

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

C. BUTUCEA<sup>1</sup> and F. COMTE<sup>2</sup>

<sup>1</sup>Laboratoire Paul Painlevé – UMR CNRS 8524, Université Lille 1, 59655 Villeneuve d'Ascq Cedex, France. E-mail: [Cristina.Butucea@math.univ-lille1.fr](mailto:Cristina.Butucea@math.univ-lille1.fr)

<sup>2</sup>MAP5, UMR CNRS 8145, Université Paris Descartes, 45 rue des Saints-Pères, 75270 Paris Cedex 06, France. E-mail: [fabienne.comte@parisdescartes.fr](mailto:fabienne.comte@parisdescartes.fr)

We consider the model  $Z_i = X_i + \varepsilon_i$ , for i.i.d.  $X_i$ 's and  $\varepsilon_i$ 's and independent sequences  $(X_i)_{i \in \mathbb{N}}$  and  $(\varepsilon_i)_{i \in \mathbb{N}}$ . The density  $f_\varepsilon$  of  $\varepsilon_1$  is assumed to be known, whereas the one of  $X_1$ , denoted by  $g$ , is unknown. Our aim is to estimate linear functionals of  $g$ ,  $\langle \psi, g \rangle$  for a known function  $\psi$ . We propose a general estimator of  $\langle \psi, g \rangle$  and study the rate of convergence of its quadratic risk as a function of the smoothness of  $g$ ,  $f_\varepsilon$  and  $\psi$ . Different contexts with dependent data, such as stochastic volatility and AutoRegressive Conditionally Heteroskedastic models, are also considered. An estimator which is adaptive to the smoothness of unknown  $g$  is then proposed, following a method studied by Laurent *et al.* (Preprint (2006)) in the Gaussian white noise model. We give upper bounds and asymptotic lower bounds of the quadratic risk of this estimator. 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 in the context of the stochastic volatility model.

*Keywords:* adaptive density estimation; ARCH models; deconvolution; linear functionals; model selection; penalized contrast; stochastic volatility model

## 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 sequences of real valued random variables. The  $X_i$  are i.i.d. with unknown density  $g$ , the  $\varepsilon_i$  are i.i.d. with known density  $f_\varepsilon$ . The Fourier transform of a function  $u \in \mathbb{L}^1(\mathbb{R})$  is denoted by  $u^*(x) = \int e^{ixt} u(t) dt$ . The smoothness of  $f_\varepsilon$  is described by parameters  $\gamma$ ,  $\alpha$ ,  $\rho$  in the following assumption:

$$\text{There exist non-negative numbers } \kappa_0, \kappa'_0, \gamma, \alpha \text{ and } \rho \text{ such that } f_\varepsilon^* \text{ satisfies} \tag{2}$$
$$\kappa_0(x^2 + 1)^{-\gamma/2} \exp\{-\alpha|x|^\rho\} \leq |f_\varepsilon^*(x)| \leq \kappa'_0(x^2 + 1)^{-\gamma/2} \exp\{-\alpha|x|^\rho\},$$

with  $\gamma > 1$  when  $\rho = 0$ . If either  $\alpha = 0$  or  $\rho = 0$ , we set  $(\alpha, \rho) = (0, 0)$ . Since  $f_\varepsilon$  is known, the constants  $\alpha$ ,  $\rho$ ,  $\kappa_0$ ,  $\kappa'_0$  and  $\gamma$  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 supersmooth. The standard examples for supersmooth densities are Gaussian or

Cauchy distributions (supersmooth of order  $\gamma = 0, \rho = 2$  and  $\gamma = 0, \rho = 1$ , respectively). An example of an ordinary smooth density is the Laplace distribution ( $\rho = 0 = \alpha$  and  $\gamma = 2$ ).

In this context, many papers have studied the deconvolution problem. Many different strategies have been developed in order to estimate the distribution  $g$  of the unobserved  $X_i$ , when  $g$  is assumed to belong to some smoothness class defined by

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

where  $b, a, r, L$  are some unknown non-negative numbers, such that  $b > 1/2$  when  $r = 0$ . If either  $a = 0$  or  $r = 0$ , we set  $(a, r) = (0, 0)$  and we say that the density is ordinary smooth. When both  $a, r > 0$ , we call the density supersmooth.

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  $\psi$  is a known integrable function with respect to the probability measure associated to  $g$ . To study the rates of convergence of our estimators, we have to take into account the smoothness of the function  $\psi$ . Thus  $\psi$  is assumed to satisfy:

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

The parameters  $A$  and  $R$  are non-negative real numbers, and  $B$  is non-negative or such that  $\psi^* g^*$  is integrable. In particular, they can be zero if  $g^*$  is integrable. We work under the convention that if either  $A = 0$  or  $R = 0$ , then we set  $(A, R) = (0, 0)$ .

We exhibit the whole range of the rates of convergence for estimators of the functional  $\theta(g)$ , depending on the parameters in (2)–(4). To the best of our knowledge, this general rate description is new. We also extend the result to different dependency contexts, in view of applications to particular hidden Markov models or AutoRegressive Conditionally Heteroskedastic-type models.

The upper bounds for the rates follow from a squared-bias/variance compromise. To obtain this compromise, we have to choose a smoothing parameter which depends on unknown quantities. Therefore, a data driven model selection type procedure is proposed. It is based on minimization of a penalized estimated criterion, which is different from the one intensively studied for mean integrated squared errors. The difficulty here lies in finding an adequate criterion for the setting of a linear functional and mean squared error. The proposed procedure is inspired by Laurent *et al.* [24]. We give upper bounds for this adaptive method, with particular interest in the cases where a loss in the rate appeared with respect to the non-adaptive estimator.

In the particular case of pointwise estimation, adaptive estimation in the direct problem (i.e., when the  $X_i$  are observed without noise) has been widely studied in the context of the Gaussian white noise and regression models, see, for example, Lepski [26], Tsybakov [28], Cai and Low [8,7] (for more general linear functionals), Artiles and Levit [2], Laurent *et al.* [24] and, in the context of density models, Lepski and Levit [25], Butucea [3] and Artiles [1]. For the model of Gaussian sequences Golubev and Levit [21] and Golubev [20] considered adaptive estimation of linear functionals in both direct and inverse problems. In the Gaussian white noise model Goldenshluger [18] and Goldenshluger and Pereverzev [19] considered pointwise estimation for the inverse problem on classes of functions similar to  $\mathcal{S}(b, 0, 0, L)$ . Their adaptive procedure is based on Lepski's procedure. Note also that in some particular inverse problems the pointwise adaptive estimation was solved by Klemelä and Tsybakov [22] for the Riesz transform

and by Cavalier [10] for the tomography problem. To the best of our knowledge, we present the first work on adaptive estimation of general functionals of the form  $\int \psi g$  in the context of indirect observation (1).

We do not study optimality in the very general case: this would be very technical. But we study as a first application the particular case of pointwise density deconvolution. This case corresponds to  $\psi^*(t) = e^{itx_0}$ , which satisfies (4), meaning that we can choose  $\psi$  as the Dirac measure at  $x_0$ . This makes sense in our problem because the definition of our estimator involves only  $\psi^*$ . We recover in this particular case the upper bound rates obtained by Fan [16], Cator [9], Butucea [4] and Butucea and Tsybakov [6]. Moreover, we prove the optimality in the minimax sense of the loss due to adaptation for Sobolev smooth and supersmooth densities in the presence of ordinary smooth noise and for supersmooth densities in the presence of supersmooth noise with  $r \geq \rho$  and  $0 < \rho \leq 1$  (in the case  $r < \rho$  no loss occurs, while the case  $r \geq \rho$  and  $1 < \rho < 2$  is still open). As a by-product we also prove in the last case that the rate of our estimator is optimal in the minimax sense, which was not yet known in the literature.

Our estimation method is also illustrated for the discrete stochastic volatility model, where derivatives of the Laplace transform of the volatility are estimated with good rates.

The plan of the paper is as follows: In Section 2 we define the estimators and we compute upper bounds for their mean squared error. In Section 3 the adaptive procedure is detailed. Both independent and  $\beta$ -mixing contexts are studied. In Section 4, two applications of our general results are detailed. Section 4.1 shows 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 an application to the context of the stochastic volatility model. Most proofs are gathered in Section 5.

## 2. Risk bound for the estimator

We denote by  $\langle \cdot, \cdot \rangle$  the  $\mathbb{L}^2$ -scalar product ( $\langle u, v \rangle = \int u(x)\bar{v}(x) dx$ ), 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  $u \in \mathbb{L}^1(\mathbb{R})$ :  $u^*(x) = \int e^{itx}u(t) dt$ .

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  $Z_i$ , for  $i = 1, \dots, n$  are available. In what follows we assume that

$$\begin{cases} \psi g, \psi \text{ and } \psi^* g^* \text{ belong to } \mathbb{L}_1(\mathbb{R}), \\ f_\varepsilon \text{ belongs to } \mathbb{L}_2(\mathbb{R}) \text{ and is such that } \forall x \in \mathbb{R} f_\varepsilon^*(x) \neq 0. \end{cases} \quad (5)$$

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

Moreover, we generalize these results to distributions having Fourier transform such that  $\int \psi^* g^* < \infty$ . For example, we estimate  $g(x_0)$  for some fixed  $x_0$  when we take  $\psi = \delta_{x_0}$ , the Dirac measure at  $x_0$ , having Fourier transform equal to  $\psi^*(t) = e^{-itx_0}$ . For estimating the derivatives  $g^{(k)}(x_0)$ , when they exist, we consider  $\psi$  such that  $\psi^*(t) = (-it)^k e^{-itx_0}$ .

## 2.1. The estimator

We write, using (5),  $\langle \psi, g \rangle = (1/2\pi)\langle \psi^*, g^* \rangle = (1/2\pi)\langle \psi^*, f_Z^*/f_\varepsilon^* \rangle$ . Replacing  $f_Z^*(t)$  by its empirical version  $(1/n)\sum_{k=1}^n e^{itZ_k}$  leads to the estimator

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

This estimator is explicit and seems attractive. Unfortunately, the integral diverges for many choices of  $f_\varepsilon^*$ ; for instance,  $\varepsilon$  is a Gaussian noise. To overcome such issues, we suggest regularization and take the following estimator of  $\theta(g)$ :

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

where  $m$  is an integer.

**Remark 2.1.** Let  $\hat{g}_m$  denote the projection estimator of  $g$  defined in Comte *et al.* [14]. Then we can prove that  $\hat{\theta}_m = \langle \hat{g}_m, \psi \rangle$ .

## 2.2. Risk bounds and rates for i.i.d. variables $X_i$ 's

If (5) holds and if, moreover,  $\psi^*(-\cdot)/f_\varepsilon^*$  is integrable, then  $m = +\infty$  can be chosen and the estimator  $\hat{\theta} = \hat{\theta}_\infty$  is unbiased and has a parametric rate.

Otherwise, we have  $\mathbb{E}(\theta - \hat{\theta}_m)^2 = b^2(\hat{\theta}_m) + \text{Var}(\hat{\theta}_m)$  with  $b(\hat{\theta}_m) = \theta - \mathbb{E}(\hat{\theta}_m)$ . As  $\mathbb{E}(\hat{\theta}_m) = (2\pi)^{-1} \int_{|t| \leq \pi m} g^*(t) \psi^*(-t) dt$ , we obtain

$$\begin{aligned} b(\hat{\theta}_m) &= \frac{1}{2\pi} \left( \int g^*(t) \psi^*(-t) dt - \int_{|t| \leq \pi m} g^*(t) \psi^*(-t) dt \right) \\ &= \frac{1}{2\pi} \int_{|t| \geq \pi m} g^*(t) \psi^*(-t) dt. \end{aligned} \quad (8)$$

Under (5),  $b(\hat{\theta}_m)$  tends to 0 when  $m$  grows to infinity.

For the variance term, write

$$\text{Var}(\hat{\theta}_m) = \frac{1}{4\pi^2 n} \text{Var} \left( \int_{-\pi m}^{\pi m} e^{iuZ_1} \frac{\psi^*(-u)}{f_\varepsilon^*(u)} du \right).$$

First, the following bound holds:

$$\text{Var}(\hat{\theta}_m) \leq \frac{1}{4\pi^2 n} \mathbb{E} \left( \left| \int_{-\pi m}^{\pi m} e^{iuZ_1} \frac{\psi^*(-u)}{f_\varepsilon^*(u)} du \right|^2 \right) \leq \frac{1}{4\pi^2 n} \left( \int_{-\pi m}^{\pi m} \frac{|\psi^*(-u)|}{|f_\varepsilon^*(u)|} du \right)^2.$$

Next, the variance can also be bounded as follows:

$$\begin{aligned}\text{Var}(\hat{\theta}_m) &= \frac{1}{4\pi^2 n} \int_{-\pi m}^{\pi m} \int_{-\pi m}^{\pi m} (f_Z^*(u-v) - f_Z^*(u)f_Z^*(-v)) \frac{\psi^*(-u)\psi^*(v)}{f_\varepsilon^*(u)f_\varepsilon^*(-v)} du dv \\ &\leq \frac{1}{4\pi^2 n} \int_{-\pi m}^{\pi m} \int_{-\pi m}^{\pi m} \frac{\psi^*(-u)\psi^*(v)}{f_\varepsilon^*(u)f_\varepsilon^*(-v)} f_Z^*(u-v) du dv.\end{aligned}$$

We use the Cauchy–Schwarz inequality and Fubini’s theorem:

$$\begin{aligned}\text{Var}(\hat{\theta}_m) &\leq \frac{1}{4\pi^2 n} \left( \int_{-\pi m}^{\pi m} \int_{-\pi m}^{\pi m} \left| \frac{\psi^*(-u)}{f_\varepsilon^*(u)} \right|^2 |f_Z^*(u-v)| du dv \right. \\ &\quad \times \left. \int_{-\pi m}^{\pi m} \int_{-\pi m}^{\pi m} \left| \frac{\psi^*(v)}{f_\varepsilon^*(-v)} \right|^2 |f_Z^*(u-v)| du dv \right)^{1/2} \\ &\leq \frac{1}{4\pi^2 n} \int_{|u| \leq \pi m} \frac{|\psi^*(-u)|^2}{|f_\varepsilon^*(u)|^2} du \int |f_Z^*(x)| dx.\end{aligned}$$

Note that since  $\psi$  is a real valued function we have  $|\psi^*(-t)| = |\psi^*(t)|$ . As  $\int |f_Z^*(x)| dx \leq \int |f_\varepsilon^*(x)| dx$ , we have the following result:

**Proposition 2.1.** *Assume that  $C_\varepsilon = \int |f_\varepsilon^*(x)| dx < +\infty$ , and let  $\hat{\theta}_m$  be defined by (7). Then, under (5),*

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

Note that we also have  $\int |f_Z^*(x)| dx \leq \|f_\varepsilon^*\| \|g^*\| = 2\pi \|f_\varepsilon^*\| \|g\|$ , if  $f_\varepsilon$  and  $g$  are square integrable.

**Remark 2.2.** If, in addition,

$$\int |\psi^*(x)/f_\varepsilon^*(x)|^2 dx < +\infty, \quad (9)$$

then the variance of  $\hat{\theta}_m$  is of order  $1/n$  and the estimator can reach the parametric rate, for  $m$  large enough. Note that a condition like  $\int |\psi^*(x)/f_\varepsilon^*(x)| dx < \infty$  (which ensures that (6) is well defined) is generally stronger than (9) as convergence problems lie only near infinity. Moreover, such conditions are fulfilled if  $\psi^*$  decreases faster than  $f_\varepsilon^*$  near infinity, which corresponds to the intuitive idea that  $\psi$  is a smoother function than  $f_\varepsilon$ . For example, this happens if  $\psi$  is supersmooth and  $f_\varepsilon$  is ordinary smooth.

Thus we can study the rates that can be deduced from the upper bounds of Proposition 2.1, as a function of the smoothness parameters of the three functions involved,  $g$ ,  $\psi$ ,  $f_\varepsilon$ . To do so, let us assume that  $\psi$  satisfies (4), that  $g$  belongs to  $\mathcal{S}(b, a, r, L)$  as defined by (3) and that  $f_\varepsilon^*$  fulfills (2). Then, use (8), (3) and (4) to get

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

On the other hand, the noise plays an important role on the variance of the estimator:

- Case (I): If  $(\rho = R = 0, \gamma < B - 1/2)$  or  $(\rho = R > 0, \alpha = A, \gamma < B - 1/2)$  or  $(\rho = R, \alpha < A)$  or  $(\rho < R)$ , then  $\text{Var}(\hat{\theta}_m) \leq C'n^{-1}$ ;
- Case (II): If  $(\rho = R = 0, \gamma = B - 1/2)$  or  $(\rho = R > 0, \alpha = A, \gamma = B - 1/2)$ , then  $\text{Var}(\hat{\theta}_m) \leq C' \ln(m)n^{-1}$ ;
- Case (III): If  $(\rho = R = 0, \gamma > B - 1/2)$  or  $(\rho = R > 0, \alpha = A, \gamma > B - 1/2)$ , then  $\text{Var}(\hat{\theta}_m) \leq C'm^{2\gamma-2B+1}n^{-1}$ ;
- Case (IV): If  $(\rho > R)$  or  $(\rho = R > 0, \alpha > A)$ , then

$$\text{Var}(\hat{\theta}_m) \leq C'n^{-1}m^{2\gamma-2B+1-\rho+(1-\rho)_+}e^{2\alpha(\pi m)^\rho-2A(\pi m)^R}.$$

We summarize in Table 1 the scenarios that arise when one minimizes over  $m$  the sum of the upper bounds on the bias and the variance. Let  $a \vee b = \max\{a, b\}$ . Note that in cases (8) and (9) the rate is given by

$$\begin{aligned}
v_n = \min_m \left\{ C_B m^{-2b-2B+1-r \vee R} e^{-2a(\pi m)^r - 2A(\pi m)^R} \right. \\
\left. + m^{2\gamma-2B+1-\rho+(1-\rho)_+} e^{2\alpha(\pi m)^\rho - 2A(\pi m)^R} \frac{1}{n} \right\}. \quad (10)
\end{aligned}$$

These rates are strictly faster than  $(\ln(n))^{-\lambda_1}$ , that is,  $v_n = o((\ln(n))^{-\lambda_1})$  for any  $\lambda_1 > 0$ , and generally slower than  $n^{-\lambda_2}$ ,  $\lambda_2 > 0$  (negative powers of  $n$  can be obtained). For precise (but cumbersome) formulae in similar cases, we refer to Lacour [23]. We give in Section 5 the orders of the  $m$  associated to the rates.

**Remark 2.3.** Different optimal choices of  $m$  depend on the unknown parameters related to  $g$  (see Section 5.1), hence the interest in an automatic selection procedure for  $m$ .

### 2.3. Extension to mixing contexts

In view of applications, it is natural to study the robustness of our method when the variables  $X_i$  are  $\beta$ -mixing. To be more precise, two dependence contexts are considered.

**Table 1.** Upper bounds for the minimax rates of convergence,  $\delta_1 = (2\gamma - 2B + 1)/(r \vee R)$ ,  $\delta_2 = (b + B - 1/2)/(b + \gamma)$  and  $\delta_3 = (2(b + B) - 1)/\rho$

Parameters	Rates	Adaptive rates
$\rho < R$	(1) $n^{-1}$	
$\rho = R, \alpha < A$	(2) $n^{-1}$ (3) $n^{-1}$	
$(\rho = R = 0)$ or $(\rho = R > 0, \alpha = A)$	$\left\{ \begin{array}{l} \gamma < B - 1/2 \\ \gamma = B - 1/2 \left\{ \begin{array}{l} r \vee R > 0 \\ r \vee R = 0 \end{array} \right. \\ \gamma > B - 1/2 \left\{ \begin{array}{l} r \vee R > 0 \\ r \vee R = 0 \end{array} \right. \end{array} \right.$	$\left\{ \begin{array}{l} (\ln \ln n)^2 n^{-1} \\ (\ln n)^2 n^{-1} \\ (\ln \ln n) \ln n)^{\delta_1} n^{-1} \\ (n / \ln(n))^{-\delta_2} \end{array} \right.$
$(\rho = R > 0, \alpha > A)$	(8) $v_n$ in (10)	$v_n (\ln(n))^{\delta_4}, 0 \leq \delta_4 \leq 1$
$\rho > R$	$\left\{ \begin{array}{l} r \vee R > 0 \\ r \vee R = 0 \end{array} \right.$	$\left\{ \begin{array}{l} (9) v_n \text{ in (10)} \\ (10) \ln(n)^{-\delta_3}, \end{array} \right.$

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

(D2) In Model (1), the  $\varepsilon_i$  are i.i.d. and, for any given  $i$ ,  $X_i$  and  $\varepsilon_i$  are independent (but the sequences  $(X_i)$  and  $(\varepsilon_i)$  are not independent). The sequence  $(Z_i, X_i)_{i \in \mathbb{Z}}$  is strongly stationary and  $\beta$ -mixing, with  $\beta$ -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  $\beta$ -mixing Markov process. As many Markov chain models or other standard models can be proved to have such mixing properties (see Doukhan [15] for a large number 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 2.2.** 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 (7). Then

$$\begin{aligned} \mathbb{E}(\theta - \hat{\theta}_m)^2 &\leq \left( \frac{1}{2\pi} \int_{|t| \geq \pi m} |g^*(t)\psi^*(t)| dt \right)^2 \\ &+ \frac{C_\varepsilon}{4\pi^2 n} \min \left\{ \int_{-\pi m}^{\pi m} \frac{|\psi^*|^2}{|f_\varepsilon^*|^2}, \left( \int_{-\pi m}^{\pi m} \frac{|\psi^*|}{|f_\varepsilon^*|} \right)^2 \right\} \\ &+ \frac{2 \left( \int_{|t| \leq \pi m} |\psi^*(t)| dt \right)^2 \sum_{k \geq 0} \beta_k}{n}. \end{aligned} \tag{11}$$

In particular, if  $K_\psi := \int |\psi^*(t)| dt < +\infty$ , then the last term in the right-hand side of (11) is of order  $O(1/n)$ . Moreover, in any case, we have in (11),

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

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

Context (D2) is related to ARCH models. Indeed, general ARCH models can be formulated as follows: Let  $(\eta_i)$  be an i.i.d. noise sequence.

$$Y_i = \sigma_i \eta_i \quad \text{with } \sigma_i = F(\eta_{i-1}, \eta_{i-2}, \dots), \quad (12)$$

for some measurable functions  $F$ , or

$$Y_i = \sigma_i \eta_i \quad \text{with } \sigma_i = F(\sigma_{i-1}, \eta_{i-1}) \text{ and } \sigma_0 \text{ independent of } (\eta_i)_{i \geq 0}. \quad (13)$$

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  $\beta$ -mixing; we refer to Comte *et al.* [12] for a review of the examples and 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). We can prove the following result in this context:

**Proposition 2.3.** *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 (7). Then*

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

Thus, the procedure attains the rates of the independent case as soon as, for some constant  $C$ ,

$$\left( \int_{-\pi n}^{\pi n} |\psi^*| \right) \left( \int_{-\pi n}^{\pi n} \frac{|\psi^*|}{|f_\varepsilon^*|} \right) \leq C \min \left\{ \int_{-\pi n}^{\pi n} \frac{|\psi^*|^2}{|f_\varepsilon^*|^2}, \left( \int_{-\pi n}^{\pi n} \frac{|\psi^*|}{|f_\varepsilon^*|} \right)^2 \right\}.$$

This does not hold in general, but in particular cases. For instance, if  $f_\varepsilon$  satisfies (2) and if  $\psi$  satisfies (4) together with

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

with either  $\gamma > \max(B, 1)$  or  $(A > 0, \rho > 0)$ , then, under the assumptions of Proposition 2.3,

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

where  $K$  is a constant.

It follows from (16) that the rates given in Table 1 are preserved in this  $\beta$ -mixing context whenever the  $\varepsilon_i$  are supersmooth.

Taking  $\psi^*(t) = e^{itx_0}$  for any  $x_0$  (as in Section 4.1 below) allows one to provide a pointwise density estimator that retains the rate of the independent case if  $\gamma > 1$ . We recover the results obtained by the kernel estimator of van Es *et al.* [29]. Our results are more general since van Es *et al.* [29] only consider a multiplicative Gaussian noise (implying supersmooth  $\varepsilon_i$ , see Section 4.2) 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.

### 3. Adaptive estimation

Now, we provide a strategy leading to an automatic choice of  $m$ . Note that such 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  $\psi$  and  $f_\varepsilon$  are assumed to be known, these conditions can be explicitly checked.

Let us describe briefly the heuristics that follow Laurent *et al.* [24]. Let  $\theta_m = \mathbb{E}(\hat{\theta}_m) = (2\pi)^{-1} \int_{-\pi m}^{\pi m} g^*(t) \psi^*(-t) dt$ . The approximation of the bias of  $(\theta(g) - \theta_m)^2$  is obtained by replacing it by  $(\theta_j - \theta_m)^2$  for  $j \geq m$ ,  $j$  great enough, and then by  $(\hat{\theta}_j - \hat{\theta}_m)^2$ . This approximation in turn introduces a bias which must be corrected (see  $H(j, m)$  below). The variance term is replaced by a penalty function  $\text{pen}(\cdot)$  from  $\mathbb{N}$  into  $\mathbb{R}^+$ . This gives the theoretical criterion

$$\text{Crit}(m) = \sup_{j \geq m} (\theta_j - \theta_m)^2 + \text{pen}(m),$$

where  $\text{pen}(m)$  has the order of the variance term (see Section 2.2) and its empirical version is

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

where  $H(j, m)$  is an additional bias correction and  $\mathcal{M}$  is a subset of  $\mathbb{N}$ . Then, we can define

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

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

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

$$\text{pen}(m) = 4 \left( 1 + \frac{1}{a} \right) (x_m \sigma_m^2 + x_m^2 c_m^2), \quad (18)$$

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

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

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

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

We shall prove the following theorem:

**Theorem 3.1.** *Consider model (1) where  $(X_i)_{1 \leq i \leq n}$  and  $(\varepsilon_i)_{1 \leq i \leq n}$  are independent sequences of i.i.d. random variables and assume that (5) is fulfilled. Let  $\hat{\theta}_{\hat{m}}$  be defined by (7) and (17)–(19) when  $\int |\psi^*/f_\varepsilon^*| = +\infty$  and  $\int |\psi^*/f_\varepsilon^*|^2 = +\infty$ . Then there exists some positive constant  $C(a)$  depending only on some  $a > 0$ , such that*

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

where  $\omega_m^2 = \sigma_m^2 \vee c_m + 2(\sigma_m^2 \vee c_m)^2$  and  $C'$  is a constant.

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  $\text{pen}(m)$ , if  $x_m$  are chosen so that  $\sum_m e^{-x_m} \omega_m^2 = O(1/n)$ . However, as the main term in  $\text{pen}(m)$  is clearly  $x_m \sigma_m^2$ , where  $\sigma_m^2$  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  $x_m$  in order to see what loss occurs, if any, when using the adaptive procedure. We assume that  $b + B > 1$ , so that we can take  $\mathcal{M} = \{1, 2, \dots, [\sqrt{n}]\}$ , where  $[\sqrt{n}]$  is the greatest integer less than  $\sqrt{n}$ . The possible choices for  $x_m$  are discussed with respect to the upper bounds on the variance given in Section 2.2:

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

- Case (IV): We take  $x_m = 4\alpha(\pi m)^\rho$ . There is no loss in case (10) if  $\rho > 0$ ,  $r = R = 0$ . In the two other cases, (8) and (9), a loss in the variance occurs. If the bias is dominating (if  $r > \rho$ ), there is no loss in the rate. Otherwise, as the optimal  $m$  is less than  $(\ln(n)/C)^{1/\rho}$ , for some  $C > 0$ , the loss in the rate is at most of order  $O(\ln(n))$ . Note that the rate being faster than logarithmic in this case, the loss remains negligible with respect to the rate.

The adaptive rates are given in the last column of Table 1. Let us emphasize that the rates presented in both the second and third columns of Table 1 are new in such a general setup.

Moreover, if we want to extend the adaptive result to the mixing case, we can use the Bernstein inequality given in Doukhan [15] or in Butucea and Neumann [5] 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  $f_\varepsilon$  satisfying (2) and  $\psi$  satisfying (4) and (15) with either  $\gamma > \max(B, 1)$  or  $A, \rho > 0$ , and assume in both cases that  $\beta_k \leq e^{-ck}$  for any  $k \in \mathbb{N}$ . Then if (5) is fulfilled and if  $\int |\psi^*(t)| dt < +\infty$ ,  $\int |\psi^*/f_\varepsilon^*| = +\infty$  and  $\int |\psi^*/f_\varepsilon^*|^2 = +\infty$ , the result of Theorem 3.1 for  $\hat{\theta}_n^m$  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  $\sigma_m^2, \sigma_{j,m}^2$  multiplied by 2.*

Clearly, the constant  $c$  appearing in the  $c_m, c_{m,j}$  is unknown, but these terms have in general negligible orders when compared to the  $\sigma_m^2, \sigma_{j,m}^2$ . In that case, these terms can be omitted in the definition of the estimator and the procedure does not depend on the mixing coefficients (see the example in Section 4.2).

## 4. Applications

### 4.1. Pointwise estimation

Pointwise estimation of  $g$ , also called pointwise deconvolution, is a particular case of our general setting and the most studied example in the literature. In this section, we give a full description of minimax and adaptive rates.

We check that our estimation procedure attains the minimax and adaptive rates (when known) in this context and that it provides the rates for the other setups. Very few results are available on the optimality of the rates in the adaptive setup and we prove here such results.

Let  $\Lambda = [\underline{b}, \bar{b}] \times [\underline{a}, \bar{a}] \times [\underline{r}, \bar{r}] \times [\underline{L}, \bar{L}] \subset [0, \infty) \times [0, \infty) \times (0, 2] \times (0, \infty)$  be a set of parameters  $\lambda = (b, a, r, L)$ . We shall denote by  $\varphi_n$  the minimax rate of convergence over the class  $\mathcal{S}(\lambda)$ ; see, for example, Butucea [4] for a definition. We shall say that an estimator is adaptive minimax over the family of classes  $\mathcal{S}(\lambda)$ ,  $\lambda \in \Lambda$ , if it attains the minimax rate  $\varphi_n$  uniformly in  $\lambda$ .

It is not always possible to attain the minimax rate uniformly over a set of parameters  $\Lambda$ . It may happen that there is a loss in the rate due to adaptation, see Lepski [26]. We shall say that an estimator is adaptive for the adaptive rate  $\phi_n$  if it attains this rate uniformly in  $\lambda$  over  $\Lambda$  and if, moreover, the lower bounds hold for this rate uniformly in  $\lambda$  over  $\Lambda$ . For a definition, see Butucea [3].

For pointwise estimation of  $g$ , we can take  $\psi(x) = \delta_{\{x_0\}}(x)$  for any given  $x_0$ , where  $\delta_{\{x_0\}}$  is the Dirac measure at  $x_0$ . This implies  $\psi^*(t) = e^{itx_0}$  and  $|\psi^*(t)| = 1$ . Therefore, the rates correspond

to the particular case  $B = A = R = 0$  in (4) and in Table 1. They are summarized more simply in Table 2. Our procedures attain the rates already found in pointwise deconvolution and cover all other previously unknown setups.

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

$$m_n^{2b+2\gamma+(1-\rho)+} \exp\{2\alpha(\pi m_n)^\rho + 2a(\pi m_n)^r\} = O(n). \quad (20)$$

Consequently, the rate of  $\hat{g}_{m_n}$  is not easy to give explicitly and depends on the ratio  $r/\rho$ . If  $r/\rho$  or  $\rho/r$  belongs to  $]k/(k+1); (k+1)/(k+2)[$  with integer  $k$ , the rate of convergence can be expressed as a function of  $k$ . For explicit formulae for the rates, see Lacour [23].

These rates are known to be optimal in the minimax sense as indicated in Table 2. The case  $r = 0$  is studied by Fan [16], the case  $r = 0, \rho > 0$  by Cator [9] and the case  $r > 0, \rho = 0$  by Butucea [4]. The rate in the case  $r > 0, \rho > 0, \gamma = 0$  is proven optimal in the minimax sense in Butucea and Tsybakov [6] for  $r \leq \rho$ . By using their construction and by following the same proof, we get near optimality (within a log factor) in the case  $r > \rho$ .

Very few results on adaptive pointwise estimation are available. We use  $|\psi^*(x)| = 1$  in the procedure described in Section 3, with  $c_m \leq \sigma_m^2$  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  $f_\varepsilon$  is ordinary smooth, the choice  $x_m = (2\gamma + 3) \ln(m)$  suits and if  $f_\varepsilon$  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

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

The rates still correspond to the particular case  $B = A = R = 0$  in (4) which are summarized in Table 2.

Let us mention that in the cases  $\rho > 0, \alpha > 0$  and  $r > 0, a > 0$  (i.e., both  $f_\varepsilon$  and  $g$  are supersmooth), then  $x_m$  is of order  $m^\rho$ . There is no loss due to adaptation if  $r < \rho$  as noticed earlier by Butucea and Tsybakov [6], but, surprisingly, we notice a loss of order  $[\ln(n)]^{\rho/r}$  if  $r > \rho$  associated to a rate faster than any power of logarithm. If  $r = \rho$ , the loss is logarithmic and the rate polynomial.

The previously defined estimator  $\hat{\theta}_{m_n}$  with  $m_n$  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  $\phi_n$ , that is, it may happen that  $\varphi_n = o(\phi_n)$ . Therefore, we check that the loss with respect to the minimax rate, when it occurs, is unavoidable.

**Theorem 4.1.** *The rates  $\phi_n$  defined in Table 2 are adaptive rates and when either  $\rho = 0$  or ( $r \geq \rho > 0$  and  $\rho \leq 1$ ) the loss with respect to the minimax rate which appears (compare in Table 2,  $\varphi_n^2$  and  $\phi_n^2$ ) is optimal, that is, it satisfies the following lower bounds:*

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

**Table 2.** Choice of  $m_n$  for pointwise deconvolution and corresponding rates under assumptions (2) and (3). Adaptive rates for comparison.  $B_m$  is abbreviated for  $m^{-2b+1-r} \exp(-2\alpha(\pi m)^r)$  and  $V_m$  for  $m^{2\gamma+1-\rho+(1-\rho)+} \exp(2\alpha(\pi m)^\rho)/n$

	$\rho = 0$ ordinary smooth	$\rho > 0$ supersmooth
$r = 0$ Sob.(b)	$\pi m_n = n^{1/(2b+2\gamma)}$ $\varphi_n^2 = O(n^{-(2b-1)/(2b+2\gamma)})$ <i>minimax rate</i> (Fan [16])  $\phi_n^2 = O((n/\ln(n))^{-(2b-1)/(2b+2\gamma)})$ <i>adaptive rate</i> (NEW)	$\pi m_n = [\ln(n)/(2\alpha + 1)]^{1/\rho}$ $\varphi_n^2 = O((\ln(n))^{-(2b-1)/\rho})$ <i>minimax rate</i> (Fan [16])  $\phi_n^2 = O((\ln(n))^{-(2b-1)/\rho})$ <i>adaptive minimax rate (no loss)</i> (Cator [9])
$r > 0$ $C^\infty$	$\pi m_n = [\ln(n)/2b]^{1/r}$ $\varphi_n^2 = O(\frac{\ln(n)^{(2\gamma+1)/r}}{n})$ <i>minimax rate</i> (Butucea [3])  $\phi_n^2 = O(\frac{\ln \ln(n) \ln(n)^{(2\gamma+1)/r}}{n})$ <i>adaptive rate</i> (NEW)	$m_n$ solution of (20) $= \ln(n) - (\ln \ln(n))^2$ $\varphi_n^2 = O(B_{m_n})$ : <i>minimax rate</i> if $r < \rho$ $\hookrightarrow$ (Butucea and Tsybakov [6]) $\varphi_n^2 = O(V_{m_n})$ : <i>minimax rate</i> if $r \geq \rho$ , $\rho \leq 1$ (NEW) $\phi_n^2 = O(m_n^{\rho I(r \geq \rho)} \varphi_n^2)$ <i>adaptive minimax rate</i> if $r < \rho$ (no loss) $\hookrightarrow$ (Butucea and Tsybakov [6]) <i>adaptive rate</i> if $r \geq \rho$ , $\rho \leq 1$ (NEW)

for  $n$  large enough, where the infimum is taken over all possible estimators  $\theta_n$ , under the additional hypothesis that the noise density is three-times continuously differentiable and

$$\text{for polynomial noise } |f'_\varepsilon(u)| \leq C \frac{1}{|u|^{\gamma+1}}, \quad \text{as } |u| \rightarrow \infty \tag{21}$$

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

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

**Remark 4.1.** Note that the adaptive property of  $\hat{\theta}_m$  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 [6]).

## 4.2. Stochastic volatility model

In this section, we consider the discrete time stochastic volatility model. Let  $\eta_i$  be an i.i.d. centered noise process,  $\mathbb{E}(\eta_i^2) = 1$  and let  $V_i$  be a sequence of positive random variables. Assume

that we observe  $U_1, \dots, U_n$ , where

$$U_i = \sqrt{V_i} \eta_i, \quad i = 1, \dots, n. \quad (23)$$

Then the conditional variance of  $U_i$  given  $V_i$  equals  $V_i$  which explains that  $V_i$  is called the volatility process. In many contexts, this process is the process of interest. We assume moreover,  $(V_i)$  and  $(\eta_i)$  are independent and  $(V_i)$  is a stationary  $\beta$ -mixing process with  $\beta$ -mixing coefficients denoted by  $(\beta_k)$ . When this model is obtained as the discretization of a set of continuous time stochastic differential equations,  $V_i$  is indeed geometrically  $\beta$ -mixing, and  $\eta_i \sim \mathcal{N}(0, 1)$ ; see Comte and Genon-Catalot [13].

Model (23) is also considered in this form by van Es *et al.* [29] among others, under the assumption  $\eta_i \sim \mathcal{N}(0, 1)$ . Setting

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

allows us to write (23) in the form (1). Then, we note that if  $\eta_1 \sim \mathcal{N}(0, 1)$ ,

$$f_\varepsilon^*(x) = \frac{2^{ix}}{\sqrt{\pi}} \Gamma(1 + ix) \quad \text{and} \quad |f_\varepsilon^*(x)| \sim_{|x| \rightarrow +\infty} \sqrt{2/e} e^{-\pi|x|/2}, \quad (24)$$

by using the Stirling formula  $\Gamma(z) \sim_{|z| \rightarrow +\infty} \sqrt{2\pi} z^{z-1/2} e^{-z}$ . We recognize (2) with  $\gamma = 0$ ,  $\alpha = \pi/2$  and  $\rho = 1$ .

Applying the results of Section 4.1 in the mixing context (D1) (see Proposition 2.2 and Corollary 3.1), we deduce that, if  $V$  is geometrically  $\beta$ -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 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 the examples in Comte and Genon-Catalot [13]). Therefore, we recover as a particular case, and substantially improve the result of van Es *et al.* [29], 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 [17] or Chaleyat-Maurel and Genon-Catalot [11]) and, in particular, it is natural to consider more general types of distributions for  $\eta$ . For instance, we suppose now that  $\eta^2$  follows a Gamma distribution, that is,  $f_{\eta^2}(x) = (e^{-x} x^{p-1} / \Gamma(p)) \mathbf{1}_{x>0}$ . In that case, we find

$$f_\varepsilon^*(x) = \frac{\Gamma(ix + p)}{\Gamma(p)} \quad \text{and} \quad |f_\varepsilon^*(x)| \sim_{|x| \rightarrow +\infty} \frac{\sqrt{2\pi} e^{-p}}{\Gamma(p)} |x|^{p-1/2} e^{-\pi|x|/2}, \quad (25)$$

that is,  $\varepsilon$  is supersmooth with  $\gamma = p - 1/2$ ,  $\alpha = \pi/2$  and  $\rho = 1$  in (2). The Gaussian case corresponds to  $p = 1/2$ . Let us recall that the Laplace transform  $Lu$  of a real valued function

$u$  is defined by  $Lu(x) = \int e^{-xt}u(t) dt$  as soon as it exists, and the Laplace transform of a non-negative random value  $Y$  is defined by  $\mathbb{E}(e^{-\lambda Y})$ . In this context, let  $\pi$  denote the density of  $V_1$ , and consider that we are interested in estimating the Laplace transform of  $V_1$ . In fact, our general method provides an estimator of  $h(\lambda) = -(L\pi)'(\lambda) = \mathbb{E}(V_1 e^{-\lambda V_1})$ , that is, minus the derivative of the Laplace transform of  $\pi$ . 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  $\lambda > 0$ ,

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

and

$$\psi_\lambda^*(x) = \lambda^{-1-ix} \Gamma(1+ix) \sim_{|x| \rightarrow +\infty} \frac{\sqrt{2\pi}}{e\lambda} \sqrt{|x|} e^{-\pi|x|/2}, \quad (26)$$

(i.e.,  $B = 1/2$ ,  $A = \pi/2$  and  $R = 1$  in (4)). Let us define

$$\hat{h}_m(\lambda) = \frac{1}{2\pi n} \sum_{k=1}^n \int_{|t| \leq \pi m} e^{itZ_k} \frac{\psi_\lambda^*(t)}{f_\varepsilon^*(t)} dt \quad (27)$$

with  $f_\varepsilon^*$  and  $\psi_\lambda^*$  given by (25) and (26). Then, taking into account the orders of  $f_\varepsilon^*$  and  $\psi_\lambda^*$ , we obtain, by applying inequality (11) of Proposition 2.2 and if  $p \neq 3/2$ :

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

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

**Proposition 4.1.** *Consider model (23) with (D1), (25) and (26). Assume that  $(X_k) = (\ln(V_k))$  is  $\beta$ -mixing with  $\sum_k \beta_k < +\infty$ , then  $\hat{h}_m$  defined by (27) satisfies, for  $\lambda > 0$ ,*

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

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

In other words, using Table 1 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  $g$  is.

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  $x_m = 4 \ln(m)$ :

**Proposition 4.2.** *Consider the stochastic volatility model (23) with (D1), (25) and (26). Assume that  $(X_i)$  is geometrically  $\beta$ -mixing and consider  $\hat{h}_m$  defined by (27), with  $\hat{m}$  defined by (17). For*

any  $\lambda > 0$ , and  $p \leq 3/2$

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

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. Indeed for  $\eta \sim \mathcal{N}(0, 1)$ ,  $\mathbb{E}(e^{i\sqrt{2\lambda}U_1}) = \mathbb{E}[\mathbb{E}(e^{i\sqrt{2\lambda}V_1}\eta_1 | V_1)] = \mathbb{E}(e^{-\lambda V_1})$ . Therefore the Laplace transform of  $\pi$ ,  $L\pi(\lambda)$  can be directly estimated by an empirical mean of the  $\exp(i\sqrt{2\lambda}U_k)$ , 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  $p$ . But the last strategy here exploits an additional assumption ( $\eta$  is Gaussian) which the general methods do not take into account.

## 5. Proofs

### 5.1. Selected $m$ for Table 1

The squared bias variance compromise is performed via the following choices of  $m$ , denoted by  $m_n$ , in the cases enumerated in Table 1:

- (1) (2) and (3) (a) Case  $r \vee R = 0$ ,  $m_n = O(n^{1/(2b+2B-1)})$  as  $2b - 1 > 0$  when  $r = 0$ .  
 (b) Case  $r \vee R > 0$ ,  $\pi m_n = (\ln(n)/C)^{1/(r \vee R)}$  for some  $C \geq A + a$ .
- (4) Optimal  $m_n$  is such that  $2a(\pi m_n)^r + 2A(\pi m_n)^R = \ln(n) - (2b + 2B - 1) \ln(m_n)$ . Take, e.g.,  $\pi m_n = (\ln(n)/C)^{1/(r \vee R)}$  with sufficiently large  $C > 0$ .
- (5) Take  $m_n = O(n^{1/(2b+2B-1)})$ .
- (6) Optimal  $m_n$  is such that  $2a(\pi m_n)^r + 2A(\pi m_n)^R = \ln(n) - (2b + 2\gamma) \ln(m_n)$ , which gives  $\pi m_n = (\ln(n)/(2a) - A/a(\ln(n)/(2a))^{R/r} - (b + \gamma)/(ar) \ln(\ln(n)))^{1/r}$  if  $r \geq R$  and exchange  $R$  and  $r$  in the last expression if  $R > r$ . For an easier choice, take, for example,  $m_n = (\ln(n)/C)^{1/(r \vee R)}$  for  $C > 0$  large enough.
- (7)  $m_n = O(n^{1/(2b+2\gamma)})$ ,  $b + \gamma > 0$ . (8) and (9) already discussed.
- (10) The optimal  $m_n$  is  $\pi m_n = (\ln(n)/(2\alpha) - (b + \gamma)/(\alpha\rho) \ln(\ln(n)))^{1/\rho}$ . For a simpler form it is sufficient to take, for example,  $\pi m_n = (\ln(n)/(4\alpha))^{1/\rho}$ .

The parameters  $a, b, r$  of the unknown function appear several times to select  $m_n$ . As  $g$  is unknown, and thus  $a, b, r$  are unknown, it is not possible to select  $m_n$  in all the cases where the rate is slower than the parametric rate  $n^{-1}$ .

## 5.2. Proof of Theorem 3.1

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

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

We define

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

and

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

It is sufficient to prove the following theorem:

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

$$\begin{aligned} \mathbb{E}[(\hat{\theta}_m - \theta)^2] &\leq C(a) (\text{Crit}(m_{\text{opt}}) + \Xi(m_{\text{opt}})) \\ &\quad + C(a) \left( \sum_{m \in \mathcal{M}} e^{-x_m} \omega_m^2 + \sum_{j \geq m_{\text{opt}}} e^{-x_{j,m_{\text{opt}}}} \omega_{j,m}^2 + \frac{1}{n} \right), \end{aligned}$$

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

First, note that Theorem 5.1 implies Theorem 3.1. Indeed, note that for  $j \geq m$ , we have  $\sigma_{m,j}^2 \leq \sigma_j^2$  and  $c_{m,j} \leq c_j$ . Therefore, choosing  $x_{m,j} = x_j$  implies that

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

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

Now we establish the following lemma:

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

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

**Proof.** Recall that the Bernstein inequality for a sum  $S_n = \sum_{k=1}^n Y_k$  of i.i.d. random variables  $Y_k$  having  $\text{var}(Y_1) \leq v^2$  and  $\|Y_1\|_\infty \leq 1/a$  states that

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

We put for  $j \geq m$

$$Y_k = Y_k(j, m) = \frac{1}{2\pi} \int_{\pi m \leq |t| \leq \pi j} e^{itZ_k} \frac{\psi_\varepsilon^*(t)}{f_\varepsilon^*(t)} dt. \quad (28)$$

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

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

Now, from the simple fact that  $(x+y)^2 \leq (1+1/a)x^2 + (1+a)y^2$  for any real numbers  $x, y$ , we deduce by setting  $u = y$  and  $v = x + y$  that  $(v-u)^2 \geq (1/(1+1/a))v^2 - (1+a)/(1+1/a)u^2$ . Use also the fact that  $(A+B)^2 \leq 2(A^2+B^2)$  for any real numbers  $A, B$ , to obtain

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

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

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

To conclude we write

$$\begin{aligned} & \mathbb{P}\left(\widehat{\text{Crit}}(m) > (1+a)\text{Crit}(m) + 4\left(1 + \frac{1}{a}\right)(x + x^2)\right) \\ & \leq \mathbb{P}\left\{\exists j \geq m, j \in \mathcal{M}, (\hat{\theta}_j - \hat{\theta}_m)^2 - H(j, m) \geq (1+a)(\theta_j - \theta_m)^2 + 4\left(1 + \frac{1}{a}\right)(x + x^2)\right\} \\ & \leq 2 \sum_{j \geq m, j \in \mathcal{M}} e^{-x_{j,m}} e^{-x/(\sigma_{j,m}^2 \vee c_{j,m})}. \end{aligned}$$

This ends the proof of Lemma 5.1. □

Now we follow the steps of the proof of Laurent *et al.* [24].

- We first consider the case where  $\hat{m} \leq m_{\text{opt}}$ . Following the same lines of proof, we get

$$\begin{aligned} & \mathbb{P}\left(\frac{1}{2}(\hat{\theta}_{\hat{m}} - \theta(g))^2 > (1+a)\text{Crit}(m_{\text{opt}}) + 4\left(1 + \frac{1}{a}\right)(x + x^2)\right. \\ & \quad \left. + \sup_{j \leq m_{\text{opt}}} H(m_{\text{opt}}, j) + (\hat{\theta}_{m_{\text{opt}}} - \theta(g))^2 + \frac{1}{n} \cap \{\hat{m} \leq m_{\text{opt}}\}\right) \\ & \leq \sum_{j \geq m_{\text{opt}}} e^{-x_{j, m_{\text{opt}}}} e^{-x/(\sigma^2 j, m_{\text{opt}} \vee c_{j, m_{\text{opt}}})}. \end{aligned} \quad (29)$$

- Now we consider the case  $\hat{m} > m_{\text{opt}}$ . We apply the 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 5.1. We obtain, for all  $m \in \mathcal{M}$ ,

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

This implies that

$$\begin{aligned} & \mathbb{P}\left((\hat{\theta}_{\hat{m}} - \theta(g))^2 \geq (1+a)(\theta_{\hat{m}} - \theta(g))^2 + 4\left(1 + \frac{1}{a}\right)(x + x^2) + \text{pen}(\hat{m})\right) \\ & \leq \sum_{m \in \mathcal{M}} 2e^{-x_m} e^{-x/(\sigma_m^2 \vee c_m)}. \end{aligned}$$

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

$$\begin{aligned} & \mathbb{P}\left((\hat{\theta}_{\hat{m}} - \theta(g))^2 \geq (1+a)(\theta_{\hat{m}} - \theta(g))^2 + 4\left(1 + \frac{1}{a}\right)(x + x^2) + \widehat{\text{Crit}}(m_{\text{opt}}) + \frac{1}{n}\right) \\ & \leq \sum_{m \in \mathcal{M}} 2e^{-x_m} e^{-x/(\sigma_m^2 \vee c_m)}. \end{aligned}$$

If  $\hat{m} > m_{\text{opt}}$ , then  $(\hat{\theta}_m - \theta(g))^2 \leq \sup_{j \geq m_{\text{opt}}} (\theta_j - \theta(g))^2$  and we apply Lemma 5.1 with  $m = m_{\text{opt}}$ .

This yields

$$\begin{aligned} \mathbb{P}\left(\left(\hat{\theta}_{\hat{m}} - \theta(g)\right)^2 \geq (1+a)\left(\sup_{j \geq m_{\text{opt}}} (\theta_j - \theta(g))^2 + 8\left(1 + \frac{1}{a}\right)(x + x^2)\right.\right. \\ \left.\left. + (1+a)\text{Crit}(m_{\text{opt}}) + \frac{1}{n} \cap \{\hat{m} > m_{\text{opt}}\}\right)\right) \\ \leq \sum_{m \in \mathcal{M}} 2e^{-x_m} e^{-x/(\sigma_m^2 \vee c_m)} + \sum_{j \geq m_{\text{opt}}} 2e^{-x_{j,m_{\text{opt}}}} e^{-x/(\sigma_{j,m_{\text{opt}}}^2 \vee c_{j,m_{\text{opt}}})}. \end{aligned} \quad (30)$$

Let

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

and

$$X = (\hat{\theta}_{\hat{m}} - \theta(g))^2, \quad Y = 2(\hat{\theta}_{m_{\text{opt}}} - \theta(g))^2.$$

It follows from (29) and (30) that, for all  $x > 0$ ,

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

We write that  $\mathbb{E}(X) = \mathbb{E}(X\mathbf{I}_{X \geq Y + C_{m_{\text{opt}}}}) + \mathbb{E}(X\mathbf{I}_{X \leq Y + C_{m_{\text{opt}}}}) \leq \mathbb{E}[(X - Y - C_{m_{\text{opt}}})_+] + \mathbb{E}(Y + C_{m_{\text{opt}}})$ .

Then, setting  $C_a = 24(1 + 1/a)$  and  $Z = X - Y - C_{m_{\text{opt}}}$

$$\begin{aligned} \mathbb{E}[Z_+] &= \int_0^{+\infty} \mathbb{P}(Z > t) dt = C_a \left( \int_0^1 \mathbb{P}(Z > C_a u) du + \int_1^\infty \mathbb{P}(Z > C_a u) du \right) \\ &= C_a \left( \int_0^1 \mathbb{P}(Z > C_a(u \vee u^2)) du + 2 \int_1^\infty \mathbb{P}(Z > C_a(v \vee v^2)) v dv \right), \end{aligned}$$

$$\begin{aligned} \mathbb{E}[(X - Y - C_{m_{\text{opt}}})_+] &\leq C_a \sum_{m \in \mathcal{M}} 2e^{-x_m} (\sigma_m^2 \vee c_m + 2(\sigma_m^2 \vee c_m)^2) \\ &\quad + C_a \sum_{j \geq m_{\text{opt}}} 2e^{-x_{j,m_{\text{opt}}}} (\sigma_{j,m_{\text{opt}}}^2 \vee c_{j,m_{\text{opt}}} + 2(\sigma_{j,m_{\text{opt}}}^2 \vee c_{j,m_{\text{opt}}})^2) \\ &= C_a \left( \sum_{m \in \mathcal{M}} 2e^{-x_m} \omega_m^2 + \sum_{j \geq m_{\text{opt}}} 2e^{-x_{j,m_{\text{opt}}}} \omega_{j,m_{\text{opt}}}^2 \right). \end{aligned}$$

The end of the proof is the same as in Laurent *et al.* [24].

### 5.3. Proof of Proposition 2.2

The same decomposition of the risk and upper bound for the bias hold, as in Section 2.2. Only the variance has to be re-examined. The basic idea is that, for  $k \neq \ell$ ,  $\text{cov}(e^{itZ_k}, e^{isZ_\ell}) = f_\varepsilon^*(t) f_\varepsilon^*(-s) \text{cov}(e^{itX_k}, e^{isX_\ell})$  by conditioning on  $(X_k, X_\ell)$ . The additional trick is the standard covariance inequality for  $\beta$ -mixing variables (see, e.g., Doukhan [15]), which implies that  $|\text{cov}(e^{itX_k}, e^{isX_\ell})| \leq \beta_{|k-\ell|}$ .

$$\begin{aligned} \text{Var}(\hat{\theta}_m) &= \frac{1}{4\pi^2 n^2} \sum_{k, \ell=1, k \neq \ell}^n \int_{-\pi m}^{\pi m} \int_{-\pi m}^{\pi m} \text{cov}(e^{itX_k}, e^{isX_\ell}) \psi^*(t) \psi^*(-s) ds dt \\ &\quad + \frac{1}{4\pi^2 n^2} \sum_{k=1}^n \int_{-\pi m}^{\pi m} \int_{-\pi m}^{\pi m} \text{cov}(e^{itZ_k}, e^{isZ_k}) \frac{\psi^*(t)}{f_\varepsilon^*(t)} \frac{\psi^*(-s)}{f_\varepsilon^*(-s)} ds dt. \end{aligned} \quad (31)$$

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

$$\begin{aligned} &\frac{2}{4\pi^2 n^2} \sum_{1 \leq k < \ell \leq n} \int_{-\pi m}^{\pi m} \int_{-\pi m}^{\pi m} |\text{cov}(e^{itX_1}, e^{isX_{\ell-k}})| |\psi^*(t) \psi^*(-s)| ds dt \\ &\leq \frac{1}{2\pi n} \sum_{k=1}^n \beta_k \left( \int_{-\pi m}^{\pi m} |\psi^*(t)| dt \right)^2. \end{aligned}$$

This gives the result.

### 5.4. Proof of Proposition 2.3

Under (D2), we only obtain that for  $k < \ell$ ,  $\text{cov}(e^{itZ_k}, e^{isZ_\ell}) = f_\varepsilon^*(-s) \text{cov}(e^{itZ_k}, e^{isX_\ell})$  by conditioning on  $(X_\ell)$ . The covariance inequality for  $\beta$ -mixing variables (see, e.g., Doukhan [15]) still applies (but to the variables  $(X_k, Z_k)$  and  $(X_\ell, Z_\ell)$ ) and implies that  $|\text{cov}(e^{itZ_k}, e^{isX_\ell})| \leq \beta_{|k-\ell|}$ . Then (31) remains true but leads, for the bound of the modulus of the last term, to:

$$\begin{aligned} &\frac{2}{4\pi^2 n^2} \sum_{1 \leq k < \ell \leq n} \int_{-\pi m}^{\pi m} \int_{-\pi m}^{\pi m} |\text{cov}(e^{itZ_1}, e^{isX_{\ell-k}})| \left| \frac{\psi^*(t)}{f_\varepsilon^*(t)} \psi^*(-s) \right| ds dt \\ &\leq \frac{1}{2\pi n} \sum_{k=1}^n \beta_k \left( \int_{-\pi m}^{\pi m} |\psi^*(t)| dt \right) \left( \int_{-\pi m}^{\pi m} \left| \frac{\psi^*(t)}{f_\varepsilon^*(t)} \right| dt \right). \end{aligned}$$

This gives inequality (14).

For the proof of (16), the result follows from the inequality

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

and the fact that the new mixing term is always negligible with respect to the independent variance term if  $\varepsilon$  is supersmooth (case A,  $\rho > 0$ ). If  $\varepsilon$  is ordinary smooth, then we only have to study when  $m^{(-B+1)_+ + \gamma - B + 1}$  is less than  $m^{2\gamma - 2B + 1}$ , which occurs as soon as  $\gamma > \max(B, 1)$ .

### 5.5. Proof of Corollary 3.1

The main difference with respect to the proof of Theorem 3.1 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  $\beta_k \leq e^{-ck}$ ), we get from Theorem 4, page 36 in Doukhan [15] that

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

with  $\|Y_1\|_\infty \leq 1/a$  and  $(1/q)\text{Var}(\sum_{k=1}^q Y_k) \leq \tilde{v}^2$ .

In all cases,  $|\mathcal{M}| \leq n$ , 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 (28) with  $c_{j,m}$  and  $c_m$  the same as previously multiplied by  $2 \ln(n)/c$ . This gives  $\tilde{c}_{j,m} = (2 \ln(n)/2)c_{j,m}$  and  $\tilde{c}_m = (2 \ln(n)/c)c_m$ . At last, it follows from the above computation of  $\text{Var}(\tilde{\theta}_m)$  that the new variance terms denoted by  $\tilde{\sigma}_{j,m}^2, \tilde{\sigma}_m^2$  can be bounded under (D1) by

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

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

### 5.6. Proof of Theorem 4.1

We describe first the general procedure for proving the theorem and postpone details of constructions and proofs to Section 5.6. As the adaptation loss is different according to whether  $r = 0$  or  $r \neq 0$ , respectively  $\rho = 0$  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  $r \geq \rho$ . We take  $b = 0$  without loss of generality, in order to simplify polynomial factors in our explicit constructions.

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

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

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

$$\begin{aligned} & \inf_{\theta_n} \sup_{\lambda \in \Lambda} \sup_{g \in \mathcal{S}(\lambda)} \phi_{n,\lambda}^{-2} \mathbb{E}_g [|\theta_n - \theta(g)|^2] \\ & \geq \inf_{\theta_n} \max \left\{ \phi_{n,\lambda}^{-2} \mathbb{E}_{g_0} [|\theta_n - \theta(g_0)|^2], \phi_{n,\underline{\lambda}}^{-2} \mathbb{E}_{g_{1,n}} [|\theta_n - \theta(g_{1,n})|^2] \right\} \\ & \geq \inf_{T_n} \max \left\{ q_n^2 \mathbb{E}_{g_0} [T_n^2], \mathbb{E}_{g_{1,n}} [T_n - G(0, m)/\phi_{n,\underline{\lambda}}]^2 \right\}, \end{aligned}$$

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

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 [28] we can deduce that, if  $|G(0, m)/\phi_{n,\underline{\lambda}}| \geq c > 0$  and if for some fixed  $0 < \epsilon < 1$  and  $\tau > 0$

$$\mathbb{P}_1 \left( \frac{d\mathbb{P}_0}{d\mathbb{P}_1} \geq \tau \right) \geq 1 - \epsilon, \quad (32)$$

then

$$\inf_{T_n} \max \left\{ q_n^2 \mathbb{E}_0 [T_n^2], \mathbb{E}_1 [T_n - G(0, m)/\phi_{n,\underline{\lambda}}]^2 \right\} \geq \frac{\tau q_n^2 \epsilon^2 c^4 (1 - \epsilon)^2}{\tau q_n^2 \epsilon^2 c^2 + (1 - \epsilon)^2 c^2}. \quad (33)$$

If we can choose  $\tau = \tau_n$  such that  $\tau_n q_n^2 \rightarrow \infty$  with  $n$ , then the bound from below in (33) 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 (33) may provide the exact asymptotic constant in case  $c \rightarrow 1$  and  $\mathbb{P}_1(d\mathbb{P}_0/d\mathbb{P}_1 \geq \tau_n) \rightarrow 1$  as  $n \rightarrow \infty$ .

In order to deal with (32), we proceed as follows:

$$\begin{aligned} \mathbb{P}_1 \left( \frac{d\mathbb{P}_0}{d\mathbb{P}_1} \geq \tau \right) &= \mathbb{P}_1 \left( \prod_{i=1}^n \frac{g_0 \star f_\epsilon}{g_{1,n} \star f_\epsilon}(Y_i) \geq \tau \right) = \mathbb{P}_1 \left( \sum_{i=1}^n \ln \left( 1 - \frac{G(\cdot - x_0) \star f_\epsilon}{g_{1,n} \star f_\epsilon}(Y_i) \right) \geq \ln(\tau) \right) \\ &= \mathbb{P}_1 \left( \frac{\sum_{i=1}^n Z_{i,n} - n\mathbb{E}_1(Z_{1,n})}{(n\text{Var}_1(Z_{1,n}))^{1/2}} \geq \frac{\ln(\tau) - n\mathbb{E}_1(Z_{1,n})}{(n\text{Var}_1(Z_{1,n}))^{1/2}} \right), \end{aligned}$$

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

$$U_n := \frac{\sum_{i=1}^n Z_{i,n} - n\mathbb{E}_1(Z_{1,n})}{(n\text{Var}_1(Z_{1,n}))^{1/2}}.$$

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

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

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

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

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.

Now, we study in more detail the different cases.

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

Let us choose  $g_0$  in the class  $\mathcal{S}(\bar{b}, \bar{L}/2)$  such that  $g_0 > 0$  and  $g_0(x) \geq c|x|^{-2}$  as  $|x| \rightarrow \infty$ . We choose next the function  $G$  such that  $G(x, m) = m^{-\underline{b}+1/2} G(mx)$  and with  $G^*$  at least three-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,  $m = (c_0 \ln(n)/n)^{-1/(2\underline{b}+2\underline{\gamma})}$ . Note that  $G^*(0) = \int G = 0$ . First,  $g_{1,n}$  is a positive function with an integral equal to 1 and it belongs to  $\mathcal{S}(\underline{b}, \bar{L})$ . Indeed, for each fixed  $x$  we have  $G(x, m) \rightarrow 0$  when  $n \rightarrow \infty$  and as  $G^*$  is three times continuously differentiable that means  $|G(x, m)| \leq O(|x|^{-3}) = o(g_0(x))$  as  $|x| \rightarrow \infty$ , giving that  $g_{1,n} \geq 0$  for  $n$  large enough. Moreover,

$$\begin{aligned} \left( \int |g_{1,n}^*(u)|^2 |u|^{2\underline{b}} du \right)^{1/2} &\leq \left( \int |g_0^*(u)|^2 |u|^{2\underline{b}} du \right)^{1/2} \\ &\quad + m^{-\underline{b}-1/2} \left( \int_{1/4 \leq |u|/m \leq 1} |G^*(u/m)|^2 |u|^{2\underline{b}} du \right)^{1/2} \\ &\leq \sqrt{2\pi \bar{L}/2} + \frac{C}{c} \left( \int_{1/4}^1 \frac{|u|^{2\underline{b}}}{(1+u^{2\underline{b}})^2} du \right)^{1/2} \leq (2\pi \bar{L})^{1/2}, \end{aligned}$$

for  $c > 0$  large enough. Second,

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

We shall prove that (32) holds with  $\tau = n^{-(2\underline{\gamma}+1)/(2\underline{b}+2\underline{\gamma})}$  and together with the fact that

$$\begin{aligned} \tau q_n^2 &= \tau \frac{\phi_{n, \underline{b}}^2}{\phi_{n, \bar{b}}^2} = \tau \left( \frac{\ln(n)}{n} \right)^{-(2\underline{\gamma}+1)(\bar{b}-\underline{b})/((2\underline{b}+2\underline{\gamma})(2\bar{b}+2\underline{\gamma}))} \\ &= (\ln(n))^{-(2\underline{\gamma}+1)(\bar{b}-\underline{b})/((2\underline{b}+2\underline{\gamma})(2\bar{b}+2\underline{\gamma}))} n^{(2\underline{\gamma}+1)/(2\underline{b}+2\underline{\gamma})(\underline{b}+\underline{\gamma})/(\bar{b}+\underline{\gamma})} \end{aligned}$$

tends to infinity, with  $n$ , the proof of (33) and hence of the theorem is finished.

We can prove that for each  $x_0$

$$\sup_x \frac{|[G(m(\cdot - x_0)) \star f_\varepsilon](x)|}{f_0^Z(x)} = o(1), \quad \text{as } n \rightarrow \infty, \quad (34)$$

therefore  $f_{1,n}^Z(x) = f_0^Z(x)(1 + o(1))$ , where  $o(1) \rightarrow 0$ ,  $n \rightarrow \infty$  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 [6],  $f_0^Z(x) = g_0 \star f_\varepsilon(x) \geq C_2/x^2$ , as  $|x| \geq M$ .

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

$$\begin{aligned} \chi^2(f_0^Z, f_{1,n}^Z) &= m^{-2b+1} \int \frac{[G(m(\cdot - x_0)) \star f_\varepsilon]^2(x)}{f_{1,n}^Z(x)} dx \\ &\leq m^{-2b+1} \left( c_2 \int_{|x| \leq M} [G(m(\cdot - x_0)) \star f_\varepsilon]^2(x) dx \right. \\ &\quad \left. + c_3 \int_{|x| > M} |x|^2 [G(m(\cdot - x_0)) \star f_\varepsilon]^2(x) dx \right), \end{aligned}$$

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

$$\begin{aligned} T_1 &\leq m^{-2b-1} \frac{c_2}{2\pi} \int \left| G^* \left( \frac{u}{m} \right) f_\varepsilon^*(u) \right|^2 du \\ &\leq c_4 m^{-2b-1} \int_{m/4}^m \frac{1}{|u|^{2\gamma}} du \leq c_5 m^{-2b-2\gamma} \leq c_6 \frac{c_0 \ln(n)}{n}. \end{aligned} \quad (35)$$

For  $T_2$  we follow the similar proof in Butucea and Tsybakov [6] and use condition (21) to get

$$\begin{aligned} T_2 &\leq m^{-2b+1} \frac{c_3}{2\pi} \int \left| \frac{\partial}{\partial u} \left( \frac{1}{m} G^* \left( \frac{u}{m} \right) f_\varepsilon^*(u) \right) \right|^2 du \\ &\leq c_6 m^{-2b-1} m^{-2\gamma} = o(T_1), \quad n \rightarrow \infty. \end{aligned} \quad (36)$$

Therefore, from (35) and (36) 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 (34) 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

$$\begin{aligned} \mathbb{E}_1[Z_{1,n}] &= \int \ln \left( 1 - \frac{[G(\cdot, m) \star f_\varepsilon](x - x_0)}{f_{1,n}^Z(x)} \right) f_{1,n}^Z(x) dx \\ &\geq - \int [G(\cdot, m) \star f_\varepsilon](x - x_0) dx - \int \frac{[G(\cdot, m) \star f_\varepsilon]^2(x - x_0)}{f_{1,n}^Z(x)} dx \\ &\geq -\chi^2(f_0^Z, f_{1,n}^Z) \geq -\kappa_n, \end{aligned}$$

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,

$$\begin{aligned} \text{Var}_1(Z_{1,n}) &\leq \mathbb{E}_1(Z_{1,n}^2) = \int \ln^2 \left( 1 - \frac{[G(\cdot, m) \star f_\varepsilon](x - x_0)}{f_{1,n}^Z(x)} \right) f_{1,n}^Z(x) dx \\ &\leq \int \frac{[G(\cdot, m) \star f_\varepsilon]^2(x - x_0)}{f_{1,n}^Z(x)^2} \left( 1 + \frac{[G(\cdot, m) \star f_\varepsilon]^2(x - x_0)}{f_{1,n}^Z(x)} \right)^2 f_{1,n}^Z(x) dx \\ &\leq c_\nu \chi^2(f_0^Z, f_{1,n}^Z) \leq c_\nu \kappa_n, \end{aligned}$$

as by (34):  $\sup_x |f_0^Z(x)/f_{1,n}^Z(x)|$  is bounded from above by some constant depending only on  $g_0$  and  $f_\varepsilon$ . By similar calculations, we also check that

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

and that

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

as  $n \rightarrow \infty$  and since  $\underline{b} > 1/2$ . Next we apply Lyapunov's central limit theorem for triangular arrays, see Petrov [27], to get  $\mathbb{P}_1(U_n \geq u_n) \geq 1 - \epsilon$ , as, when  $n \rightarrow +\infty$ ,

$$0 \geq u_n = \frac{\ln(\tau) + \kappa_n}{\sqrt{c_\nu \kappa_n}} = \frac{-(2\gamma + 1)/(2\underline{b} + 2\gamma) + c_\chi c_0}{\sqrt{c_\nu c_\chi c_0}} \sqrt{\ln(n)} \rightarrow -\infty.$$

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

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

three-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,  $g_{1,n}(x) = g_0(x) + \sqrt{c_0 \ln \ln n / nm^{\gamma+1/2}} G(m(x-x_0))$ , where  $m$  is such that

$$c_0 \frac{\ln \ln n}{n} m^{2\gamma+r-1} \exp(2\underline{a}(\pi m)^{\underline{r}}) \leq 2\pi \bar{L}/2. \quad (37)$$

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 (37), we get that  $g_{1,n}$  belongs to  $S(\underline{a}, \underline{r}, \bar{L})$  for any  $a \geq \underline{a}$ . Next,  $|g_{1,n}(x_0) - g_0(x_0)|/\phi_{n,\underline{a},\underline{r}} = c_0|G(0)| > 0$  and we get, in the same way as for case 1,

$$\begin{aligned} \chi^2(f_0^Z, f_{1,n}^Z) &= c_0 \frac{\ln \ln n}{n} m^{2\gamma+1} \int \frac{[G(m(\cdot-x_0)) \star f_\varepsilon]^2(x)}{f_{1,n}^Z(x)} dx \\ &\leq c_0 \frac{\ln \ln n}{n} m^{2\gamma+1} c_1 \int [G(m(\cdot-x_0)) \star f_\varepsilon]^2(x) dx (1+o(1)) \\ &\leq c_0 c_\chi \frac{\ln \ln n}{n} =: \kappa_n. \end{aligned}$$

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

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

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

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

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

(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  $S(a, \bar{r}, \bar{L}/2)$  such that  $g_0 > 0$  and  $g_0(x) \geq c|x|^{-2}$  as  $|x| \rightarrow \infty$ . Let also  $G$  be a function such that  $G^*$  is three-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

$$g_{1,n}^*(u) = g_0^*(u) + c_0 \frac{e^{-\alpha(\pi m)^\rho}}{\sqrt{n}} m^{\rho-1/2} e^{2\alpha|u|^\rho} G^*(|u|^\rho - (\pi m)^\rho) e^{iu x_0},$$

where  $m$  is the solution of the equation

$$2\underline{a}(\pi m)^r + 2\alpha(\pi m)^\rho = \log n - (\log \log n)^2. \quad (38)$$

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

$$\begin{aligned} & (2\pi n)^{-1} \int c_0^2 e^{-2\alpha(\pi m)^\rho} m^{2\rho-1} e^{4\alpha|u|^\rho} |G^*(|u|^\rho - (\pi m)^\rho)|^2 e^{-2\underline{a}|u|^r} du \\ & \leq (2\pi n)^{-1} c_0^2 m^{2\rho-1} e^{-2\alpha(\pi m)^\rho} \int_{\pi/4 \leq |u|^\rho - (\pi m)^\rho \leq 3\pi/4} e^{4\alpha|u|^\rho + 2\underline{a}|u|^r} du \\ & \leq (2\pi n)^{-1} c_0^2 c_1 m^{2\rho-1} e^{-2\alpha(\pi m)^\rho} (\pi m)^{1-r} e^{4\alpha(\pi m)^\rho + 2\underline{a}(\pi m)^r} \\ & \leq c_0^2 c_2 n^{-1} m^{2\rho-r} e^{2\underline{a}(\pi m)^r + 2\alpha(\pi m)^\rho}, \end{aligned}$$

which tends to 0 when  $m$  is defined by (38). Next,

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

and we can check similarly to Butucea and Tsybakov [6] that for  $m$  solution of (38) this sequence is equivalent to  $\phi_{n,\underline{a},r}$  when  $n \rightarrow \infty$ . Finally

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

say  $T_1 + T_2$ . Then

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

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

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

for  $\rho \leq 1$  and  $n$  large enough. Thus  $\chi^2(f_0^Z, f_{1,n}^Z) \leq c_0^2 c_\chi (\pi m)^\rho / n =: \kappa_n$ .

Let  $c_0$  be small such that  $c_0^2 c_\chi < 2\alpha$  and let  $\xi$  and  $\tau$  be defined by  $c_0^2 c_\chi < \xi < 2\alpha$  and  $\tau = e^{-\xi(\pi \ln(n)/(2\underline{a}))^{\rho/\underline{\tau}}}$ . We have  $\tau \phi_{n,\underline{a},\underline{\tau}}^2 / \phi_{n,\bar{a},\bar{\tau}}^2 \geq (\ln(n))^A \exp((-\xi + 2\alpha)(\ln(n)/(2\underline{a}))^{\rho/\underline{\tau}} + B(\ln(n))^C)$  tends to infinity for some real numbers  $A, B, C$ , as  $C < \rho/\underline{\tau}$  and  $\xi < 2\alpha$ . We check that Lyapunov's theorem holds and that

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

with  $n$ , as  $m$  defined by (38) is larger than  $(\ln(n)/(2\underline{a}))^{1/\underline{\tau}}$ .

The proof that  $\varphi_n$  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

$$g_{1,n}^*(u) = g_0^*(u) + c_0 \frac{e^{-\alpha(\pi m)^\rho}}{\sqrt{n}} m^{(\rho-1)/2} e^{2\alpha|u|^\rho} G^*(|u|^\rho - (\pi m)^\rho) e^{iu x_0},$$

where  $m$  is the solution of equation (38).

This gives the rate  $|g_{1,n}(x_0) - g_0(x_0)| \geq c_0 c_3 m^{-(\rho-1)/2} e^{\alpha(\pi m)^\rho} / \sqrt{n}$ , which is equivalent to  $V_{\tilde{m}}$  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  $\varphi_n$  is a minimax rate of convergence for  $r \geq \rho$ ,  $\rho \leq 1$ .

## References

- [1] Artiles Martínez, L.M. (2001). Adaptive minimax estimation in classes of smooth functions. Ph.D. thesis. Utrecht University, Netherlands.
- [2] Artiles Martínez, L.M. and Levit, B.Y. (2003). Adaptive estimation of analytic functions on an interval. *Math. Methods Statist.* **12** 62–94. [MR1990514](#)
- [3] Butucea, C. (2001). Exact adaptive pointwise estimation on Sobolev classes of densities. *ESAIM Probab. Statist.* **5** 1–31. [MR1845320](#)
- [4] Butucea, C. (2004). Deconvolution of supersmooth densities with smooth noise. *Canad. J. Statist.* **32** 181–192. [MR2064400](#)
- [5] Butucea, C. and Neumann, M.H. (2005). Exact asymptotics for estimating the marginal density of discretely observed diffusion processes. *Bernoulli* **11** 411–444. [MR2146889](#)
- [6] Butucea, C. and Tsybakov, A.B. (2007). Sharp optimality for density deconvolution with dominating bias. I and II. *Theory Probab. Appl.* **52** 24–39. [MR2354572](#)

- [7] Cai, T.T. and Low, M.G. (2005). Adaptive estimation of linear functionals under different performance measures. *Bernoulli* **11** 341–358. [MR2132730](#)
- [8] Cai, T.T. and Low, M.G. (2005). On adaptive estimation of linear functionals. *Ann. Statist.* **33** 2311–2343. [MR2211088](#)
- [9] Cator, E.A. (2001). Deconvolution with arbitrarily smooth kernels. *Statist. Probab. Lett.* **54** 205–214. [MR1858635](#)
- [10] Cavalier, L. (2001). On the problem of local adaptive estimation in tomography. *Bernoulli* **7** 63–78. [MR1811744](#)
- [11] Chaleyat-Maurel, M. and Genon-Catalot, V. (2006). Computable infinite-dimensional filters with applications to discretized diffusion processes. *Stochastic Process. Appl.* **116** 1447–1467. [MR2260743](#)
- [12] Comte, F., Dedecker, J. and Taupin, M.-L. (2008). Adaptive density estimation for general ARCH models. *Econometric Theory* **24** 1628–1662.
- [13] Comte, F. and Genon-Catalot, V. (2006). Penalized projection estimator for volatility density. *Scand. J. Statist.* **33** 875–893. [MR2300921](#)
- [14] Comte, F., Rozenholc, Y. and Taupin, M.-L. (2006). Penalized contrast estimator for adaptive density deconvolution. *Canad. J. Statist.* **34** 431–452. [MR2328553](#)
- [15] Doukhan, P. (1994). *Mixing*. Properties and examples. New York: Springer. [MR1312160](#)
- [16] Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19** 1257–1272. [MR1126324](#)
- [17] Genon-Catalot, V. and Kessler, M. (2004). Random scale perturbation of an AR(1) process and its properties as a nonlinear explicit filter. *Bernoulli* **10** 701–720. [MR2076070](#)
- [18] Goldenshluger, A. (1999). On pointwise adaptive nonparametric deconvolution. *Bernoulli* **5** 907–925. [MR1715444](#)
- [19] Goldenshluger, A. and Pereverzev, S.V. (2003). On adaptive inverse estimation of linear functionals of Hilbert scales. *Bernoulli* **9** 783–807. [MR2047686](#)
- [20] Golubev, G.K. (2004). The method of risk envelopes in the estimation of linear functionals. *Problemy Peredachi Informatsii* **40** 58–72. [MR2099020](#)
- [21] Golubev, Y. and Levit, B. (2004). An oracle approach to adaptive estimation of linear functionals in a Gaussian model. *Math. Methods Statist.* **13** 392–408. [MR2126747](#)
- [22] Klemelä, J. and Tsybakov, A.B. (2004). Exact constants for pointwise adaptive estimation under the Riesz transform. *Probab. Theory Related Fields* **129** 441–467. [MR2128241](#)
- [23] Lacour, C. (2006). Rates of convergence for nonparametric deconvolution. *C. R. Math. Acad. Sci. Paris* **342** 877–882. [MR2224640](#)
- [24] Laurent, B., Ludeña, C. and Prieur, C. (2008). Adaptive estimation of linear functionals by model selection. *Electron. J. Stat.* **2** 993–1020.
- [25] Lepski, O.V. and Levit, B.Y. (1998). Adaptive minimax estimation of infinitely differentiable functions. *Math. Methods Statist.* **7** 123–156. [MR1643256](#)
- [26] Lepskiĭ, O.V. (1990). A problem of adaptive estimation in Gaussian white noise. *Teor. Veroyatnost. i Primenen.* **35** 459–470. [MR1091202](#)
- [27] Petrov, V.V. (1995). *Limit Theorems of Probability Theory*. New York: The Clarendon Press, Oxford Univ. Press. [MR1353441](#)
- [28] Tsybakov, A.B. (1998). Pointwise and sup-norm sharp adaptive estimation of functions on the Sobolev classes. *Ann. Statist.* **26** 2420–2469. [MR1700239](#)
- [29] Van Es, B., Spreij, P. and van Zanten, H. (2005). Nonparametric volatility density estimation for discrete time models. *J. Nonparametr. Statist.* **17** 237–251. [MR2112523](#)

*Received February 2007 and revised May 2008*