*Bernoulli* **11**(2), 2005, 309–340

# Minimax estimation of the noise level and of the deconvolution density in a semiparametric convolution model

CRISTINA BUTUCEA[1,3] and CATHERINE MATIAS[2]

[1]*Modal'X, Université Paris X, Bât. G, 200, avenue de la République, 92001 Nanterre Cedex, France. E-mail: butucea@ccr.jussieu.fr*
[2]*CNRS, Laboratoire de Statistique et Génome, Tour Evry 2, 523 place des Terrasses de l'Agora, 91000 Evry, France. E-mail: matias@genopole.cnrs.fr. Research done while at 3.*
[3]*Laboratoire de Probabilités et Modèles Aléatoires, Université Paris VI, 175 rue du Chevaleret, 75013 Paris, France*

We consider a semiparametric convolution model where the noise has known Fourier transform which decays asymptotically as an exponential with unknown scale parameter; the deconvolution density is less smooth than the noise in the sense that the tails of the Fourier transform decay more slowly, ensuring the identifiability of the model. We construct a consistent estimation procedure for the noise level and prove that its rate is optimal in the minimax sense. Two convergence rates are distinguished according to different smoothness properties for the unknown density. If the tail of its Fourier transform does not decay faster than exponentially, the asymptotic optimal rate and exact constant are evaluated, while if it does not decay faster than polynomially, this rate is evaluated up to a constant. Moreover, we construct a consistent estimator of the unknown density, by using a plug-in method in the classical kernel estimation procedure. We establish that the rates of estimation of the deconvolution density are slower than in the case of an entirely known noise distribution. In fact, nonparametric rates of convergence are equal to the rate of estimation of the noise level, and we prove that these rates are minimax. In a few specific cases the plug-in method converges at even slower rates.

*Keywords:* analytic densities; deconvolution; $L_2$ risk; minimax estimation; noise level; pointwise risk; semiparametric model; Sobolev classes; supersmooth densities

## 1. Introduction

Let us consider the observations $Y_i$, $i = 1, \ldots, n$, such that

$$Y_i = X_i + \sigma \varepsilon_i,$$

where $X_i$ and $\varepsilon_i$ are independent and identically distributed real-valued random variables, the two sequences $\{X_i\}$ and $\{\varepsilon_i\}$ being independent of each other. Two components are unknown in this model: the common law of $X_i$ having probability density $f$ (with respect to the Lebesgue measure on $\mathbb{R}$) and characteristic function $\Phi$, and the scale parameter $\sigma > 0$. The

1350–7265 

variables $\varepsilon_i$ have a known $s$-exponential distribution, that is, a known density function $f^\varepsilon$ having a Fourier transform $\Phi^\varepsilon$ such that, for large enough $|u|$,

$$b\mathrm{e}^{-|u|^s} \leqslant |\Phi^\varepsilon(u)| \leqslant B\mathrm{e}^{-|u|^s}, \tag{1}$$

for some known $s > 0$ and fixed constants $b$, $B > 0$.

In the more classical deconvolution problem the distribution of the noise is supposed to be completely known (the law as well as the scale parameter). In this case, minimax rates of convergence are described in the literature for various associations of smoothness classes for the unknown density (Hölder, Sobolev, Besov or analytic functions) and global behaviours of the errors' law. Even if the noise law is entirely known, estimators behave differently whether the characteristic function of the noise decays polynomially or exponentially asymptotically. There has been a huge amount of literature since the paper by Caroll and Hall (1988). For the $s$-exponential error distribution case, exact asymptotic rates were computed: pointwise and $L_2$ rates for periodic Sobolev densities by Efromovich (1997), $L_2$ rates for bivariate circular structural models with Sobolev and analytic estimated densities by Goldenshluger (2002), and both pointwise and $L_2$ (faster than usual) rates by Butucea and Tsybakov (2004), for classes of supersmooth densities (which can be infinitely differentiable, analytic on a strip around the real axis, or analytic on the complex plane). Estimation of the Sobolev density with $s$-exponential, entirely known noise has been done adaptively by Goldenshluger (1999).

Here, the deconvolution density and the scale parameter of the noise are unknown. In Section 2, we are interested in recovering the scale of the noise $\sigma > 0$. Indeed, the assumption of a completely known noise distribution is rather unrealistic from a practical point of view. Therefore, evaluating the scale parameter of the noise can help to cope with the situation where this assumption is not satisfied. Moreover, in Section 3, we use it as a preliminary step in the nonparametric deconvolution problem of estimating the unknown density when the scale is unknown.

The estimation of the scale parameter and of the unknown density has already been considered by Matias (2002) in the case of Gaussian errors and a large collection of density functions, densities 'without Gaussian component'. The estimators of the scale parameter were based on Fourier or Laplace transforms and they were proven to be consistent over certain subclasses. Lower bounds of order $1/\log n$ were found for both estimation problems. Matias (2002) noted that estimation of the nonparametric density is more difficult (larger lower bounds of order $1/\log n$) when the scale parameter is unknown than in the classical deconvolution problem.

The problem of noise-level estimation in a convolution model has been formulated by Matias (2002) in relation to error-in-variables nonlinear regression. More generally, in physics and biology error-in-variables models are widely used. Our paper allows a convolution model to be used where the scale parameter for the noise is unknown.

A similar problem was considered by Lindsay (1986) for a mixture of exponential families with applications to Bayesian statistics. Among other results, he considers an infinitely divisible mixing density, with unknown parameters which are recovered via least-squares estimation. This problem is similar to noise-level evaluation in our model where the deconvolution density is in a parametric exponential family. Thus, our results extend this estimation to nonparametric main densities.

Similarly to Zhang (1990), we can regard this model as a mixing model of location families. Zhang (1990) considers location (in $\theta$) families $f^{\varepsilon}(\cdot - \theta)$ with mixing density $f(\theta)$. The observations $Y_i$, $i = 1, \ldots, n$, have density $\int f^{\varepsilon}(\cdot - \theta) f(\theta) \mathrm{d}\theta$ and the mixing density $f$ is estimated. More generally, in our model the location families $f^{\varepsilon}((\cdot - \theta)/\sigma)/\sigma$ have an unknown scale parameter $\sigma$ which we estimate together with the mixing density $f$.

In multidimensional deconvolution problems with Gaussian errors, Koltchinskii (2000) suggested an estimator of the covariance matrix of the Gaussian errors in order to obtain the geometric structure of the support of the deconvolution density.

In this paper, we propose a new estimation algorithm for the scale parameter, prove its consistency and compute the upper bounds of its mean squared error in several different set-ups. Moreover, we prove that the rates obtained are optimal by giving the corresponding minimax lower bounds.

We solve the problem of estimating the noise level in two possible set-ups, with respectively the following assumptions:

***Assumption A.*** *We suppose that the unknown density belongs to the class $\mathcal{A}(\alpha, r)$ of densities whose Fourier transform decays asymptotically slower than some exponential*

$$|\Phi(u)| \geqslant c\mathrm{e}^{-\alpha|u|^r}, \qquad |u| \text{ large enough},$$

*with known parameters $\alpha > 0$ and $r \in (0, s)$, and some arbitrary constant $c > 0$.*

***Assumption B.*** *The unknown density is in the class $\mathcal{B}(\beta)$ of densities having Fourier transform decaying asymptotically slower than some polynomial*

$$|\Phi(u)| \geqslant c|u|^{-\beta}, \qquad |u| \text{ large enough},$$

*with known parameter $\beta > 1$ and an arbitrary constant $c > 0$.*

Under either Assumption A or Assumption B, the model is identifiable. In fact, considering Fourier transforms, we obtain

$$-\alpha|u|^{r-s} \leqslant \frac{\log|\Phi(u)|}{|u|^s} \leqslant 0, \qquad \text{for } |u| \text{ large, under Assumption A,}$$

$$\frac{-\beta\log|u| + \log c}{|u|^s} \leqslant \frac{\log|\Phi(u)|}{|u|^s} \leqslant 0, \qquad \text{for } |u| \text{ large, under Assumption B}$$

(recall that the parameter $s$ is defined in (1)). So that $\lim_{|u|\to\infty}|u|^{-s}\log|\Phi(u)| = 0$. Since the Fourier transform $\Phi^Y(u)$ of the distribution of the observations equals the product $\Phi(u)\Phi^{\varepsilon}(\sigma u)$, and using (1), we obtain:

$$\lim_{|u|\to\infty} \frac{\log|\Phi^Y(u)|}{|u|^s} = \lim_{|u|\to\infty} \frac{\log|\Phi^{\varepsilon}(\sigma u)|}{|u|^s} = -\sigma^s.$$

Consequently, the distribution of the observations $Y_i$ determines uniquely the scale parameter $\sigma$ and then also the density $f$. This establishes the identifiability of the model.

We remark that information on the nonparametric density as well as on the unknown noise level must be retrieved from the same sample of $Y_i$s. We need to estimate first the scale parameter of the noise. It is important that the deconvolution density be significantly less smooth than the noise. We note that if the deconvolution density becomes smoother than the noise the parameter is non-identifiable. Moreover, our results give faster rates when the noise is significantly smoother than the deconvolution density and slower rates when the noise is smooth but behaves similarly to this density.

These rates are overall slower when compared to classical parametric estimation. This is not surprising in this semiparametric model where we distinguish a parametric component from a nonparametric unknown function.

We then establish that the rate of convergence of our estimator is optimal with respect to the minimax risk. This is done under the additional assumption that the noise has a stable distribution.

***Assumption S.*** *The noise $\varepsilon$ has stable distribution denoted $S(1, s, \nu, \mu)$, with scale parameter fixed to* 1*, self-similarity index $s \in (0, 2]$, symmetry parameter $\nu \in [-1, 1]$ and location $\mu \in \mathbb{R}$.*

See Section 3.1 for more details on stable laws and Zolotarev (1986) for a complete overview of the subject.

In fact, we only need the exact expression for the function $|\Phi^\varepsilon|$, which is very simple under Assumption S since $|\Phi^\varepsilon(u)| = e^{|u|^s}$. For simplicity of notation, we decide to fix the noise so that it is exactly distributed according to a stable law which corresponds to the parameter $s$ belonging in the interval (0, 2]. This assumption is not very restrictive, since most examples encountered are in this range, and could be relaxed at this point. We note that the rates of convergence of our estimator are sharp minimax under Assumption A and nearly sharp under Assumption B.

In Section 3, we study pointwise and $L_2$ rates of convergence for the estimation of the deconvolution density, in the presence of unknown noise level, regarded as a nuisance parameter. These rates are significantly slower than the rates obtained in classical deconvolution problems by Efromovich (1997) and Butucea and Tsybakov (2004), and are equal to (or even slower than) the rates of estimation of the noise level. The estimators are classical kernel estimators where we plug in the estimated value of the underlying parameter. This implies a study of uniform empirical processes explaining the loss of performance of this estimator. Because of this loss in the rate, lower bounds cannot be directly deduced from the results cited above. As in the classical nonparametric minimax theory, we construct a pair of couples $(\sigma_1, f_1)$ and $(\sigma_2, f_2)$ so that for fixed $x \in \mathbb{R}$, the densities $f_1(x)$ and $f_2(x)$ are as far apart as possible (or, when dealing with $L_2$ risks, the distance $\|f_1 - f_2\|_2$ is the greatest possible) under the restriction that the corresponding likelihoods $f_1 * (f^\varepsilon(\cdot/\sigma_1)/\sigma_1)$ and $f_2 * (f^\varepsilon(\cdot/\sigma_2)/\sigma_2)$ are close in $\chi^2$-distance. Note that the choice of the parameters is natural for our problem and quite simple (see Section 5.2). This shows that the loss in the rate is unavoidable. Note also that these results agree with the lower bounds established by Matias (2002) in the case $s = 2$ and are more precise as we study here the rates of convergence when the unknown signal belongs to the classes $\mathcal{A}(\alpha, r)$ (with $0 < r < s$) and $\mathcal{B}(\beta)$ ($\beta > 1$), whereas Matias' lower bounds concern every density with non-Gaussian component.

In the rest of the paper, we first define the estimation method for the scale parameter and study its consistency, for the defined parameters $\alpha > 0$ and $s > r > 0$ under Assumption A, or $\beta > 1$ and $s > 0$ under Assumption B (Section 2). We also establish the optimality with respect to the minimax risk of our estimator under the additional Assumption S.

Then, using a plug-in method combined with the natural kernel deconvolution technique, we construct an estimator of the deconvolution density and study its pointwise and $L_2$ rates of convergence, under the additional Assumption S (Sections 3.2–3.5). In order to obtain these rates, we need to add assumptions on the largest smoothness the estimated density may have as in classical deconvolution problem (see the definitions of the sets $\mathcal{S}(\alpha', R, L)$ and $W(\beta', L)$ given by (9) and (10)). We also prove that the estimator constructed is optimal in the minimax setting.

# 2. Estimation of the noise level

## 2.1. Noise-level evaluation algorithm

The estimator we propose is defined implicitly via the following criterion. Let us first estimate the characteristic function of the observed variables $\Phi^Y(u) = \mathbb{E}\{\exp(\mathrm{i}uY)\}$ by using the given sample:

$$\hat{\Phi}_n^Y(u) = \frac{1}{n}\sum_{k=1}^{n} \mathrm{e}^{\mathrm{i}uY_k}.$$

Remark that in the following $\mathbb{P}$, $\mathbb{E}$ and var denote, respectively, the probability $\mathbb{P}_{\sigma,f}$ the expectation $\mathbb{E}_{\sigma,f}$ and the variance $\mathrm{var}_{\sigma,f}$ with respect to the probability when the underlying unknown parameters are $\sigma > 0$ and $f$ a density in the class.

Let us next consider the function

$$\hat{F}_n(\tau, u) = \hat{\Phi}_n^Y(u)\mathrm{e}^{(\tau u)^s}, \tag{2}$$

for $\tau$, $u > 0$ and fixed known $s > 0$ (given by (1)). Our estimator $\hat{\sigma}_n$ of $\sigma$ is defined by

$$\hat{\sigma}_n = \hat{\sigma}_n(Y_1, \ldots, Y_n) = \inf\{\tau : \tau > 0, |\hat{F}_n(\tau, u_n)| \geqslant 1\}, \tag{3}$$

for some positive sequence $u_n \to \infty$ well chosen, as described later (see Propositions 1 and 2).

This construction is based on the observation that $|\hat{F}_n(\tau, u)|$ is an unbiased estimator of $|F(\tau, u)|$, where

$$F(\tau, u) = \Phi^Y(u)\mathrm{e}^{(\tau u)^s}. \tag{4}$$

We have $|F(\tau, u)| = O(1)|\Phi(u)|e^{(\tau^s - \sigma^s)u^s}$ for large enough $|u|$, so that this quantity converges, when $u \to \infty$, either to 0 when $\tau \leqslant \sigma$ or to $\infty$ when $\tau > \sigma$.

Note that, for the convergence of this estimation method it is sufficient to assume (1) on the distribution of the noise, but we prove the convergence of the plug-in estimator for $f$ and its optimality (Section 3.2) under the additional assumption that the noise has a stable distribution (i.e. Assumption S).

## 2.2. Consistency and optimality

For underlying deconvolution densities satisfying Assumption A we establish local consistency of our procedure $\hat{\sigma}$, that is, when $\sigma$ varies in a neighbourhood of some fixed $\sigma_0$. Note that $\delta > 0$ is not necessarily small at this stage and that only $\sigma_0 + 2\delta$ appears in the procedure, that is a strict upper bound of the open set where $\sigma$ takes values. We consider this particular upper bound in view of local minimax results of Theorem 1, but it is easy to see how arbitrary choices of the compact set for $\sigma$ and of its strict upper bound lead to the corresponding corollary.

**Proposition 1.** *Fix $\sigma_0 > 0$, $\delta > 0$ such that $\sigma_0 > \delta$, and a neighbourhood $\mathcal{V}(\sigma_0) = \mathcal{V}_\delta(\sigma_0) = (\sigma_0 - \delta, \sigma_0 + \delta)$. Under Assumption A, consider the sequence of parameters $u_n = (\sigma_0 + 2\delta)^{-1}(\log n/2)^{1/s}$, the estimator $\hat{\sigma}_n$ defined by* (3) *and the rate*

$$\varphi_{n,\delta} = \frac{\alpha}{s}\frac{(\sigma_0 + 2\delta)^{1-r+s}}{(\sigma_0 - \delta)^s}\left(\frac{\log n}{2}\right)^{r/s-1}.$$

*Then, for all $\sigma \in \mathcal{V}(\sigma_0)$, for all $f \in \mathcal{A}(\alpha, r)$, and large enough $n$, we have*

$$\mathbb{P}(|\hat{\sigma}_n - \sigma| \geqslant \varphi_{n,\delta}) \leqslant O(1)\exp\left\{\frac{2\alpha}{(\sigma_0 + 2\delta)^r}\left(\frac{\log n}{2}\right)^{r/s}\right\}\left(\frac{1}{n}\right)^{1-\sigma^s/(\sigma_0+2\delta)^s}.$$

The next theorem gives local sharp (or exact, i.e. the asymptotic value of the risk is 1) minimax rate of convergence. We note that the estimator of $\sigma$ depends on a strict upper bound (say, $\sigma_0 + 2\delta$) of the neighbourhood $\mathcal{V}_\delta(\sigma_0)$, which tends to $\sigma_0$ as the neighbourhood shrinks to $\{\sigma_0\}$.

**Theorem 1.** *For all fixed $\sigma_0 > 0$, under Assumption A and for $\hat{\sigma}_n$ defined as in Proposition 1, consider the rate*

$$\varphi_n = \frac{\alpha}{s\sigma_0^{r-1}}\left(\frac{\log n}{2}\right)^{r/s-1}.$$

*Then, for any neighbourhood $\mathcal{V}(\sigma_0)$,*

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\sup_{\sigma\in\mathcal{V}(\sigma_0)}\sup_{f\in\mathcal{A}(\alpha,r)}\varphi_n^{-2}\mathbb{E}(|\hat{\sigma}_n - \sigma|^2) \leqslant 1,$$

*and, under the additional Assumption S,*

$$\lim_{\delta\to 0}\liminf_{n\to\infty}\inf_{\sigma_n}\sup_{\sigma\in\mathcal{V}(\sigma_0)}\sup_{f\in\mathcal{A}(\alpha,r)}\varphi_n^{-2}\mathbb{E}(|\sigma_n-\sigma|^2)\geqslant 1,$$

*where the infimum is taken over arbitrary estimators $\sigma_n$ of $\sigma$.*

Under Assumption B, the noise is much smoother than the deconvolution density, and this explains the faster rate in this case. Note that the class of densities allowed here is also much smaller than in the previous set-up.

**Proposition 2.** *Fix $\sigma_0>0$, $\delta>0$ such that $\sigma_0>\delta$, and a neighbourhood $\mathcal{V}(\sigma_0)=\mathcal{V}_\delta(\sigma_0)=(\sigma_0-\delta;\sigma_0+\delta)$. Under Assumption B, consider the sequence of parameters $u_n=(\sigma_0+2\delta)^{-1}(\log n/2)^{1/s}$, and the rate*

$$\psi_{n,\delta}=\frac{2\beta}{s^2}\frac{(\sigma_0+2\delta)^{s+1}}{(\sigma_0-\delta)^s}\frac{\log\log n}{\log n}.$$

*Then, for all $\sigma\in\mathcal{V}(\sigma_0)$, for all $f\in\mathcal{B}(\beta)$ and large enough n, we have*

$$\mathbb{P}(|\hat\sigma_n-\sigma|\geqslant\psi_{n,\delta})\leqslant O(1)(\log n)^{2\beta/s}\left(\frac{1}{n}\right)^{1-\sigma^s/(\sigma_0+2\delta)^s}.$$

Minimax lower bounds in the next theorem are exact for $s=1$ and nearly exact otherwise, since $|s-1|\leqslant 1$ and $2\beta>2$.

**Theorem 2.** *For all fixed $\sigma_0>0$, under Assumption B and for $\hat\sigma_n$ defined as in Proposition 2, consider the rate*

$$\psi_n=\frac{2\beta\sigma_0}{s^2}\frac{\log\log n}{\log n}.$$

*Then, for any $\beta>1$ and for any neighbourhood $\mathcal{V}(\sigma_0)$,*

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\sup_{\sigma\in\mathcal{V}(\sigma_0)}\sup_{f\in\mathcal{B}(\beta)}\psi_n^{-2}\mathbb{E}(|\hat\sigma_n-\sigma|^2)\leqslant 1,$$

*and, under the additional Assumption S,*

$$\lim_{\delta\to 0}\liminf_{n\to\infty}\inf_{\sigma_n}\sup_{\sigma\in\mathcal{V}(\sigma_0)}\sup_{f\in\mathcal{B}(\beta)}\psi_n^{-2}\mathbb{E}(|\sigma_n-\sigma|^2)\geqslant\left(1-\frac{|s-1|}{2\beta}\right)^2,$$

*where the infimum is taken over arbitrary estimators $\sigma_n$ of $\sigma$.*

The proofs can be found in Section 4 for the upper bounds and in Section 5 for the lower bounds.

For the purposes of practical implementation we may use an immediate consequence of Theorem 1 (or Theorem 2). This is a global version of the minimax upper bounds, where the unknown parameter is supposed to belong to some compact set in $\mathbb{R}_+^*$ and the estimation algorithm is based only on a strict upper bound $\Sigma$ of this set.

**Corollary 1.** *Suppose $\sigma$ is in some bounded set $\Theta$, $0 < \inf \Theta < \sup \Theta < \Sigma$. Under Assumption A, consider*

$$u_n = \left(\frac{\log n}{2\Sigma^s}\right)^{r/s} \quad \text{and} \quad \varphi_n(\Sigma) = \frac{\alpha}{s\sigma^{s-1}} \left(\frac{\log n}{2\Sigma^s}\right)^{r/s-1}.$$

*Then we have*

$$\limsup_{n\to\infty} \sup_{\sigma\in\Theta} \sup_{f\in\mathcal{A}(\alpha,r)} \varphi_n^{-2}(\Sigma)\mathbb{E}(|\hat{\sigma}_n - \sigma|^2) \leqslant 1.$$

**Corollary 2.** *Suppose $\sigma$ is in some bounded set $\Theta$, $0 < \inf \Theta < \sup \Theta < \Sigma$. Under Assumption B, consider*

$$u_n = \left(\frac{\log n}{2\Sigma^s}\right)^{r/s} \quad \text{and} \quad \psi_n(\Sigma) = \frac{2\beta\Sigma^s}{s^2\sigma^{s-1}} \frac{\log \log n}{\log n}.$$

*Then we have*

$$\limsup_{n\to\infty} \sup_{\sigma\in\Theta} \sup_{f\in\mathcal{B}(\beta)} \psi_n^{-2}(\Sigma)\mathbb{E}(|\hat{\sigma}_n - \sigma|^2) \leqslant 1.$$

A simulation study was performed on the estimator of $\sigma$ with tuning $u_n$ as defined in the corollaries. As we might expect, the quality of estimation is improved when the upper bound $\Sigma$ is closer to the true underlying parameter $\sigma$. Moreover, in the proof of Propositions 1 and 2 we see that the probability that $\hat{\sigma}_n$ overestimates $\sigma$ dominates the probability that it underestimates it. Therefore, in practice it is useful to reiterate the procedure and take the former $\hat{\sigma}_n$ as $\Sigma_1$ in the next step and obtain a new estimator $\hat{\sigma}_n^{(1)}$, etc. Results become robust with respect to the a priori large and difficult choice of $\Sigma$.

# 3. Estimation of the density $f$

Our estimator $\hat{\sigma}_n$ of the noise level, defined by (3), leads to a natural estimator of the deconvolution density, using a kernel estimator combined with a plug-in method. In this section, we establish the rate of convergence of this estimator and prove its optimality. In the following, $C > 0$ denotes a large enough constant.

## 3.1. Preliminaries on stable distributions

We consider a noise $\varepsilon$ having a stable distribution denoted by $S(1, s, \nu, \mu)$, with scale parameter fixed at 1, self-similarity index $s \in (0, 2]$, symmetry parameter $\nu \in [-1, 1]$ and location $\mu \in \mathbb{R}$ (Assumption S). In our model the noise is multiplied by an unknown scale parameter $\sigma > 0$. By Zolotarev (1986), $\sigma\varepsilon$ has also a stable law whose explicit Fourier transform is given by

$$\Phi^\varepsilon(\sigma u) = \begin{cases} \exp\{-\sigma^s|u|^s(1 - \mathrm{i}\nu\,\mathrm{sgn}(u)\tan(\pi s/2)) + \mathrm{i}u\sigma\mu\}, & s \neq 1 \\ \exp\left\{-\sigma|u|\left(1 + \mathrm{i}\nu\,\mathrm{sgn}(u)\frac{2}{\pi}\log|u|\right) + \mathrm{i}u\sigma\left(\mu - \nu\frac{2}{\pi}\log\sigma\right)\right\}, & s = 1. \end{cases} \quad (5)$$

Note that $|\Phi^\varepsilon(\sigma u)| = \mathrm{e}^{-\sigma^s|u|^s}$. Moreover, a sum of independent copies of a stable law with the same self-similarity index $s$ is distributed as a stable law with the same parameter $s$. Indeed, for $\sigma_1, \sigma_2 > 0$,

$$\Phi^\varepsilon(\sigma_1 u)\Phi^\varepsilon(\sigma_2 u) = \Phi^\varepsilon(\sigma u)\mathrm{e}^{\mathrm{i}ua},$$

for any values of the parameters $s$ and $\nu$, where $\sigma_1^s + \sigma_2^s = \sigma^s$ and

$$a = \begin{cases} \mu(\sigma_1 + \sigma_2 - \sigma), & s \neq 1, \\ \frac{2}{\pi}\nu(\sigma_1\log(\sigma_1/\sigma) + \sigma_2\log(\sigma_2/\sigma)), & s = 1. \end{cases}$$

Define moreover the parameter

$$\tilde{s} = \begin{cases} s \vee 1, & \text{if } \mu \neq 0, \\ s, & \text{if } \mu = 0. \end{cases} \quad (6)$$

This parameter will be useful since it is related to the behaviour in a neighbourhood of zero of the function $\Phi^\varepsilon(\sigma\cdot)$. Its role will be clearer in the proofs of the following theorems. Note that when the location parameter $\mu$ differs from 0, we can write the model as $Y = X + \sigma(\varepsilon_0 + \mu)$, with noise $\varepsilon_0$ having stable law located at 0, which means centred if it has finite expectation ($s \geqslant 1$). This expression shows that the role of the known location parameter $\mu$ cannot be neglected, as the model does not simply write $Y = X + \sigma\varepsilon_0 + \mu$.

## 3.2. Plug-in deconvolution density estimator

We will now describe the estimation procedure. Consider the kernel $k_n$ defined by its Fourier transform

$$\Phi^{k_n}(u) = \{\Phi^\varepsilon(h_n^{-1}u)\}^{-1}1_{|u|\leqslant 1}, \quad (7)$$

where $h_n$ is some positive sequence of numbers decreasing to zero. The kernel estimator of the unknown density $f$ is given by

$$\hat{f}_{n,\hat{\sigma}_n}(x) = \frac{1}{n\hat{\sigma}_n h_n}\sum_{i=1}^n k_n\left(\frac{Y_i - x}{\hat{\sigma}_n h_n}\right). \quad (8)$$

In order to obtain an upper-bound for the pointwise risk of our estimator, we need to restrict ourselves to densities belonging to bounded function spaces: classes of supersmooth densities or Sobolev balls. Let us denote

$$\mathcal{S}(\alpha', R, L) = \left\{ f;\ f \text{ is a density and } \int |\Phi(u)|^2 e^{2\alpha'|u|^R} du \leqslant L^2 \right\}, \tag{9}$$

$$W(\beta', L) = \{ f;\ f \text{ is a density and } \int |\Phi(u)|^2 (1 + |u|^{2\beta'}) du \leqslant L^2 \}, \tag{10}$$

where $\alpha'$, $R$, $L > 0$ and $\beta' > \frac{1}{2}$.

Three cases occur, according to whether the unknown density $f$ belongs to $\mathcal{A}(\alpha, r) \cap \mathcal{S}(\alpha', R, L)$, which is non-empty for $R < r$ or $\{R = r \text{ and } \alpha' < \alpha\}$; to $\mathcal{A}(\alpha, r) \cap W(\beta', L)$; or to $\mathcal{B}(\beta) \cap W(\beta', L)$, which is non-empty when $\beta > \beta' + \frac{1}{2}$. Note that in the third case, we automatically obtain that $\beta > 1$. Note also that the intersection $\mathcal{B}(\beta) \cap \mathcal{S}(\alpha', R, L)$ is always empty.

## 3.3. Pointwise rates and optimality

The main goal of this section is to tune the bandwidth $h_n$ in (8) and compute associated pointwise minimax convergence rates associated with each set-up.

**Theorem 3.** *For all fixed $\sigma_0 > 0$, under the assumptions and notation of Theorem 1 and under Assumption S, consider the kernel estimator $\hat{f}_{n,\hat{\sigma}_n}$ defined by* (7) *and* (8) *with bandwidth*

$$h_n = \left\{ \frac{(\sigma_0 + \delta)^R}{\alpha'} \left(1 - \frac{r}{s}\right) \log\log n - \frac{(\sigma_0 + \delta)^R}{\alpha'} \frac{1 - R}{2R} \log\log\log n \right\}^{-1/R}.$$

*Then, for any neighbourhood $\mathcal{V}(\sigma_0)$ of $\sigma_0$ and for any $x$ in $\mathbb{R}$, we have*

$$\limsup_{n\to\infty} \sup_{\sigma \in \mathcal{V}(\sigma_0)} \sup_{f \in \mathcal{A}(\alpha,r)\cap\mathcal{S}(\alpha',R,L)} \varphi_n^{-2} \mathbb{E}(|\hat{f}_{n,\hat{\sigma}_n}(x) - f(x)|^2) \leqslant C < \infty$$

*and*

$$\liminf_{n\to\infty} \inf_{f_n} \sup_{\sigma \in \mathcal{V}(\sigma_0)} \sup_{f \in \mathcal{A}(\alpha,r)\cap\mathcal{S}(\alpha',R,L)} \varphi_n^{-2} \mathbb{E}(|f_n(x) - f(x)|^2) \geqslant c > 0,$$

*where the infimum is taken over arbitrary estimators $f_n$ of $f$.*

Sobolev classes of deconvolution densities are considered in next two theorems. Note that we obtain the same rate as for estimating $\sigma$ in the case $\beta' > \tilde{s} + \frac{1}{2}$, where the parameter $\tilde{s}$ is defined by (6). The rates of convergence can be even slower when $\beta' \leqslant \tilde{s} + \frac{1}{2}$. Upper bounds for the risks in such specific cases are discussed in the next subsection, but corresponding lower bounds are not proven.

**Theorem 4.** *For all fixed $\sigma_0 > 0$, under the assumptions and notation of Theorem 1, under Assumption S, and assuming $\beta' > \tilde{s} + \frac{1}{2}$, consider the kernel estimator $\hat{f}_{n,\hat{\sigma}_n}$ defined by* (7) *and* (8) *with bandwidth $h_n = (\log n)^{2(r/s-1)/(2\beta'-1)}$. Then, for any neighbourhood $\mathcal{V}(\sigma_0)$ of $\sigma_0$ and any $x$ in $\mathbb{R}$, we have*

$$\limsup_{n\to\infty} \sup_{\sigma\in\mathcal{V}(\sigma_0)} \sup_{f\in\mathcal{A}(\alpha,r)\cap W(\beta',L)} \varphi_n^{-2}\mathbb{E}(|\hat{f}_{n,\hat{\sigma}_n}(x) - f(x)|^2) \leqslant C < \infty$$

*and*

$$\liminf_{n\to\infty} \inf_{f_n} \sup_{\sigma\in\mathcal{V}(\sigma_0)} \sup_{f\in\mathcal{A}(\alpha,r)\cap W(\beta',L)} \varphi_n^{-2}\mathbb{E}(|f_n(x) - f(x)|^2) \geqslant c > 0,$$

*where the infimum is taken over arbitrary estimators $f_n$ of $f$.*

**Theorem 5.** *For all fixed $\sigma_0 > 0$, under the assumptions and notation of Theorem 2, under Assumption S, and assuming $\beta' > \tilde{s} + \frac{1}{2}$, consider the kernel estimator $\hat{f}_{n,\hat{\sigma}_n}$ defined by* (7) *and* (8) *with bandwidth $h_n = (\log\log n/\log n)^{2/(2\beta'-1)}$. Then, for any neighbourhood $\mathcal{V}(\sigma_0)$ of $\sigma_0$ and any $x$ in $\mathbb{R}$, we have*

$$\limsup_{n\to\infty} \sup_{\sigma\in\mathcal{V}(\sigma_0)} \sup_{f\in\mathcal{B}(\beta)\cap W(\beta',L)} \psi_n^{-2}\mathbb{E}(|\hat{f}_{n,\hat{\sigma}_n}(x) - f(x)|^2) \leqslant C < \infty$$

*and*

$$\liminf_{n\to\infty} \inf_{f_n} \sup_{\sigma\in\mathcal{V}(\sigma_0)} \sup_{f\in\mathcal{B}(\beta)\cap W(\beta',L)} \psi_n^{-2}\mathbb{E}(|f_n(x) - f(x)|^2) \geqslant c > 0,$$

*where the infimum is taken over arbitrary estimators $f_n$ of $f$.*

Here we use the same kernel estimators as in Butucea and Tsybakov (2004) and plug the preliminary estimator $\hat{\sigma}_n$ into the $\sigma$-dependent bandwidth. Fortunately, the deconvolution kernel can be made free of $\sigma$, and we finally obtain a kernel estimator with data-dependent bandwidth. Thus, we prove that the global estimation risk is at most that of the estimation of the noise level (the slowest).

The proofs for the upper bounds can be found in Section 4. They are based on the convergence of $\hat{\sigma}_n$ to $\sigma$. We evaluate the uniform risk for some parameter in a neighbourhood of $\sigma$ using maximal inequalities for empirical processes in order to treat the uniform stochastic term. Next, we prove that the probability that $\hat{\sigma}_n$ is outside the neighbourhood of $\sigma$ is small enough to make this part of the risk even smaller. This idea was previously used by Butucea (2001) for a density estimator adaptive to the unknown smoothness of the density.

## 3.4. Specific cases

Particular cases for the upper bounds in Theorems 4 and 5 are treated in Table 1. Indeed, when $\beta' \leqslant \tilde{s} + \frac{1}{2}$, we obtain losses which seem inevitable. The proof of these results is given in the respective proofs of the upper bounds of Theorems 4 and 5, when $\beta' < \tilde{s} + \frac{1}{2}$ and when $s = 1$ and $\nu = 0$. The other proofs are immediate consequences of the expressions appearing in the term denoted by $T_{11}$ and are omitted.

Remember that the parameter $\tilde{s} = s$ if the noise is located at $\mu = 0$ but that $\tilde{s} = s \vee 1$ if $\mu \neq 0$, and that $\mu = 0$ cannot be assumed without restriction.

Note also that in the borderline case of $\beta' = \tilde{s} + \frac{1}{2}$, optimal bandwidths are the same as

**Table 1.** Upper bounds for the quadratic pointwise risk when $f \in \mathcal{A}(\alpha, r) \cap W(\beta', L)$, with $h_n = (\log n)^{(r/s-1)/\tilde{s}}$, and $f \in \mathcal{B}(\beta) \cap W(\beta', L)$, with $h_n = (\log\log n/\log n)^{1/\tilde{s}}$, respectively

| | $\beta' = \tilde{s} + \frac{1}{2}$ | | $\beta' < \tilde{s} + \frac{1}{2}$ | |
|---|---|---|---|---|
| $(s = 1, \nu \neq 0)$ | $\varphi_n^2 \log^3(1/h_n)$ | $\psi_n^2 \log^3(1/h_n)$ | $\varphi_n^2 \log^2(1/h_n)$ | $\psi_n^2 \log^2(1/h_n)$ |
| $s \neq 1$ or $(s = 1, \nu = 0)$ | $\varphi_n^2 \log(1/h_n)$ | $\psi_n^2 \log(1/h_n)$ | $\varphi_n^{(2\beta'-1)/2\tilde{s}}$ | $\psi_n^{(2\beta'-1)/2\tilde{s}}$ |

those in Theorem 4, or Theorem 5. The rates are lowered by $\log(1/h_n) = C \log\log n$ at some power, in these cases, where $C > 0$ is some constant.

## 3.5. Rates for $L_2$ risk and optimality

In this subsection we consider global $L_2$ risk for estimating the deconvolution density. In the main cases, rates are the same as for the pointwise risk. A proof is briefly sketched at the very end of the paper.

**Theorem 6.** *For all fixed $\sigma_0 > 0$, consider the kernel estimator $\hat{f}_{n,\hat{\sigma}_n}$ defined by* (7) *and* (8) *with bandwidth $h_n$.*

(i) *Under the assumptions of Theorem 3, take*

$$h_n = \left( \frac{(\sigma_0 + \delta)^R}{\alpha'} \log(1/\varphi_n) \right)^{-1/R}$$

*and consider $\mathcal{F} = \mathcal{A}(\alpha, r) \cap \mathcal{S}(\alpha', R, L)$ and $v_n = \varphi_n$,*

(ii) *Under assumptions of Theorem 4 and $\beta' > \tilde{s}$, take $h_n = \varphi_n^{1/\beta'}$ and consider $\mathcal{F} = \mathcal{A}(\alpha, r) \cap W(\beta', L)$ and $v_n = \varphi_n$,*

(iii) *Under assumptions of Theorem 5 and $\beta' > \tilde{s}$, take $h_n = \psi_n^{1/\beta'}$ and consider $\mathcal{F} = \mathcal{B}(\beta) \cap W(\beta', L)$ and $v_n = \psi_n$.*

*Then, for any neighbourhood $\mathcal{V}(\sigma_0)$ of $\sigma_0$, we have in each different set-up*

$$\limsup_{n\to\infty} \sup_{\sigma \in \mathcal{V}(\sigma_0)} \sup_{f \in \mathcal{F}} v_n^{-2} \mathbb{E}(\|\hat{f}_{n,\hat{\sigma}_n} - f\|_2^2) \leqslant C < \infty$$

*and*

$$\liminf_{n\to\infty} \inf_{f_n} \sup_{\sigma \in \mathcal{V}(\sigma_0)} \sup_{f \in \mathcal{F}} v_n^{-2} \mathbb{E}(\|f_n - f\|_2^2) \geqslant c > 0,$$

*where the infimum is taken over arbitrary estimators $f_n$ of $f$.*

In some specific cases, the rates can be even slower and we do not prove the associated lower bounds. In set-up (ii) of Theorem 6, with $\beta' < \tilde{s}$, the same estimator with bandwidth $h_n = \varphi_n^{1/\tilde{s}}$ attains the rate $v_n = \varphi_n^{\beta'/\tilde{s}}$. In set-up (iii) of the same theorem, with $\beta' < \tilde{s}$, take

$h_n = \psi_n^{1/\tilde{s}}$ to attain the rate $v_n = \psi_n^{\beta'/\tilde{s}}$. As in the case of pointwise risk, we have logarithmic losses when $\beta' = \tilde{s}$.

# 4. Proofs: upper bounds

**Lemma 1.** *For any* $\sigma > 0$ *and any density* $f$*, we have for all* $\tau, u > 0$,

$$\mathbb{E}(\hat{F}_n(\tau, u)) = F(\tau, u) \quad \textit{and} \quad \mathrm{var}(\hat{F}_n(\tau, u)) \leqslant \frac{e^{2(\tau u)^s}}{n},$$

*where* $\hat{F}_n$ *and* $F$ *are defined by* (2) *and* (4).

The proof of this lemma is trivial and therefore omitted.

***Proof of Proposition 1 and Theorem 1 (upper bound).*** Consider the probability of the event $\{|\hat{\sigma}_n - \sigma| \geqslant \varphi_{n,\delta}\}$ and split it into two terms:

$$\mathbb{P}(|\hat{\sigma}_n - \sigma| \geqslant \varphi_{n,\delta}) = \mathbb{P}(\hat{\sigma}_n \geqslant \sigma + \varphi_{n,\delta}) + \mathbb{P}(\hat{\sigma}_n \leqslant \sigma - \varphi_{n,\delta}) = T_1 + T_2.$$

By definition of the estimator $\hat{\sigma}_n$, we bound the first term

$$T_1 \leqslant \mathbb{P}(|\hat{F}_n(\sigma + \varphi_{n,\delta}, u_n)| \leqslant 1) \leqslant \mathbb{P}(|F(\sigma + \varphi_{n,\delta}, u_n)| \leqslant 1 + M) + \Delta_M, \tag{11}$$

for some arbitrary $M > 0$ and $\Delta_M$ defined as

$$\Delta_M = \mathbb{P}(|\hat{F}_n(\sigma + \varphi_{n,\delta}, u_n) - F(\sigma + \varphi_{n,\delta}, u_n)| \geqslant M).$$

Note that

$$\Delta_M \leqslant \frac{1}{M^2}\mathbb{E}(|\hat{F}_n(\sigma + \varphi_{n,\delta}, u_n) - F(\sigma + \varphi_{n,\delta}, u_n)|^2) = \frac{1}{M^2}\mathrm{var}(\hat{F}_n(\sigma + \varphi_{n,\delta}, u_n)).$$

But Lemma 1 leads to

$$\Delta_M \leqslant \frac{e^{2(\sigma+\varphi_{n,\delta})^s u_n^s}}{nM^2}. \tag{12}$$

Note also that

$$\begin{aligned}\mathbb{P}(|F(\sigma + \varphi_{n,\delta}, u_n)| \leqslant 1 + M) &= \mathbb{P}(|\Phi^Y(u_n)| \exp\{(\sigma + \varphi_{n,\delta})^s u_n^s\} \leqslant 1 + M) \\ &= \mathbb{P}(|\Phi(u_n)| \exp[\{(\sigma + \varphi_{n,\delta})^s - \sigma^s\}u_n^s] \leqslant 1 + M) \\ &= \mathbb{P}(|\Phi(u_n)| \exp\{s\varphi_{n,\delta}\sigma^{s-1}u_n^s(1 + o(1))\} \leqslant 1 + M).\end{aligned}$$

With no loss of generality, we have restricted ourselves here to the case $|\Phi^\varepsilon(u)| = e^{-|u|^s}$, for large enough $|u|$. A slight adaptation in the following choice of the parameter $M$ is needed in a more general context. Since Assumption A ensures that, for large enough $n$, $|\Phi(u_n)| \geqslant c\exp\{-\alpha u_n^r\}$, we obtain that

$$\mathbb{P}(|F(\sigma + \varphi_{n,\delta}, u_n)| \leqslant 1 + M) \leqslant \mathbb{P}(c\exp\{-\alpha u_n^r + s\varphi_{n,\delta}\sigma^{s-1}u_n^s(1 + o(1))\} \leqslant 1 + M).$$

With our choice of the parameters $u_n$ and $\varphi_{n,\delta}$, we have

$$\lim_{n\to\infty} (-\alpha u_n^r + s\varphi_{n,\delta}\sigma^{s-1}u_n^s(1+o(1))) = +\infty;$$

then we choose

$$M = \tfrac{1}{2}c\exp\{-\alpha u_n^r + s\varphi_{n,\delta}\sigma^{s-1}u_n^s\},$$

and obtain that $\mathbb{P}(|F(\sigma+\varphi_{n,\delta}, u_n)| \leqslant 1+M)$ is zero for large enough $n$. Combining this with (11), (12) and the choice of $M$, we obtain that, for large enough $n$,

$$\begin{aligned} T_1 &\leqslant \frac{4}{nc^2}\exp\{2(\sigma+\varphi_{n,\delta})^s u_n^s + 2\alpha u_n^r - 2s\sigma^{s-1}\varphi_{n,\delta}u_n^s\} \\ &\leqslant \frac{4}{nc^2}\exp\{2\sigma^s u_n^s + 2\alpha u_n^r + o(\varphi_{n,\delta}u_n^s)\}, \end{aligned}$$

which converges to zero with our choice of the parameters $u_n$ and $\varphi_{n,\delta}$.

Consider now the second term:

$$T_2 = \mathbb{P}(\hat{\sigma}_n \leqslant \sigma - \varphi_{n,\delta}) \leqslant \mathbb{P}(|\hat{F}_n(\sigma-\varphi_{n,\delta}, u_n)| \geqslant 1),$$

by definition of the estimator $\hat{\sigma}_n$. Note that

$$T_2 \leqslant \mathbb{E}|\hat{F}_n(\sigma-\varphi_{n,\delta}, u_n)|^2 = \operatorname{var}(\hat{F}_n(\sigma-\varphi_{n,\delta}, u_n)) + |F(\sigma-\varphi_{n,\delta}, u_n)|^2.$$

Since

$$|F(\sigma-\varphi_{n,\delta}, u_n)| = |\Phi(u_n)|\exp\{-\sigma^s u_n^s + (\sigma-\varphi_{n,\delta})^s u_n^s\} \leqslant \exp\{-s\sigma^{s-1}\varphi_{n,\delta}u_n^s(1+o(1))\},$$

and using the fact that by Lemma 1,

$$\operatorname{var}(\hat{F}_n(\sigma-\varphi_{n,\delta}, u_n)) \leqslant \frac{\exp\{2(\sigma-\varphi_{n,\delta})^s u_n^s\}}{n},$$

our choice of the parameters $u_n$ and $\varphi_{n,\delta}$ gives that $T_2$ converges also to zero as $n$ tends to infinity and even faster than the upper bound of $T_1$. In conclusion, the quantity

$$\begin{aligned} \mathbb{P}(|\hat{\sigma}_n - \sigma| \geqslant \varphi_{n,\delta}) &\leqslant \frac{4(1+o(1))}{nc^2}\exp\{2\sigma^s u_n^s + 2\alpha u_n^r + o(\varphi_{n,\delta}u_n^s)\} \\ &\leqslant O(1)\exp\left\{\log n\left(-1 + \left(\frac{\sigma}{\sigma_0+2\delta}\right)^s + \frac{\alpha}{(\sigma_0+2\delta)^{r-s}}\left(\frac{\log n}{2}\right)^{r/s-1}\right)\right\} \end{aligned}$$

converges to zero as $n$ tends to infinity. Moreover, note that for all $\sigma$ in $\mathcal{V}(\sigma_0)$, and large enough $n$,

$$\mathbb{E}(|\hat{\sigma}_n - \sigma|^2) = \int_0^{+\infty} \mathbb{P}(|\hat{\sigma}_n - \sigma|^2 \geqslant t)\mathrm{d}t$$

$$= \int_0^{\varphi_{n,\delta}^2} \mathbb{P}(|\hat{\sigma}_n - \sigma|^2 \geqslant t)\mathrm{d}t + \int_{\varphi_{n,\delta}^2}^{2(\sigma_0+\delta)^2} \mathbb{P}(|\hat{\sigma}_n - \sigma|^2 \geqslant t)\mathrm{d}t$$

$$\leqslant \varphi_{n,\delta}^2 + 2(\sigma_0+\delta)^2 \mathbb{P}(|\hat{\sigma}_n - \sigma| \geqslant \varphi_{n,\delta}).$$

By the previous statement, the second term on the right-hand side is negligible compared to $\varphi_{n,\delta}^2$. Finally, $\varphi_{n,\delta}/\varphi_n \to 1$ when $\delta \to 0$, giving the desired result. □

***Proof of Proposition 2 and of Theorem 2 (upper bound).*** The beginning of the proof follows the same lines and we establish that

$$\mathbb{P}(|\hat{\sigma}_n - \sigma| \geqslant \psi_{n,\delta}) = \mathbb{P}(\hat{\sigma}_n \geqslant \sigma + \psi_{n,\delta}) + \mathbb{P}(\hat{\sigma}_n \leqslant \sigma - \psi_{n,\delta}) = T_1 + T_2,$$

with

$$T_1 \leqslant \mathbb{P}(|F(\sigma + \psi_{n,\delta}, u_n)| \leqslant 1 + M) + \frac{\mathrm{e}^{2(\sigma+\psi_{n,\delta})^s u_n^s}}{nM^2}.$$

Assumption B ensures that, for large enough $n$,

$$|\Phi(u_n)| \geqslant c|u_n|^{-\beta},$$

leading to

$$\mathbb{P}(|F(\sigma + \psi_{n,\delta}, u_n)| \leqslant 1 + M) \leqslant \mathbb{P}(c \exp\{-\beta \log u_n + s\psi_{n,\delta}\sigma^{s-1}u_n^s(1+o(1))\} \leqslant 1 + M).$$

With our choice of the parameters $u_n$ and $\psi_{n,\delta}$, we have

$$\lim_{n\to\infty} -\beta \log u_n + s\psi_{n,\delta}\sigma^{s-1}u_n^s(1+o(1)) = +\infty;$$

then we choose

$$M = \frac{c}{2} \exp\{s\psi_{n,\delta}\sigma^{s-1}u_n^s - \beta \log u_n\},$$

and obtain that $\mathbb{P}(|F(\sigma + \psi_{n,\delta}, u_n)| \leqslant 1 + M)$ is null for large enough $n$. Combining with the bound on $T_1$ we obtain that, for large enough $n$,

$$T_1 \leqslant \frac{4}{nc^2} \exp\{2(\sigma + \psi_{n,\delta})^s u_n^s - 2s\sigma^{s-1}\psi_{n,\delta}u_n^s + 2\beta \log u_n\}$$

$$\leqslant \frac{4}{c^2} \exp\{-\log n + 2\sigma^s u_n^s + 2\beta \log u_n + o(\psi_{n,\delta}u_n^s)\},$$

which converges to zero with our choice of the parameters $u_n$ and $\psi_{n,\delta}$. The rest of the proof is exactly the same as in the preceding theorem. □

***Proof of Theorem 3 (upper bound).*** Fix $\sigma$ in $\mathcal{V}(\sigma_0)$ and $f$ in $\mathcal{A}(\alpha, r) \cap \mathcal{S}(\alpha', R, L)$. Denote the following neighbourhood of $\sigma$ by $\mathcal{U}(\sigma) = (\sigma - \varphi_{n,\delta}; \sigma + \varphi_{n,\delta})$. The idea of the proof is that, $\hat{\sigma}_n$ being convergent to $\sigma$, we study separately the uniform behaviour of the kernel estimator when $\hat{\sigma}_n$ is in a neighbourhood of the true value or not. For the first part we use the bias–variance decomposition and treat the uniform variance with maximal inequality for empirical processes. Then we prove that the small probability of $\hat{\sigma}_n$ being outside the neighbourhood makes the global estimation risk even smaller. We split the risk of our estimator into two terms:

$$\mathbb{E}(|\hat{f}_{n,\hat{\sigma}_n}(x) - f(x)|^2) = \mathbb{E}(|\hat{f}_{n,\hat{\sigma}_n}(x) - f(x)|^2 1_{\hat{\sigma}_n \in \mathcal{U}(\sigma)}) + \mathbb{E}(|\hat{f}_{n,\hat{\sigma}_n}(x) - f(x)|^2 1_{\hat{\sigma}_n \notin \mathcal{U}(\sigma)}) = T_1 + T_2. \quad (13)$$

We consider the first term:

$$\begin{aligned} T_1 &\leqslant \mathbb{E}\left( \sup_{\tau \in \mathcal{U}(\sigma)} |\hat{f}_{n,\tau}(x) - f(x)|^2 \right) \\ &\leqslant 2 \sup_{\tau \in \mathcal{U}(\sigma)} |\mathbb{E}\hat{f}_{n,\tau}(x) - f(x)|^2 + 2\mathbb{E}\left( \sup_{\tau \in \mathcal{U}(\sigma)} |\hat{f}_{n,\tau}(x) - \mathbb{E}\hat{f}_{n,\tau}(x)|^2 \right) \\ &\leqslant 2T_{11} + 2T_{12}. \end{aligned} \quad (14)$$

The term $T_{11}$ is the maximal bias term over $\mathcal{U}(\sigma)$. Note that

$$\mathbb{E}\hat{f}_{n,\tau}(x) = \frac{1}{\tau h_n} \int k_n \left( \frac{u-x}{\tau h_n} \right) f^Y(u) \mathrm{d}u = \frac{1}{2\pi} \int \Phi^{k_n}(\tau h_n t) \mathrm{e}^{-\mathrm{i}xt} \Phi(t) \Phi^{\varepsilon}(\sigma t) \mathrm{d}t$$

(remember that $\mathbb{E}$ is shorthand for $\mathbb{E}_{\sigma,f}$ the expectation when the unknown parameters are $\sigma$ and $f$), so that we obtain

$$\begin{aligned} T_{11} &= \sup_{\tau \in \mathcal{U}(\sigma)} \left| \frac{1}{2\pi} \int \mathrm{e}^{-\mathrm{i}xt} \Phi(t) (\Phi^{\varepsilon}(\tau t)^{-1} \Phi^{\varepsilon}(\sigma t) 1_{|t| \leqslant 1/(\tau h_n)} - 1) \mathrm{d}t \right|^2 \\ &\leqslant \frac{1}{4\pi^2} \left( \int |\Phi(t)|^2 \mathrm{e}^{2\alpha'|t|^R} \mathrm{d}t \right) \sup_{\tau \in \mathcal{U}(\sigma)} \int_{|t| \leqslant 1/(\tau h_n)} \mathrm{e}^{-2\alpha'|t|^R} |\Phi^{\varepsilon}(\tau t)^{-1} \Phi^{\varepsilon}(\sigma t) - 1|^2 \mathrm{d}t \\ &\quad + \sup_{\tau \in \mathcal{U}(\sigma)} \frac{1}{4\pi^2} \left( \int_{|t| > 1/(\tau h_n)} |\Phi(t)| \mathrm{d}t \right)^2. \end{aligned}$$

By assumption, $f$ belongs to $\mathcal{S}(\alpha', R, L)$ so that

$$\left( \int_{|t| > 1/(\tau h_n)} |\Phi(t)| \mathrm{d}t \right)^2 \leqslant L^2 \int_{|t| > 1/(\tau h_n)} \mathrm{e}^{-2\alpha'|t|^R} \mathrm{d}t \leqslant \frac{L^2 \tau^{1-R} h_n^{1-R}}{\alpha' R} \exp\left( \frac{-2\alpha'}{\tau^R h_n^R} \right) (1 + o(1)),$$

so that

$$T_{11} \leqslant \frac{L^2}{4\pi^2} \sup_{\tau \in \mathcal{U}(\sigma)} \int_{|t| \leqslant 1/(\tau h_n)} e^{-2\alpha'|t|^R} |\Phi^\varepsilon(\tau t)^{-1} \Phi^\varepsilon(\sigma t) - 1|^2 dt$$
$$+ \frac{L^2 \sigma^{1-R}}{4\pi^2 \alpha' R} h_n^{1-R} \exp\left(\frac{-2\alpha'}{\sigma^R h_n^R}\right)(1 + o(1)).$$

According to (5),

$$\Phi^\varepsilon(\tau t)^{-1} \Phi^\varepsilon(\sigma t) = \begin{cases} \exp\{(\tau^s - \sigma^s)|t|^s(1 - i\nu\, \mathrm{sgn}(t) \tan(\pi s/2)) - it\mu(\tau - \sigma)\}, & \text{if } s \neq 1, \\ \exp\left\{(\tau - \sigma)|t|\left(1 + i\nu\, \mathrm{sgn}(t) \frac{2}{\pi} \log|t|\right) - it\mu(\tau - \sigma) + it\nu \frac{2}{\pi}(\tau \log \tau - \sigma \log \sigma)\right\}, & \text{if } s = 1. \end{cases}$$

Write $\tau = \sigma + a$ with $|a| \leqslant \varphi_{n,\delta}$ and $|t| \leqslant 1/(\tau h_n)$ such that $a|t|^s = o(1)$. We obtain that

$$\Phi^\varepsilon(\tau t)^{-1} \Phi^\varepsilon(\sigma t) = \begin{cases} \exp\{sa\sigma^{s-1}|t|^s(1 + o(1))(1 - i\nu\, \mathrm{sgn}(t) \tan(\pi s/2)) - it\mu a\}, & \text{if } s \neq 1, \\ \exp\left\{(a|t|\left(1 + i\nu\, \mathrm{sgn}(t) \frac{2}{\pi} \log|t|) - it\mu a + it\nu \frac{2}{\pi} a\right)(1 + o(1))\right\}, & \text{if } s = 1, \end{cases}$$

which leads to

$$|\Phi^\varepsilon(\tau t)^{-1} \Phi^\varepsilon(\sigma t) - 1| = \begin{cases} O(1)\varphi_{n,\delta}|t|^s + O(1)\varphi_{n,\delta}\mu|t| = O(1)\varphi_{n,\delta}|t|^{\tilde{s}}, & \text{if } s \neq 1. \\ O(1)\varphi_{n,\delta}|t|(1 + \nu \log|t|), & \text{if } s = 1. \end{cases} \tag{15}$$

Returning to the upper bound on $T_{11}$, we obtain that

$$T_{11} \leqslant O(1)\left(\int |t|^{2\tilde{s}}(1 + \nu \log|t|)^2 e^{-2\alpha'|t|^R} dt\right)\varphi_{n,\delta}^2 + O(1) h_n^{1-R} \exp\left(\frac{-2\alpha'}{\sigma^R h_n^R}\right).$$

The bandwidth $h_n$ is the largest possible such that $T_{11}$ is not larger than the inevitable (large enough) loss of $\varphi_{n,\delta}^2$ . We see later that all other terms in the decomposition of $T_1$ and $T_2$ in (13) are much smaller, because $R \leqslant r < s$. Thus:

$$T_{11} \leqslant O(\varphi_{n,\delta}^2). \tag{16}$$

Return now to inequality (14) and consider the second term $T_{12}$. We have

$$|\hat{f}_{n,\tau}(x) - \mathbb{E}\hat{f}_{n,\tau}(x)|^2 = \frac{1}{4\pi^2}\left|\int \Phi^{k_n}(u\tau h_n)e^{iux}(\hat{\Phi}_n^Y(u) - \Phi^Y(u))du\right|^2$$

$$\leqslant \frac{1}{4\pi^2 n}\left(\sup_{|u|\leqslant(\tau h_n)^{-1}} |\mathbb{G}_n g_u|\right)^2 \left(\int_{|u|\leqslant(\tau h_n)^{-1}} e^{\tau^s|u|^s}du\right)^2,$$

where $\mathbb{G}$ is the empirical process associated to the measure $\mathbb{P} = \mathbb{P}_{\sigma,f}$, which means that $\mathbb{G}(g) = n^{-1/2}\sum_{i=1}^n(g(Y_i) - \mathbb{P}g)$ and the function $g_u : y \mapsto e^{iuy}$. Finally,

$$T_{12} \leqslant \frac{(\sigma h_n)^{2(s-1)}e^{2/h_n^s}}{4\pi^2 n}\left(\sup_{|u|\leqslant((\sigma-\varphi_{n,\delta})h_n)^{-1}} |\mathbb{G}_n g_u|\right)^2 (1 + o(1)). \tag{17}$$

We now use a maximal inequality to control the norm of the empirical process. The following notation can be found in more detail in van der Vaart and Wellner (1996). We consider the class of functions $\mathcal{F}_n$ defined by $\{g_u;\ |u| \leqslant ((\sigma - \varphi_{n,\delta})h_n)^{-1}\}$. The complexity of this family lies in its entropy, defined through the bracketing numbers for this class. This class satisfies,

$$\forall u, s, x \in \mathbb{R}, \qquad |g_u(x) - g_s(x)| \leqslant 2|x| \times |u - s|,$$

so that Theorem 2.7.11 in van der Vaart and Wellner (1996) applies with $F(x) = 2|x|$ and gives that the bracketing numbers for the class $\mathcal{F}_n$ (which means the minimal number of brackets of size $\epsilon$ needed to cover $\mathcal{F}_n$) are controlled by the covering numbers of $I_n = [-((\sigma - \varphi_{n,\delta})h_n)^{-1}, ((\sigma - \varphi_{n,\delta})h_n)^{-1}]$ (i.e the minimal number of balls of radius $\epsilon$ needed to cover this interval). We have

$$N_{[\,]}(2\epsilon\|F\|_{L_2(Q)};\ \mathcal{F}_n;\ L_2(Q)) \leqslant N(\epsilon;\ I_n;\ |\cdot|),$$

where $Q$ is any discrete probability measure such that $\|F\|_{L_2(Q)} > 0$. But it is easy to bound the covering numbers for $I_n$:

$$N(\epsilon;\ I_n;\ |\cdot|) \leqslant \frac{2}{\epsilon}((\sigma - \varphi_{n,\delta})h_n)^{-1}.$$

Using the fact that the covering number $N(\epsilon\|F\|_{L_2(Q)};\ \mathcal{F}_n;\ L_2(Q))$ is bounded by the bracketing number $N_{[\,]}(2\epsilon\|F\|_{L_2(Q)};\ \mathcal{F}_n;\ L_2(Q))$, we finally obtain

$$N(\epsilon\|F\|_{L_2(Q)};\ \mathcal{F}_n;\ L_2(Q)) \leqslant \frac{2}{\epsilon}((\sigma - \varphi_{n,\delta})h_n)^{-1}. \tag{18}$$

Let us define the entropy of this class by the formula

$$J(1, \mathcal{F}_n) = \sup_Q \int_0^1 \{1 + \log N(\epsilon\|F\|_{L_2(Q)};\ \mathcal{F}_n;\ L_2(Q))\}^{1/2}d\epsilon, \tag{19}$$

where the supremum is taken over all discrete probability measures $Q$. Then, Theorem 2.14.1 in van der Vaart and Wellner (1996) applies (since $\mathcal{F}_n$ is a measurable class of measurable functions with measurable envelope $F$) and gives that

$$\mathbb{E}\left\{\left(\sup_{|u|\leqslant((\sigma-\varphi_{n,\delta})h_n)^{-1}}|\mathbb{G}_n g_u|\right)^2\right\}\leqslant c\|F\|^2_{L_2(\mathbb{P})}J(1,\mathcal{F}_n)^2,$$

where $c$ is an absolute constant and $\mathbb{P}=\mathbb{P}_{\sigma,f}$. Combining the definition of the entropy (19) with inequality (18), we obtain that there exists some constant $\kappa$ such that

$$\mathbb{E}\left\{\left(\sup_{|u|\leqslant((\sigma-\varphi_{n,\delta})h_n)^{-1}}|\mathbb{G}_n g_u|\right)^2\right\}\leqslant \kappa|\log(h_n)|(1+o(1)).$$

Returning to the bound (17), we obtain

$$T_{12}=\frac{O(1)}{n}h_n^{2s-2}\mathrm{e}^{2/h_n^s}|\log(h_n)|. \tag{20}$$

Combining inequalities (14), (16) and (20), and the definition of the rate $\varphi_{n,\delta}$, gives

$$T_1\leqslant O(\varphi^2_{n,\delta}). \tag{21}$$

Return to the expression of the risk (13) and consider the second term:

$$\begin{aligned}T_2&=\mathbb{E}(|\hat{f}_{n,\hat{\sigma}_n}(x)-f(x)|^2 1_{\hat{\sigma}_n\notin\mathcal{U}(\sigma)})\\&\leqslant 2\mathbb{E}\{(\|\hat{f}_{n,\hat{\sigma}_n}\|^2_\infty+\|f\|^2_\infty)1_{\hat{\sigma}_n\notin\mathcal{U}(\sigma)}\}\end{aligned}$$

But we know that

$$\|\hat{f}_{n,\hat{\sigma}_n}\|^2_\infty\leqslant\frac{1}{|\hat{\sigma}_n|^2}\times\frac{h_n^{2(s-1)}\mathrm{e}^{2/h_n^s}}{\pi^2 s^2}(1+o(1)),$$

and that

$$\mathbb{E}\left(\frac{1}{|\hat{\sigma}_n|^2}1_{\hat{\sigma}_n\notin\mathcal{U}(\sigma)}\right)=\frac{1}{\sigma^2}\mathbb{P}(\hat{\sigma}_n\notin\mathcal{U}(\sigma))(1+r),$$

where

$$\begin{aligned}|r|&=\sigma^2 P(\hat{\sigma}_n\notin\mathcal{U}(\sigma))^{-1}|\mathbb{E}\{(\hat{\sigma}_n^{-2}-\sigma^{-2})1_{\hat{\sigma}_n\notin\mathcal{U}(\sigma)}\}|\\&\leqslant\sigma^2\mathbb{E}|\hat{\sigma}_n^{-2}-\sigma^{-2}|^2\\&\leqslant\sigma^{-2}\mathbb{E}\{\hat{\sigma}_n^{-4}|\hat{\sigma}_n^2-\sigma^2|^2\}.\end{aligned}$$

But $\hat{\sigma}_n$ converges in probability to $\sigma>0$, so that $\hat{\sigma}_n^{-4}$ is bounded in probability and, finally,

$$|r|=O(1)\mathbb{E}|\hat{\sigma}_n^2-\sigma^2|^2=o(1).$$

We obtain

$$T_2\leqslant 2\left(\frac{h_n^{2(s-1)}\mathrm{e}^{2/h_n^s}}{\pi^2 s^2\sigma^2}(1+o(1))+\|f\|^2_\infty\right)\mathbb{P}(\hat{\sigma}_n\notin\mathcal{U}(\sigma)).$$

As $f$ belongs to $\mathcal{A}(\alpha, r)$, Proposition 1 gives that

$$T_2 \leqslant O(1) h_n^{2(s-1)} \mathrm{e}^{2/h_n^s} \exp\left\{ \frac{2\alpha}{(\sigma_0 + 2\delta)^r} \left( \frac{\log n}{2} \right)^{r/s} \right\} \left( \frac{1}{n} \right)^{1 - \sigma^s/(\sigma_0 + 2\delta)^s}. \tag{22}$$

Combining inequalities (13), (21) and (22) gives the desired result. □

***Proof of Theorem 4 (upper bounds).*** Fix $\sigma$ in $\mathcal{V}(\sigma_0)$ and $f$ in $\mathcal{A}(\alpha, r) \cap W(\beta', L)$. We only sketch the proof as it follows the same lines and notation as the proof of Theorem 3 (upper bound). Here the term $T_{11}$ is written as

$$\begin{aligned} T_{11} &= \sup_{\tau \in \mathcal{U}(\sigma)} \left| \frac{1}{2\pi} \int \mathrm{e}^{-\mathrm{i}xt} \Phi(t) (\Phi^\varepsilon(\sigma t) \Phi^\varepsilon(\tau t)^{-1} \mathbb{1}_{|t| \leqslant 1/(\tau h_n)} - 1) \mathrm{d}t \right|^2 \\ &\leqslant \frac{1}{4\pi^2} \left( \int |\Phi(t)|^2 (1 + |t|^{2\beta'}) \mathrm{d}t \right) \sup_{\tau \in \mathcal{U}(\sigma)} \int_{|t| \leqslant 1/(\tau h_n)} (1 + |t|^{2\beta'})^{-1} |\Phi^\varepsilon(\sigma t) \Phi^\varepsilon(\tau t)^{-1} - 1|^2 \mathrm{d}t \\ &\quad + \sup_{\tau \in \mathcal{U}(\sigma)} \frac{1}{4\pi^2} \left( \int_{|t| > 1/(\tau h_n)} |\Phi(t)| \mathrm{d}t \right)^2. \end{aligned}$$

But here $f$ belongs to $W(\beta', L)$, so that

$$\left( \int_{|t| > 1/(\tau h_n)} |\Phi(t)| \mathrm{d}t \right)^2 \leqslant L^2 \int_{|t| \geqslant 1/(\tau h_n)} (1 + |t|^{2\beta'})^{-1} \,\mathrm{d}t \leqslant \frac{2}{2\beta' - 1} L^2 \tau^{2\beta' - 1} h_n^{2\beta' - 1} (1 + o(1)).$$

In the same way as we established the bound (16) using the expressions given in equality (15), we obtain

$$T_{11} \leqslant O(1) \varphi_{n,\delta}^2 \int_{|t| \leqslant 1/(\tau h_n)} \frac{|t|^{2\tilde{s}} (1 + \nu \log |t| \mathbb{1}_{s=1})^2}{1 + |t|^{2\beta'}} \,\mathrm{d}t + O(1) h_n^{2\beta' - 1}.$$

In the case $\beta' > \tilde{s} + \frac{1}{2}$, we bound $\int_{|t| \leqslant 1/(\tau h_n)} |t|^{2\tilde{s}} (1 + \nu \log |t|)^2 / (1 + |t|^{2\beta'}) \mathrm{d}t$ by the constant limit. The choice of the bandwidth $h_n = (\log n)^{2(r/s-1)/(2\beta'-1)}$ is the largest such that

$$T_{11} \leqslant O(\varphi_{n,\delta}^2) = O(1)(\log n)^{2(r/s-1)}.$$

In the case $\beta' < \tilde{s} + \frac{1}{2}$, and when $s = 1$ and $\nu = 0$, we evaluate the rate of divergence of the integral in the bound of $T_{11}$ and obtain a global slower rate of convergence. Here,

$$T_{11} \leqslant O(1) h_n^{2\beta' - 2\tilde{s} - 1} \varphi_{n,\delta}^2 + O(1) h_n^{2\beta' - 1}.$$

This bound is optimized for $h_n = (\log n)^{(r/s-1)/\tilde{s}}$, giving the global rate of order

$$T_{11} \leqslant O(1) (\varphi_{n,\delta})^{(2\beta'-1)/\tilde{s}} = O(1) (\log n)^{(r/s-1)(2\beta'-1)/\tilde{s}}.$$

More generally, when $\beta' \leqslant \tilde{s} + \frac{1}{2}$, we can evaluate the rate of divergence of the remaining integral in every case, which gives the more precise results presented in Section 3.4. The rates obtained differ from the previous ones only by powers of $\log \log n$.

The controls of the terms $T_{12}$ and $T_2$ remain valid and they are much smaller than $T_{11}$ giving the global rate, and we automatically obtain different results. □

***Proof of Theorem 5 (upper bounds).*** Fix $\sigma$ in $\mathcal{V}(\sigma_0)$ and $f$ in $\mathcal{B}(\beta) \cap W(\beta', L)$. The first part of the proof of Theorem 4 (upper bounds) applies and we just replace $\varphi_{n,\delta}$ by $\psi_{n,\delta}$:

$$T_{11} \leqslant O(1)\psi_{n,\delta}^2 \int_{|t| \leqslant 1/(\tau h_n)} \frac{|t|^{2\tilde{s}}(1 + \nu \log |t| \mathbb{1}_{s=1})^2}{1 + |t|^{2\beta'}} \, \mathrm{d}t + O(1) h_n^{2\beta'-1}.$$

In a similar manner, in the case $\beta' > \tilde{s} + \frac{1}{2}$, we bound the previous integral by its constant limit and choose $h_n = (\log\log n/\log n)^{2/(2\beta'-1)}$ as large as possible such that

$$T_{11} \leqslant O(\psi_{n,\delta}^2) = O(1)\left(\frac{\log\log n}{\log n}\right)^2.$$

In the other cases, a loss in rate is inevitable. When $\beta' < \tilde{s} + \frac{1}{2}$ and if $s = 1$ and $\nu = 0$, we obtain

$$T_{11} \leqslant O(1) h_n^{2\beta'-2\tilde{s}-1} \psi_{n,\delta}^2 + O(1) h_n^{2\beta'-1}.$$

The optimal bandwidth is $h_n = (\log\log n/\log n)^{1/\tilde{s}}$, giving a slower risk rate of order

$$T_{11} \leqslant O((\psi_{n,\delta})^{(2\beta'-1)/\tilde{s}}) = O(1)\left(\frac{\log\log n}{\log n}\right)^{(2\beta'-1)/\tilde{s}}.$$

More generally, evaluating the rate of divergence of the integral appearing in the bound of $T_{11}$ gives the remaining cases.

The control of the term $T_{12}$ remains valid. The control of the last term, $T_2$, follows the same lines, leading to

$$T_2 \leqslant 2\left(\|f\|_\infty^2 + \frac{h_n^{2(s-1)} \mathrm{e}^{2/h_n^s}}{\pi^2 s^2 \sigma^2}(1 + o(1))\right)\mathbb{P}(\hat{\sigma}_n \notin \mathcal{U}(\sigma)).$$

As $f$ belongs to $\mathcal{B}(\beta)$, Theorem 2 gives that

$$T_2 \leqslant O(1) h_n^{2(s-1)} \mathrm{e}^{2/h_n^s} (\log n)^{2\beta/s} \left(\frac{1}{n}\right)^{1-\sigma^s/(\sigma_0+2\delta)^s},$$

and we obtain the desired results. Indeed, under the respective hypotheses with chosen bandwidths, $T_{12}$ and $T_2$ converge to 0 faster than $T_{11}$. □

# 5. Proofs: lower bounds

## 5.1. Prerequisites

The proofs of the lower bounds in all these theorems are based on suitable choices of two models with convenient parameters being as far from each other as possible, such that the convolution models are close in $\chi^2$-distance.

The following proposition is the main tool in the proof of our lower bounds and can be found in Butucea and Tsybakov (2004). The notation $\chi^2(P, Q)$ denotes the $\chi^2$-distance between the probabilities $P$ and $Q$:

$$\chi^2(P, Q) = \begin{cases} \displaystyle\int \left(\frac{\mathrm{d}P}{\mathrm{d}Q} - 1\right)^2 \mathrm{d}Q, & \text{if } P \ll Q, \\ +\infty, & \text{otherwise.} \end{cases}$$

**Proposition 3.** *Let $\mathcal{P}_\Theta = \{\mathbb{P}_\theta; \theta \in \Theta\}$ be a family of models. Assume that there exist $\theta_1$ and $\theta_2$ in $\Theta$ with $|\theta_2 - \theta_1| \geqslant 2s_n > 0$ such that the probability measures $\mathbb{P}_1 = \mathbb{P}_{\theta_1}$ and $\mathbb{P}_2 = \mathbb{P}_{\theta_2}$ satisfy*

$$\mathbb{P}_1 \ll \mathbb{P}_2 \quad \text{and} \quad \chi^2(\mathbb{P}_1^{\otimes n}, \mathbb{P}_2^{\otimes n}) \leqslant K^2 < \infty.$$

*Then we have*

$$\liminf_{n\to\infty} \inf_{\hat{\theta}_n} s_n^{-2} \max\{\mathbb{E}_1(|\hat{\theta}_n - \theta_1|^2), \mathbb{E}_2(|\hat{\theta}_n - \theta_2|^2)\} \geqslant (1 - K)^2(1 - \sqrt{K})^2,$$

*where the infimum is over any estimator $\hat{\theta}_n$ of the underlying parameter and this bound is actually arbitrary close to 1 for $K$ small enough.*

Now, the previous lower bounds are established by the construction, in each different case (density in $\mathcal{A}(\alpha, r)$ or $\mathcal{B}(\beta)$), of two particular models $\mathbb{P}_1 = \mathbb{P}_{\sigma_1, f_1}$ and $\mathbb{P}_2 = \mathbb{P}_{\sigma_2, f_2}$, with $\chi^2$ distance converging to zero. Note that

$$\sup_\sigma \sup_f s_n^{-2} \mathbb{E}(|\hat{\sigma}_n - \sigma|^2) \geqslant s_n^{-2} \max\{\mathbb{E}_1(|\hat{\sigma}_n - \sigma_1|^2), \mathbb{E}_2(|\hat{\sigma}_n - \sigma_2|^2)\},$$

$$\sup_\sigma \sup_f s_n^{-2} \mathbb{E}(d^2(\hat{f}_n, f)) \geqslant s_n^{-2} \max\{\mathbb{E}_1(d^2(\hat{f}_n, f_1)), \mathbb{E}_2(d^2(\hat{f}_n, f_2)\},$$

where $d(f, g)$ denotes pointwise absolute difference $|f(x_0) - g(x_0)|$ or $\mathbb{L}_2$ norm $\|f - g\|_2$ of two arbitrary functions $f$ and $g$, respectively. Proposition 3 for the particular models constructed in the proof entails the results. Note that the different rates of convergence $s_n$ correspond to half of the distance between the parameters $\sigma_1$ and $\sigma_2$, when estimating the scale; and to $d(f_1, f_2)/2$, when estimating the unknown density.

## 5.2. Construction

We construct two models which are close in $\chi^2$-distance but come from parameters which are far enough from each other. They are used throughout the proofs of the lower bounds with suitable choices of densities and scale parameters, under Assumptions A and B, respectively.

Let us fix the scale parameter $\sigma_0$ and a symmetric density $f_1$ in the class we consider, having Fourier transform $\Phi_1$. The first model has deconvolution density $f_1$ and scale parameter $\sigma_1 = (1 + t)^{1/s}\sigma_0$. In this model, observations $Y_1, \ldots, Y_n$ have density $f_1^Y(x) =$

$f_1 * \{f^{\varepsilon}(\cdot/\sigma_1)/\sigma_1\}(x)$ and Fourier transform $\Phi_1^Y(u) = \Phi_1(u)\Phi^{\varepsilon}(\sigma_1 u)$. Recall that the noise has stable density $S(1, s, \nu, \mu)$.

Consider next a perturbation of this model, having scale parameter $\sigma_2 = (1-t)^{1/s}\sigma_0$ and a deconvolution density $f_2$ defined by its Fourier transform,

$$\Phi_2(u) = \Phi_1(u)\left\{\Phi^{\varepsilon}((2t)^{1/s}\sigma_0 u)\mathrm{e}^{-\mathrm{i}ua_t}k^*\left(\frac{u}{M}\right) + 1 - k^*\left(\frac{u}{M}\right)\right\}, \tag{23}$$

where the auxiliary function $k$ has Fourier transform $k^*$, $M = M_n$ is some sequence of positive numbers, and the real-valued function $a_t$ is defined by the relations

$$a_t = \begin{cases} \mu(\sigma_2 + (2t)^{1/s}\sigma_0 - \sigma_1), & \text{if } s \neq 1, \\ -\nu\sigma_0 \dfrac{2}{\pi}\left\{(1-t)\log\left(\dfrac{1-t}{2}\right) + (1+t)\log\left(\dfrac{1+t}{2}\right)\right\}, & \text{if } s = 1. \end{cases} \tag{24}$$

Indeed, by (5) and a simple computation,

$$\Phi^{\varepsilon}(\sigma_2 u)\Phi^{\varepsilon}((2t)^{1/s}\sigma_0 u) = \Phi^{\varepsilon}(\sigma_1 u)\mathrm{e}^{\mathrm{i}ua_t}.$$

We denote in this case $f_2^Y(x) = f_2 * \{f^{\varepsilon}(\cdot/\sigma_2)/\sigma_2\}(x)$ and $\Phi_2^Y(u) = \Phi_2(u)\Phi^{\varepsilon}(\sigma_2 u)$.

This construction is actually based on Fourier transforms $\Phi_1$, $\Phi_2$ which have the same behaviour for large values of $u$, so that they belong to the same class of densities. Moreover, the resulting models $\Phi_1^Y$, $\Phi_2^Y$ coincide (in absolute value) on a large interval around 0 in order to obtain models close together in $\chi^2$-distance. By Proposition 3, the rate of convergence is given, for small $t$, by

$$|\sigma_1 - \sigma_2| = \frac{2\sigma_0}{s}t(1 + o(1)),$$

when we estimate the scale parameter $\sigma$, and by the differences

$$|f_1(x) - f_2(x)| = \frac{1}{2\pi}\left|\int \mathrm{e}^{-\mathrm{i}xu}(\Phi_1 - \Phi_2)(u)\mathrm{d}u\right| \quad \text{or} \quad \|f_1 - f_2\|_2 = (2\pi)^{-1/2}\|\Phi_1 - \Phi_2\|_2$$

when we are interested in pointwise or $\mathbb{L}_2$ estimation of $f$.

Let us proceed to the proof of the lower bounds via an auxiliary result.

**Lemma 2.** *Let $g$ and $h$ be two non-negative functions such that $g$ has a unique mode and $\int h(x)\mathrm{d}x = c > 0$. Then the convolution product $g * h$ satisfies*

$$g * h(x) \geqslant \frac{c}{2}\min\{g(x+A), g(x-A)\},$$

*for some large enough $A > 0$. If $g$ is symmetric the lower bound becomes $g(|x| + A)c/2$.*

***Proof.*** It is obvious that for some $A > 0$ large enough, $\int_{-A}^{A} h(u)\mathrm{d}u \geqslant c/2$ and

$$g * h(x) \geqslant \int_{-A}^{A} g(x-u)h(u)\mathrm{d}u \geqslant \min\{g(x+A), g(x-A)\}\int_{-A}^{A} h(u)\mathrm{d}u,$$

which concludes the proof. □

Throughout the following sections, $C$ denotes some positive constant.

### 5.3. Deconvolution densities in the class $\mathcal{A}(\alpha, r)$

We now particularize the choice of the function $f_1$ to deal with the case of deconvolution densities belonging to $\mathcal{A}(\alpha, r)$.

**Lemma 3.** *Consider the function $\Phi_1(u) = \mathrm{e}^{-\alpha|u|^r}$ which is the Fourier transform of a symmetric stable density $f_1$ in the class $\mathcal{A}(\alpha, r)$. There exists a kernel $k$ such that:*

(a) *$k$ is an even function;*
(b) *the Fourier transform $k^*$ has a support included in $[-2, 2]$;*
(c) *for all $u$ in $[-1, 1]$, $k^*(u) = 1$;*
(d) *$k^*$ is four times continuously differentiable on $\mathbb{R}$ (i.e. $\mathcal{C}^4$).*

*Consider the function $\Phi_2$ defined by (23). Then $\Phi_2$ is the Fourier transform of a density $f_2$ included in $\mathcal{A}(\alpha, r)$ for all large enough $M$ and small enough $t > 0$.*

***Proof.*** Without loss of generality, we assume that $\sigma_0 = 1$.

Let us construct a function $k$ with the desired properties. Consider the function $g(x) = \sin x/(\pi x)$, with $g^*(u) = 1_{|u|\leqslant 1}$. Next, consider successive convolutions of $g^*$ with itself, say $g^{*32}$ having support on $[-32, 32]$ and being four-times continuously differentiable, corresponding to a positive density function $g^{32}(x)$. Let us rescale this function $G^*(u) = g^{*32}(u/32)/32$ and finally integrate $G^*$ as follows:

$$k^*(u) = \int_{u-3/2}^{u+3/2} 2G^*(2v)\mathrm{d}v.$$

Remember that $f_2$ denotes the function $f_2(x) = (2\pi)^{-1} \int \exp(-\mathrm{i}ux)\Phi_2(u)\mathrm{d}u$. Since $\Phi_2(0) = 1$, we know that $\int f_2(x)\mathrm{d}x = 1$. Our purpose is to establish that $f_2$ is a positive function (and then a density function). The fact that $f_2$ belongs to the class $\mathcal{A}(\alpha, r)$ is a direct consequence of the construction of $\Phi_2$ (see equation (23)), since the kernel $k$ has a Fourier transform boundedly supported.

The argument for the positivity of $f_2$ consists of two steps. First, we prove that the uniform distance $\|f_2 - f_1\|_\infty$ converges to zero as $t$ tends to zero and $M$ tends to infinity, and then ($f_1$ being strictly positive; see Zolotarev 1986, Remark 4 after Theorem 2.2.3), for all fixed compact $K$ in $\mathbb{R}$, small enough $t$ and large enough $M$, we obtain $f_2(x) > 0$ for all $x$ in $K$ . The second step is to establish that, for large enough $|x|$, we have

$$f_2(x) \geqslant \frac{C}{|x|^{r+1}} + \frac{O(1)}{|x|^3}, \tag{25}$$

for some constant $C > 0$, and since $r < s \leqslant 2$, we conclude that, for large enough $|x|$, the function $f_2$ is positive.

Let us establish the first step. Note that

$$\Phi_2(u) = \Phi_1(u)\Phi^\varepsilon((2t)^{1/s}u)\mathrm{e}^{-\mathrm{i}ua_t} + \Phi_1(u)\left\{1 - k^*\left(\frac{u}{M}\right)\right\}(1 - \Phi^\varepsilon((2t)^{1/s}u)\mathrm{e}^{-\mathrm{i}ua_t}),$$

and consequently

$$f_2(x) = \left[f_1 * \left\{\frac{1}{(2t)^{1/s}} f^\varepsilon\left(\frac{\cdot}{(2t)^{1/s}}\right)\right\}\right](x + a_t)$$
$$+ \frac{1}{2\pi}\int \mathrm{e}^{-\mathrm{i}ux}\Phi_1(u)\left\{1 - k^*\left(\frac{u}{M}\right)\right\}(1 - \Phi^\varepsilon((2t)^{1/s}u)\mathrm{e}^{-\mathrm{i}ua_t})\mathrm{d}u. \qquad (26)$$

Using the fact that the kernel $k$ satisfies

$$\left|1 - k^*\left(\frac{u}{M}\right)\right| \leqslant 1_{|u|\geqslant M},$$

the second term on the right-hand side of (26) is bounded by

$$\left|\frac{1}{2\pi}\int \mathrm{e}^{-\mathrm{i}ux}\Phi_1(u)\left\{1 - k^*\left(\frac{u}{M}\right)\right\}(1 - \Phi^\varepsilon((2t)^{1/s}u)\mathrm{e}^{-\mathrm{i}ua_t})\mathrm{d}u\right| \leqslant \frac{1}{\pi}\int_{|u|\geqslant M} \mathrm{e}^{-\alpha|u|^r}\mathrm{d}u,$$

which is $O(M^{1-r}\mathrm{e}^{-\alpha M^r})$, uniformly in $x$, and then converges to zero as $M$ tends to infinity. Now let us denote by $f_t^\varepsilon$ the scale-transformed function $(2t)^{-1/s}f^\varepsilon((2t)^{-1/s}\cdot)$, so that the first term in (26) is the convolution of the continuous and bounded function $f_1$ with $f_t^\varepsilon$, combined with a translation by $a_t$. We obtain that

$$\|f_1 * f_t^\varepsilon(\cdot + a_t) - f_1\|_\infty \leqslant \|f_1 * f_t^\varepsilon - f_1\|_\infty + \|f_1 * f_t^\varepsilon - f_1 * f_t^\varepsilon(\cdot + a_t)\|_\infty$$
$$= o(1) + \|f_1 * f_t^\varepsilon - f_1 * f_t^\varepsilon(\cdot + a_t)\|_\infty,$$

as $t$ tends to zero, by properties of approximate convolution identities. Now using the fact that $a_t \to 0$ (see (24)) and that $f_t^\varepsilon$ (and thus also the convolution product) is continuously differentiable, with uniformly bounded derivative, we have

$$\|f_1 * f_t^\varepsilon - f_1 * f_t^\varepsilon(\cdot + a_t)\|_\infty = O(a_t) = o(1),$$

as $t$ tends to zero. Returning to (26), we obtain that

$$\|f_2 - f_1\|_\infty = o(1), \qquad \text{as } M \to \infty \text{ and } t \to 0.$$

Denote by $\Psi$ the function

$$\Psi(u) = \Phi_1(u)\left\{1 - k^*\left(\frac{u}{M}\right)\right\}(1 - \Phi^\varepsilon((2t)^{1/s}u)\mathrm{e}^{-\mathrm{i}ua_t}),$$

in such a way, that according to (26),

$$f_2(x) = (f_1 * f_t^\varepsilon)(x + a_t) + \frac{1}{2\pi}\int \mathrm{e}^{-\mathrm{i}ux}\Psi(u)\mathrm{d}u.$$

Using the fact that $\Psi$ is three times continuously differentiable, identically equal to zero on $[-2M, 2M]$ and vanishes at infinity in the same way as its derivatives, integration by parts establishes that

$$\left|\int \mathrm{e}^{-\mathrm{i}ux}\Psi(u)\mathrm{d}u\right| = \left|\int \mathrm{e}^{-\mathrm{i}ux}\frac{\Psi^{(3)}(u)}{(-\mathrm{i}x)^3}\mathrm{d}u\right| = \frac{O(1)}{|x|^3},$$

since we have $\|\Psi^{(3)}\|_1 < \infty$. This means that

$$f_2(x) = (f_1 * f_t^\varepsilon)(x + a_t) + \frac{O(1)}{|x|^3}.$$

We now apply Lemma 2 with the densities $f_1$ and $f_t^\varepsilon$, the former being a symmetric function with unique mode at zero, which gives

$$f_2(x) \geqslant \frac{1}{2} f_1(|x| + A) + \frac{O(1)}{|x|^3},$$

for some large enough $A > 0$. Since $\Phi_1(u) = \mathrm{e}^{-\alpha|u|^r}$ with $r < 2$, we know that the asymptotic behaviour of $f_1(x)$ is $C/|x|^{r+1}$ for some positive constant $C$ (if $r = 1$, this is the Cauchy distribution; for $r \neq 1$, see Zolotarev 1986, equations (2.4.8) and (2.5.4)). Finally, we obtain equality (25) and conclude the proof. □

***Proof of Theorem 1 (lower bound).*** Consider, for arbitrary small $\delta > 0$, the sequences of positive numbers

$$t = t_n = \sqrt{1-\delta}\left\{\frac{\alpha}{\sigma_0^r}\left(\frac{\log n}{2}\right)^{r/s-1} - \frac{7-s}{s}\frac{\log\log n}{\log n}\right\} \quad \text{and} \quad M = M_n = \left(\frac{\log n}{2\sigma_0^s}\right)^{1/s}. \quad (27)$$

Note that, for $n$ large enough, $\sigma_1, \sigma_2 \in \mathcal{V}(\sigma_0)$. Now the densities $f_1$ and $f_2$ constructed in Lemma 3 with the preceding choice of parameters belong to the class $\mathcal{A}(\alpha, r)$ (for large enough $n$). Then, applying Proposition 3, we need to control the distance $\chi^2(\mathbb{P}_1^{\otimes n}, \mathbb{P}_2^{\otimes n}) = n\chi^2(f_1^Y, f_2^Y)$. Write

$$n\chi^2(f_1^Y, f_2^Y) = n\int \frac{|f_1^Y(y) - f_2^Y(y)|^2}{f_1^Y(y)}\,\mathrm{d}y,$$

and use Lemma 2 and the relation $f_1^Y(x) = f_1 * [f^\varepsilon(\cdot/\sigma_1)/\sigma_1](x)$ to bound this expression:

$$n\chi^2(f_1^Y, f_2^Y) \leqslant n\int \frac{|f_1^Y(y) - f_2^Y(y)|^2}{f_1(|y| + A)}\,\mathrm{d}y,$$

for some large enough $A$. Now split this integral into two terms and use the fact that $f_1$ is a strictly positive function, with behaviour $O(1/|x|^{r+1})$ at infinity, to obtain

$$n\chi^2(f_1^Y, f_2^Y) \leqslant nC_1\int_{|y|\leqslant A} |f_1^Y(y) - f_2^Y(y)|^2\,\mathrm{d}y + nC_2\int_{|y|>A} |y|^{r+1}|f_1^Y(y) - f_2^Y(y)|^2\,\mathrm{d}y \quad (28)$$

for some positive constants $C_1$ and $C_2$.

Consider the first term on the right-hand side:

$$T_1 = nC_1\int_{|y|\leqslant A} |f_1^Y(y) - f_2^Y(y)|^2\,\mathrm{d}y \leqslant nC_1\int |f_1^Y(y) - f_2^Y(y)|^2\,\mathrm{d}y = \frac{nC_1}{2\pi}\|\Phi_1^Y - \Phi_2^Y\|_2^2.$$

By definition,

$$\Phi_2^Y(u) = \Phi_2(u)\Phi^\varepsilon(\sigma_2 u) = \Phi_1(u)\Big\{\Phi^\varepsilon((2t)^{1/s}\sigma_0 u)\mathrm{e}^{-\mathrm{i}ua_t} k^*\Big(\frac{u}{M}\Big) + 1 - k^*\Big(\frac{u}{M}\Big)\Big\}\Phi^\varepsilon(\sigma_2 u)$$

and

$$\Phi_1^Y(u) = \Phi_1(u)\Phi^\varepsilon(\sigma_1 u) = \Phi_1(u)\Phi^\varepsilon((2t)^{1/s}\sigma_0 u)\Phi^\varepsilon(\sigma_2 u)\mathrm{e}^{-\mathrm{i}ua_t},$$

so that

$$\begin{aligned} |\Phi_1^Y(u) - \Phi_2^Y(u)| &= |\Phi_1(u)\Phi^\varepsilon(\sigma_2 u)\Big\{1 - k^*\Big(\frac{u}{M}\Big)\Big\}(1 - \Phi^\varepsilon((2t)^{1/s}\sigma_0 u)\mathrm{e}^{-\mathrm{i}ua_t})| \qquad (29)\\ &\leqslant 2\mathrm{e}^{-\alpha|u|^r - (1-t)\sigma_0^s|u|^s}1_{|u|>M}. \end{aligned}$$

Then

$$T_1 \leqslant \frac{2nC_1}{\pi}\int_{|u|>M} \mathrm{e}^{-2\alpha|u|^r - 2(1-t)\sigma_0^s|u|^s}\mathrm{d}u = O(nM^{1-s}\mathrm{e}^{-2\alpha M^r - 2(1-t)\sigma_0^s M^s}).$$

But $M = (\log n/2\sigma_0^s)^{1/s}$ and, by our choice of $t$, given in (27), we have $T_1 = o(1)$.

Let us deal with the second term appearing on the right-hand side in (28):

$$T_2 = nC_2\int_{|y|>A} |y|^{r+1}|f_1^Y(y) - f_2^Y(y)|^2\,\mathrm{d}y \leqslant nC_2\int |y|^4|f_1^Y(y) - f_2^Y(y)|^2\,\mathrm{d}y.$$

By Parseval's equality and since $(\Phi_1^Y - \Phi_2^Y)(u)$ is $\mathcal{C}^4$ on its support $\{|u| \geqslant M\}$,

$$T_2 \leqslant \frac{nC_2}{2\pi}\int |(\Phi_1^Y - \Phi_2^Y)''(u)|^2\mathrm{d}u,$$

and according to the expression (29) for the difference $\Phi_1^Y - \Phi_2^Y$, we bound this term by

$$T_2 \leqslant nC_2'\int_{|u|\geqslant M} |u|^6\mathrm{e}^{-2\alpha|u|^r - 2(1-t)\sigma_0^s|u|^s}\mathrm{d}u = O(nM^{7-s}\mathrm{e}^{-2\alpha M^r - 2(1-t)\sigma_0^s M^s}),$$

and conclude in exactly the same way that $T_2 = o(1)$.

Then, using Proposition 3,

$$\inf_{\hat\sigma_n}\sup_{f,\sigma}\varphi_n^{-2}\mathbb{E}\{|\hat\sigma_n - \sigma|^2\} \geqslant (1-\delta)\frac{\inf_{\hat\sigma_n}\max_{i=1,2}\mathbb{E}\{|\hat\sigma_n - \sigma_i|^2\}}{(\sigma_0 t/s)^2} \geqslant 1 - \delta,$$

for arbitrary small $\delta > 0$, hence the theorem. □

***Proof of Theorem 3 (lower bound).*** The proof uses the same construction as for the Theorem 1 (lower bound). We therefore use the notation and reasoning introduced in Lemma 3. We again apply Proposition 3 for functions $f_1$ and $f_2$. Indeed, $\Phi_1(u) = \exp(-\alpha|u|^r)$ and thus $f_1$ belongs to $\mathcal{A}(\alpha, r) \cap \mathcal{S}(\alpha', R, L)$ if $R < r$ or if $R = r$ and $\alpha' < \alpha$. We have already seen that $f_2$ is in $\mathcal{A}(\alpha, r)$. Let us remark that $|\Phi_2(u)| \leqslant |\Phi_1(u)|$ and then $f_2$ belongs to $\mathcal{S}(\alpha', R, L)$, too.

As we have already verified that $n\chi^2(f_1^Y, f_2^Y) = o(1)$, when $n \to \infty$, for $t$ and $M$ given by (27), it is enough (by Proposition 3) to obtain a lower bound of $|f_1(x) - f_2(x)|$. Without loss of generality we can evaluate:

$$|f_1(0) - f_2(0)| = \frac{1}{2\pi}\left|\int(\Phi_1(u) - \Phi_2(u))\mathrm{d}u\right|$$

$$= \frac{1}{2\pi}\left|\int \Phi_1(u)k^*\left(\frac{u}{M}\right)(1 - \Phi^\varepsilon((2t)^{1/s}\sigma_0 u)\mathrm{e}^{-\mathrm{i}ua_t})\mathrm{d}u\right|.$$

Using the definition of the characteristic functions of stable laws, we get that the real part

$$\mathrm{Re}(\Phi^\varepsilon((2t)^{1/s}\sigma_0 u)\mathrm{e}^{-\mathrm{i}ua_t}) = \mathrm{e}^{-2t\sigma_0^s|u|^s}\cos(R(t, u)),$$

for some function $R(t, u)$. This leads to the lower bound

$$|f_1(0) - f_2(0)| \geqslant \frac{1}{2\pi}\left|\int \mathrm{e}^{-\alpha|u|^r}k^*\left(\frac{u}{M}\right)\{1 - \mathrm{e}^{-2t\sigma_0^s|u|^s}\cos(R(t, u))\}\mathrm{d}u\right|$$

$$\geqslant \frac{1}{2\pi}\int \mathrm{e}^{-\alpha|u|^r}k^*\left(\frac{u}{M}\right)(1 - \mathrm{e}^{-2t\sigma_0^s|u|^s})\mathrm{d}u$$

$$\geqslant \frac{1}{2\pi}\left(\int_{|u|\leqslant 1}\mathrm{e}^{-\alpha|u|^r}(1 - \mathrm{e}^{-2t\sigma_0^s|u|^s})\mathrm{d}u + \int_{1<|u|\leqslant M}\mathrm{e}^{-\alpha|u|^r}(1 - \mathrm{e}^{-2t\sigma_0^s})\mathrm{d}u\right),$$

as $k^*$ is a positive function and for large enough $M$. Finally, both terms above are of order $O(t)$:

$$|f_1(0) - f_2(0)| \geqslant \frac{1}{2\pi}\int_{|u|\leqslant 1}\mathrm{e}^{-\alpha|u|^r}2t\sigma_0^s|u|^s\mathrm{d}u(1 + o(1)) + \frac{2t\sigma_0^s}{2\pi}\int_{1<|u|\leqslant M}\mathrm{e}^{-\alpha|u|^r}\mathrm{d}u$$

$$\geqslant Ct \geqslant C(\log n)^{r/s-1},$$

and the proof is complete. □

***Proof of Theorem 4 (lower bounds).*** The same construction of functions $f_1$ and $f_2$ remains valid for the model. Indeed, the deconvolution densities are supersmooth so that they belong to the Sobolev class $W(\beta', L)$ as well. Thus we obtain the lower rate of convergence $\varphi_n$ whatever the value of $\beta'$. But this rate is nearly optimal (because of a logarithmic loss) in the case $\beta' = \tilde{s} + \frac{1}{2}$ and too small (this bound is too low) in the case $\beta' < \tilde{s} + \frac{1}{2}$, where the rate of our estimator is $\varphi_n^{(2\beta'-1)/2\tilde{s}}$. □

## 5.4. Deconvolution densities in the class $\mathcal{B}(\beta)$

Those proofs will follow the same lines as the ones concerning deconvolution densities in $\mathcal{A}(\alpha, r)$. We choose a new function $f_1$ belonging to the class $\mathcal{B}(\beta)$ such that the resulting function $f_2$ (defined via its Fourier transform $\Phi_2$ and equation (23)) also belongs to the set $\mathcal{B}(\beta)$.

**Lemma 4.** *Let* $\Phi_1(u) = \frac{1}{2}(1+u^2)^{-\beta/2} + \frac{1}{2}\mathrm{e}^{-|u|/2}$, *with* $\beta > 1$. *This function is the Fourier transform of some density* $f_1$ *in the class* $\mathcal{B}(\beta)$. *Use the kernel* $k$ *constructed in Lemma 3 and define the function* $\Phi_2$ *by equation* (23). *Then* $\Phi_2$ *is the Fourier transform of a density* $f_2$ *included in* $\mathcal{B}(\beta)$ *for small enough* $t > 0$ *and large enough* $M$.

***Proof.*** First, let us prove that the function $f_1$ defined by

$$f_1(x) = \frac{1}{2\pi}\int \mathrm{e}^{-\mathrm{i}ux}\Phi_1(u)\mathrm{d}u = \frac{1}{2}(g_1(x) + g_2(x))$$

where

$$g_1(x) = \frac{1}{2\pi}\int \frac{\mathrm{e}^{-\mathrm{i}ux}}{(1+u^2)^{\beta/2}}\,\mathrm{d}u \quad \text{and} \quad g_2(x) = \frac{1}{\pi(1+x^2)},$$

is a positive and integrable function, and thus is a density, as by Parseval's equality $\int f_1(x)\mathrm{d}x = \Phi_1(0) = 1$. Indeed, $g_2$ is the density of the Cauchy law, and we have

$$g_1(x) = \frac{1}{\pi}\int_0^{+\infty} \frac{\cos(ux)}{(1+u^2)^{\beta/2}}\,\mathrm{d}u.$$

Using formulae 3.771.2, 8.432.3 and 8.334.3 in Gradshteyn and Ryzhik (2000), we obtain that, for any $x \geqslant 0$, $g_1$ is given by

$$g_1(x) = (\Gamma(\beta/2))^{-2}\left(\frac{x}{2}\right)^{\beta-1}\int_1^{+\infty} \mathrm{e}^{-xt}(t^2-1)^{\beta/2-1}\mathrm{d}t,$$

and thus is a positive function on $\mathbb{R}^+$, which is also an even function. Moreover, according to formulae 3.771.2 and 8.451.6 in Gradshteyn and Ryzhik (2000), we have

$$g_1(x) \underset{+\infty}{\sim} Cx^{\beta/2-1}\mathrm{e}^{-x},$$

for some positive constant $C$, and thus an integrable function. This establishes that $f_1$ is a density function on $\mathbb{R}$.

The rest of the proof follows the same lines as the one for Lemma 3. Establishing the positivity of $f_2$, the first step involves a bound on the quantity

$$\left|\frac{1}{2\pi}\int \mathrm{e}^{-\mathrm{i}ux}\Phi_1(u)\left(1 - k^*\left(\frac{u}{M}\right)\right)(1 - \Phi^\varepsilon((2t)^{1/s}u)\mathrm{e}^{-iua_t})\mathrm{d}u\right|$$

$$\leqslant \frac{1}{2\pi}\int_{|u|\geqslant M}(1+u^2)^{-\beta/2}\mathrm{d}u + \frac{1}{2\pi}\int_{|u|\geqslant M}\mathrm{e}^{-|u|/2}\,\mathrm{d}u$$

which is $O(M^{-\beta+1})$ and converges to zero, uniformly in $x$, as $M$ tends to infinity (and for $\beta > 1$). The second step is proved exactly in the same way, as the asymptotic behaviour of $f_1$ is given by the Cauchy density $g_2$ and it is of order $\pi^{-1}x^{-2}$. □

***Proof of Theorem 2 (lower bound).*** Here, the notation is the same as used in Lemma 4, and the proof follows the same lines as the proof of Theorem 1 (lower bound). Indeed, for arbitrary small $\delta > 0$, consider the parameters

$$t = \sqrt{1-\delta}\,\frac{2\beta - |s-1|}{s}\,\frac{\log\log n}{\log n} \quad \text{and} \quad M = \left(\frac{\log n}{2\sigma_0^s}\right)^{1/s}, \tag{30}$$

and the functions $f_1$ and $f_2$ corresponding to this choice, in Lemma 4.

According to this lemma, the functions $f_1$ and $f_2$ belong to $\mathcal{B}(\beta)$ (for large enough $n$). The control of the $\chi^2$-distance between the laws induced by $f_1$ and $f_2$ is established exactly in the same way as in Theorem 1 (lower bound), where now the asymptotic behaviour of the function $f_1$ is $O(x^{-2})$. The first term $T_1$ is controlled by $O(n)\|\Phi_1^Y - \Phi_2^Y\|_2^2$ and

$$|(\Phi_1^Y - \Phi_2^Y)(u)| \leqslant O(1)\frac{1_{|u|>M}}{(1+u^2)^{\beta/2}}\exp\left(-(1-t)\sigma_0^s|u|^s\right).$$

This gives

$$T_1 = O(n)M^{-2\beta+1-s}\mathrm{e}^{-2\sigma_0^s M^s + 2t\sigma_0^s M^s}. \tag{31}$$

On the other hand,

$$T_2 = O(n)\int_{|y|>A}|y|^2|f_1^Y(y) - f_2^Y(y)|^2\mathrm{d}y = O(n)\|(\Phi_1^Y - \Phi_2^Y)'\|_2^2.$$

We write first

$$(\Phi_1^Y - \Phi_2^Y)(u) = \Phi_1(u)(\Phi^\varepsilon(\sigma_1 u) - \Phi^\varepsilon(\sigma_2 u))\left(1 - k^*\left(\frac{u}{M}\right)\right),$$

to see that the function is continuously differentiable on its support $\{|u| > M\}$. Now,

$$\begin{aligned}|(\Phi_1^Y - \Phi_2^Y)'(u)| &\leqslant O(1)|\Phi_1(u)||(\Phi^\varepsilon(\sigma_1 u) - \Phi^\varepsilon(\sigma_2 u))'|\left(1 - k^*\left(\frac{u}{M}\right)\right)\\ &\leqslant O(1)|u|^{-\beta+(s-1)_+}\mathrm{e}^{-(1-t)\sigma_0^s|u|^s}1_{|u|>M},\end{aligned}$$

where $a_+$ denotes the positive part of a real $a$. Then

$$T_2 = O(n)\int_{|u|>M}|u|^{-2\beta+2(s-1)_+}\mathrm{e}^{-2(1-t)\sigma_0^s|u|^s}\mathrm{d}u = O(M^{-2\beta+2(s-1)_+ + 1-s}\mathrm{e}^{-2(1-t)\sigma_0^s M^s}). \tag{32}$$

From (31) and (32) we deduce that

$$n\chi^2(f_1^Y, f_2^Y) \leqslant O(n)M^{-2\beta+|s-1|}\mathrm{e}^{-2(1-t)\sigma_0^s M^s}.$$

Finally, the $\chi^2$-distance goes to 0 when $n \to \infty$, for $M$ and $t$ in (30). Thus

$$\inf_{\hat{\sigma}_n}\sup_{f,\sigma}\psi_n^{-2}\mathbb{E}\{|\hat{\sigma}_n - \sigma|^2\} \geqslant (1-\delta)\left(1 - \frac{|s-1|}{2\beta}\right)^2,$$

for arbitrary small $\delta > 0$, and this concludes the proof. □

***Proof of Theorem 5 (lower bound).*** We use the construction given in the proof of Theorem 2 (lower bound). As in this proof, $n\chi^2(f_1^Y, f_2^Y)$ goes to 0 when $n \to \infty$, for $M$ and $t$ in (30). Then we need to bound the difference $|f_1(0) - f_2(0)|$ from below:

$$\begin{aligned} |f_1(0) - f_2(0)| &= \frac{1}{2\pi}\left|\int \Phi_1(u) k^*\left(\frac{u}{M}\right)(1 - \Phi^\varepsilon((2t)^{1/s}u)\mathrm{e}^{-iua_t})\mathrm{d}u\right| \\ &\geqslant Ct\int_{|u|\leqslant 1} \frac{|u|^s}{(1+u^2)^{\beta/2}}\mathrm{d}u + C\int_{1<|u|\leqslant M} \frac{t\mathrm{d}u}{(1+u^2)^{\beta/2}} \\ &\geqslant Ct \geqslant C\frac{\log\log n}{\log n} \end{aligned}$$

and the integrals are convergent whenever $\beta > \beta' + \frac{1}{2}$. □

***Sketch of proof of Theorem 6.*** In the upper bounds, the $L_2$ bias changes (see, for example, $T_{11}$ in the proof of Theorem 3), since

$$\sup_{\sigma\in\mathcal{V}(\sigma_0)} \sup_{f\in S(\alpha',R,L)} \sup_{\tau\in\mathcal{U}(\sigma)} \frac{1}{2\pi}\int_{|t|>1/(\tau h_n)} |\Phi(t)|^2\mathrm{d}t \leqslant L\exp\left(-\frac{2\alpha}{(\sigma_0+\delta)^R h_n^R}\right),$$

and

$$\sup_{\sigma\in\mathcal{V}(\sigma_0)} \sup_{f\in W(\beta',L)} \sup_{\tau\in\mathcal{U}(\sigma)} \frac{1}{2\pi}\int_{|t|>1/(\tau h_n)} |\Phi(t)|^2\mathrm{d}t \leqslant L(\sigma_0+\delta)^{2\beta'} h_n^{2\beta'},$$

thus giving slightly different bandwidths $h_n$ and rates for the case $\beta' \leqslant \tilde{s}$.

As for the lower bounds, precisely the same constructions provide the lower bounds since the rate is now given by $\|f_1 - f_2\|_2 = (2\pi)^{-1/2}\|\Phi_1 - \Phi_2\|_2$. □

## Acknowledgements

The authors are grateful to A. Tsybakov and to F. Ruymgaart for initiating this work. We would also like to thank an anonymous referee for his/her thoughtful remarks.

## References

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

Butucea, C. and Tsybakov, A. (2004) Sharp optimality for density deconvolution with dominating bias. arXiv: math.ST/0409471.

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

Efromovich, S. (1997) Density estimation for the case of supersmooth measurement error. *J. Amer. Statist. Assoc.*, **92**, 526–535.

Goldenshluger, A. (1999) On pointwise adaptive nonparametric deconvolution, *Bernoulli*, **5**, 907–925.
Goldenshluger, A. (2002) Density deconvolution in the circular structural model. *J. Multivariate Anal.*, **81**, 360–375.
Gradshteyn, I. and Ryzhik, I. (2000) *Table of Integrals, Series, and Products*, 6th edition. San Diego, CA: Academic Press.
Koltchinskii, V.I. (2000) Empirical geometry of multivariate data: a deconvolution approach, *Ann. Statist.*, **28**, 591–629.
Lindsay, B.G. (1986) Exponential family mixture models (with least-squares estimators). *Ann. Statist.*, **14**, 124–137.
Matias, C. (2002) Semiparametric deconvolution with unknown noise variance, *ESAIM Probab. Statist.*, **6**, 271–292.
van der Vaart, A.W. and Wellner, J.A. (1996) *Weak Convergence and Empirical Processes*. New York: Springer-Verlag.
Zhang, C.-H. (1990) Fourier methods for estimating mixing densities and distributions, *Ann. Statist.*, **18**, 806–831.
Zolotarev, V. (1986) *One-Dimensional Stable Distributions*, Transl. Math. Monogr. 65. Providence, RI: American Mathematical Society.

Received February 2003 and revised July 2004