \overline{L}/2)$ such that $g_0 > 0$ and $g_0(x) \geq c|x|^{-2}$ as $|x| \to \infty$. Let also G be a function such that G^* is 3-times continuously differentiable with a bounded first derivative and having the property

$$I(\pi/2 \leq |u| \leq 3\pi/4) \leq G^*(u) \leq I(\pi/4 \leq |u| \leq \pi).$$

Next, define $g_{1,n}$ via its Fourier transform

$$g_{1,n}^*(u) = g_0^*(u) + c_0 \frac{e^{-\alpha(\pi m)^\rho}}{\sqrt{n}} m^{\rho - 1/2} e^{2\alpha|u|^\rho} G^*\left(|u|^\rho - (\pi m)^\rho\right) e^{iux_0},$$

where m is solution of the equation

$$2\underline{a}(\pi m)^{\underline{r}} + 2\alpha(\pi m)^\rho = \log n - (\log \log n)^2. \tag{42}$$

We stress the fact that m is no longer a scaling parameter of the function G in this construction.

Again, as previously, we can check that $g_{1,n}$ is a proper probability density, as soon as n is large enough, and that for some $M > 0$ we have $f_{1,n}^Z(x) \geq C|x|^{-2}$ for all $|x| \geq M$.

Let us check that $g_{1,n}$ belongs to $S(\underline{a}, \underline{r}, \overline{L})$. It is enough to bound from above

$$\frac{1}{2\pi} \int c_0^2 \frac{e^{-2\alpha(\pi m)^\rho}}{n} m^{2\rho - 1} e^{4\alpha|u|^\rho} |G^*\left(|u|^\rho - (\pi m)^\rho\right)|^2 e^{2\underline{a}|u|^{\underline{r}}} du$$

$$\leq c_0^2 m^{2\rho - 1} \frac{e^{-2\alpha(\pi m)^\rho}}{2\pi n} \int_{\pi/4 \leq |u|^\rho - (\pi m)^\rho \leq 3\pi/4} e^{4\alpha|u|^\rho + 2\underline{a}|u|^{\underline{r}}} du$$

$$\leq c_0^2 c_1 m^{2\rho - 1} \frac{e^{-2\alpha(\pi m)^\rho}}{2\pi n} (\pi m)^{1-\underline{r}} e^{4\alpha(\pi m)^\rho + 2\underline{a}(\pi m)^{\underline{r}}}$$

$$\leq c_0^2 c_2 n^{-1} m^{2\rho - \underline{r}} e^{2\underline{a}(\pi m)^{\underline{r}} + 2\alpha(\pi m)^\rho}$$

28

which tends to 0 when m is defined by (42).

Next,

$$
\begin{aligned}
|g_{1,n}(x_0) - g_0(x_0)| &= \frac{1}{2\pi}\left|\int c_0 \frac{e^{-\alpha(\pi m)^\rho}}{\sqrt{n}} m^{\rho-1/2} e^{2\alpha|u|^\rho} G^*\left(|u|^\rho - (\pi m)^\rho\right) du\right| \\
&\geq c_0 m^{\rho-1/2}\frac{e^{-\alpha(\pi m)^\rho}}{2\pi\sqrt{n}}\int_{\pi/2\leq|u|^\rho-(\pi m)^\rho\leq\pi} e^{2\alpha|u|^\rho} du \\
&\geq c_0 c_3 m^{1/2}\frac{e^{\alpha(\pi m)^\rho}}{2\pi\sqrt{n}}
\end{aligned}
$$

and we can check similarly to Butucea and Tsybakov [7] that for m solution of (42) this sequence is equivalent to $\psi_{n,\underline{a},r}$ when $n \to \infty$.

Finally

$$
\begin{aligned}
\chi^2(f_0^Z, f_{1,n}^Z) &= c_0^2 \int \frac{[(g_{1,n} - g_0) \star f_\varepsilon]^2(x)}{f_{1,n}^Z(x)}dx \\
&\leq c_0^2\left\{\int_{|x|\leq M}[(g_{1,n} - g_0) \star f_\varepsilon]^2(x)dx + \int_{|x|>M} x^2[(g_{1,n} - g_0) \star f_\varepsilon]^2(x)dx\right\},
\end{aligned}
$$

say $T_1 + T_2$. Then

$$
\begin{aligned}
T_1 &\leq c_0^2 c_4 \frac{e^{-2\alpha(\pi m)^\rho}}{n} m^{2\rho-1} \int |G^*(|u|^\rho - (\pi m)^\rho) f_\varepsilon^*(u)|^2 du \\
&\leq c_0^2 c_5 \frac{e^{-2\alpha(\pi m)^\rho}}{n} m^{2\rho-1} \int_{\pi/4\leq|u|^\rho-(\pi m)^\rho\leq 3\pi/4} e^{2\alpha|u|^\rho} du \\
&= c_0^2 c_6 \frac{(\pi m)^\rho}{n}.
\end{aligned}
$$

Moreover, under the additionnal assumption (22) that $|\partial f_\varepsilon^*(u)/\partial u| \leq O(1)|u|^{\rho-1}\exp(-\alpha|u|^\rho)$ as $|u| \to \infty$,

$$
\begin{aligned}
T_2 &\leq c_0^2 c_7 \frac{e^{-2\alpha(\pi m)^\rho}}{n} m^{2\rho-1} \int \left|\frac{\partial}{\partial u}[G^*(|u|^\rho - (\pi m)^\rho) f_\varepsilon^*(u)]\right|^2 du \\
&\leq c_8 \frac{e^{-2\alpha(\pi m)^\rho}}{n} m^{2\rho-1} \int_{\pi/4\leq|u|^\rho-(\pi m)^\rho\leq 3\pi/4} |u|^{2(\rho-1)} e^{2\alpha|u|^\rho} du \\
&\leq c_9 \frac{(\pi m)^{3\rho-2}}{n} = o(T_1),
\end{aligned}
$$

for $\rho \leq 1$ and n large enough. Thus

$$
\chi^2(f_0^Z, f_{1,n}^Z) \leq c_0^2 c_\chi \frac{(\pi m)^\rho}{n} =: \kappa_n.
$$

Let c_0 be small such that $c_0^2 c_\chi < 2\alpha$ and let ξ and τ be defined by

$$
c_0^2 c_\chi < \xi < 2\alpha \quad \text{and} \quad \tau = e^{-\xi(\pi \ln(n)/(2\underline{a}))^{\rho/r}}.
$$

We have

$$\tau\psi^2_{n,\underline{a},\underline{r}}/\psi^2_{n,\overline{a},\overline{r}} \geq (\ln(n))^A \exp\left((-\xi+2\alpha)\left(\frac{\ln(n)}{2\underline{a}}\right)^{\rho/\underline{r}} + B(\ln(n))^C\right) \to \infty$$

for some real numbers A, B C, as $C < \rho/\underline{r}$ and $\xi < 2\alpha$. We check that Lyapounov's theorem holds and that

$$u_n = \frac{-\ln(\tau) + n\kappa_n}{\sqrt{c_v n\kappa_n}} = \frac{-\xi(\pi\ln(n)/(2\underline{a}))^{\rho/\underline{r}} + c_0^2 c_\chi(\pi m)^\rho}{c_0\sqrt{c_v c_\chi}(\pi m)^{\rho/2}} \to -\infty$$

with n, as m defined by (42) is larger than $(\ln(n)/(2\underline{a}))^{1/\underline{r}}$.

The proof that φ_n is the minimax rate of estimation in this case repeats the proof of 3 with modified choice of $g_{1,n}$ via its Fourier transform

$$g^*_{1,n}(u) = g^*_0(u) + c_0\frac{e^{-\alpha(\pi m)^\rho}}{\sqrt{n}}m^{(\rho-1)/2}e^{2\alpha|u|^\rho}G^*\left(|u|^\rho - (\pi m)^\rho\right)e^{iux_0},$$

where m is solution of the equation (42). This gives the rate

$$|g_{1,n}(x_0) - g_0(x_0)| \geq c_0 c_3 m^{-(\rho-1)/2}\frac{e^{\alpha(\pi m)^r}}{\sqrt{n}},$$

which is equivalent to $V_{\breve{m}}$ for n large enough and

$$n\chi^2(f_0^Z, f_{1,n}^Z) \leq c_0^2 c_6 + c_9 m^{2\rho-2} \leq c_0^2 c_\chi.$$

Thus, the rate φ_n is a minimax rate of convergence for $r \geq \rho$, $\rho \leq 1$. $\qquad\square$

References

[1] M. Abramowitz and I. A. Stegun. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, volume 55 of *National Bureau of Standards Applied Mathematics Series*. For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C., 1964.

[2] L. M. Artiles Martínez. Adaptive minimax estimation in classes of smooth functions. 2001. Ph. D. Thesis.

[3] L. M. Artiles Martínez and B. Y. Levit. Adaptive estimation of analytic functions on an interval. *Math. Methods Statist.*, 12(1):62–94, 2003.

[4] C. Butucea. Exact adaptive pointwise estimation on Sobolev classes of densities. *ESAIM Probab. Statist.*, 5:1–31 (electronic), 2001.

[5] C. Butucea. Deconvolution of supersmooth densities with smooth noise. *Canad. J. Statist.*, 32(2):181–192, 2004.

[6] C. Butucea and M. Neumann. Exact asymptotics for estimating the marginal density of discretely observed diffusion processes. *Bernoulli*, 11(3):411–444, 2005.

[7] C. Butucea and A. B. Tsybakov. Sharp optimality for density deconvolution with dominating bias. i and ii. *Theory Probab. Appl.*, 52(1):–, 2007.

[8] T. T. Cai and M. G. Low. Adaptive estimation of linear functionals under different performance measures. *Bernoulli*, 11(2):341–358, 2005.

[9] T. T. Cai and M. G. Low. On adaptive estimation of linear functionals. *Ann. Statist.*, 33(5):2311–2343, 2005.

[10] R. J. Carroll and P. Hall. Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.*, 83(404):1184–1186, 1988.

[11] L. Cavalier. On the problem of local adaptive estimation in tomography. *Bernoulli*, 7:63–78, 2001.

[12] M. Chaleyat-Maurel and V. Genon-Catalot. Computable infinite-dimensional filters with applications to discretized diffusion processes. *Stochastic Process. Appl.*, 116(10):1447–1467, 2006.

[13] F. Comte, J. Dedecker, and M.-L. Taupin. Adaptive density estimation for general arch models. *Working Paper MAP5 2006-10*, (http://www.math-info.univ-paris5.fr/map5/publis/titres06.html), 2006.

[14] F. Comte and V. Genon-Catalot. Penalized projection estimator for volatility density. *Scand. J. Statist.*, 33(4):875–893, 2006.

[15] F. Comte, Y. Rozenholc, and M.-L. Taupin. Penalized contrast estimator for adaptive density deconvolution. *Can. J. Stat.*, 34(3):431–452, 2006.

[16] P. Doukhan. *Mixing*, volume 85 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1994. Properties and examples.

[17] J. Fan. On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.*, 19(3):1257–1272, 1991.

[18] V. Genon-Catalot and M. Kessler. Random scale perturbation of an AR(1) process and its properties as a nonlinear explicit filter. *Bernoulli*, 10(4):701–720, 2004.

[19] G. K. Golubev. The method of risk envelopes in the estimation of linear functionals. *Problemy Peredachi Informatsii*, 40(1):58–72, 2004.

[20] Y. Golubev and B. Levit. An oracle approach to adaptive estimation of linear functionals in a Gaussian model. *Math. Methods Statist.*, 13(4):392–408 (2005), 2004.

[21] J. Klemelä and A. B. Tsybakov. Exact constants for pointwise adaptive estimation under the Riesz transform. *Probab. Theory Related Fields*, 129(3):441–467, 2004.

[22] C. Lacour. Rates of convergence for nonparametric deconvolution. *C. R. Math. Acad. Sci. Paris*, 342(11):877–882, 2006.

[23] B. Laurent, C. Ludena, and C. Prieur. Adaptive estimation of linear functionals by model selection. *Preprint LSP-2006-05, http://www.lsp.ups-tlse.fr/*, 2006.

[24] O. V. Lepski. A problem of adaptive estimation in Gaussian white noise. *Teor. Veroyatnost. i Primenen.*, 35(3):459–470, 1990.

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

[26] Y. Meyer. *Ondelettes et opérateurs. I.* Actualités Mathématiques. [Current Mathematical Topics]. Hermann, Paris, 1990. Ondelettes. [Wavelets].

[27] M. Pensky and B. Vidakovic. Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Statist.*, 27(6):2033–2053, 1999.

[28] V. V. Petrov. *Limit theorems of probability theory*, volume 4 of *Oxford Studies in Probability*. The Clarendon Press Oxford University Press, New York, 1995. Sequences of independent random variables, Oxford Science Publications.

[29] L. Stefanski and R. J. Carroll. Deconvoluting kernel density estimators. *Statistics*, 21(2):169–184, 1990.

[30] A. B. Tsybakov. Pointwise and sup-norm sharp adaptive estimation of functions on the Sobolev classes. *Ann. Statist.*, 26(6):2420–2469, 1998.

[31] B. Van Es, P. Spreij, and H. van Zanten. Nonparametric volatility density estimation for discrete time models. *J. Nonparametr. Stat.*, 17(2):237–251, 2005.