Centre de Recherche en Économie et Statistique
CREST

Série des Documents de Travail

n° 2013-31

Adaptive Density Estimation in Deconvolution Problems with Unknown Error Distribution

J. KAPPUS[1]
G. MABON[2]

December, 2013

[1] Institut für Mathematik, Universität Rostock, Germany. Email : johanna.kappus@uni-rostock.de
[2] CREST and MAP5, Université Paris Descartes (France). Email : gwennaelle.mabon@ensae.fr

Adaptive density estimation in deconvolution problems with unknown error distribution

J. KAPPUS[*] and G. MABON[†,‡]

[*]Institut für Mathematik, Universität Rostock
18051 Rostock, Germany
e-mail: johanna.kappus@uni-rostock.de

[†]CREST
3 avenue Pierre Larousse
92245 Malakoff, France
e-mail: gwennaelle.mabon@ensae.fr

[‡]MAP5, Université Paris Descartes,
45 rue des Saints-Pères
75006 Paris, France

December 9, 2013

Abstract

A density deconvolution problem with unknown distribution of the errors is considered. To make the target density identifiable, one has to assume that some additional information on the noise is available. We consider two different models: the framework where some additional sample of the pure noise is available, as well as the repeated observation model, where the contaminated random variable of interest can be observed repeatedly. We introduce kernel estimators and present upper risk bounds. The focus of this work lies on the optimal data driven choice of the smoothing parameter using a penalization strategy.

Keywords. Adaptive estimation. Deconvolution. Density estimation. Mean square risk. Nonparametric methods. Replicate observations.

1 Introduction

We consider a density deconvolution model. We are given independent copies of X, perturbed by an additional noise:

$$Y_j = X_j + \varepsilon_j, \quad j = 1, \dots, n. \tag{1}$$

It is assumed that X has a square integrable density f with respect to the Lebesgue measure. Our target is to recover f from the data $(Y_j)_{1\leq j\leq n}$.

This setting is classical in nonparametric statistics and there exists a large amount of literature on the subject. Rates of convergence and their optimality have been studied, for example, in Carroll and Hall (1988), Stefanski (1990), Stefanski and Carroll (1990), Fan (1991), Efromovich (1997) and Delaigle and Gijbels (2004) for kernel estimators, Butucea (2004), Butucea and Tsybakov (2008a) and Butucea and Tsybakov (2008b) for studies of rate optimality in the minimax sense. However, those authors have worked under the assumption that the distribution of the noise is known, which is often not realistic in applications.

In the present paper, we assume that the distribution of the noise is unknown. To make the target density identifiable, one has to assume that some additional information on the errors is available. We discuss two different models. The setting which is more frequently studied in the literature is that, in addition to (1), a preliminary sample $\varepsilon_{-M}, \dots, \varepsilon_{-1}$ of the pure noise, independent of the Y_j is available. In applications, this model plays a role if the $(\varepsilon_{-j})_{1\leq j\leq M}$ are measurement errors due to the measuring device. The distribution of the noise can then be recovered, in a first calibration step, from repeated measurements in absence of the signal X. Typical domains of application are optics, astrophysics or spectrometry. Another important case is the model of repeated observations, where a variable X_j can be observed repeatedly, with independent errors. In this case, we have in addition to (1)

$$Y_{n+j,k} = X_{n+j} + \varepsilon_{n+j,k}, \quad j = 1, \dots, 2M \quad \text{and} \quad k = 1, 2. \tag{2}$$

Considering this model can be motivated from applications in medicine. Comte et al. (2013) discuss a repeated measurement model in the estimation of onset of pregnancies. Applications in the field of economics are discussed in Bonhomme and Robin (2010).

Rates of convergence have been presented in Neumann (1997) and, more recently, in Johannes (2009), or Meister (2009) under the assumption that observations of the pure noise are feasible. For rate results in a repeated observations model, see Li and Vuong (1998), Neumann (2007), Delaigle et al. (2008) and Comte et al. (2013).

In the present paper, our main interest lies in the adaptive choice of the smoothing parameter. So far, the adaptive bandwidth selection has mostly been studied in a deconvolution model with known error distribution, see for example, Pensky and Vidakovic (1999) for wavelet strategy, Comte et al. (2006), Butucea and Comte (2009) for projection strategies, or Meister (2009). Delaigle et al. (2008) propose an adaptive bandwidth selector in the repeated observation model, but do not study its theoretical properties. The rigorous study of adaptive procedures in a deconvolution model with unknown errors has only recently been addressed. We are aware of the work by Comte and Lacour (2011), Johannes and Schwarz (2012), Kappus (2014) and Dattner et al. (2013).

Contrarily to the last mentioned authors, who consider adaptive quantile estimation in the spirit of Lepski's method, we propose, in the present paper, a model selection approach. The essential difference lies in the fact that the problem is considered from a non-asymptotic perspective. We refer to Massart (2003) for a systematic comparison of both methods. Moreover, we allow, in the present paper, for a much broader class of error distributions, since we do not assume that the tails of the characteristic function decay polynomially.

In comparison with Comte and Lacour (2011) or Johannes and Schwarz (2012), we do not impose any a priori semi parametric assumptions on the shape of the characteristic function of the noise, such as polynomial or exponential decay or boundedness from above and below by some monotone function. Compared to Comte and Lacour (2011), we also relax the hypothesis that $M \geq n^{1+\eta}$, $\eta > 0$. Our strategy hence allows a fully general treatment under minimal assumptions. The approach presented in this paper compares to the results achieved in Kappus (2014) for an estimation problem for discretely observed Lévy processes.

This paper is organized as follows: in Section 2, we give the notations, specify the statistical model and estimation procedure and present upper risk bounds. In Section 3, we introduce alternative estimators which enables us to propose a new adaptive procedure by penalization in the context of deconvolution under weak assumptions. Besides the theoretical properties of the adaptive estimators (one for each model) are studied. In Section 4, we lead a study of the alternative estimators through simulation experiments. We try out the influence of the sample size of the noise. Numerical results are then presented. The procedure shows good performances even when there is a small number of replications. Section 5 gives final remarks on the proposed adaptive procedure and other foreseeable field of application. All proofs are postponed to Section 6.

2 Statistical model, estimation procedure and risk bounds

In the present section, we fix the statistical model and assumptions, introduce the estimator and recall, for the readers convenience, the non asymptotic risk bounds which have been presented in earlier publications on the subject. But first let us fix some notations.

Notations. For two real numbers a and b, we denote $a \vee b = \max(a, b)$ and $a \wedge b = \min(a, b)$. For two functions $\varphi, \psi : \mathbb{R} \to \mathbb{C}$ belonging to $\mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$, we denote $\|\varphi\|$ the $\mathbb{L}^2$ norm of φ defined by $\|\varphi\|^2 = \int_{\mathbb{R}} |\varphi(x)|^2 \,\mathrm{d}x$, $\langle \varphi, \psi \rangle$ the scalar product between φ and ψ defined by $\langle \varphi, \psi \rangle = \int_{\mathbb{R}} \varphi(x)\overline{\psi(x)} \,\mathrm{d}x$. The Fourier transform φ^* is defined by

$$\varphi^*(x) = \int e^{ixu} \varphi(u) \,\mathrm{d}u.$$

Besides, if φ^* belongs to $\mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$, then the function φ is the inverse Fourier transform of φ^* and can be written $\varphi(x) = 1/(2\pi) \int e^{-ixu} \varphi^*(u) \,\mathrm{d}u$. Lastly the convolution product $*$ is defined as $(\varphi * \psi)(x) = \int \varphi(x - u)\psi(u) \,\mathrm{d}u$.

2.1 Statistical model

Under Model (1) and independence assumptions, it is clear that $f_Y = f * f_\varepsilon$ which implies $f_Y^* = f^* f_\varepsilon^*$. Moreover, we make the following assumptions:

(A1) $\forall x \in \mathbb{R}$, $f_\varepsilon^*(x) \neq 0$ and $f_\varepsilon \in \mathbb{L}^2(\mathbb{R})$.

(A2) ε is symmetric.

Under **(A1)**, we have the equality $f^* = f_Y^*/f_\varepsilon^*$. If f_ε^* is known, we can estimate f^* with $\hat{f}_Y^*/f_\varepsilon^*$ where $\hat{f}_Y^*$ is an estimator obtained directly from the data with a simple plug-in estimator. We should only apply the inverse Fourier transform to get an estimate of f. Nevertheless, it happens that $1/f_\varepsilon^*$ is not integrable so we cannot compute the inverse Fourier transform. It is necessary to apply some regularization, for example, a spectral cutoff estimator. In this particular case, the estimator of f would be $1/(2\pi)\int_{|u|\leq \pi m} e^{-iux} f_Y^*(u)/f_\varepsilon^*(u)\,\mathrm{d}u$. We can notice that this estimator corresponds both to a kernel estimator built with a sinc kernel (Butucea (2004)) or to a projection type estimator as in Comte et al. (2006).

In the present case the error distribution is assumed to be unknown. To make the problem identifiable, some additional information on the noise is required. We consider two different models. The case which is more commonly studied in the literature is when there exists a preliminary sample of the noise $(\varepsilon_{-j})_{1\leq j\leq M}$, independent of the (Y_j). Then we can compute a plug-in estimator of f_ε^*

$$\forall x \in \mathbb{R},\ \hat{f}^*_{\varepsilon,\mathrm{NS}}(x) = \frac{1}{M}\sum_{j=1}^{M} e^{ix\varepsilon_{-j}} \tag{3}$$

and the following estimator of f_Y^*

$$\forall x \in \mathbb{R},\ \hat{f}^*_{Y,\mathrm{NS}}(x) = \frac{1}{n}\sum_{j=1}^{n} e^{ixY_j}. \tag{4}$$

Thereafter, we denote this model: (NS) model where (NS) is set for noise sample.

Secondly the identification of f is also possible in the so called repeated observation model, where the contaminated random variable can be measured repeatedly. In this case, we observe, in addition to the original sample

$$Y_{n+j,k} = X_{n+j} + \varepsilon_{n+j,k}, \quad j = 1, \dots, 2M \quad \text{and} \quad k = 1, 2,$$

where the $(X_{n+j})_{1\leq j\leq 2M}$ are *i.i.d.* copies of X_1 and the $(\varepsilon_{n+j,k})_{1\leq j\leq 2M}$ are *i.i.d.* copies of ε_1. The sequences of $(X_{n+j})_{1\leq j\leq 2M}$ and $(\varepsilon_{n+j,k})_{1\leq j\leq 2M}$ are mutually independent and independent of $(X_j)_{1\leq j\leq n}$ and $(\varepsilon_{n+j,k})_{1\leq j\leq n}$. Under assumption **(A2)**, we have the following equalities

$$\forall x \in \mathbb{R},\ \mathbb{E}\left[e^{ix(Y_{n+j,1}-Y_{n+j,2})}\right] = \mathbb{E}\left[e^{ix(\varepsilon_{n+j,1}-\varepsilon_{n+j,2})}\right] = \left|\mathbb{E}\left[e^{ix\varepsilon_{n+j,1}}\right]\right|^2 = \left(\mathbb{E}\left[e^{ix\varepsilon_{n+j,1}}\right]\right)^2 = \left(f_\varepsilon^*(x)\right)^2.$$

Without assumption **(A2)**, the model is much more complicated and requires a completely different approach since f_ε^* is not a real positive function anymore. We can now define the following estimator of $(f_\varepsilon^*)^2$

$$\forall x \in \mathbb{R},\ \widehat{f_\varepsilon^{*2}}(x) = \left(\frac{1}{M}\sum_{j=1}^{M}\cos\left(x(Y_{n+j,1}-Y_{n+j,2})\right)\right)_+, \tag{5}$$

and the following estimator of f_Y^*

$$\forall x \in \mathbb{R},\ \hat{f}^*_{Y,\mathrm{RD}}(x) = \frac{1}{n+M}\left(\sum_{j=1}^{n} e^{ixY_j} + \sum_{j=M+1}^{2M} e^{ixY_{n+j,2}}\right). \tag{6}$$

We have split $(Y_{n+j,k})_{1\leq j\leq 2M,\,k=1,2}$ into two independent samples of size M. The observations $(Y_j)_{1\leq j\leq n}$ and $(Y_{n+j,2})_{M+1\leq j\leq 2M}$ are used to build an estimator f_Y^*, see formula (6). From the observations $(Y_{n+j,k})_{1\leq j\leq M,\,k=1,2}$, we build an estimator of f_ε^*, see formula (5). Thereafter, we denote this model: (RD) model where (RD) is set for replicate or repeated data. In this model, the estimator of f_Y^* is slightly different from the previous model since we take into account the information brought by the replicate data.

2.2 Estimation procedure and upper bound

Here the noise is unknown and has to be estimated as aforesaid in this section. Nonetheless it is needed to prevent the denominator from becoming too small with a regularization of the characteristic function in the denominator following ideas presented in Neumann (1997), see also Comte and Lacour (2011) or Comte et al. (2013). Delaigle et al. (2008) propose a different regularization procedure with a ridge parameter ρ added to the denominator. However, this requires a careful discussion of the choice of the ridge-parameter, while our regularized version of the empirical characteristic function is completely data driven.

So to prevent the denominator from becoming too small, we regularize the characteristic function in the denominator as in Comte and Lacour (2011) for the (NS) model and Comte et al. (2013) for the (RD) model as follows

$$\frac{1}{\tilde{\varphi}(x)} = \frac{\mathbb{1}\left\{|\hat{f}^*_{\varepsilon,\text{NS}}(x)| \geq M^{-1/2}\right\}}{\hat{f}^*_{\varepsilon,\text{NS}}(x)} \quad (\text{NS}) \quad \text{or} \quad \frac{1}{\tilde{\varphi}(x)} := \frac{\mathbb{1}\left\{|\widehat{f^{*2}_\varepsilon}(x)| \geq M^{-1/2}\right\}}{\sqrt{\widehat{f^{*2}_\varepsilon}(x)}} \quad (\text{RD}). \tag{7}$$

Given a kernel function f and smoothing parameter m, the corresponding spectral cutoff estimator is

$$\hat{f}_{m,M,n}(x) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} e^{-iux} \frac{\hat{f}^*_Y(u)}{\tilde{\varphi}(u)} \,\mathrm{d}u, \tag{8}$$

where $\hat{f}^*_Y$ is an estimator of f^*_Y either in the (NS) model as defined by (4) or in the (RD) model as defined by (6). Thus the following bounds can be given on the mean squared error:

Proposition 2.1.
• *In presence of a noise sample, under **(A1)**, for $\hat{f}_{m,M,n}$ defined by (8) there exists an universal positive constant C such as*

$$\mathbb{E}\left\|f - \hat{f}_{m,M,n}\right\|^2 \leq \|f - f_m\|^2 + C\left(\frac{1}{n}\int_{-\pi m}^{\pi m} |f^*_\varepsilon(u)|^{-2}\,\mathrm{d}u + \frac{1}{M}\int_{-\pi m}^{\pi m} \frac{|f^*(u)|^2}{|f^*_\varepsilon(u)|^2}\,\mathrm{d}u\right). \tag{9}$$

• *In the repeated measurement model, under **(A1)** and **(A2)**, for $\hat{f}_{m,M,n}$ defined by (8) there exists an universal positive constant C' such as*

$$\mathbb{E}\left\|f - \hat{f}_{m,M,n}\right\|^2 \leq \|f - f_m\|^2 + C'\left(\frac{1}{n+M}\int_{-\pi m}^{\pi m} |f^*_\varepsilon(u)|^{-2}\,\mathrm{d}u + \frac{1}{M}\int_{-\pi m}^{\pi m} \frac{|f^*(u)|^2}{|f^*_\varepsilon(u)|^2(|f^*_\varepsilon(u)|^2 \vee M^{-1/2})}\,\mathrm{d}u\right). \tag{10}$$

The first two terms of right-hand side of Equations (9) and (10) correspond to the usual terms when the distribution of the errors is known (see Comte et al. (2006)): a squared bias term and a bound on a variance. The last term is due to the estimation of f^*_ε which here depends on the considered model. For the proofs we refer to Comte and Lacour (2011) and Comte et al. (2013).

For the rates of convergence, we refer to Lacour (2006) for a complete study of the known-error case, Comte and Lacour (2011) for the (NS) model. For the (RD) model, we refer to Delaigle et al. (2008) who show that in statistical deconvolution problems there is no first-order loss of performance when estimating the error density with repeated data when compared to the known-error case under the assumption that "f is smoother than half a derivative of f_ε". Finally, in that model, we also refer to Comte et al. (2013) who complete the theoretical study of Delaigle et al. (2008) by studying this time the integrated risk.

The idea now for the adaptive estimation is to find a penalty term which would have the same order as the bound on a variance. The following section will show that we can obtain an adaptive procedure for the (NS) and (RD) model under weak assumptions.

3 Adaptive estimation procedure

For the adaptive estimation procedure, we follow the steps of Kappus (2014) by introducing alternative estimators (one for each case aforementioned) of f^*_ε which can be controlled uniformly on the real line which brings a model selection procedure with very weak assumptions.

For $\delta > 0$, let us introduce the weight function w defined as

$$\forall x \in \mathbb{R},\, w(x) = (\log(e + |x|))^{-\frac{1}{2}-\delta}$$

which has originally been proposed in Neumann and Reiß (2009). Their results play an important role for the arguments developed in Kappus (2014). Introducing the logarithmic factor allows to apply concentration inequalities of Talagrand type. For a positive constant κ to be specified, we set the following threshold

$$k_M(x) = \kappa\,(\log M)^{1/2}\, w(x)^{-1} M^{-1/2} \tag{11}$$

which will prevent the estimator of characteristic function f_ε^* from being too small. We use exactly here the same idea as Neumann (1997). This being said, we can introduce the estimators considered in this work: in the (NS) case, we define an alternative estimator of f_ε^*

$$\tilde{f}^*_{\varepsilon,\mathrm{NS}}(x) = \begin{cases} \hat{f}^*_{\varepsilon,\mathrm{NS}}(x) & \text{if} \quad |\hat{f}^*_{\varepsilon,\mathrm{NS}}(x)| \geq k_M(x), \\ k_M(x) & \text{otherwise.} \end{cases} \tag{12}$$

Similarly, in the (RD) case, we propose an alternative estimator of f_ε^*

$$\tilde{f}^*_{\varepsilon,\mathrm{RD}}(x) = \begin{cases} \hat{f}^*_{\varepsilon,\mathrm{RD}}(x) = \sqrt{\widehat{f^{*2}_\varepsilon}(x)} & \text{if} \quad \widehat{(f^*_\varepsilon)^2}(x) \geq k_M(x), \\ \sqrt{k_M(x)} & \text{otherwise.} \end{cases} \tag{13}$$

So using Fourier transform, we can estimate f_m respectively for the (NS) and (RD) model as follows

$$\hat{f}_{m,\mathrm{NS}}(x) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} e^{-ixu} \frac{\hat{f}^*_{Y,\mathrm{NS}}(u)}{\tilde{f}^*_{\varepsilon,\mathrm{NS}}(u)}\, \mathrm{d}u, \tag{14}$$

$$\hat{f}_{m,\mathrm{RD}}(x) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} e^{-ixu} \frac{\hat{f}^*_{Y,\mathrm{RD}}(u)}{\tilde{f}^*_{\varepsilon,\mathrm{RD}}(u)}\, \mathrm{d}u. \tag{15}$$

Given a finite collection $\mathcal{M}_n = \{1, \ldots, n\}$ of cutoff parameters our target is to realize the optimal data driven choice of the smoothing parameter $\hat{m}$. Ideally, $\hat{m}$ should outbalance the bias and variance term displayed in Equations (9) and (10). In a deconvolution problem with perfectly known error distribution, this would mean that $\hat{m}$ should mimic the oracle choice

$$m^* = \underset{m \in \mathcal{M}_n}{\operatorname{argmin}} \left\{ -\|f_m\|^2 + \frac{1}{n}\frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{(1 - |f_Y^*(u)|^2)}{|f_\varepsilon^*(u)|^2}\, \mathrm{d}u \right\}.$$

In the present framework, considerations are even more involved since the characteristic function in the denominator is unknown and the variance is hence not feasible to actually compute. Following the model selection paradigm, see Birgé (1999), Birgé and Massart (1997) or Massart (2003), we select $\hat{m}$ as the minimizer of a penalized criterion,

$$\hat{m} = \underset{m \in \mathcal{M}_n}{\operatorname{argmin}} \left\{ -\|\hat{f}_m\|^2 + \widehat{\mathrm{pen}}(m) \right\}.$$

The penalty term should be chosen large enough to annihilate the fluctuation of $\hat{f}_m$ around its target, but on the other hand, should ideally not be much larger than the variance. In a model selection problem with known variance, the penalty term is deterministic, which is no longer the case in the present situation. For the (NS) model, Comte and Lacour (2011) propose an adaptive procedure with the following random penalty

$$\widehat{\mathrm{pen}}(m) = 128 \left(\frac{\log\left(\int_{-\pi m}^{\pi m} |\tilde{\varphi}(u)|^{-2}\, \mathrm{d}u \right)}{\log(1+m)} \right)^2 \frac{\int_{-\pi m}^{\pi m} |\tilde{\varphi}(u)|^{-2}\, \mathrm{d}u}{n} \tag{16}$$

but their procedure is only valid when the sample size of the noise ε is greater than that of the observations. We will present an adaptive procedure released of that assumption. Moreover it does not take into account the second term on the bound of the variance $\int_{-\pi m}^{\pi m} \frac{|f^*(u)|^2}{|f_\varepsilon^*(u)|^2} \mathrm{d}u$ since it cannot be computed directly. For the (RD) model, Comte et al. (2013) investigate the theoretical deterministic penalty

$$\mathrm{pen}(m) = K_0 \left(\frac{\log\left(\int_{-\pi m}^{\pi m} |f_\varepsilon^*(u)|^{-2} \, \mathrm{d}u \right)}{\log(1+m)} \right)^2 \frac{\int_{-\pi m}^{\pi m} |f_\varepsilon^*(u)|^{-2} \, \mathrm{d}u}{n+M}$$

but results for the true stochastic penalty are not presented in that paper.

Now let us define the deterministic and stochastic terms respectively for the (NS) and (RD) model appearing in the bounds on the variance with the alternate estimators introduced in the current section. We first define the classical bound appearing in the known-error case. The only difference is the introduction of the function w essential for the adaptive procedure we propose.

$$\Delta(m) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w(u)^{-2}}{|f_\varepsilon^*(u)|^2} \, \mathrm{d}u, \hat{\Delta}_{\mathrm{NS}}(m) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w(u)^{-2}}{|\tilde{f}_{\varepsilon,\mathrm{NS}}^*(u)|^2} \, \mathrm{d}u, \hat{\Delta}_{\mathrm{RD}}(m) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w(u)^{-2}}{|\tilde{f}_{\varepsilon,\mathrm{RD}}^*(u)|^2} \, \mathrm{d}u.$$

We now define terms due to the estimation of f_ε^* which differ according to the model

$$\Delta_{\mathrm{NS}}^f(m) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w(u)^{-2} |f^*(u)|^2}{|f_\varepsilon^*(u)|^2} \, \mathrm{d}u \quad \text{and} \quad \Delta_{\mathrm{RD}}^f(m) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w(u)^{-2} |f^*(u)|^2}{|f_\varepsilon^*(u)|^2 (|f_\varepsilon^*(u)|^2 \vee k_M(u))} \, \mathrm{d}u,$$
$$\hat{\Delta}_{\mathrm{NS}}^f(m) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w(u)^{-2} |\hat{f}_{Y,\mathrm{NS}}^*(u)|^2}{|\tilde{f}_{\varepsilon,\mathrm{NS}}^*(u)|^4} \, \mathrm{d}u \quad \text{and} \quad \hat{\Delta}_{\mathrm{RD}}^f(m) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w(u)^{-2} |\hat{f}_{Y,\mathrm{RD}}^*(u)|^2}{|\tilde{f}_{\varepsilon,\mathrm{RD}}^*(u)|^6} \, \mathrm{d}u.$$

Thus, for both models, we define empirical penalties as

$$\begin{aligned} \widehat{\mathrm{pen}}_{\mathrm{NS}}(m) &= \widehat{\mathrm{pen}}_{1,\mathrm{NS}}(m) + \widehat{\mathrm{pen}}_{2,\mathrm{NS}}(m) \\ &= 12\lambda_1^2(m, \hat{\Delta}_{\mathrm{NS}}(m)) \frac{\hat{\Delta}_{\mathrm{NS}}(m)}{n} + 16\kappa^2 \log(Mm) \frac{\hat{\Delta}_{\mathrm{NS}}^f(m)}{M} \end{aligned} \tag{17}$$

$$\begin{aligned} \widehat{\mathrm{pen}}_{\mathrm{RD}}(m) &= \widehat{\mathrm{pen}}_{1,\mathrm{RD}}(m) + \widehat{\mathrm{pen}}_{2,\mathrm{RD}}(m) \\ &= 12\lambda_2^2(m, \hat{\Delta}_{\mathrm{RD}}(m)) \frac{\hat{\Delta}_{\mathrm{RD}}(m)}{n+M} + 16\kappa^2 \log(Mm) \frac{\hat{\Delta}_{\mathrm{RD}}^f(m)}{M} \end{aligned} \tag{18}$$

where κ is defined in Equation (11) and deterministic penalties

$$\begin{aligned} \mathrm{pen}_{\mathrm{NS}}(m) &= \mathrm{pen}_{1,\mathrm{NS}}(m) + \mathrm{pen}_{2,\mathrm{NS}}(m) \\ &= 12\lambda_1^2(m, \Delta_{\mathrm{NS}}(m)) \frac{\Delta(m)}{n} + 16\kappa^2 \log(Mm) \frac{\Delta_{\mathrm{NS}}^f(m)}{M}, \end{aligned} \tag{19}$$

$$\begin{aligned} \mathrm{pen}_{\mathrm{RD}}(m) &= \mathrm{pen}_{1,\mathrm{RD}}(m) + \mathrm{pen}_{2,\mathrm{RD}}(m) \\ &= 12\lambda_2^2(m, \Delta_{\mathrm{RD}}(m)) \frac{\Delta(m)}{n+M} + 16\kappa^2 \log(Mm) \frac{\Delta_{\mathrm{RD}}^f(m)}{M} \end{aligned} \tag{20}$$

with

$$\lambda_1(m, D) = \max\left\{ \sqrt{8 \log\left(1 + Dm^2\right)}, \frac{16\sqrt{2}}{3\sqrt{n}} \log\left(1 + Dm^2\right) \right\},$$
$$\lambda_2(m, D) = \max\left\{ \sqrt{8 \log\left(1 + Dm^2\right)}, \frac{16\sqrt{2}}{3\sqrt{n+M}} \log\left(1 + Dm^2\right) \right\}.$$

We can notice that both penalties are split between a part coming from the estimation of f_Y^* and another from the estimation of f_ε^* which is consistent with the upper bound given in Proposition 2.1.

We can also see that for both models in the terms $\hat{\Delta}(m)$ and $\hat{\Delta}^f(m)$ we do not lose track of the inverse Fourier transform of the target density f by estimating it by $\hat{f}_Y^*/\tilde{f}_\varepsilon^*$ which is a difficulty as showed in Section 6. Whereas in the penalty (16) of Comte and Lacour (2011) only the equivalent of $\hat{\Delta}(m)$ is kept. The term $\hat{\Delta}^f(m)$ is in fact included in $\hat{\Delta}(m)$ since in their model they suppose that $M \geq n^{1+\eta}$, $\eta > 0$ which allows them them to bound $\hat{\Delta}^f(m)/M$ by $\hat{\Delta}(m)/n$. Besides their considered model collection only includes the $m's$ such as the stochastic term $\hat{\Delta}(m)/n$ do not blow up. As a consequence, their collection of models is random. This is a drawback we can get rid of with our method. In this paper we consider a deterministic model collection.

Our procedure enables us to overcome the difficulty of keeping $\hat{\Delta}^f(m)$ in the penalty and we also do not need to assume a particular model collection.

Then, we select the cutoff parameters $\hat{m}_{\text{NS}}$ and $\hat{m}_{\text{RD}}$ as minimizers of the following penalized criteria

$$\hat{m}_{\text{NS}} = \underset{m \in \mathcal{M}_n}{\operatorname{argmin}} \left\{ -\left\| \hat{f}_{m,\text{NS}} \right\|^2 + \widehat{\text{pen}}_{\text{NS}}(m) \right\}, \tag{21}$$

$$\hat{m}_{\text{RD}} = \underset{m \in \mathcal{M}_n}{\operatorname{argmin}} \left\{ -\left\| \hat{f}_{m,\text{RD}} \right\|^2 + \widehat{\text{pen}}_{\text{RD}}(m) \right\}. \tag{22}$$

Finally we can derive, for the corresponding estimator $\hat{f}_{\hat{m}}$ the following non-asymptotic risk bound for a collection model $\mathcal{M}_n = \{1, \dots, n\}$ and hence the main result of this section:

Theorem 3.1.
• *In the* (NS) *model, consider under* ***(A1)****,* $\hat{f}_{\hat{m}_{\text{NS}},\text{NS}}$ *defined by (14) and (21). Then there are positive constants* C^{ad} *and* C *such that*

$$\mathbb{E}\|f - \hat{f}_{\hat{m}_{\text{NS}},\text{NS}}\|^2 \leq C^{ad} \inf_{m \in \mathcal{M}_n} \left\{ \|f - f_m\|^2 + \text{pen}_{\text{NS}}(m) \right\} + C \left(\frac{1}{n} + \frac{1}{M} \right). \tag{23}$$

• *In the* (RD) *model, consider under* ***(A1)*** *and* ***(A2)****,* $\hat{f}_{\hat{m}_{\text{RD}},\text{RD}}$ *defined by (15) and (22). Then there are positive constants* C^{ad} *and* C *such that*

$$\mathbb{E}\|f - \hat{f}_{\hat{m}_{\text{RD}},\text{RD}}\|^2 \leq C^{ad} \inf_{m \in \mathcal{M}_n} \left\{ \|f - f_m\|^2 + \text{pen}_{\text{RD}}(m) \right\} + C \left(\frac{1}{n+M} + \frac{1}{M} \right) \tag{24}$$

The rates of convergence in deconvolution problems are classically intricate and depend on the regularity types of the function f under estimation and the noise density f_ε^*. To compute them, one has to study the orders of the bias and variance terms in inequalities (9) or (10) and to look for an optimal value of m in function of n. Even in the known-error case, to show the optimality of a kernel-type estimator Butucea and Tsybakov (2008a) work with an implicit equation satisfied by the optimal bandwidth. But once this is done, it is clear that the choices of an optimal m depend on unknown quantities and thus cannot be implemented. That is why inequalities (23) or (24) are of high interest: the compromise is automatically made and completely data driven in an almost non-asymptotic setting. Rates of convergence are reached by themselves without being specified in the framework. Yet we may wonder if there is a cost due to adaptation here. To answer that question, right-hand sides of Equations (9) and (23) have to be compared. More precisely, since the bias term $\|f - f_m\|^2$ is unchanged and $n^{-1} + M^{-1}$ are negligible terms, it comes down to compare $\text{pen}_{\text{NS}}(m)$ with

$$\frac{\Delta(m)}{n} + \frac{\log M}{M} \Delta_{\text{NS}}^f(m).$$

Clearly the difference lies in the logarithmic terms, and thus, the loss is negligible. Moreover we cannot emphasize enough that our procedure is very general and the assumptions very weak. Obviously, the same kind of remarks holds for Equation (24) compared to Equation (10) for the (RD) model. In this case, we must also point out that no other adaptive procedure exists for this problem. This is another quality of our general methodology.

4 Simulation study

The whole implementation is conducted using R software. The integrated squared error $\|f - \hat{f}_{\hat{m}}\|^2$ (generally speaking) is computed via a standard approximation and discretization (over 300 points) of the integral on an interval of $\mathbb{R}$ denoted by I. Then the mean integrated squared error (MISE) $\mathbb{E}\|f - \hat{f}_{\hat{m}}\|^2$ is computed as the empirical mean of the approximated ISE over 100 simulation samples.

Practical estimation procedure. The adaptive procedure is implemented as follows for both cases with their respective estimators :

- ▷ For $m \in \mathcal{M}_n = \{m_1, \dots, m_n\}$, compute $-\|\hat{f}_m\|^2 + \widehat{\text{pen}}(m)$.
- ▷ Choose $\hat{m}$ such as $\hat{m} = \underset{m\in\mathcal{M}_n}{\text{argmin}} \left\{-\|\hat{f}_m\|^2 + \widehat{\text{pen}}(m)\right\}$.
- ▷ And compute $\hat{f}_{\hat{m}}(x) = \displaystyle\int_{-\pi\hat{m}}^{\pi\hat{m}} e^{-ixu} \frac{\hat{f}_Y^*(u)}{\tilde{f}_\varepsilon^*(u)} du$.

We remind the reader once again that Riemann's sums are used to approximate all the integrals. The penalties are chosen according to Theorem 3.1 and as in Comte et al. (2007) we consider that m can be fractional by takingthe following collection model

$$\mathcal{M}_n = \left\{m = \frac{k}{10}, \quad 1 \le k \le 25\right\},$$

$$\widehat{\text{pen}}_{\text{NS}}(m) = \kappa_{1,\text{NS}} \frac{\log\left(1 + \hat{\Delta}_{\text{NS}}(m)m^2\right)\hat{\Delta}_{\text{NS}}(m)}{n} + \kappa_{2,\text{NS}} \frac{\log(Mm)\hat{\Delta}^f_{\text{NS}}(m)}{M},$$

$$\widehat{\text{pen}}_{\text{RD}}(m) = \kappa_{1,\text{RD}} \frac{\log\left(1 + \hat{\Delta}_{\text{RD}}(m)m^2\right)\hat{\Delta}_{\text{RD}}(m)}{n+M} + \kappa_{2,\text{RD}} \frac{\log(Mm)\hat{\Delta}^f_{\text{RD}}(m)}{M}$$

Influence of M on the estimation. We compute different estimators of the signal for different values of M and consider different signal densities and two noises. Following Comte et al. (2006), we study the following densities on the interval I:

- ▷ Standard Gaussian distribution, $I = [-4, 4]$.
- ▷ Laplace distribution, $f(x) = e^{-\sqrt{2}|x|}/\sqrt{2}$, $I = [-5, 5]$.
- ▷ Mixed Gamma distribution : $X = W/\sqrt{5.48}$, with $W \sim 0.4\Gamma(5, 1) + 0.6\Gamma(13, 1)$, $I = [-1.5, 26]$.
- ▷ Cauchy distribution : $f(x) = \left(\pi(1 + x^2)\right)^{-1}$, $I = [-10, 10]$.

We consider the two following noise densities with same variance 1/10. The first one is a *supersmooth* density which means that its Fourier transform has an exponential decay. The second one is an *ordinary smooth* density which means that its Fourier transform has a polynomial decay.

Gaussian noise : $f_\varepsilon(x) = \dfrac{1}{\sigma\sqrt{2\pi}} \exp\left(-\dfrac{x^2}{2\sigma^2}\right)$, $f_\varepsilon^*(x) = \exp\left(-\dfrac{x^2}{2\sigma^2}\right)$, with $\sigma = 1/\sqrt{10}$.
The gaussian noise is a *supersmooth* noise with $\gamma = 0$, $\delta = 2$ and $\mu = \sigma^2/2$.

Laplace noise : $f_\varepsilon(x) = \dfrac{1}{2\sigma} \exp\left(-\dfrac{|x|}{\sigma}\right)$, $f_\varepsilon^*(x) = \dfrac{1}{1 + \sigma^2 x^2}$, with $\sigma = 1/2\sqrt{5}$.
The Laplace noise is an *ordinary smooth* noise with $\delta = \mu = 0$, $\gamma = 2$.

All the densities are normalized with unit variance except the Cauchy density. We want to study the influence of the relationship of n and M on the estimation of f_ε^* in the (NS) and (RD) cases. For both cases, we consider different values of n and values of $M = \sqrt{n}$ and $M = n$.

It is worth mentioning that the calibration of the four constants were done with intensive simulations. We have $\kappa_{1,\text{NS}} = \kappa_{1,\text{RD}} = 1$ and $\kappa_{2,\text{NS}} = \kappa_{2,\text{RD}} = 1.5$.

Results. The results of the simulations are given in Tables 1 and 2. For both tables, the values of the MISE are multiplied by 100 for each case and computed from 100 simulated data sets. We also give the medians. Table 1 illustrates the case where we can recover a preliminary sample of the noise ε named in this paper (NS) model. The results are very close to those of Comte and Lacour (2011) except when the signal is a Laplace one. Nevertheless our procedure is equivalent or better for other cases. The main improvement

		$n=100$		$n=250$		$n=500$	
		Lap.	Gaus.	Lap.	Gaus.	Lap.	Gaus.
Laplace	f_ε known	2.508	2.480	1.634	1.700	1.204	1.279
		(2.257)	(2.348)	(1.484)	(1.584)	(1.156)	(1.217)
	$M=\lfloor\sqrt{n}\rfloor$	4.182	4.229	2.774	2.795	1.842	1.927
		(4.582)	(4.731)	(2.716)	(2.685)	(1.791)	(1.749)
	$M=n$	2.510	2.508	1.631	1.690	1.205	1.277
		(2.257)	(2.352)	(1.489)	(1.567)	(1.156)	(1.217)
Mixed Gamma	f_ε known	0.842	0.967	0.419	0.417	0.243	0.244
		(0.699)	(0.809)	(0.375)	(0.377)	(0.243)	(0.221)
	$M=\lfloor\sqrt{n}\rfloor$	1.038	1.004	0.612	0.631	0.289	0.306
		(1.085)	(1.074)	(0.448)	(0.533)	(0.268)	(0.265)
	$M=n$	0.842	0.956	0.419	0.417	0.243	0.244
		(0.699)	(0.809)	(0.375)	(0.377)	(0.223)	(0.221)
Cauchy	f_ε known	0.970	1.060	0.420	0.480	0.261	0.267
		(0.887)	(0.903)	(0.373)	(0.371)	(0.227)	(0.228)
	$M=\lfloor\sqrt{n}\rfloor$	1.169	1.105	0.715	0.696	0.341	0.359
		(1.262)	(1.119)	(0.468)	(0.464)	(0.225)	(0.236)
	$M=n$	0.971	1.053	0.420	0.481	0.262	0.273
		(0.887)	(0.880)	(0.373)	(0.371)	(0.227)	(0.228)
Gaussian	f_ε known	0.613	0.606	0.297	0.272	0.178	0.192
		(0.398)	(0.395)	(0.241)	(0.223)	(0.147)	(0.157)
	$M=\lfloor\sqrt{n}\rfloor$	1.140	1.128	0.665	0.646	0.302	0.305
		(0.962)	(0.961)	(0.530)	(0.471)	(0.287)	(0.277)
	$M=n$	0.610	0.592	0.301	0.274	0.178	0.189
		(0.398)	(0.373)	(0.241)	(0.214)	(0.147)	(0.153)

Table 1: Results of simulation as MISE $\mathbb{E}\left(\|f-\hat{f}_{\hat{m}_{\mathrm{NS},\mathrm{NS}}}\|^2\right)\times 100$ averaged over 100 samples. In brackets we give the median of the MISE also averaged over 100 samples.

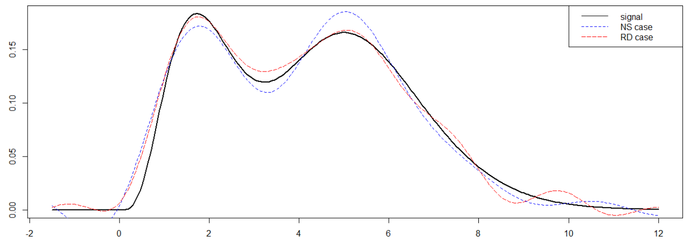
of this procedure is the fact that we reach the same performances as Comte and Lacour (2011) for $M=n$ when they need to take $M=n^2$. That is why we did not consider that case. Besides the risk decreases more rapidly when M and n increase. In many cases estimating the Fourier transform of the noise f_ε^* reduces the risk compared to knowing the density of the noise, a fact already pointed in Comte and Lacour (2011). This can be explained by the fact that an additional regularization of the characteristic function of the noise comes in. Concerning the medians of the MISE, they are always lower than the means of the MISE.

In Table 2 when f_ε^* is known, it is noticeable that the estimation of the signal density is better than in Table 1. This can be explained by the improvement of the estimation f_Y using one of the two samples of replicate data. This fact is well illustrated when the signal is a gaussian. The values of the MISE do not decrease as fast as in the (NS) case with n. But they decrease faster with M. They are pretty much split by two except for the Laplace noise. When $M=n$ the estimation is better in the (RD) case. Nevertheless when $M=\sqrt{n}$, the procedure of the (NS) case is more effective.

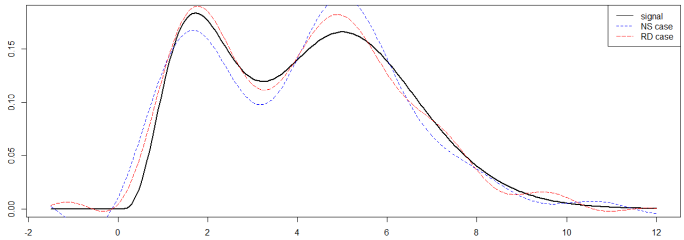
Figures 1 and 2 illustrate the estimation of a mixed Gamma using both penalized estimators for $n=M=250$ and $n=M=500$. The estimation is made with additional Gaussian and Laplace noises. The bimodal specificity of the density is well described. Moreover the precision increases with the sample size. Also the Laplace noise is closer to the target distribution than with the Gaussian noise. Indeed the degree of ill-posedness depends on the decay of f_ε^*. The faster f_ε^* decays, the more complicated it is to recover f_Y from the data. As aforesaid, the Fourier transform of a Laplace density has a polynomial decay whereas a Gaussian density has an exponential decay. So Figures 1 and 2 describe that phenomenon.

		$n = 100$		$n = 250$		$n = 500$	
		Lap.	Gaus.	Lap.	Gaus.	Lap.	Gaus.
Laplace	f_ε known	2.229	2.268	1.454	1.515	1.064	1.156
		(2.105)	(2.178)	(1.431)	(1.463)	(1.100)	(1.128)
	$M = \lfloor\sqrt{n}\rfloor$	3.994	3.927	2.978	3.024	2.258	2.307
		(3.894)	(4.143)	(2.951)	(2.926)	(2.137)	(2.195)
	$M = n$	2.520	2.615	1.777	1.845	1.354	1.481
		(2.494)	(2.711)	(1.745)	(1.825)	(1.366)	(1.500)
Mixed Gamma	f_ε known	0.508	0.551	0.265	0.274	0.159	0.165
		(0.475)	(0.523)	(0.249)	(0.245)	0.158	0.154
	$M = \lfloor\sqrt{n}\rfloor$	1.067	1.076	0.906	0.905	0.701	0.692
		(1.065)	(1.065)	(0.931)	(0.918)	(0.892)	(0.889)
	$M = n$	0.459	0.471	0.263	0.268	0.211	0.212
		(0.417)	(0.404)	(0.237)	(0.244)	(0.208)	(0.206)
Cauchy	f_ε known	0.564	0.584	0.369	0.341	0.190	0.203
		(0.517)	(0.525)	(0.360)	(0.302)	(0.185)	(0.207)
	$M = \lfloor\sqrt{n}\rfloor$	1.079	1.035	0.746	0.738	0.579	0.615
		(0.947)	(0.964)	(0.736)	(0.728)	(0.517)	(0.544)
	$M = n$	0.576	0.569	0.352	0.364	0.281	0.294
		(0.527)	(0.517)	(0.311)	(0.320)	(0.261)	(0.290)
Gaussian	f_ε known	0.349	0.360	0.164	0.147	0.079	0.084
		(0.252)	(0.270)	(0.118)	(0.106)	(0.072)	(0.067)
	$M = \lfloor\sqrt{n}\rfloor$	1.067	1.100	0.709	0.707	0.415	0.403
		(0.969)	(0.960)	(0.749)	(0.741)	(0.290)	(0.269)
	$M = n$	0.537	0.537	0.264	0.255	0.143	0.149
		(0.494)	(0.461)	(0.222)	(0.225)	(0.139)	(0.141)

Table 2: Results of simulation as MISE $\mathbb{E}\left(\|f - \hat{f}_{\hat{m}_{\mathrm{RD,RD}}}\|^2\right) \times 100$ averaged over 100 samples. In brackets we give the median of the MISE also averaged over 100 samples.

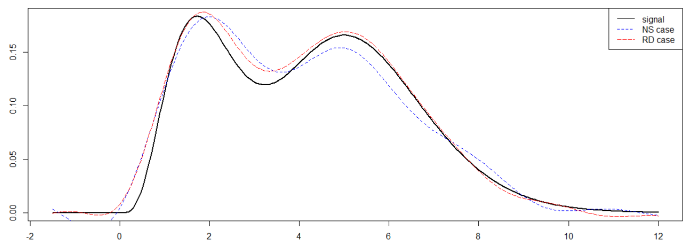


(a) Laplace noise

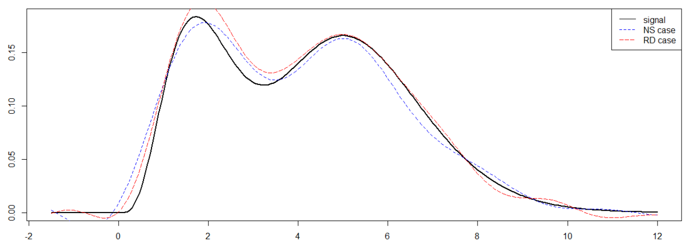


(b) Gaussian noise

Figure 1: Estimations for $n = M = 250$ of a mixed Gamma (bold black line), in blue dashed line for the NS estimator and in red longdashed for the RD estimator.



(a) Laplace noise



(b) Gaussian noise

Figure 2: Estimations for $n = M = 500$ of a mixed Gamma (bold black line), in blue dashed line for the NS estimator and in red longdashed for the RD estimator.

5 Concluding remarks

This paper deals with adaptive deconvolution estimation of a density when the noise density is unknown. We have considered two cases: one where a preliminary sample of the noise can be observed and another one where the variable of interest X can be observed repeatedly with independent errors. For both models, we have proposed a theoretical adaptive procedure which automatically makes a data driven bias-variance compromise. Moreover, it allows us to not specify rates of convergence since they are mechanically reached. This procedure enables us to treat the problem of adaptive estimation in repeated observation model which is completely new. Its advantage is to be very general under weak assumptions. Indeed, that procedure takes into account cases where there can be small number of replications which matches realistic applications as in medicine or economics. Besides of its theoretical properties, our procedure has showed good performances in simulation.

At last, we think that our procedure can be extended to the density estimation of a random effect in linear mixed-effects model. Indeed we are aware of the work of Comte and Samson (2012) who proposed an adaptive procedure based on deconvolution methods in the unknown-error case which is not optimal and Dion (2013) who used a Lepski's method in the known-error case. In that model, the noise can also be recovered by successive difference similarly to the repeated model but the characteristic function of the noise would be raised to a greater power. We may then propose in the same spirit an adaptive procedure for the random effect in linear mixed-effects model.

6 Proofs

In this section, we mainly present simultaneously the proofs for the (NS) and (RD) since the arguments are completely analogous.

6.1 Proof of Theorem 3.1

Before proving any result, let us introduce some notations: for $k > m$,

$$\begin{aligned}
\hat{\Delta}(m,k) &= \hat{\Delta}(k) - \hat{\Delta}(m),\\
\hat{\Delta}^f_{\text{NS}}(m,k) &= \hat{\Delta}^f_{\text{NS}}(k) - \hat{\Delta}^f_{\text{NS}}(m),\\
\hat{\Delta}^f_{\text{RD}}(m,k) &= \hat{\Delta}^f_{\text{RD}}(k) - \hat{\Delta}^f_{\text{RD}}(m).
\end{aligned}$$

Moreover,

$$\begin{aligned}
\widehat{\text{pen}}_{\text{NS}}(m,k) &:= 12\hat{\lambda}_1^2(m,k)\frac{\hat{\Delta}_{\text{NS}}(m,k)}{n} + 16\kappa^2\log(M(k-m))\frac{\hat{\Delta}^f_{\text{NS}}(m,k)}{M},\\
\widehat{\text{pen}}_{\text{RD}}(m,k) &:= 12\hat{\lambda}_2^2(m,k)\frac{\hat{\Delta}_{\text{RD}}(m,k)}{M+n} + 16\kappa^2\log(M(k-m))\frac{\hat{\Delta}^f_{\text{RD}}(m,k)}{M}
\end{aligned}$$

with

$$\begin{aligned}
\hat{\lambda}_1(m,k) &= \max\left\{\sqrt{8\log\left(1+\hat{\Delta}_{\text{NS}}(m,k)(k-m)^2\right)}, \frac{16\sqrt{2}}{3\sqrt{n}}\log\left(1+\hat{\Delta}_{\text{NS}}(m,k)(k-m)^2\right)\right\}\\
\hat{\lambda}_2(m,k) &= \max\left\{\sqrt{8\log\left(1+\hat{\Delta}_{\text{RD}}(m,k)(k-m)^2\right)}, \frac{16\sqrt{2}}{3\sqrt{n+M}}\log\left(1+\hat{\Delta}_{\text{RD}}(m,k)(k-m)^2\right)\right\}.
\end{aligned}$$

Now we can start the proof of Theorem 3.1. The proof is here formulated simultaneously for both models. Thus thereafter for the sake of clarity all subscripts relative to one particular model are dropped.

We denote by $m^\star$ the oracle cutoff defined by

$$m^\star = \operatorname*{argmin}_{m\in\mathcal{M}_n}\left\{-\|f_m\|^2 + \text{pen}(m)\right\}.$$

We have

$$\left\|f - \hat{f}_{\hat{m}}\right\|^2 \le 2\left\|f - \hat{f}_{m^\star}\right\|^2 + 2\left\|\hat{f}_{m^\star} - \hat{f}_{\hat{m}}\right\|^2$$

• Let us notice on the set $G = \{\hat{m} \le m^\star\}$:

$$\left\|\hat{f}_{m^\star} - \hat{f}_{\hat{m}}\right\|^2 \mathbb{1}_G = \left(\left\|\hat{f}_{m^\star}\right\|^2 - \left\|\hat{f}_{\hat{m}}\right\|^2\right)\mathbb{1}_G.$$

Besides according to the definition of $\hat{m}$, one has the following inequality:

$$-\left\|\hat{f}_{\hat{m}}\right\|^2 + \widehat{\text{pen}}(\hat{m}) \leq -\left\|\hat{f}_{m^\star}\right\|^2 + \widehat{\text{pen}}(m^\star), \tag{25}$$

which implies

$$-\left\|\hat{f}_{\hat{m}}\right\|^2 \leq -\left\|\hat{f}_{m^\star}\right\|^2 + \widehat{\text{pen}}(m^\star).$$

Thus

$$\left\|\hat{f}_{m^\star} - \hat{f}_{\hat{m}}\right\|^2 \mathbb{1}_G = \left(\left\|\hat{f}_{m^\star}\right\|^2 - \left\|\hat{f}_{\hat{m}}\right\|^2\right) \mathbb{1}_G \leq \widehat{\text{pen}}(m^\star).$$

Taking expectation, we apply the following Lemma

Lemma 6.1. *For a random penalty noted* $\widehat{\text{pen}}$ *defined by (17) or (18) and a deterministic penalty noted* pen *defined by (19) or (20), there is a positive constant C such that for any arbitrary $m \in \mathcal{M}_n$*

$$\mathbb{E}\left[\widehat{\text{pen}}(m)\right] \leq C\text{pen}(m). \tag{26}$$

It yields for some positive constant C

$$\begin{aligned}
\mathbb{E}\left[\left\|f - \hat{f}_{\hat{m}}\right\|^2 \mathbb{1}_G\right] &\leq 2\mathbb{E}\left[\left\|f - \hat{f}_{m^\star}\right\|^2\right] + 2\mathbb{E}\left[\widehat{\text{pen}}(m^\star)\right] \\
&\leq 2\left\|f - f_{m^\star}\right\|^2 + 2C\text{pen}(m^\star).
\end{aligned}$$

We just proved the desired result on G

$$\mathbb{E}\left[\|f - \hat{f}_{\hat{m}}\|^2 \mathbb{1}_G\right] \leq C \inf_{m\in\mathcal{M}_n} \left\{\|f - f_m\|^2 + \text{pen}(m)\right\}. \tag{27}$$

- We now consider the set $G^c = \{\hat{m} > m^\star\}$.

$$\begin{aligned}
\left\|\hat{f}_{\hat{m}} - \hat{f}_{m^\star}\right\|^2 \mathbb{1}_{G^c} &= \left(\left\|\hat{f}_{\hat{m}} - \hat{f}^*_{m^\star}\right\|^2 - 6\left\|f_{\hat{m}} - f_{m^\star}\right\|^2 - \frac{1}{2}\widehat{\text{pen}}(m^\star, \hat{m})\right)\mathbb{1}_{G^c} \\
&\quad + \left(6\left\|f_{\hat{m}} - f_{m^\star}\right\|^2 + \frac{1}{2}\widehat{\text{pen}}(m^\star, \hat{m})\right)\mathbb{1}_{G^c} \\
&\leq \sup_{\substack{k\geq m^\star \\ k\in\mathcal{M}_n}} \left\{\left\|\hat{f}_k - \hat{f}_{m^\star}\right\|^2 - 6\left\|f_k - f_{m^\star}\right\|^2 - \frac{1}{2}\widehat{\text{pen}}(m^\star, k)\right\}_+ \\
&\quad + 6\|f - f_{m^\star}\|^2 + \frac{1}{2}\sum_{\substack{k\geq m^\star \\ k\in\mathcal{M}_n}} \widehat{\text{pen}}(m^\star, k)\mathbb{1}\left\{\hat{m} = k\right\}
\end{aligned} \tag{28}$$

Let us first notice the following inequality

$$\forall k > m, \quad \widehat{\text{pen}}(m, k) \leq \widehat{\text{pen}}(k). \tag{29}$$

Besides by definition of $\hat{m}$ (see Equation (21) or (22)), on the set $\{\hat{m} = k\} \cap G^c$ and applying Equation (25), one has

$$\begin{aligned}
&\frac{1}{2}\left(\widehat{\text{pen}}(k) - \widehat{\text{pen}}(m^\star)\right) \leq \left\|\hat{f}_{\hat{m}} - \hat{f}_{m^\star}\right\|^2 - \frac{1}{2}\left(\widehat{\text{pen}}(k) - \widehat{\text{pen}}(m^\star)\right) \\
&\frac{1}{2}\widehat{\text{pen}}(k) \leq \left\|\hat{f}_{\hat{m}} - \hat{f}_{m^\star}\right\|^2 - \frac{1}{2}\widehat{\text{pen}}(m^\star, k) + \frac{1}{2}\widehat{\text{pen}}(m^\star) \\
&\qquad \leq \left(\left\|\hat{f}_{\hat{m}} - \hat{f}_{m^\star}\right\|^2 - 6\left\|f_{\hat{m}} - f_{m^\star}\right\|^2 - \frac{1}{2}\widehat{\text{pen}}(m^\star, k)\right) + 6\left\|f_{\hat{m}} - f_{m^\star}\right\|^2 + \frac{1}{2}\widehat{\text{pen}}(m^\star) \\
&\qquad \leq \left(\left\|\hat{f}_{\hat{m}} - \hat{f}_{m^\star}\right\|^2 - 6\left\|f_{\hat{m}} - f_{m^\star}\right\|^2 - \frac{1}{2}\widehat{\text{pen}}(m^\star, k)\right) + 6\left\|f^* - f^*_{m^\star}\right\|^2 + \frac{1}{2}\widehat{\text{pen}}(m^\star)
\end{aligned} \tag{30}$$

Now using Inequalities (29) and (30)

$$\frac{1}{2}\sum_{\substack{k\geq m^\star\\ k\in\mathcal{M}_n}}\widehat{\mathrm{pen}}(m^\star,k)\leq \sup_{\substack{k\geq m^\star\\ k\in\mathcal{M}_n}}\left\{\left\|\hat{f}_k-\hat{f}_{m^\star}\right\|^2-6\left\|f^*_{\hat{m}}-f_{m^\star}\right\|^2-\frac{1}{2}\widehat{\mathrm{pen}}(m^\star,k)\right\}_+ + 6\left\|f-f_{m^\star}\right\|^2+\frac{1}{2}\widehat{\mathrm{pen}}(m^\star).$$

From Inequality (28), we now have

$$\left\|\hat{f}_{\hat{m}}-\hat{f}_{m^\star}\right\|^2\mathbb{1}_{G^c}\leq 2\sup_{\substack{k\geq m^\star\\ k\in\mathcal{M}_n}}\left\{\left\|\hat{f}_k-\hat{f}_{m^\star}\right\|^2-6\left\|f_k-f_{m^\star}\right\|^2-\frac{1}{2}\widehat{\mathrm{pen}}(m^\star,k)\right\}_+ +12\|f-f_{m^\star}\|^2+\frac{1}{2}\widehat{\mathrm{pen}}(m^\star). \tag{31}$$

Taking expectation the first summand is negligible by applying the following Proposition.

Proposition 6.2. *Under (**A1**) and (**A2**), there is a positive constant C such that for any arbitrary $m\in\mathcal{M}_n$*

$$\mathbb{E}\left[\sup_{\substack{k\geq m\\ k\in\mathcal{M}_n}}\left\{\left\|\hat{f}_{k,\mathrm{NS}}-\hat{f}_{m,\mathrm{NS}}\right\|^2-6\left\|f_k-f_m\right\|^2-\frac{1}{2}\widehat{\mathrm{pen}}_{\mathrm{NS}}(m,k)\right\}_+\right]\leq\frac{C}{n}+\frac{C}{M} \tag{32}$$

$$\mathbb{E}\left[\sup_{\substack{k\geq m\\ k\in\mathcal{M}_n}}\left\{\left\|\hat{f}_{k,\mathrm{RD}}-\hat{f}_{m,\mathrm{RD}}\right\|^2-6\left\|f_k-f_m\right\|^2-\frac{1}{2}\widehat{\mathrm{pen}}_{\mathrm{RD}}(m,k)\right\}_+\right]\leq\frac{C}{n+M}+\frac{C}{M} \tag{33}$$

Finally we have

$$\mathbb{E}\left[\left\|f-\hat{f}_{\hat{m}}\right\|^2\mathbb{1}_{G^c}\right]\leq C\left(\|f-f_{m^\star}\|^2+\mathrm{pen}(m^\star)\right)+C'\left(\frac{1}{n}+\frac{1}{M}\right) \tag{34}$$

$$\mathbb{E}\left[\left\|f-\hat{f}_{\hat{m}}\right\|^2\mathbb{1}_{G^c}\right]\leq C\left(\|f-f_{m^\star}\|^2+\mathrm{pen}(m^\star)\right)+C'\left(\frac{1}{n+M}+\frac{1}{M}\right). \tag{35}$$

This combining with (27) complete the proof.

□

6.2 Proof of Lemma 6.1

Before proving Lemma 6.1, we first need to prove two auxiliary lemmas. In the sequel, C will always denote some universal positive constant, but the value may vary from line to line.

Lemma 6.3. *For an estimator of f^*_ε defined by (12) or (13), assume $\kappa>\sqrt{c_1p}$. Let $\tau\geq 2\kappa$ and $x\geq 1$. Then for some positive constant C*

$$\mathbb{P}\left[\exists u\in\mathbb{R}:|\tilde{f}^*_{\varepsilon,\mathrm{NS}}(u)-f^*_\varepsilon(u)|>\tau\left(\log(Mx)\right)^{1/2}w(u)^{-1}M^{-1/2}\right]\leq Cx^{-p}M^{-p}$$

$$\mathbb{P}\left[\exists u\in\mathbb{R}:|\tilde{f}^*_{\varepsilon,\mathrm{RD}}(u)^2-f^*_\varepsilon(u)^2|>\tau\left(\log(Mx)\right)^{1/2}w(u)^{-1}M^{-1/2}\right]\leq Cx^{-p}M^{-p}$$

Proof. We give the proof for a $\tilde{f}^*_\varepsilon$ which can be equal to $\tilde{f}^*_{\varepsilon,\mathrm{NS}}$ or $\tilde{f}^{*2}_{\varepsilon,\mathrm{RD}}$. In consequence for the second case we need to consider f^{*2}_ε instead of f^*_ε as follows

$$\left|\tilde{f}^*_\varepsilon(u)-f^*_\varepsilon(u)\right|\leq\left|\tilde{f}^*_\varepsilon(u)-\hat{f}^*_\varepsilon(u)\right|+\left|\hat{f}^*_\varepsilon(u)-f^*_\varepsilon(u)\right|\leq 2k_M(u)+\left|\hat{f}^*_\varepsilon(u)-f^*_\varepsilon(u)\right|.$$

Using the previous inequality and applying Lemma A.3, we have

$$\begin{aligned}
&\mathbb{P}\left[\exists u\in\mathbb{R}:|\tilde{f}^*_\varepsilon(u)-f^*_\varepsilon(u)|>\tau\left(\log(Mx)\right)^{1/2}w(u)^{-1}M^{-1/2}\right]\\
&\leq\mathbb{P}\left[\exists u\in\mathbb{R}:|\hat{f}^*_\varepsilon(u)-f^*_\varepsilon(u)|+2k_M(u)>\tau\left(\log(Mx)\right)^{1/2}w(u)^{-1}M^{-1/2}\right]\\
&\leq\mathbb{P}\left[\exists u\in\mathbb{R}:|\hat{f}^*_\varepsilon(u)-f^*_\varepsilon(u)|>(\tau-2\kappa)\left(\log(Mx)\right)^{1/2}w(u)^{-1}M^{-1/2}\right]\\
&\leq Cx^{-p}M^{-p}.
\end{aligned}$$

□

Lemma 6.4. *In the situation of the preceding Lemma*

$$\mathbb{P}\left[\exists u \in \mathbb{R} : \left|\tilde{f}^*_{\varepsilon,\mathrm{RD}}(u) - f^*_\varepsilon(u)\right| \mathbb{1}\left\{\tilde{f}^*_{\varepsilon,\mathrm{RD}}(u) < f^*_\varepsilon(u)\right\} > \frac{\tau(\log(Mx))^{1/2} w(u)^{-1} M^{-1/2}}{\tilde{f}^*_{\varepsilon,\mathrm{RD}}(u)}\right] \leq C x^{-p} M^{-p}$$

Proof. This is a direct consequence of Lemma 6.3 using the fact that for $x, y \geq 0$, $|\sqrt{x} - \sqrt{y}| \leq \frac{|x-y|}{2\sqrt{x \wedge y}}$ holds. ☐

Proof of Lemma 6.1. Here all subscripts relative to one of the two considered models are dropped since the argument for both models are totally alike.

For $q = 1/2$ or 1, using Cauchy-Schwarz's inequality, we have

$$\mathbb{E}\left[\log^q\left(1 + \hat{\Delta}(m)m^2\right)\hat{\Delta}(m)\right] \leq \sqrt{\mathbb{E}\left[\log^{2q}\left(1 + \hat{\Delta}(m)m^2\right)\right]\mathbb{E}\left[\hat{\Delta}^2(m)\right]}$$

Let $A_p(u)$ be defined as in Lemma A.4

$$\begin{aligned}
\hat{\Delta}(m) &= \frac{1}{2\pi}\int_{-\pi m}^{\pi m} \frac{w(u)^{-2}}{|\tilde{f}^*_\varepsilon(u)|^2}\,\mathrm{d}u \\
&= \frac{1}{2\pi}\int_{-\pi m}^{\pi m} w(u)^{-2}\left|\frac{1}{\tilde{f}^*_\varepsilon(u)} - \frac{1}{f^*_\varepsilon(u)} + \frac{1}{f^*_\varepsilon(u)}\right|^2 \mathrm{d}u \\
&\leq 2\Delta(m) + \frac{1}{\pi}\int_{-\pi m}^{\pi m} w(u)^{-2}\left|\frac{1}{\tilde{f}^*_\varepsilon(u)} - \frac{1}{f^*_\varepsilon(u)}\right|^2 \mathrm{d}u \\
&\leq 2\Delta(m) + 2\Delta(m)\sup_{u\in\mathbb{R}} A_1(u) \\
&\leq 2\Delta(m)(1 + \sup_{u\in\mathbb{R}} A_1(u))
\end{aligned}$$

and applying Lemma A.4

$$\mathbb{E}\left[\hat{\Delta}^2(m)\right] \leq 4C\Delta^2(m).$$

By Jensen inequality (since log is concave)

$$\begin{aligned}
\mathbb{E}\left[\log^{2q}\left(1 + \hat{\Delta}(m)m^2\right)\right] &\leq \log^{2q}\left(\mathbb{E}\left[1 + \hat{\Delta}(m)m^2\right]\right) \\
&\leq \log^{2q}\left(1 + \mathbb{E}\left[\hat{\Delta}(m)\right]m^2\right) \\
&\leq \log^{2q}\left(1 + 2\Delta(m)\left(1 + \mathbb{E}\left[\sup_{u\in\mathbb{R}} A_1(u)\right]\right)m^2\right) \\
&\leq \log^{2q}\left(1 + 2\Delta(m)\left(1 + C\right)m^2\right) \\
&\leq C\log^{2q}\left(1 + \Delta(m)m^2\right)
\end{aligned}$$

So

$$\mathbb{E}\left[\log^q\left(1 + \hat{\Delta}(m)m^2\right)\hat{\Delta}(m)\right] \leq C\log^q\left(1 + \Delta(m)m^2\right)\Delta(m)$$

which means $\mathbb{E}[\widehat{\mathrm{pen}}_1(m)] \leq C\mathrm{pen}_1(m)$. Consider now $\widehat{\mathrm{pen}}_2(m)$. Another application of Lemma A.4 yields

$$\begin{aligned}
&\frac{1}{M}\mathbb{E}\left[\frac{1}{2\pi}\int_{-\pi m}^{\pi m} \frac{w(u)^{-2}|\hat{f}^*_Y(u)|^2}{|\tilde{f}^*_\varepsilon(u)|^4}\,\mathrm{d}u\right] \\
&\leq \frac{2}{M}\mathbb{E}\left[\frac{1}{2\pi}\int_{-\pi m}^{\pi m} \frac{w(u)^{-2}|f^*_Y(u)|^2}{|\tilde{f}^*_\varepsilon(u)|^4}\,\mathrm{d}u\right] + \frac{2}{M}\mathbb{E}\left[\frac{1}{2\pi}\int_{-\pi m}^{\pi m} \frac{w(u)^{-2}|\hat{f}^*_Y(u) - f^*_Y(u)|^2}{|\tilde{f}^*_\varepsilon(u)|^4}\,\mathrm{d}u\right] \\
&\leq \frac{4}{M}\mathbb{E}\left[\frac{1}{2\pi}\int_{-\pi m}^{\pi m} \frac{w(u)^{-2}|f^*(u)|^2}{|f^*_\varepsilon(u)|^2}\,\mathrm{d}u\right]\mathbb{E}\left[1 + \sup_{u\in\mathbb{R}} A_2(u)\right] + \frac{2}{M}\mathbb{E}\left[\frac{1}{2\pi}\int_{-\pi m}^{\pi m} \frac{w(u)^{-2}|\hat{f}^*_Y(u) - f^*_Y(u)|^2}{|\tilde{f}^*_\varepsilon(u)|^4}\,\mathrm{d}u\right] \\
&\leq \frac{4}{M}\Delta^f(m)\,\mathbb{E}\left[1 + \sup_{u\in\mathbb{R}} A_2(u)\right] + \frac{2}{M}\mathbb{E}\left[\frac{1}{2\pi}\int_{-\pi m}^{\pi m} \frac{w(u)^{-2}|\hat{f}^*_Y(u) - f^*_Y(u)|^2}{|\tilde{f}^*_\varepsilon(u)|^4}\,\mathrm{d}u\right]
\end{aligned}$$

We use the fact that $\tilde{f}_\varepsilon^*(u) \geq M^{-1/2}(\log M)^{1/2} w(u)^{-1}$ as well as the independence of $\hat{f}_Y^*$ and $\tilde{f}_\varepsilon^*$ to find

$$\mathbb{E}\left[\frac{1}{2\pi}\int_{-\pi m}^{\pi m} \frac{w(u)^{-2}|\hat{f}_Y^*(u) - f_Y^*(u)|^2}{|\tilde{f}_\varepsilon^*(u)|^4}\,\mathrm{d}u\right] \leq \mathbb{E}\left[\sup_{u\in\mathbb{R}} |\hat{f}_Y^*(u) - f_Y^*(u)|^2 w(u)^2\right] \mathbb{E}\left[\frac{1}{2\pi}\int_{-\pi m}^{\pi m} \frac{w(u)^{-2}}{|\tilde{f}_\varepsilon^*(u)|^2}\,\mathrm{d}u\right]$$

Thanks to Theorem 5.1 in Neumann and Reiß (2009), for some positive constant C,

$$\mathbb{E}\left[\sup_{u\in\mathbb{R}} |\hat{f}_Y^*(u) - f_Y^*(u)|^2 w(u)^2\right] \leq \frac{C}{n}.$$

Applying the same arguments as for the bounding of $\mathbb{E}\left[\widehat{\mathrm{pen}}_1(m)\right]$, we get

$$\mathbb{E}\left[\frac{1}{2\pi}\int_{-\pi m}^{\pi m} \frac{w(u)^{-2}}{|\tilde{f}_\varepsilon^*(u)|^2}\,\mathrm{d}u\right] \leq C\Delta(m).$$

This completes the proof. □

6.3 Proof of Proposition 6.2

For $k > m$, let us introduce the following notation : $A(m,k) := \{u \in \mathbb{R}, |u| \in [\pi m, \pi k]\}$. Similarly to the previous section, we need to prove some auxilary lemmas before proving Proposition 6.2.

Lemma 6.5. *There is a positive constant C such that for any arbitrary $m \in \mathcal{M}_n$*

$$\mathbb{E}\left[\sup_{k\geq m}\left\{\frac{1}{2\pi}\int_{A(m,k)} \frac{|\hat{f}_{Y,\mathrm{NS}}^*(u) - f_Y^*(u)|^2}{|\tilde{f}_{\varepsilon,\mathrm{NS}}^*(u)|^2}\,\mathrm{d}u - \frac{1}{12}\widehat{\mathrm{pen}}_{1,\mathrm{NS}}(m,k)\right\}_+\right] \leq \frac{C}{n} \tag{36}$$

$$\mathbb{E}\left[\sup_{k\geq m}\left\{\frac{1}{2\pi}\int_{A(m,k)} \frac{|\hat{f}_{Y,\mathrm{RD}}^*(u) - f_Y^*(u)|^2}{|\tilde{f}_{\varepsilon,\mathrm{RD}}^*(u)|^2}\,\mathrm{d}u - \frac{1}{12}\widehat{\mathrm{pen}}_{1,\mathrm{RD}}(m,k)\right\}_+\right] \leq \frac{C}{n+M} \tag{37}$$

Proof. Once again the proof is given simultaneously for both models. Thus thereafter for the sake of clarity all subscripts relative to the (NS) model or the (RD) model are dropped again.

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\substack{k\geq m\\ k\in\mathcal{M}_n}}\left\{\frac{1}{2\pi}\int_{A(m,k)} \frac{|\hat{f}_Y^*(u) - f_Y^*(u)|^2}{|\tilde{f}_\varepsilon^*(u)|^2}\,\mathrm{d}u - \frac{1}{12}\widehat{\mathrm{pen}}_1(m,k)\right\}_+\right] \\
&\leq \mathbb{E}\left[\sum_{\substack{k\geq m\\ k\in\mathcal{M}_n}}\left\{\mathbb{E}\left[\frac{1}{2\pi}\int_{A(m,k)} \frac{|\hat{f}_Y^*(u) - f_Y^*(u)|^2}{|\tilde{f}_\varepsilon^*(u)|^2}\,\mathrm{d}u - \frac{1}{12}\widehat{\mathrm{pen}}_1(m,k)\right\}_+ \Big|\tilde{f}_\varepsilon^*\right]\right] \\
&\leq \mathbb{E}\left[\sum_{\substack{k\geq m\\ k\in\mathcal{M}_n}} \frac{1}{2\pi}\int_{A(m,k)} \frac{1}{2\pi}\mathbb{E}\left[\left\{\frac{|\hat{f}_Y^*(u) - f_Y^*(u)|^2}{|\tilde{f}_\varepsilon^*(u)|^2} - \frac{1}{12}\frac{\hat{\lambda}^2(m,k)}{n|\tilde{f}_\varepsilon^*(u)|^2}\right\}_+ \Big|\tilde{f}_\varepsilon^*\right]\mathrm{d}u\right]
\end{aligned}$$

Now $\hat{f}_Y^*(u)/\tilde{f}_\varepsilon^*(u)$ (conditional on $\tilde{f}_\varepsilon^*(u)$) the sum of n independent and identically distributed random variables with variance $v^2 \leq 1/|\tilde{f}_\varepsilon^*(u)|^2$ which are surely bounded by $2/\tilde{f}_\varepsilon^*(u)$. Thus Lemma A.1 gives

$$\begin{aligned}
&\mathbb{E}\left[\left\{\frac{|\hat{f}_Y^*(u) - f_Y^*(u)|^2}{|\tilde{f}_\varepsilon^*(u)|^2} - \frac{1}{12}\frac{\hat{\lambda}^2(m,k)}{n|\tilde{f}_\varepsilon^*(u)|^2}\right\}_+ \Big|\tilde{f}_\varepsilon^*\right] \\
&\leq 32\frac{n^{-1}}{|\tilde{f}_\varepsilon^*(u)|^2}\exp\left(-\hat{\lambda}^2(m,k)\right) + 128\sqrt{2}\frac{n^{-2}}{|\tilde{f}_\varepsilon^*(u)|^2}\exp\left(-n^{1/2}\hat{\lambda}(m,k)\right) \\
&\leq 32\frac{n^{-1}}{|\tilde{f}_\varepsilon^*(u)|^2}(k-m)^{-2}\hat{\Delta}(m,k)^{-1} + 128\sqrt{2}\frac{n^{-2}}{|\tilde{f}_\varepsilon^*(u)|^2}(k-m)^{-2}\hat{\Delta}(m,k)^{-1}
\end{aligned}$$

where we used the fact that

$$\hat{\lambda}(m,k) \leq \max\left\{\sqrt{8\log\left(1+\hat{\Delta}(m,k)(k-m)^2\right)}, \frac{16\sqrt{2}}{3\sqrt{n}}\log\left(1+\hat{\Delta}(m,k)(k-m)^2\right)\right\}.$$

We have thus shown for a universal positive constant C that for any $m,k\in\mathcal{M}_n$

$$\begin{aligned}
&\int_{A(m,k)} \mathbb{E}\left[\left\{\frac{|\hat{f}_Y^*(u)-f_Y^*(u)|^2}{|\tilde{f}_\varepsilon^*(u)|^2} - \frac{1}{12}\frac{\hat{\lambda}_1^2(m)}{n|\tilde{f}_\varepsilon^*(u)|^2}\right\}_+ \left|\tilde{f}_\varepsilon^*\right|\right] \mathrm{d}u \\
&\leq \frac{C}{n}(k-m)^{-2}\hat{\Delta}(m,k)^{-1}\int_{A(m,k)}\frac{\mathrm{d}u}{|\tilde{f}_\varepsilon^*(u)|^2} \\
&\leq \frac{C}{n}(k-m)^{-2}
\end{aligned}$$

Finally

$$\mathbb{E}\left[\sup_{\substack{k\geq m\\ k\in\mathcal{M}_n}}\left\{\frac{1}{2\pi}\int_{A(m,k)}\frac{|\hat{f}_Y^*(u)-f_Y^*(u)|^2}{|\tilde{f}_\varepsilon^*(u)|^2}\mathrm{d}u - \frac{1}{6}\widehat{\mathrm{pen}}_1(m,k)\right\}_+\right] \leq \frac{C}{n}$$

or

$$\mathbb{E}\left[\sup_{\substack{k\geq m\\ k\in\mathcal{M}_n}}\left\{\frac{1}{2\pi}\int_{A(m,k)}\frac{|\hat{f}_Y^*(u)-f_Y^*(u)|^2}{|\tilde{f}_\varepsilon^*(u)|^2}\mathrm{d}u - \frac{1}{6}\widehat{\mathrm{pen}}_1(m,k)\right\}_+\right] \leq \frac{C}{n+M}$$

$\square$

Proof of Proposition 6.2. We give the proof for a $\tilde{f}_\varepsilon^*$ which can be equal to $\tilde{f}_{\varepsilon,\mathrm{NS}}^*$ or $\tilde{f}_{\varepsilon,\mathrm{RD}}^{*2}$. In consequence, once again all subscripts relative to one model are dropped.

Applying Plancherel's formula we get

$$\begin{aligned}
\left\|\hat{f}_k - \hat{f}_m\right\|^2 &= \frac{1}{2\pi}\left\|\hat{f}_k^* - \hat{f}_m^*\right\|^2 \\
&= \frac{1}{2\pi}\int_{A(m,k)}\frac{|\hat{f}_Y^*(u))|^2}{|\tilde{f}_\varepsilon^*(u)|^2}\mathrm{d}u \\
&= \frac{1}{2\pi}\int_{A(m,k)}\frac{|\hat{f}_Y^*(u))|^2}{|\tilde{f}_\varepsilon^*(u)|^2}\mathbb{1}\left\{|\tilde{f}_\varepsilon^*(u)| > |f_\varepsilon^*(u)|\right\}\mathrm{d}u \\
&\quad + \frac{1}{2\pi}\int_{A(m,k)}\frac{|\hat{f}_Y^*(u))|^2}{|\tilde{f}_\varepsilon^*(u)|^2}\mathbb{1}\left\{|\tilde{f}_\varepsilon^*(u)| \leq |f_\varepsilon^*(u)|\right\}\mathrm{d}u
\end{aligned} \tag{38}$$

Let us consider the first term of Equation (38)

$$\frac{1}{2\pi}\int_{A(m,k)}\frac{|\hat{f}_Y^*(u))|^2}{|\tilde{f}_\varepsilon^*(u)|^2}\mathbb{1}\left\{|\tilde{f}_\varepsilon^*(u)| > |f_\varepsilon^*(u)|\right\}\mathrm{d}u \leq \frac{1}{\pi}\int_{A(m,k)}\frac{|\hat{f}_Y^*(u)-f_Y^*(u)|^2}{|\tilde{f}_\varepsilon^*(u)|^2}\mathrm{d}u + \underbrace{\frac{1}{\pi}\int_{A(m,k)}\frac{|f_Y^*(u)|^2}{|f_\varepsilon^*(u)|^2}\mathrm{d}u}_{2\|f_k-f_m\|^2},$$

now the second one

$$\begin{aligned}
&\frac{1}{2\pi}\int_{A(m,k)}\frac{|\hat{f}_Y^*(u))|^2}{|\tilde{f}_\varepsilon^*(u)|^2}\mathbb{1}\left\{|\tilde{f}_\varepsilon^*(u)| \leq |f_\varepsilon^*(u)|\right\}\mathrm{d}u \\
&= \frac{1}{2\pi}\int_{A(m,k)}|\hat{f}_Y^*(u)|^2\left|\frac{1}{\tilde{f}_\varepsilon^*(u)} - \frac{1}{f_\varepsilon^*(u)} + \frac{1}{f_\varepsilon^*(u)}\right|^2\mathbb{1}\left\{|\tilde{f}_\varepsilon^*(u)| > |f_\varepsilon^*(u)|\right\}\mathrm{d}u \\
&\leq \frac{1}{\pi}\int_{A(m,k)}|\hat{f}_Y^*(u)|^2\left|\frac{1}{\tilde{f}_\varepsilon^*(u)} - \frac{1}{f_\varepsilon^*(u)}\right|^2\mathbb{1}\left\{|\tilde{f}_\varepsilon^*(u) \leq |f_\varepsilon^*(u)|\right\}\mathrm{d}u \\
&\quad + \frac{1}{\pi}\int_{A(m,k)}\left|\frac{\hat{f}_Y^*(u))}{f_\varepsilon^*(u)}\right|^2\mathbb{1}\left\{|\tilde{f}_\varepsilon^*(u) \leq |f_\varepsilon^*(u)|\right\}\mathrm{d}u.
\end{aligned} \tag{39}$$

Yet for the second term of Equation (39), we can notice

$$\frac{1}{\pi}\int_{A(m,k)}\left|\frac{\hat{f}_Y^*(u))}{f_\varepsilon^*(u)}\right|^2 \mathbb{1}\left\{|\tilde{f}_\varepsilon^*(u) \le |f_\varepsilon^*(u)|\right\} \mathrm{d}u \le \frac{2}{\pi}\int_{A(m,k)} \frac{|\hat{f}_Y^*(u) - f_Y^*(u)|^2}{|\tilde{f}_\varepsilon^*(u)|^2} \mathrm{d}u + 4\left\|f_k - f_m\right\|^2.$$

For the first term of Equation (39) we can write

$$\begin{aligned}
&\frac{1}{\pi}\int_{A(m,k)} |\hat{f}_Y^*(u)|^2 \left|\frac{1}{\tilde{f}_\varepsilon^*(u)} - \frac{1}{f_\varepsilon^*(u)}\right|^2 \mathbb{1}\left\{\ |\tilde{f}_\varepsilon^*(u) \le |f_\varepsilon^*(u)|\right\} \mathrm{d}u\\
&= \frac{1}{\pi}\int_{A(m,k)} |\hat{f}_Y^*(u)|^2 \frac{|\tilde{f}_\varepsilon^*(u) - f_\varepsilon^*(u)|^2}{|\tilde{f}_\varepsilon^*(u)|^2 |f_\varepsilon^*(u)|^2} \mathbb{1}\left\{|\tilde{f}_\varepsilon^*(u) \le |f_\varepsilon^*(u)|\right\} \mathrm{d}u\\
&\le \frac{1}{\pi}\int_{A(m,k)} |\hat{f}_Y^*(u)|^2 \frac{|\tilde{f}_\varepsilon^*(u) - f_\varepsilon^*(u)|^2}{|\tilde{f}_\varepsilon^*(u)|^4} \mathbb{1}\left\{\ |\tilde{f}_\varepsilon^*(u) \le |f_\varepsilon^*(u)|\right\} \mathrm{d}u.
\end{aligned}$$

Let us introduce the set C which equals to one of the following set depending on the model

$$C_{\text{NS}}(m,k) = \left\{\forall u \in \mathbb{R} : |\tilde{f}_\varepsilon^*(u) - f_\varepsilon^*(u)|^2 \le 4\kappa^2 \log\left(M(k-m)\right) w(u)^{-2} M^{-1}\right\}, \tag{40}$$

$$C_{\text{RD}}(m,k) = \left\{\forall u \in \mathbb{R} : |\tilde{f}_\varepsilon^*(u) - f_\varepsilon^*(u)|^2 \mathbb{1}\left\{\tilde{f}_\varepsilon^*(u) \le f_\varepsilon^*(u)\right\} \le \frac{4\kappa^2 \log\left(M(k-m)\right) w(u)^{-2} M^{-1}}{\tilde{f}_\varepsilon^*(u)^2}\right\}. \tag{41}$$

On $C(m,k)$, for example for $C = C_{\text{NS}}$ the following inequalities can be deduced

$$\begin{aligned}
&\frac{1}{2\pi}\int_{A(m,k)} |\hat{f}_Y^*(u)|^2 \frac{|\tilde{f}_\varepsilon^*(u) - f_\varepsilon^*(u)|^2}{|\tilde{f}_\varepsilon^*(u)|^4} \mathbb{1}\left\{\ |\tilde{f}_\varepsilon^*(u) \le |f_\varepsilon^*(u)|\right\} \mathrm{d}u\\
&\le 4\kappa^2 \log\left(M(k-m)\right) M^{-1} \frac{1}{2\pi}\int_{A(m,k)} \frac{w(u)^{-2}|\hat{f}_Y^*(u)|^2}{|\tilde{f}_\varepsilon^*(u)|^4} \mathbb{1}\left\{\ |\tilde{f}_\varepsilon^*(u)| \le |f_\varepsilon^*(u)|\right\} \mathrm{d}u\\
&\le 4\kappa^2 \log\left(M(k-m)\right) M^{-1} \hat{\Delta}^f(m,k)\\
&\le \frac{1}{4}\widehat{\text{pen}}_2(m,k).
\end{aligned}$$

This far we have showed that Equation (38) can be bounded as follows

$$\begin{aligned}
\left\|\hat{f}_k - \hat{f}_m\right\|^2 &\le \frac{1}{\pi}\int_{A(m,k)} \frac{|\hat{f}_Y^*(u) - f_Y^*(u)|^2}{|\tilde{f}_\varepsilon^*(u)|^2} \mathrm{d}u + \left\|f_k - f_m\right\|^2\\
&\quad + \frac{2}{\pi}\int_{A(m,k)} \frac{|\hat{f}_Y^*(u) - f_Y^*(u)|^2}{|\tilde{f}_\varepsilon^*(u)|^2} \mathrm{d}u + 4\left\|f_k - f_m\right\|^2\\
&\quad + \frac{1}{\pi}\int_{A(m,k)} |\hat{f}_Y^*(u)|^2 \frac{|\tilde{f}_\varepsilon^*(u) - f_\varepsilon^*(u)|^2}{|\tilde{f}_\varepsilon^*(u)|^4} \mathbb{1}\left\{\ |\tilde{f}_\varepsilon^*(u)| \le |f_\varepsilon^*(u)|\right\} \mathrm{d}u\, \mathbb{1}\left\{C(m,k)\right\}\\
&\quad + \frac{1}{\pi}\int_{A(m,k)} |\hat{f}_Y^*(u)|^2 \frac{|\tilde{f}_\varepsilon^*(u) - f_\varepsilon^*(u)|^2}{|\tilde{f}_\varepsilon^*(u)|^4} \mathbb{1}\left\{\ |\tilde{f}_\varepsilon^*(u)| \le |f_\varepsilon^*(u)|\right\} \mathrm{d}u\, \mathbb{1}\left\{C(m,k)^c\right\}\\
&\le \frac{3}{\pi}\int_{A(m,k)} \frac{|\hat{f}_Y^*(u) - f_Y^*(u)|^2}{|\tilde{f}_\varepsilon^*(u)|^2} \mathrm{d}u + 6\left\|f_k - f_m\right\|^2 + \frac{1}{2}\widehat{\text{pen}}_2(m,k)\\
&\quad + \frac{1}{\pi}\int_{A(m,k)} |\hat{f}_Y^*(u)|^2 \frac{|\tilde{f}_\varepsilon^*(u) - f_\varepsilon^*(u)|^2}{|\tilde{f}_\varepsilon^*(u)|^4} \mathbb{1}\left\{\ |\tilde{f}_\varepsilon^*(u)| \le |f_\varepsilon^*(u)|\right\} \mathrm{d}u\, \mathbb{1}\left\{C(m,k)^c\right\}.
\end{aligned} \tag{42}$$

Starting from Equation (42), we can now write the following inequalities

$$
\begin{aligned}
&\left\|\hat{f}_{k,\mathrm{NS}}-\hat{f}_{m,\mathrm{NS}}\right\|^2-6\|f_k-f_m\|^2-\frac{1}{2}\widehat{\mathrm{pen}}(m,k)\\
&\leq \frac{3}{\pi}\int_{A(m,k)}\frac{|\hat{f}_Y^*(u)-f_Y^*(u)|^2}{|\tilde{f}_\varepsilon^*(u)|^2}\,\mathrm{d}u-\frac{1}{2}\widehat{\mathrm{pen}}_1(m,k)\\
&\quad+6\|f_k-f_m\|^2-6\|f_k-f_m\|^2+\frac{1}{2}\widehat{\mathrm{pen}}_2(m,k)-\frac{1}{2}\widehat{\mathrm{pen}}_2(m,k)\\
&\quad+\frac{1}{\pi}\int_{A(m,k)}|\hat{f}_Y^*(u)|^2\frac{|\tilde{f}_\varepsilon^*(u)-f_\varepsilon^*(u)|^2}{|\tilde{f}_\varepsilon^*(u)|^4}\mathbb{1}\left\{\ |\tilde{f}_\varepsilon^*(u)|\leq|f_\varepsilon^*(u)|\right\}\mathrm{d}u\,\mathbb{1}\left\{C(m,k)^c\right\}\\
&\leq 6\left\{\frac{1}{2\pi}\int_{A(m,k)}\frac{|\hat{f}_Y^*(u)-f_Y^*(u)|^2}{|\tilde{f}_\varepsilon^*(u)|^2}\,\mathrm{d}u-\frac{1}{12}\widehat{\mathrm{pen}}_1(m,k)\right\}_+\\
&\quad+\frac{1}{\pi}\int_{A(m,k)}|\hat{f}_Y^*(u)|^2\frac{|\tilde{f}_\varepsilon^*(u)-f_\varepsilon^*(u)|^2}{|\tilde{f}_\varepsilon^*(u)|^4}\mathbb{1}\left\{\ |\tilde{f}_\varepsilon^*(u)|\leq|f_\varepsilon^*(u)|\right\}\mathrm{d}u\,\mathbb{1}\left\{C(m,k)^c\right\}.
\end{aligned}
$$

Taking expectation we get

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{\substack{k\geq m\\ k\in\mathcal{M}_n}}\left\{\left\|\hat{f}_k-\hat{f}_m\right\|^2-6\left\|f_k-f_m\right\|^2-\frac{1}{2}\widehat{\mathrm{pen}}(m,k)\right\}_+\right]\\
&\leq\sum_{\substack{k\geq m\\ k\in\mathcal{M}_n}}\mathbb{E}\left[\left\{\left\|\hat{f}_k-\hat{f}_m\right\|^2-6\left\|f_k-f_m\right\|^2-\frac{1}{2}\widehat{\mathrm{pen}}(m,k)\right\}_+\right]\\
&\leq\sum_{\substack{k\geq m\\ k\in\mathcal{M}_n}}\mathbb{E}\left[6\left\{\frac{1}{2\pi}\int_{A(m,k)}\frac{|\hat{f}_Y^*(u)-f_Y^*(u)|^2}{|\tilde{f}_\varepsilon^*(u)|^2}\,\mathrm{d}u-\frac{1}{12}\widehat{\mathrm{pen}}(m,k)\right\}_+\right]\\
&\quad+\frac{2}{2\pi}\sum_{\substack{k\geq m\\ k\in\mathcal{M}_n}}\mathbb{E}\left[\int_{A(m,k)}|\hat{f}_Y^*(u)|^2\frac{|\tilde{f}_\varepsilon^*(u)-f_\varepsilon^*(u)|^2}{|\tilde{f}_\varepsilon^*(u)|^4}\mathbb{1}\left\{|\tilde{f}_\varepsilon^*(u)|>|f_\varepsilon^*(u)|\right\}\mathrm{d}u\,\mathbb{1}\left\{C(m,k)^c\right\}\right].
\end{aligned}
$$

We can notice that on $C(m,k)^c$ defined by (40) following Lemma 6.3 we have

$$\mathbb{P}\left[C(m,k)^c\right]\leq M^{-3}(k-m)^{-3}$$

and we get

$$
\begin{aligned}
&\sum_{\substack{k\geq m\\ k\in\mathcal{M}_n}}\mathbb{E}\left[\int_{A(m,k)}|\hat{f}_Y^*(u)|^2\frac{|\tilde{f}_\varepsilon^*(u)-f_\varepsilon^*(u)|^2}{|\tilde{f}_\varepsilon^*(u)|^4}\mathbb{1}\left\{|\tilde{f}_\varepsilon^*(u)|\leq|f_\varepsilon^*(u)|\right\}\mathrm{d}u\,\mathbb{1}\left\{C(m,k)^c\right\}\right]\\
&\leq\sum_{\substack{k\geq m\\ k\in\mathcal{M}_n}}\mathbb{E}\left[\int_{A(m,k)}|\hat{f}_Y^*(u)|^2\frac{4|f_\varepsilon^*(u)|^2}{k_M^4(u)}\mathbb{1}\left\{|\tilde{f}_\varepsilon^*(u)|\leq|f_\varepsilon^*(u)|\right\}\mathrm{d}u\,\mathbb{1}\left\{C(m,k)^c\right\}\right]\\
&\leq\sum_{\substack{k\geq m\\ k\in\mathcal{M}_n}}\mathbb{E}\left[\int_{A(m,k)}4\kappa^{-4}(\log M)^{-2}w(u)^4M^2\,\mathrm{d}u\,\mathbb{1}\left\{C(m,k)^c\right\}\right]\\
&\leq\sum_{\substack{k\geq m\\ k\in\mathcal{M}_n}}4\kappa^{-4}(\log M)^{-2}M^2(k-m)\mathbb{P}\left[C(m,k)^c\right]\\
&\leq\sum_{\substack{k\geq m\\ k\in\mathcal{M}_n}}4\kappa^{-4}(\log M)^{-2}M^2(k-m)M^{-3}(k-m)^{-3}\\
&\leq\sum_{\substack{k\geq m\\ k\in\mathcal{M}_n}}4\kappa^{-4}M^{-1}(k-m)^{-2}\\
&\leq\frac{C}{M}
\end{aligned}
$$

and applying Lemma 6.5, we have

$$\mathbb{E}\left[\sup_{\substack{k\geq m\\ k\in\mathcal{M}_n}}\left\{\left\|\hat{f}_k-\hat{f}_m\right\|^2-6\left\|f_k-f_m\right\|^2-\frac{1}{2}\widehat{\mathrm{pen}}(m,k)\right\}_+\right]\leq\frac{C}{n}+\frac{C}{M}$$

or

$$\mathbb{E}\left[\sup_{\substack{k\geq m\\ k\in\mathcal{M}_n}}\left\{\left\|\hat{f}_k-\hat{f}_m\right\|^2-6\left\|f_k-f_m\right\|^2-\frac{1}{2}\widehat{\mathrm{pen}}(m,k)\right\}_+\right]\leq\frac{C}{n+M}+\frac{C}{M}$$

This completes the proof.

□

ACKNOWLEDGEMENTS

The authors would like to thank F.Comte for her advice, suggestions and her understanding support during this work.

A Appendix

First we remind, for the readers convenience, some well known results.

Lemma A.1. *Let $X_1,\ldots,X_n$ be i.i.d. random variables with $\mathbb{V}\mathrm{ar}[X_1]\leq v^2$ and suppose that almost surely $\|X_1\|_\infty\leq B$. Let $S_n=1/n\sum_{j=1}^n(X_j-\mathbb{E}[X_1])$. Let $\mathbb{E}|S_n|\leq H$. Then*

$$\mathbb{E}\left[\left\{|S_n|^2-H^2\right\}_+\right]\leq 32\frac{v^2}{n}\exp\left(-n\frac{H^2}{8v^2}\right)+128\sqrt{2}\frac{B^2}{n^2}\exp\left(-n\frac{H}{\frac{16\sqrt{2}}{3}B}\right).$$

Lemma A.2. *(Talagrand's inequality). Let I be some countable index set. For each $i\in I$, let $X_1^{(i)},\ldots,X_n^{(i)}$ be centered i.i.d. random variables, defined on the same probability space, with $\|X_1^{(i)}\|\leq B$ for some $B<\infty$. Let $v^2:=\sup_{i\in I}\mathbb{V}\mathrm{ar}(X_1)$. Then for arbitrary $\epsilon>0$, there are positive constants c_1 and $c_2=c_2(\epsilon)$ depending only on ϵ such that for any $\kappa>0$:*

$$\mathbb{P}\left[\left\{\sup_{i\in I}|S_n^{(i)}|\leq(1+\epsilon)\mathbb{E}\left[\sup_{i\in I}|S_n^{(i)}|\right]+\kappa\right\}\right]\leq 2\exp\left(-n\left(\frac{\kappa^2}{c_1v^2}\wedge\frac{\kappa}{c_2B}\right)\right).$$

A proof can be found, for example, on page 170 in Massart (2003).

Next we give some technical results which will be essential for the proofs.

Lemma A.3. *In the definition of $\tilde{f}^*_{\varepsilon,\mathrm{NS}}$ and $\tilde{f}^*_{\varepsilon,\mathrm{RD}}$, assume $\kappa>\sqrt{c_1p}$. Let $\tau\geq 2\kappa$ and $x\geq 1$. Then for some positive constant C*

$$\mathbb{P}\left[\exists u\in\mathbb{R}:|\hat{f}^*_{\varepsilon,\mathrm{NS}}(u)-f^*_\varepsilon(u)|>\tau\left(\log(Mx)\right)^{1/2}w(u)^{-1}M^{-1/2}\right]\leq Cx^{-p}M^{-p},$$
$$\mathbb{P}\left[\exists u\in\mathbb{R}:|\hat{f}^*_{\varepsilon,\mathrm{RD}}(u)^2-f^*_\varepsilon(u)^2|>\tau\left(\log(Mx)\right)^{1/2}w(u)^{-1}M^{-1/2}\right]\leq Cx^{-p}M^{-p}.$$

See Lemma 5.5 in Kappus (2014) for the proof.

Lemma A.4. *In the definition of* $\tilde{f}^*_{\varepsilon,\mathrm{NS}}$ *and* $\tilde{f}^*_{\varepsilon,\mathrm{RD}}$, *assume* $\kappa > 2\sqrt{2c_1 p}$. *Let*

$$A_p(u) := \left| \frac{1}{\tilde{f}^*_{\varepsilon,\mathrm{NS}}(u)} - \frac{1}{f^*_\varepsilon(u)} \right|^{2p} \Bigg/ \min\left(\frac{1}{f^*_\varepsilon(u)^{2p}}, \frac{k_M(u)^{2p}}{f^*_\varepsilon(u)^{4p}} \right),$$

$$B_p(u) := \left| \frac{1}{\tilde{f}^*_{\varepsilon,\mathrm{RD}}(u)^2} - \frac{1}{f^*_\varepsilon(u)^2} \right|^{2p} \Bigg/ \min\left(\frac{1}{f^*_\varepsilon(u)^{4p}}, \frac{k_M(u)^{2p}}{f^*_\varepsilon(u)^{8p}} \right).$$

Then for some $C > 0$

$$\mathbb{E}\left[\sup_{u\in\mathbb{R}} A_p(u)\right] \le C \quad \text{and} \quad \mathbb{E}\left[\sup_{u\in\mathbb{R}} B_p(u)\right] \le C.$$

See Lemma 4.2 in Kappus (2014) for the proof.

References

Birgé, L. (1999). An alternative point of view on Lepski's method. In *State of the art in probability and statistics: Festschrift for Willem R. van Zwet*, pages 113–133. IMS Lecture Notes-Monograph Series.

Birgé, L. and Massart, P. (1997). From model selection to adaptive estimation. In *Festschrift for Lucien Le Cam*, pages 55–87. Springer.

Bonhomme, S. and Robin, J.-M. (2010). Generalized non-parametric deconvolution with an application to earnings dynamics. *Review of Economic Studies*, 77(2):491–533.

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

Butucea, C. and Comte, F. (2009). Adaptive estimation of linear functionals in the convolution model and applications. *Bernoulli*, 15(1):69–98.

Butucea, C. and Tsybakov, A. (2008a). Sharp optimality in density deconvolution with dominating bias I. *Theory Proba. Appl.*, 52(1):24–39.

Butucea, C. and Tsybakov, A. (2008b). Sharp optimality in density deconvolution with dominating bias II. *Theory Proba. Appl.*, 52(2):237–249.

Carroll, R. J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *Journal of the American Statistical Association*, 83(404):1184–1186.

Comte, F. and Lacour, C. (2011). Data-driven density estimation in the presence of additive noise with unknown distribution. *Journal of the Royal Statistical Society: Series B*, 73:601–627.

Comte, F., Rozenholc, Y., and Taupin, M.-L. (2006). Penalized contrast estimator for adaptive density deconvolution. *Canadian Journal of Statistics*, 3(34):431–452.

Comte, F., Rozenholc, Y., and Taupin, M.-L. (2007). Finite sample penalization in adaptive density deconvolution. *Journal of Statistical Computation and Simulation*, 77(11):977–1000.

Comte, F. and Samson, A. (2012). Nonparametric estimation of random-effects densities in linear mixed-effects model. *Journal of Nonparametric Statistics*, 24(4):951–975.

Comte, F., Samson, A., and Stirnemann, J. (2013). Deconvolution estimation of onset of pregnancy with replicate observations. To appear in Scandinavian Journal of Statistics.

Dattner, I., Reiß, M., and Trabs, M. (2013). Adaptive quantile estimation in deconvolution with unknown error distribution. arXiv:1303.1698.

Delaigle, A. and Gijbels, I. (2004). Bootstrap bandwidth selection in kernel density estimation from a contaminated sample. *Annals of the Institute of Statistical Mathematics*, 56(1):19–47.

Delaigle, A., Hall, P., and Meister, A. (2008). On deconvolution with repeated measurements. *The Annals of Statistics*, 36(2):665–685.

Dion, C. (2013). New strategies for nonparametric estimation in a linear mixed model. *Work in progress.*

Efromovich, S. (1997). Density estimation for the case of supersmooth measurement errors. *Journal of the American Statistical Association*, 92:526–535.

Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *The Annals of Statistics*, 19(3):1257–1272.

Johannes, J. (2009). Deconvolution with unknown error distribution. *The Annals of Statistics*, 37(5a):2301–2323.

Johannes, J. and Schwarz, M. (2012). Adaptive circular deconvolution by model selection under unknown error distribution. arXiv:0912.1207v2. To appear in Bernoulli.

Kappus, J. (2014). Adaptive nonparametric estimation for Lévy processes observed at low frequency. *Stochastic Processes and their Applications*, 124:730–758.

Lacour, C. (2006). Rates of convergence for nonparametric deconvolution. *C. R. Acad. Sci. Paris*, Ser. I 342.

Li, T. and Vuong, Q. (1998). Nonparametric estimation of the measurement error model using multiple indicators. *Journal of Multivariate Analysis*, 65:139–165.

Massart, P. (2003). *Concentration Inequalities and Model selection.* Number 1896 in Lecture Notes in Mathematics. Springer.

Meister, A. (2009). *Deconvolution Problems in Nonparametric Statistics.* Number 193 in Lecture Notes in Statistics. Springer.

Neumann, M. (2007). Deconvolution from panel data with unknown error distribution. *Journal of Multivariate Analysis*, 98(10):1955–1968.

Neumann, M. and Reiß, M. (2009). Nonparametric estimation for Lévy processes from low frequency observations. *Bernoulli*, 15(1):223–248.

Neumann, M. H. (1997). On the effect of estimating the error density in nonparametric deconvolution. *Journal of Nonparametric Statistics*, 7(4):307–330.

Pensky, M. and Vidakovic, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *The Annals of Statistics*, 27(6):2033–2053.

Stefanski, L. (1990). Rates of convergence of some estimators in a class of deconvolution problems. *Statistics and Probability Letters*, 9:229–235.

Stefanski, S. and Carroll, R. (1990). Deconvoluting kernel density estimators. *Statistics*, 21:169–184.