ISSN 1066-5307, Mathematical Methods of Statistics, 2019, Vol. 28, No. 3, pp. 208–227. © Allerton Press, Inc., 2019.

Density Deconvolution with Small Berkson Errors

R. Rimal[1*] and M. Pensky[1]**

[1]*Dept. Math., Univ. Central Florida, Orlando, FL, USA*

Received December 20, 2018; in final form, August 17, 2019

Abstract—The present paper studies density deconvolution in the presence of small Berkson errors, in particular, when the variances of the errors tend to zero as the sample size grows. It is known that when the Berkson errors are present, in some cases, the unknown density estimator can be obtained by simple averaging without using kernels. However, this may not be the case when Berkson errors are asymptotically small. By treating the former case as a kernel estimator with the zero bandwidth, we obtain the optimal expressions for the bandwidth. We show that the density of Berkson errors acts as a regularizer, so that the kernel estimator is unnecessary when the variance of Berkson errors lies above some threshold that depends on the shapes of the densities in the model and the number of observations.

Keywords: density deconvolution, Berkson errors, bandwidth.

AMS 2010 Subject Classification: 62G07, 62G20.

DOI: 10.3103/S1066530719030025

1. INTRODUCTION

In many real life problems one is interested in the distribution of a certain variable which can be observed only indirectly. Mathematically, this leads to a density deconvolution problem where one needs to estimate the pdf of a variable X on the basis of observations of a surrogate variable $Y = X + \xi$ where the pdf f_ξ of ξ is known. The real life applications of this model arise in econometrics, astronomy, biometrics, medical statistics, image reconstruction (see, e.g., [2, 18], and also [3, 17] and the references therein). Density deconvolution problem was extensively studied in the last thirty years (see, e.g., [4, 5, 12, 14] among others and [17] and the references therein).

However, Berkson [1] argued that in many situations it is more appropriate to treat the true unobserved variable as being contaminated with an error itself and search for the distribution of $W = X + \eta$, where η is the so-called Berkson error with a known pdf f_η. Here, X, ξ and η are assumed to be independent. The objective is to estimate the pdf f_W of W on the basis of i.i.d. observations

$$Y_i = X_i + \xi_i, \;\; i = 1, \cdots, n, \tag{1.1}$$

where X_i and ξ_i are i.i.d. with, respectively, the pdfs f_X which is unknown and f_ξ which is known. The density f_ξ is called the error (or the blurring) density.

Estimation with Berkson errors occurs in a variety of statistics fields such as analysis of chemicals, nutritional, economics or astronomical data (see, e.g., [13, 16, 18, 20, 22] among others). For example, in occupational medicine, an important problem is the assessment of the health hazard of specific harmful substances in a working area. A modeling approach usually assumes that there is a threshold concentration, called the threshold limiting value (TLV), under which there is no risk due to the substance. Estimating the TLV is of particular interest in the industrial workplace. The classical errors in this model come from the measures of dust concentration in factories, while the Berkson errors come from the usual occupational epidemiology construct, wherein no direct measures of dust exposure are taken on individuals, but instead plant records of where they worked and for how long are used to impute some version of dust exposure (see [3]). In economics, the household income is usually not precisely

*E-mail: ramchandra@knights.ucf.edu
**E-mail: Marianna.Pensky@ucf.edu

collected due to the survey design or data sensitivity. It was described by Kim et al. [13] (see also [10]) that when the income data were collected by asking individuals which salary range categories they belong to, then the midpoint of the range interval was used in analysis. In this case, it is wise to assume that the true income fluctuates around the midpoint observation subject to errors.

Estimation with Berkson errors was studied by Carroll et al. [4], Delaigle [6, 7], Du et al. [9], Geng and Koul [10], Wang [20, 21] among others. It is well known that the presence of Berkson errors improves precision of estimation of the density function f_W in comparison to the case of $\eta = 0$. For example, Delaigle [6, 7] who studied estimation with Berkson errors noted that in the cases when the pdf f_η of Berkson errors has higher degree of smoothness than the error density f_ξ, one can obtain estimators of f_W with the parametric convergence rate.

However, in some practical situations, the Berkson errors are small. Hence the question arises whether small Berkson errors improve the estimation accuracy and how much. A similar inquiry has been recently carried out by Long et al. [16] who considered a somewhat different setting. In particular, they studied a p-dimensional version of the problem where variable X is directly observed and the objective is estimation of the pdf f_W of $W = X + \eta$ on the basis of observations $X_1, \ldots, X_n$, where the pdf f_η of η is known and variable η is small. In this formulation, the pdf f_W can be written as

$$f_W(x) = \int_{\mathbb{R}^p} f_X(x - z) f_\eta(z)\, dz$$

and can be estimated by

$$\hat{f}_W(x) = n^{-1} \sum_{i=1}^{n} f_\eta(x - X_i) \tag{1.2}$$

with the parametric error rate of Cn^{-1}. However, if $\mathrm{Var}(\eta) = \sigma^2$ is small, this rate becomes $C(\sigma)n^{-1}$ where $C(\sigma) \to \infty$ when $\sigma \to 0$, so the error of the estimator (1.2) may be very high.

To resolve this difficulty, in addition to estimator (1.2), Long et al. [16] proposed two alternative kernel estimators where the bandwidths of the kernels are chosen as $h = h_W$ or $h = h_X$, so to minimize the error of the estimator of f_W in the first case and the error of the estimator of f_X in the second case. Subsequently, the authors studied all three estimators by simulations and concluded that overall the kernel estimator with $h = h_W$ outperforms the remaining two. When the error variance σ is small, the estimator (1.2) leads to sub-optimal error rates. On the other hand, the choice of $h = h_X$ leads to oversmoothing, especially when the error variance is large. The authors do not provide a comprehensive theoretical study of the bandwidth selection in a general case. In particular, their rule-of-thumb recipe is based on the case where f_X is a Gaussian density. Furthermore, Long et al. [16] did not investigate when estimator (1.2) that corresponds to the bandwidth $h = 0$ is preferable and suggested that it is always suboptimal.

The objective of the present paper is to study the situation where both the blurring and the Berkson errors are present and, in addition, the Berkson errors η_i, $i = 1, \cdots, n$, are small. To quantify this phenomenon, we assume that the pdf f_η is of the form

$$f_\eta(x) = \sigma^{-1} g(\sigma^{-1} x), \tag{1.3}$$

where σ is small, specifically, $\sigma = \sigma_n \to 0$ as $n \to \infty$, while the variable X has a non-asymptotic scale. Specifically, we shall provide a full theoretical study of the bandwidth selection in a density deconvolution with small Berkson errors.

The setting of Long et al. [16] corresponds to the multivariate version of the problem in this paper where $\xi_i = 0$ and $f_\xi^* = 1$. We provide full theoretical treatment of the problem. In particular, we prove that one should always choose the bandwidth to minimize the error of the estimator of f_W, but in some cases this optimal bandwidth can be zero if σ lies above some threshold that depends on the shapes of the densities f_ξ, f_X and g and the sample size. In the particular case studied in [16], the latter situation would lead to the estimator of the form (1.2).

Since the setting (1.3) leads to three asymptotic parameters, n, σ and h, in order to keep the paper clear and readable, we consider a one-dimensional version of the problem. Extensions of our results to the situation of multivariate densities is a matter of future work.

In what follows, we are using the following notation. For any function f, f^* denotes its Fourier transform defined by $f^*(x) = \int_{-\infty}^{\infty} e^{ixt} f(t)\,dt$. If f is a pdf, then f^* is the characteristic function of f. We use the symbol C for a generic positive constant, which takes different values at different places and is independent of n. Also, for any positive functions $a(n)$ and $b(n)$, we write $a(n) \asymp b(n)$ if the ratio $a(n)/b(n)$ is bounded above and below by finite positive constants independent of n, and $a(n) \lesssim b(n)$ if the ratio $a(n)/b(n)$ is bounded above by finite positive constants independent of n.

The rest of the paper is organized as follows. Section 2 presents an estimator of f_W in the case of small Berkson errors. Section 3 provides an expression for the error of this estimator and also derives the optimal value of the bandwidth that depends on the shapes of the densities in the model and on the values of parameters n and σ. For some combinations of parameters, the optimal value of the bandwidth cannot be used since it depends on the unknown smoothness of the density f_X. Hence, in Section 4 we present construction of adaptive estimators using modification of the Lepski method. Section 5 is devoted to the discussion of the results of the paper. The proofs of all statements can be found in Section 6.

2. CONSTRUCTION OF THE DECONVOLUTION ESTIMATOR

Since (1.1) and $W = X + \eta$ imply that

$$f_Y^*(w) = f_X^*(w) f_\xi^*(w), \quad f_W^*(w) = f_X^*(w) f_\eta^*(w) \tag{2.1}$$

and also, due to (1.3), $f_\eta^*(w) = g^*(\sigma w)$, one obtains

$$f_W^*(w) = f_X^*(w) g^*(\sigma w) = \frac{f_Y^*(w) g^*(\sigma w)}{f_\xi^*(w)}.$$

Note that the unbiased estimator of $f_Y^*(w)$ is given by the empirical characteristic function

$$\hat{f}_Y^*(w) = n^{-1} \sum_{j=1}^{n} \exp(iwY_j). \tag{2.2}$$

If $g^*(\sigma w)/f_\xi^*(w)$ is square integrable, i.e.,

$$\rho^2(\sigma) = \int_{-\infty}^{\infty} \left| \frac{g^*(\sigma w)}{f_\xi^*(w)} \right|^2 dw < \infty, \tag{2.3}$$

then the inverse Fourier transform of $f_Y^*(w) g^*(\sigma w)/f_\xi^*(w)$ exists and $f_W(x)$ can be estimated by

$$\hat{f}_W(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-iwx)\, \frac{\hat{f}_Y^*(w) g^*(\sigma w)}{f_\xi^*(w)}\, dw. \tag{2.4}$$

If $g^*(\sigma w)/f_\xi^*(w)$ is not square integrable, one needs to obtain a kernel estimator of f_W. Construct approximations $f_{W,h}$ and $f_{W,h}^*$ of f_W and f_W^*, respectively,

$$f_{W,h}(x) = \int_{-\infty}^{\infty} \frac{1}{h} K\Big(\frac{x-w}{h}\Big) f_W(w)\, dw, \quad f_{W,h}^*(s) = K^*(sh) \frac{f_Y^*(s) g^*(\sigma s)}{f_\xi^*(s)} \tag{2.5}$$

and arrive at the estimator $\hat{f}_{W,h}^*(s)$ of $f_{W,h}^*(s)$ of the form

$$\hat{f}_{W,h}^*(s) = K^*(sh) \hat{f}_Y^*(s) g^*(\sigma s)/f_\xi^*(s)$$

where $\hat{f}_Y^*$ is defined in (2.2).

Consider the kernel function $K(x) = \sin(x)/(\pi x)$, so that $K^*(s) = I(|s| \le 1)$, where $I(A)$ denotes the indicator function of a set A. Since $K^*(s)$ is bounded and compactly supported, the inverse Fourier transform of $\hat{f}_{W,h}^*$ always exists and

$$\hat{f}_{W,h}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-ixs)\, \frac{\hat{f}_Y^*(s) K^*(sh) g^*(\sigma s)}{f_\xi^*(s)}\, ds. \tag{2.6}$$

We set $\hat{f}_{W,0}(x) \equiv \hat{f}_W(x)$.

In order to obtain an expression for the bandwidth h we introduce the following assumptions:

(A1). There exist positive numbers c_ξ and C_ξ and nonnegative numbers a, b, and d such that for any s

$$c_\xi(s^2+1)^{-\frac{a}{2}}\exp(-d|s|^b) \le |f_\xi^*(s)| \le C_\xi(s^2+1)^{-\frac{a}{2}}\exp(-d|s|^b), \tag{2.7}$$

where $b = 0$ iff $d = 0$ and $a > 0$ whenever $d = 0$.

(A2). There exist positive numbers c_g and C_g and nonnegative numbers ϑ, β, and γ such that for any s

$$c_g(s^2+1)^{-\frac{\vartheta}{2}}\exp(-\gamma|s|^\beta) \le |g^*(s)| \le C_g(s^2+1)^{-\frac{\vartheta}{2}}\exp(-\gamma|s|^\beta), \tag{2.8}$$

where $\beta = 0$ iff $\gamma = 0$ and $\vartheta > 0$ whenever $\gamma = 0$.

(A3). $f_X(s)$ belongs to the Sobolev ball

$$\mathcal{S}(k,B) = \Big\{f\colon \int_{-\infty}^{\infty} |f_X^*(s)|^2(s^2+1)^k\,ds \le B^2, k \ge 1/2\Big\}. \tag{2.9}$$

Also, since density deconvolution with Berkson errors of relatively large size has been fairly well studied, below we only study the case where σ is small, in particular, if $\gamma > 0$, $d > 0$, one has

$$\sigma < 0.5\,(d/\gamma)^{1/b}. \tag{2.10}$$

3. ESTIMATION ERROR

Table 1. The asymptotic expressions for $\Delta_2 \equiv \Delta_2(\sigma, h)$

Case	Δ_2
(I) $b = \beta = 0, \vartheta > a + \frac{1}{2}$,	$\min(h^{-(2a+1)}, \sigma^{-(2a+1)})$
(II) $b = \beta = 0, \vartheta = a + \frac{1}{2}$	$\min(h^{-(2a+1)}, \sigma^{-(2a+1)}) \max\big\{\log\big(\frac{\sigma}{h}\big), 1\big\}$
(III) $b = \beta = 0, \vartheta < a + \frac{1}{2}$,	$h^{-(2a+1)}\ \min\big\{\big(\frac{h}{\sigma}\big)^{2\vartheta}, 1\big\}$
(IV) $b = 0, \beta > 0$	$\min(h^{-(2a+1)}, \sigma^{-(2a+1)})$
(V) $\beta > b > 0, h > \big(\frac{\gamma\beta}{db}\sigma^\beta\big)^{\frac{1}{\beta-b}}$ $\beta > b > 0, h < \big(\frac{\gamma\beta}{db}\sigma^\beta\big)^{\frac{1}{\beta-b}}$	$h^{-(2a+1)+b}\exp(2dh^{-b}) \min\big\{\big(\frac{h}{\sigma}\big)^{2\vartheta}, 1\big\}$ $\times \exp\big(\kappa\sigma^{-\frac{\beta b}{\beta-b}}\big)\sigma^{\frac{\beta}{\beta-b}\cdot\frac{b-2}{2}-2\vartheta}$
(VI) $b = \beta > 0$	$h^{-(2a+1)+b}\exp(2h^{-b}(d-\gamma\sigma^b)) \min\big\{\big(\frac{h}{\sigma}\big)^{2\vartheta}, 1\big\}$
(VII) $b > 0, \beta = 0$	$h^{-(2a+1)+b}\exp(2dh^{-b}) \min\big\{\big(\frac{h}{\sigma}\big)^{2\vartheta}, 1\big\}$
(VIII) $b > \beta > 0$	$h^{-(2a+1)+b}\exp(2dh^{-b}) \min\big\{\big(\frac{h}{\sigma}\big)^{2\vartheta}, 1\big\}$

We characterize the accuracy of the estimator $\hat{f}_{W,h}$ of f_W by its Mean Integrated Squared Error (MISE)

$$\mathrm{MISE}(\hat{f}_{W,h}, f_W) = \mathbb{E}\int_{-\infty}^{\infty} |\hat{f}_{W,h}(x) - f_W(x)|^2\, dx.$$

Since, under Assumptions (2.7)–(2.9), both $\hat{f}^*_{W,h}$ and f^*_W are square integrable, by the Plancherel theorem, derive that

$$\mathrm{MISE}(\hat{f}_{W,h}, f_W) = \frac{1}{2\pi}\,\mathbb{E}\int_{-\infty}^{\infty} \frac{|g^*(\sigma s)|^2}{|f^*_\xi(s)|^2}\, |K^*(sh)\hat{f}^*_Y(s) - f^*_Y(s)|^2\, ds.$$

Therefore

$$\mathrm{MISE}(\hat{f}_{W,h}, f_W) = R_1(\hat{f}_{W,h}, f_W) + n^{-1} R_2(\hat{f}_{W,h}, f_W), \tag{3.1}$$

where

$$R_1(\hat{f}_{W,h}, f_W) = \|\mathbb{E}\hat{f}_{W,h} - f_W\|^2 = \frac{1}{2\pi}\int_{-\infty}^{\infty} |g^*(\sigma s)|^2\, |f^*_X(s)|^2\, I(|s| > h^{-1})\, ds \tag{3.2}$$

is the integrated squared bias of the estimator $\hat{f}_{W,h}$ and

$$R_2(\hat{f}_{W,h}, f_W) = n\,\mathbb{E}\|\hat{f}_{W,h} - \mathbb{E}\hat{f}_{W,h}\|^2 \le I(\sigma, h), \tag{3.3}$$

where

$$I(\sigma, h) = \frac{1}{2\pi}\int_{-1/h}^{1/h} \frac{|g^*(\sigma s)|^2}{|f^*_\xi(s)|^2}\, ds. \tag{3.4}$$

We shall be interested in the maximum value of $\mathrm{MISE}(\hat{f}_{W,h}, f_W)$ over all $f_X \in \mathcal{S}(k,B)$ where $\mathcal{S}(k,B)$ is defined in (2.9). In particular, we denote $\mathbb{E}\hat{f}_{W,h} = f_{W,h}$ and define

$$\Delta \equiv \Delta(n,\sigma,h) = \max_{f_X\in\mathcal{S}(k,B)} \mathrm{MISE}(\hat{f}_{W,h}, f_W) \quad \text{subject to} \quad f^*_W(w) = f^*_X(w) f^*_\eta(w). \tag{3.5}$$

It is easy to see that

$$\Delta \le \Delta_1 + n^{-1}\,\Delta_2, \tag{3.6}$$

where

$$\Delta_1 \equiv \Delta_1(n,\sigma,h) = \max_{f_X\in\mathcal{S}(k,B)} R_1(\hat{f}_{W,h}, f_W), \quad \Delta_2 \equiv \Delta_2(n,\sigma,h) = \max_{f_X\in\mathcal{S}(k,B)} R_2(\hat{f}_{W,h}, f_W). \tag{3.7}$$

Then the following statements hold.

Lemma 1. *Under the assumptions* (2.7)–(2.10)*, for* Δ_1 *in* (3.7)*, one has*

$$\Delta_1 \lesssim \begin{cases} \sigma^{-2\vartheta} h^{2\vartheta+2k} \exp\big(-2\gamma(\sigma/h)^\beta\big) & \text{if } h < \sigma, \\ h^{2k} & \text{if } h \ge \sigma. \end{cases} \tag{3.8}$$

Lemma 2. *If* $\beta > b > 0$*, denote*

$$\kappa = \Big(\frac{db}{\gamma\beta}\Big)^{\frac{b}{\beta-b}} \Big[\frac{d(\beta-b)}{b}\Big] > 0. \tag{3.9}$$

Then, under the assumptions (2.7)–(2.10)*, the expressions for* Δ_2 *defined in* (3.7)*, are given in Table* 1.

Table 2. The optimal values h_{opt} of the bandwidth h and the corresponding expressions for $\Delta(n,\sigma,h)$ defined in (3.6). Here, μ_1 and μ_2 are given by (3.10)

Case	$\Delta(n,\sigma,h_{opt})$	condition	h_{opt}
(I) $b=\beta=0$,	$n^{-1}\sigma^{-(2a+1)}$	$\sigma > n^{-\frac{1}{2k+2a+1}}$	0
$\vartheta > a+\frac{1}{2}$	$n^{-\frac{2k}{2k+2a+1}}$	$\sigma \le n^{-\frac{1}{2k+2a+1}}$	$n^{-\frac{1}{2k+2a+1}}$
(II) $b=\beta=0$	$n^{-1}\sigma^{-(2a+1)}\log n$	$\sigma > n^{-\frac{1}{2k+2a+1}}$	$n^{-\frac{1}{2k+2a+1}}$
$\vartheta = a+\frac{1}{2}$	$n^{-\frac{2k}{2k+2a+1}}$	$\sigma \le n^{-\frac{1}{2k+2a+1}}$	$n^{-\frac{1}{2k+2a+1}}$
(III) $b=\beta=0$,	$\sigma^{-2\vartheta}n^{-\frac{2\vartheta+2k}{2k+2a+1}}$	$\sigma > n^{-\frac{1}{2k+2a+1}}$	$n^{-\frac{1}{2k+2a+1}}$
$\vartheta < a+\frac{1}{2}$	$n^{-\frac{2k}{2k+2a+1}}$	$\sigma \le n^{-\frac{1}{2k+2a+1}}$	$n^{-\frac{1}{2k+2a+1}}$
(IV) $b=0, \beta>0$	$n^{-1}\sigma^{-(2a+1)}$	$\sigma > n^{-\frac{1}{2k+2a+1}}$	0
	$n^{-\frac{2k}{2k+2a+1}}$	$\sigma \le n^{-\frac{1}{2k+2a+1}}$	$n^{-\frac{1}{2k+2a+1}}$
(V) $\beta > b > 0$	$n^{-1}\exp\left(\kappa\sigma^{\frac{-\beta b}{\beta-b}}\right)\sigma^{\frac{\beta(b-2)}{2(\beta-b)}-2\vartheta}$	$\sigma > \mu_1$	0
	$(\log n)^{-\frac{2k}{b}}$	$\sigma \le \mu_1$	μ_1
(VI) $b=\beta>0$	$\sigma^{-2\vartheta}(\log n)^{-\frac{2\vartheta+2k}{b}}\exp\left(-2\gamma\sigma^{\beta}(\log n)^{\frac{\beta}{b}}\right)$	$\sigma > \mu_1$	μ_1
	$(\log n)^{-\frac{2k}{b}}$	$\sigma \le \mu_1$	μ_2
(VII) $b>0, \beta=0$	$(\log n)^{-\frac{(2\vartheta+2k)}{b}}\sigma^{-2\vartheta}$	$\sigma > \mu_1$	μ_1
	$(\log n)^{-\frac{2k}{b}}$	$\sigma \le \mu_1$	μ_1
(VIII) $b>\beta>0$	$\sigma^{-2\vartheta}(\log n)^{\frac{(1+2a-2\vartheta)}{b}-1}$	$\sigma > \mu_1$	μ_1
	$(\log n)^{-\frac{2k}{b}}$	$\sigma \le \mu_1$	μ_1

Observe that in every case, the expression for the variance depends not only on the values of h, σ and n but also on their mutual relationship. Also, the bias term $\Delta_1(\sigma,h)$ is an increasing function of h while the variance term $\Delta_2(\sigma,h)$ is a decreasing function of h, so the optimal value $h=h_{\rm opt}$ is such that $\Delta_1(\sigma,h)\asymp n^{-1}\,\Delta_2(\sigma,h)$. Theorem 1 below presents the optimal expressions h_{opt} for the bandwidth h as well as the corresponding values for the risk $\Delta(n,\sigma,h_{opt})$, where $\Delta(n,\sigma,h)$ is defined in (3.5).

Theorem 1. *Let conditions* (2.7)–(2.10) *hold. Then, the asymptotic values of*

$$h_{opt} = \arg\min_h[\Delta(n,\sigma,h)]$$

and also of $\Delta(n,\sigma,h_{opt})$ *are provided in Table* 2. *Here,*

$$\begin{aligned} \mu_1 = \mu_1(n) \; &= \; \left[\frac{1}{2d}\left(\log n + \left(\frac{b-2a-1}{b}\right)\log\log n\right)\right]^{-\frac{1}{b}}, \\ \mu_2 = \mu_2(n) \; &= \; \left[\frac{1}{2(d-\gamma\sigma^b)}\left(\log n + \left(\frac{b-2a-1}{b}\right)\log\log n\right)\right]^{-\frac{1}{b}}. \end{aligned} \tag{3.10}$$

4. ADAPTIVE ESTIMATION USING LEPSKI'S METHOD

Note that although Theorem 1 provides the optimal values for the bandwidth and the corresponding convergence rates, in practice, we can use those values only in cases V–VIII, since in cases I–IV the value of the optimal bandwidth h_{opt} depends on the smoothness parameter k of the unknown density f_X. Moreover, in cases I and IV the optimal bandwidth is zero if $\sigma > n^{-\frac{1}{2k+2a+1}}$, where the threshold value

$n^{-\frac{1}{2k+2a+1}}$ itself depends on the unknown value of k. In order to resolve this difficulty, we use a novel modification of the Lepski method for construction of adaptive estimators (see, e.g., [15, 12]).

Below we consider the cases I—IV, for which the optimal value h_{opt} depends on the unknown parameter k. To start with, note that, by Lemma 1, if $h_{opt} = 0$, as it happens in the cases I and IV, one has

$$\Delta(n, \sigma, 0) \asymp \Delta(n, \sigma, n^{-1}).$$

Moreover, if $\sigma \le n^{-\frac{1}{2a+1}} < n^{-\frac{1}{2k+2a+1}}$, then $h_{opt} > 1/n$.

In order to replace the unknown value of h_{opt} by its estimated value, we use the variance term given by

$$\begin{aligned} D(n, \sigma, h) &= \|\hat{f}_{W,h}(x) - f_{W,h}(x)\|^2 = \tfrac{1}{2\pi}\, \|\hat{f}^*{}_{W,h}(x) - f^*{}_{W,h}(x)\|^2 \\ &= \tfrac{1}{2\pi} \int_{-1/h}^{1/h} \frac{|g^*(\sigma s)|^2}{|f_\xi^*(s)|^2}\, |\hat{f}_Y^*(s) - f_Y^*(s)|^2\, ds. \end{aligned}$$

If $h \ge 1/n$, then it is easy to see that

$$D(n, \sigma, h) \le \max_{|s| \le n} |\hat{f}_Y^*(s) - f_Y^*(s)|^2\, I(\sigma, h),$$

where $I(\sigma, h)$ is defined in (3.4).

Recall also that the value h_{opt} is such that it minimizes the sum of $\Delta_1(n, \sigma, h) + n^{-1}\Delta_2(n, \sigma, h)$ where, under the assumptions A1—A3, the first term is growing polynomially in h while the second is decreasing polynomially in h. Therefore

$$\Delta(n, \sigma, h_{opt}) \asymp \Delta_1(n, \sigma, h_{opt}) \asymp n^{-1}\, I(n, \sigma, h_{opt}). \tag{4.1}$$

Consider the sets

$$\mathcal{J} = \{1, 2, 3, \ldots, j_{\max}\} \quad \text{and} \quad \mathcal{H} = \{h = 2^{-j}, j \in \mathcal{J}\} \tag{4.2}$$

and denote

$$j_{\max} = \min\Big(\frac{\log n}{2a+1}, \log\Big(\frac{1}{\sigma}\Big)\Big). \tag{4.3}$$

Let $q > 0$ be such that $\mathbb{E}(|X_1|^q) \le C_q < \infty$ and

$$C(\tau, q) \ge 8\,\sqrt{2\tau(q+1) + 6q + 2}/\sqrt{q}. \tag{4.4}$$

Define a set in the sample space

$$\Omega_{\sigma,n} = \begin{cases} \{w\colon \|\hat{f}_{W,\sigma} - f_{W,\frac{1}{n}}\| \ge 4C(\tau, q)\sqrt{n^{-1}\, I(\sigma, 1/n) \log n}\} & \text{for cases I and IV,} \\ \emptyset \text{ (the empty set)} & \text{for cases II and III.} \end{cases} \tag{4.5}$$

Then the following statement holds.

Theorem 2. *Let conditions* (2.7)—(2.10) *hold with* $b = 0$ (*cases* I—IV) *and* $\tau \ge 4$. *Define*

$$\hat{h} = \begin{cases} 1/n, & \text{if } w \in \Omega_{\sigma,n} \\ \max\{h \in \mathcal{H}\colon \|\hat{f}_{W,h} - f_{W,\tilde{h}}\| \le 4C(\tau, q)\sqrt{\frac{\log n}{n} I(\sigma, \tilde{h})} \quad \text{for any } \tilde{h} \le h, \tilde{h} \in \mathcal{H}\}, & \text{if } w \notin \Omega_{\sigma,n} \end{cases}$$

Then

$$\mathbb{E}\|\hat{f}_{W,\hat{h}} - f_W\|^2 \lesssim \Delta(n, \sigma, h_{opt})\, \log n. \tag{4.6}$$

5. DISCUSSION

In the present paper, our main goal was to justify the choice of a bandwidth in deconvolution problems with small Berkson errors. To the best of our knowledge, our paper is the first paper which carries out a comprehensive theoretical study of density deconvolution with Berkson errors when Berkson errors are asymptotically small.

In particular, we refined the conclusion of Long et al. (2016) and studied the relationship between the three parameters: the bandwidth h, the sample size n and the standard deviation of the Berkson errors σ. As Theorem 1 above shows, the expressions for the optimal bandwidth are always chosen to minimize the error in the estimator of the density of interest f_W. In particular, if $h = 0$ is possible, one should choose this value as long as the Berkson errors are not too small, i.e., σ lies above some threshold level that depends on the shapes of the densities and the number of observations n.

In order to uncover the reason for this, compare expressions (2.4) and (2.6) and observe that $g^*(\sigma s)$ in (2.4) acts as a kernel function g with the bandwidth $h = \sigma$. If σ is large enough (i.e., $\sigma > h_{opt}$, where h_{opt} is the value of h that achieves the best bias-variance balance), then convolution with g leads to sufficient regularization and no kernel estimation is necessary. However, if $\sigma < h_{opt}$, then one needs additional kernel smoothing with $h > \sigma$.

The setting of [16] corresponds to cases I, II, III, and IV in Tables 1 and 2 with $a = b = 0$. If $\vartheta > 1/2$, then h_{opt} is zero if σ is large enough and h_{opt} is of the order $n^{-1/(2k+1)}$ (where k is the degree of smoothness of the density f_X of the measurements) otherwise. The choice depends on the relationship between parameters σ, n and k. Since k is unknown, we construct adaptive estimators of f_W using a novel modification of Lepski method. Indeed, one cannot use the traditional Lepski method since the value of the optimal bandwidth depends on the relationship between σ and the unknown threshold $n^{-1/(2k+2a+1)}$. Hence our paper presents a non-trivial extension of the Lepski technique.

Note that we did not consider the case of multivariate density functions. This extension is fairly straightforward but rather cumbersome. We shall leave this case for the future investigation.

6. PROOFS

6.1. Proofs of the Statements in the Paper

Proof of Lemma 1. Since for any $f_X \in S(k, B)$ one has

$$\begin{aligned}
\Delta_1 &= \max_{f_X \in \mathcal{S}(k,B)} \|\mathbb{E}\hat{f}_{W,h} - f_W\|^2 \\
&= \max_{f_X \in \mathcal{S}(k,B)} \frac{1}{2\pi} \int_{|s|>1/h} |g^*(\sigma s)|^2 |f_X^*(s)|^2 \, ds \\
&= \max_{f_X \in \mathcal{S}(k,B)} \frac{1}{\pi} \int_{\frac{1}{h}}^{\infty} |g^*(\sigma s)|^2 |f_X^*(s)|^2 \, ds \\
&\le \max_{f_X \in \mathcal{S}(k,B)} \frac{2C_g}{\pi} \int_{\frac{1}{h}}^{\infty} (\sigma^2 s^2 + 1)^{-\vartheta} \exp(-2\gamma |s|^\beta \sigma^\beta) \frac{(s^2+1)^k}{(s^2+1)^k} |f_X^*(s)|^2 \, ds \\
&\le \frac{2C_g\, B^2}{\pi} \max_{s \ge \frac{1}{h}} [(\sigma^2 s^2 + 1)^{-\vartheta} \exp(-2\gamma |s|^\beta \sigma^\beta)] \, (h^{-2} + 1)^{-k},
\end{aligned}$$

hence

$$\Delta_1 \lesssim \min\Big\{ \Big(\frac{h}{\sigma}\Big)^{2\vartheta}, 1 \Big\} h^{2k} \exp\Big(-2\gamma \Big(\frac{\sigma}{h}\Big)^\beta \Big)$$

which implies (3.8).

Proof of Lemma 2. Note that the variance term is given by

$$\Delta_2 \le \frac{1}{2\,\pi} \int_{-\infty}^{\infty} \frac{|g^*(\sigma s)|^2}{|f_\xi^*(s)|^2} I(|s| < h^{-1}) \, ds$$

$$\leq \frac{C_g}{c_\xi} \int_0^{\frac{1}{h}} (\sigma^2 s^2 + 1)^{-\vartheta} (s^2+1)^a \exp(-2\gamma |s|^\beta \sigma^\beta + 2d|s|^b)\, ds.$$

Using the change of variables $s = z/h$ we obtain

$$\Delta_2 \lesssim h^{-(2a+1)} V(\sigma, h) \quad \text{with} \quad V(\sigma,h) = \int_0^1 P(z \mid \sigma, h) \exp\{\phi(z|\sigma,h)\}\, dz, \tag{6.1}$$

where

$$\phi(z \mid \sigma, h) = 2dz^b h^{-b} - 2\gamma z^\beta \sigma^\beta h^{-\beta}, \quad P(z \mid \sigma, h) = (\sigma^2 z^2 h^{-2} + 1)^{-\vartheta} (z^2 + h^2)^a. \tag{6.2}$$

For the cases when $b = 0$ (cases I–IV), one can obtain an asymptotic expression for Δ_2 using direct calculations. If $b > 0$ and $d \geq 0$, one needs to apply Lemma 6. Denote by z_0 and z_h, respectively, the point where $\phi(z \mid \sigma, h)$ attains its global maximum on the interval $[0, 1]$ and its critical point:

$$z_0 \equiv z_0(\sigma, h) = \underset{z \in [0,1]}{\operatorname{argmax}}\, \phi(z \mid \sigma, h), \quad z_h = (db\,(\gamma\beta)^{-1}\, \sigma^{-\beta})^{\frac{1}{\beta - b}} h. \tag{6.3}$$

Since $z_h > 0$, there are two possible cases here: $z_h \in (0, 1]$ and $z_h > 1$. If $z_h \in (0, 1]$, then $z_0 = z_h$, $\phi'(z_0) = 0$ and $\phi''(z_0) < 0$. If $z_h > 1$, then $z_0 = 1$ and $\phi'(z_0) = \phi'(1) > 0$.

Hence Lemma 6 and formula (6.1) yield that, for small values of h and σ,

$$h^{2a+1}\, \Delta_2 \lesssim \begin{cases} \frac{\exp\{\phi(z_h|\sigma,h)\} P(z_h|\sigma,h)}{\sqrt{|\phi''(z_h|\sigma,h)|}} & \text{if} \quad z_0 = z_h, \\ \frac{\exp\{\phi(1|\sigma,h)\} P(1|\sigma,h)}{\phi'(1|\sigma,h)} & \text{if} \quad z_0 = 1. \end{cases} \tag{6.4}$$

Here

$$\begin{aligned} &\phi(1 \mid \sigma, h) = 2dh^{-b} - 2\gamma\sigma^\beta h^{-\beta}, \;\; \phi'(1 \mid \sigma, h) = 2(dbh^{-b} - \gamma\beta\sigma^\beta h^{-\beta}), \\ &P(1 \mid \sigma, h) \asymp (\sigma^2 h^{-2} + 1)^{-\vartheta}, \qquad P(z_h \mid \sigma, h) = (\sigma^2 z_h^2 h^{-2} + 1)^{-\vartheta} (z_h^2 + h^2)^a. \end{aligned} \tag{6.5}$$

Below we consider various cases.

Cases I, II, III: $b = \beta = 0$. Note that

$$I(\sigma, h) = \frac{2{C_g}^2}{{c_\xi}^2} \int_0^{\frac{1}{h}} (\sigma^2 s^2 + 1)^{-\vartheta} (s^2+1)^a\, ds = \frac{2{C_g}^2}{{c_\xi}^2 h} \int_0^1 (\sigma^2 z^2 h^{-2} + 1)^{-\vartheta}\, (z^2 h^{-2} + 1)^a\, dz. \tag{6.6}$$

If $h \geq \sigma$, then $\sigma^2 z^2 h^{-2} + 1 \in (1, 2)$ and $I(\sigma, h) \leq \frac{2^{1-\vartheta} {C_g}^2}{{c_\xi}^2} h^{-(2a+1)}$.

If $h < \sigma$, then, by the change of variables $\sigma s = u$ in (6.6), we obtain

$$I(\sigma, h) = \frac{2{C_g}^2}{{c_\xi}^2 \sigma} \int_0^{\frac{\sigma}{h}} (u^2+1)^{-\vartheta}\, (u^2 \sigma^{-2} + 1)^a\, du \leq \frac{2{C_g}^2 C_a}{{c_\xi}^2} \sigma^{-(2a+1)} \int_0^{\frac{\sigma}{h}} \frac{u^{2a}}{(u^2+1)^\vartheta}\, du.$$

Hence

$$I(\sigma, h) \leq \frac{2{C_g}^2 C_a}{{c_\xi}^2} \min(h^{-(2a+1)}, \sigma^{-(2a+1)}) \Delta_{h\sigma},$$

where

$$\Delta_{h,\sigma} = \begin{cases} 1 & \text{if } \vartheta > a + 1/2, \\ \max\left\{ \log\left(\frac{\sigma}{h}\right), 1 \right\} & \text{if } \vartheta = a + 1/2, \\ \max\left\{ 1, \left(\frac{\sigma}{h}\right)^{2a - 2\vartheta + 1} \right\} & \text{if } \vartheta < a + 1/2. \end{cases} \tag{6.7}$$

Case IV: $b = 0, \beta > 0$. In this case,

$$\Delta_2 \asymp h^{-1} \int_0^1 (\sigma^2 z^2 h^{-2} + 1)^{-\vartheta} (z^2 h^{-2} + 1)^a \exp(-2\gamma\sigma^\beta z^\beta h^{-\beta}) \, dz.$$

If $h > \sigma$, then the argument of the exponent is bounded above and $\Delta_2 \asymp h^{-2a-1}$. If $h < \sigma$, then by changing variables $u = 2\gamma(\sigma z/h)^\beta$ we obtain

$$\Delta_2 \asymp \sigma^{-1} \int_0^\infty \Big(\Big(\frac{u}{2\gamma}\Big)^{\frac{2}{\beta}} + 1\Big)^{-\vartheta} \Big(\frac{1}{\sigma^{2a}}\Big(\frac{u}{2\gamma}\Big)^{\frac{2a}{\beta}} + 1\Big) \exp(-u) u^{\frac{1}{\beta}-1} \, du \asymp \sigma^{-(2a+1)}.$$

Hence

$$\Delta_2 \asymp \min(h^{-(2a+1)}, \sigma^{-(2a+1)}).$$

Case V: $\beta > b > 0$. In this case $\rho^2(\sigma) = \infty$ in (2.3), so that $h > 0$. The expression for the variance is given by (6.1) with $\phi(z|\sigma, h)$ defined in (6.2). Let z_h be given by (6.3). It is easy to check that

$$z_h = (db\,(\gamma\beta)^{-1}\,\sigma^{-\beta})^{\frac{1}{\beta-b}} h \asymp \sigma^{-\frac{\beta}{\beta-b}} h. \tag{6.8}$$

It is easy to check that $\phi''(z_h \mid \sigma, h) < 0$, so that z_h is the local maximum. Now consider two cases.

(a) If $h > \big(\frac{\gamma\beta}{db}\sigma^\beta\big)^{\frac{1}{\beta-b}}$, then $z_h > 1$. Hence $\phi(z \mid \sigma, h)$ does not have a local maximum on $[0, 1]$ and it attains its global maximum at $z_0 = 1$. Then $2dh^{-b} > \phi(1 \mid \sigma, h) = 2dh^{-b} - 2\gamma\sigma^\beta h^{-\beta} > 2dh^{-b}(1 - b/\beta)$. Moreover, since $\beta > b$ and $h > \big(\frac{\gamma\beta}{db}\sigma^\beta\big)^{\frac{1}{\beta-b}} > \sigma$, one has $2dbh^{-b} > 2\gamma\beta\sigma^\beta h^{-\beta}$, which yields

$$\phi'(1 \mid \sigma, h) = 2dbh^{-b} - 2\gamma\beta\sigma^\beta h^{-\beta} = 2dbh^{-b}\Big(1 - \frac{\gamma\beta}{db}\sigma^\beta h^{b-\beta}\Big) \asymp h^{-b}.$$

Plugging these expressions into the second equation of (6.4) and using (6.5), we obtain

$$\Delta_2 \asymp h^{-(2a+1)} \min\{(h\sigma^{-1})^{2\vartheta}, 1\} \exp(2dbh^{-b})\, h^b \asymp h^{b-2a-1} \exp(2dh^{-b}).$$

(b) If $h < \big(\frac{\gamma\beta}{db}\sigma^\beta\big)^{\frac{1}{\beta-b}}$, then z_h is given by formula (6.8) and $z_0 = z_h < 1$. Hence Δ_2 is given by the first expression in formula (6.4)

$$\Delta_2 \asymp \frac{\exp(\phi(z_h \mid \sigma, h))}{\sqrt{|\phi''(z_h \mid \sigma, h)|}}\, h^{-(2a+1)} (\sigma^2 z_h^2 h^{-2} + 1)^{-\vartheta} (z_h^2 + h^2)^a. \tag{6.9}$$

Note that, due to $\beta > b > 0$, $\frac{\beta^2}{\beta-b} > \frac{\beta b}{\beta-b}$ and $\beta - \frac{\beta^2}{\beta-b} = -\frac{\beta b}{\beta-b}$, one has

$$\phi(z_h \mid \sigma, h) = \frac{2d}{h^b}\Big(\frac{db}{\gamma\beta}\sigma^{-\beta}\Big)^{\frac{b}{\beta-b}} h^b - \frac{2\gamma\sigma^\beta}{h^\beta}\Big(\frac{db}{\gamma\beta}\sigma^{-\beta}\Big)^{\frac{\beta}{\beta-b}} h^\beta = \kappa\sigma^{-\frac{\beta b}{\beta-b}},$$

where κ is a positive constant defined in (3.9). Also

$$\phi''(z_h \mid \sigma, h) = \frac{2}{z_h^2}\Big(\frac{db(b-1)z_h^b}{h^b} - \frac{\gamma\beta(\beta-1)\, z_h^\beta \sigma^\beta}{h^\beta}\Big) = \frac{2db(b-\beta)z_h^{b-2}}{h^b} \asymp \frac{z_h^{b-2}}{h^b}.$$

Then plugging $\phi(z_h \mid \sigma, h)$ and $\phi''(z_h \mid \sigma, h)$ into (6.9) we obtain

$$\Delta_2 \asymp \exp\Big(\kappa\sigma^{-\frac{\beta b}{\beta-b}}\Big)\sigma^{\frac{\beta(b-2)}{2(\beta-b)} - 2\vartheta}.$$

Case VI: $b = \beta > 0$. In this case $\rho^2(\sigma) = \infty$ in (2.3), so that $h > 0$. Moreover, since $\phi(z \mid \sigma, h) = 2z^b\, h^{-b}(d - \gamma\sigma^b)$ where, due to condition (2.10), $d - \gamma\sigma^b > 0$, $z_0 = 1$ is the non-local maximum of $\phi(z \mid \sigma, h)$. Then the second expression in formula (6.4) yields

$$\Delta_2 \lesssim \frac{\exp(\phi(1 \mid \sigma, h))}{\phi'(1 \mid \sigma, h)}\, h^{-(2a+1)} (\sigma^2 h^{-2} + 1)^{-\vartheta}. \tag{6.10}$$

Using (6.5) with $\beta = b$, we derive

$$\Delta_2 \lesssim h^{b-(2a+1)} \min\left(\left(\frac{h}{\sigma}\right)^{2\vartheta}, 1\right) \exp(2h^{-b}(d - \gamma\sigma^b)).$$

Case VII: $b > 0, \beta = \gamma = 0$. In this case, $z_0 = 1$ is the non-local maximum of $\phi(z \mid \sigma, h)$ and (6.5) yields $\phi(1|\sigma, h) = 2dh^{-b}$ and $\phi'(1 \mid \sigma, h) = 2dbh^{-b}$. Plugging those expressions into (6.10), we derive

$$\Delta_2 \lesssim \min\left(\left(\frac{h}{\sigma}\right)^{2\vartheta}, 1\right) h^{b-(2a+1)} \exp(2dh^{-b}).$$

Case VIII: $b > \beta > 0$. In this case $\rho^2(\sigma) = \infty$ in (2.3), so that $h > 0$. Also, it is easy to check that although $z_h \in (0, 1)$, one has $\phi''(z_h \mid \sigma, h) > 0$, so z_h is the local minimum. It is easy to see that $z_0 = 1$ and $\phi(1 \mid \sigma, h) = 2dh^{-b}(1 - \gamma d^{-1}\sigma^{\beta} h^{b-\beta}) \asymp 2dh^{-b}$. Moreover, $\phi'(1 \mid \sigma, h) = 2h^{-b}(db - \gamma\beta\sigma^{\beta} h^{b-\beta}) \asymp h^{-b}$, so formula (6.10) yields

$$\Delta_2 \lesssim h^{b-(2a+1)} \min\left(\left(\frac{h}{\sigma}\right)^{2\vartheta}, 1\right) \exp(2dh^{-b}).$$

Proof of Theorem 1. Consider various cases.

Cases I, II, III: $b = \beta = 0$. One has

$$\Delta \lesssim \min\{(h\,\sigma^{-1})^{2\vartheta}, 1\} h^{2k} + n^{-1} \min(h^{-(2a+1)}, \sigma^{-(2a+1)}) \Delta_{h\sigma}, \tag{6.11}$$

where $\Delta_{h,\sigma}$ is defined in (6.7).

Case I: $b = \beta = 0$, $\vartheta > a + 1/2$. In this case $\rho^2(\sigma) < \infty$ and $h = 0$ is possible. If $h = 0$, then $\Delta = O(\sigma^{-(2a+1)} n^{-1})$. If $h \neq 0$, then choose $h \geq \sigma$, so that $\Delta_1(\sigma, h) \lesssim h^{2k}$, $\Delta_2(\sigma, h) \lesssim h^{-(2a+1)}$. Then $h_{opt} \asymp n^{-\frac{1}{2k+2a+1}}$ and $\Delta_1(\sigma, h_{opt}) + n^{-1}\,\Delta_2(\sigma, h_{opt}) \lesssim n^{-\frac{2k}{2k+2a+1}}$. Choose $h = h_{opt}$ if $h_{opt} \geq \sigma$, i.e., if $n^{-\frac{1}{2k+2a+1}} \geq \sigma$. We obtain

$$\Delta \asymp \begin{cases} n^{-1}\,\sigma^{-(2a+1)}, & h_{opt} = 0 \quad\quad \text{if } \sigma > n^{-\frac{1}{2k+2a+1}}, \\ n^{-\frac{2k}{2k+2a+1}}, & h_{opt} = n^{-\frac{1}{2k+2a+1}} \quad \text{if } \sigma \leq n^{-\frac{1}{2k+2a+1}}. \end{cases}$$

Case II: $b = \beta = 0$, $\vartheta = a + \frac{1}{2}$. Here, Δ is given by (6.11), where $\Delta_{h\sigma} = \max\{\log(\sigma/h), 1\}$. If $h < \sigma$, then $\Delta \lesssim \sigma^{-2\vartheta} h^{2\vartheta+2k} + \sigma^{-(2a+1)} n^{-1} \log(\sigma/h)$. Setting $\sigma^{-2\vartheta} h^{2\vartheta+2k} = \sigma^{-(2a+1)} n^{-1} \log(\sigma/h)$ leads to

$$h_{opt} \asymp n^{-\frac{1}{2k+2a+1}}, \quad \Delta \lesssim n^{-1}\,\sigma^{-(2a+1)}\,\log n.$$

Note that $h_{opt} < \sigma$ if and only if $n^{-\frac{1}{2k+2a+1}} < \sigma$. Now, consider the case when $h \geq \sigma$.

Then by (6.11), $\Delta \lesssim n^{-\frac{2k}{2k+2a+1}}$ if $n^{-\frac{1}{2k+2a+1}} \geq \sigma$. Hence

$$\Delta \asymp \begin{cases} \frac{\sigma^{-(2a+1)}}{n} \log n, & h_{opt} = n^{-\frac{1}{2k+2a+1}} \quad \text{if } \sigma > n^{-\frac{1}{2k+2a+1}}, \\ n^{-\frac{2k}{2k+2a+1}}, & h_{opt} \asymp n^{-\frac{1}{2k+2a+1}} \quad \text{if } \sigma \leq n^{-\frac{1}{2k+2a+1}}. \end{cases}$$

Case III: $b = \beta = 0$, $\vartheta < a + \frac{1}{2}$. First, consider the case when $h < \sigma$. Then, by (6.11) and (6.7), obtain

$$\Delta \lesssim \sigma^{-2\vartheta} h^{2\vartheta+2k} + \sigma^{-(2\vartheta)} n^{-1} h^{2\vartheta-2a-1}.$$

Setting $\sigma^{-2\vartheta} h^{2\vartheta+2k} = \sigma^{-(2\vartheta)} n^{-1} h^{2\vartheta-2a-1}$, obtain $h_{opt} \asymp n^{-\frac{1}{2k+2a+1}}$ and $\Delta \lesssim \sigma^{-2\vartheta} n^{-\frac{2\vartheta+2k}{2k+2a+1}}$. Also note that $h_{opt} < \sigma$ if and only if $\sigma > n^{-\frac{1}{2k+2a+1}}$. Now, consider the case when $h \geq \sigma$. Then (6.11) and (6.7), imply that $\Delta \asymp n^{-\frac{2k}{2k+2a+1}}$ if $n^{-\frac{1}{2k+2a+1}} \geq \sigma$. Hence

$$\Delta \asymp \begin{cases} \sigma^{-2\vartheta} n^{-\frac{2\vartheta+2k}{2k+2a+1}}, & h_{opt} = n^{-\frac{1}{2k+2a+1}} \quad \text{if } \sigma > n^{-\frac{1}{2k+2a+1}}, \\ n^{-\frac{2k}{2k+2a+1}}, & h_{opt} = n^{-\frac{1}{2k+2a+1}} \quad \text{if } \sigma \leq n^{-\frac{1}{2k+2a+1}}. \end{cases}$$

Case IV: $b = 0, \beta > 0$. In this case $\rho^2(\sigma) < \infty$ and $h = 0$ is possible. Consider the case $h < \sigma$. Then

$$\Delta_1(\sigma, h) \lesssim \sigma^{(-2\vartheta)} h^{2\vartheta+2k} \exp\Big(-2\gamma\Big(\frac{\sigma}{h}\Big)^{\beta}\Big), \quad \Delta_2(\sigma, h) \lesssim \sigma^{-(2a+1)}.$$

If $h < \sigma$, then $h_{opt} = 0$ and $\Delta \asymp n^{-1}\sigma^{-(2a+1)}$. If $h > \sigma$, then $\Delta_1(\sigma, h) \leq h^{2k}$ and $\Delta_2(\sigma, h) \lesssim h^{-(2a+1)}$. Therefore, $h_{opt} \asymp n^{-\frac{1}{2k+2a+1}}$ and $\Delta \lesssim n^{\frac{-2k}{2k+2a+1}}$. Observing that $h_{opt} \geq \sigma$ if $\sigma \leq n^{-\frac{1}{2k+2a+1}}$, we obtain

$$\Delta \asymp \begin{cases} n^{-1}\sigma^{-(2a+1)} & h_{opt} = 0 \quad \text{if } \sigma > n^{-\frac{1}{2k+2a+1}}, \\ n^{-\frac{2k}{2k+2a+1}} & h_{opt} = n^{-\frac{1}{2k+2a+1}} \quad \text{if } \sigma \leq n^{-\frac{1}{2k+2a+1}}. \end{cases}$$

Case V: $\beta > b > 0$. In this case $\rho^2(\sigma) < \infty$ and $h = 0$ is possible. The bias is given by (3.8) and

$$\Delta_2 \lesssim \begin{cases} n^{-1} h^{b-2a-1} \exp(2dh^{-b}) & \text{if } h > \big(\frac{\gamma\beta}{db}\sigma^{\beta}\big)^{\frac{1}{\beta-b}}, \\ n^{-1} \exp\big(\kappa\sigma^{\frac{-\beta b}{\beta-b}}\big)\sigma^{\frac{\beta}{\beta-b}\cdot\frac{b-2}{2}-2\vartheta} & \text{if } h < \big(\frac{\gamma\beta}{db}\sigma^{\beta}\big)^{\frac{1}{\beta-b}}. \end{cases}$$

If $h = 0$, then $\Delta \asymp n^{-1}\exp\big(\kappa\sigma^{\frac{-\beta b}{\beta-b}}\big)\sigma^{\frac{\beta}{\beta-b}\cdot\frac{b-2}{2}-2\vartheta}$. If $h > 0$, then one needs $h > \sigma \gtrsim \big(\frac{\gamma\beta}{db}\sigma^{\beta}\big)^{\frac{1}{\beta-b}}$ and $\Delta \asymp h^{2k} + n^{-1}h^{b-2a-1}\exp(2dh^{-b})$. Choosing h such that $h^{2k} = n^{-1}h^{b-2a-1}\exp(2dh^{-b})$, arrive at

$$(2dh^{-b})^{\frac{2a+2k+1-b}{b}} \exp(2dh^{-b}) = (2d)^{\frac{2a+2k+1-b}{b}}\, n \tag{6.12}$$

and, by Lemma 7, obtain $h_{opt} = \mu_1(n)$, where $\mu_1(n)$ is defined in (3.10), and, hence, $\Delta \asymp (\log n)^{-\frac{2k}{b}}$. Therefore

$$\Delta \asymp \begin{cases} n^{-1}\exp\big(\kappa\sigma^{\frac{-\beta b}{\beta-b}}\big)\sigma^{\frac{\beta(b-2)}{2(\beta-b)}-2\vartheta}, & h_{opt} = 0, \quad \text{if } \sigma > \mu_1(n), \\ (\log n)^{-\frac{2k}{b}}, & h_{opt} = \mu_1(n), \quad \text{if } \sigma \leq \mu_1(n), \end{cases}$$

where $\mu_1(n)$ is given by (3.10).

Case VI: $b = \beta > 0, h > 0$. Note that, due to (2.10), one has $\sigma < (d\gamma^{-1})^{\frac{1}{b}}$. Consider two cases. If $h < \sigma$, then

$$\Delta_1(\sigma, h) \lesssim \sigma^{-2\vartheta} h^{2\vartheta+2k} \exp\big(-2\gamma(\sigma/h)^{\beta}\big), \quad \Delta_2(\sigma, h) \lesssim h^{(b+2\vartheta-2a-1)}\sigma^{-2\vartheta}\exp(2h^{-b}(d-\gamma\sigma^b)).$$

Then the bias-variance balance is achieved when

$$h^{(b-2k-2a-1)}\exp(2h^{-b}(d-\gamma\sigma^b) + 2\gamma\sigma^b h^{-b}) = n,$$

which leads to (6.12) and, hence, $h_{opt} = \mu_1(n)$, where $\mu_1(n)$ is defined in (3.10). Therefore $h_{opt} \asymp (\log n)^{-\frac{1}{b}}$ and hence

$$\Delta \lesssim \sigma^{-2\vartheta}(\log n)^{-\frac{2\vartheta+2k}{b}} \exp\big(-2\gamma\sigma^{\beta}(\log n)^{\frac{\beta}{b}}\big).$$

If $h \geq \sigma$, then $\Delta \lesssim h^{2k} + n^{-1} h^{b-(2a+1)} \exp(2h^{-b}(d-\gamma\sigma^b))$ and the bias-variance balance is achieved when $h^{2k} \asymp n^{-1} h^{b-(2a+1)} \exp(2h^{-b}(d-\gamma\sigma^b))$. Then, by Lemma 7, we derive that $h_{opt} = \mu_2(n)$, where $\mu_2(n)$ is defined in (3.10), and $\Delta \lesssim (\log n)^{-\frac{2k}{b}}$. Hence

$$\Delta \lesssim \begin{cases} \sigma^{(-2\vartheta)} (\log n)^{-\frac{2\vartheta+2k}{b}} \exp\big(-2\gamma\sigma^{\beta}(\log n)^{\frac{\beta}{b}}\big), & h_{opt} = \mu_1(n), \ \text{if } \sigma > \mu_1(n), \\ (\log n)^{-\frac{2k}{b}}, & h_{opt} = \mu_2(n), \ \text{if } \sigma \leq \mu_1(n), \end{cases}$$

where $\mu_1(n)$ and $\mu_2(n)$ are given by (3.10).

Case VII: $b > 0$, $\beta = 0$. If $h < \sigma$, then

$$\Delta \lesssim \sigma^{-2\vartheta} h^{2\vartheta+2k} + n^{-1}\, \sigma^{-2\vartheta}\, h^{2\vartheta-2a+b-1} \exp(2dh^{-b}).$$

Setting $\sigma^{-2\vartheta} h^{2\vartheta+2k} = n^{-1}\, \sigma^{-2\vartheta}\, h^{2\vartheta-2a+b-1} \exp(2dh^{-b})$, we arrive at (6.12) and $h_{opt} = \mu_1(n)$ where $\mu_1(n)$is defined in (3.10). Hence $h_{opt} \asymp (\log n)^{-1/b}$ and $\Delta \lesssim (\log n)^{-\frac{2\vartheta+2k}{b}}\, \sigma^{-2\vartheta}$, provided $\sigma > \mu_1(n)$.

If $h \geq \sigma$, then

$$\Delta \lesssim h^{2k} + n^{-1} h^{b-2a-1} \exp(2dh^{-b}). \tag{6.13}$$

Setting $h^{2k} \approx n^{-1} h^{b-2a-1} \exp(2dh^{-b})$, arrive at (6.12), so that $h_{opt} = \mu_1(n) \asymp (\log n)^{-1/b}$ and $\Delta \lesssim (\log n)^{-2k/b}$ if $\sigma \leq \mu_1(n)$. Hence

$$\Delta \asymp \begin{cases} (\log n)^{-\frac{2\vartheta+2k}{b}} \sigma^{-2\vartheta}, & h_{opt} = \mu_1(n), \ \text{if } \sigma > \mu_1(n), \\ (\log n)^{-\frac{2k}{b}}, & h_{opt} = \mu_1(n), \ \text{if } \sigma \leq \mu_1(n), \end{cases}$$

where $\mu_1(n)$is defined in (3.10).

Case VIII: $b > \beta > 0$. If $h \leq \sigma$, then

$$\Delta(\sigma, h) \lesssim \sigma^{-2\vartheta} h^{2\vartheta+2k} \exp(-2\gamma\sigma^{\beta} h^{-\beta}) + n^{-1} h^{2\vartheta+b-(2a+1)} \sigma^{-2\vartheta} \exp(2dh^{-b}).$$

Then the minimum of $\Delta(\sigma, h)$ is attained if $n \asymp h^{b-(2a+1)-2k} \exp(2dh^{-b} + 2\gamma\sigma^{\beta} h^{-\beta})$. Note that, due to $\sigma^{\beta} < (d/\gamma)\, h^{-(b-\beta)}$, $b > \beta$ and $\sigma < 1$, one has $2dh^{-b} > 2\gamma\sigma^{\beta} h^{-\beta}$. Therefore we arrive at (6.12), so that $h_{opt} \asymp (\log n)^{-1/b}$ and $\Delta \lesssim \sigma^{-2\vartheta} (\log n)^{\frac{(1+2a-2\vartheta)}{b} - 1}$.

If $h > \sigma$, then $\Delta \lesssim h^{2k} + n^{-1} h^{b-(2a+1)} \exp(2dh^{-b})$, which coincides with (6.13) and we obtain the same expressions for h_{opt} and Δ as in that case. Hence

$$\Delta \asymp \begin{cases} \sigma^{-2\vartheta} (\log n)^{\frac{(1+2a-2\vartheta)}{b} - 1}, & h_{opt} = \mu_1(n), \ \text{if } \sigma > \mu_1(n), \\ (\log n)^{-\frac{2k}{b}}, & h_{opt} = \mu_1(n), \ \text{if } \sigma \leq \mu_1(n), \end{cases}$$

where $\mu_1(n)$is defined in (3.10).

Proof of Theorem 2. Observe that

$$\mathbb{E}\|\hat{f}_{W,\hat{h}} - f_W\|^2 = \widetilde{\Delta}_1 + \widetilde{\Delta}_2 + \widetilde{\Delta}_3, \tag{6.14}$$

where

$$\begin{aligned} \widetilde{\Delta}_1 &= \mathbb{E}\big[\|\hat{f}_{W,\frac{1}{n}} - f_W\|^2 I(w \in \Omega_{\sigma,n})\big]\, I\big(\sigma > n^{-\frac{1}{2a+1}}\big), \\ \widetilde{\Delta}_2 &= \sum_{j=1}^{j_{opt}} \mathbb{E}\big[\|\hat{f}_{W,h} - f_W\|^2 I(\hat{h} = h = 2^{-j}) I(w \notin \Omega_{\sigma,n} \text{ or } \sigma \leq n^{-\frac{1}{2a+1}})\big], \\ \widetilde{\Delta}_3 &= \sum_{j=j_{opt}+1}^{j_{\max}} \mathbb{E}\|\hat{f}_{W,h} - f_W\|^2\, I(\hat{h} = h = 2^{-j})\, I(w \notin \Omega_{\sigma,n} \text{ or } \sigma \leq n^{-\frac{1}{2a+1}}). \end{aligned}$$

We start with construction of an upper bound for $\widetilde{\Delta}_1$. Consider the cases I and IV, since, otherwise, $\widetilde{\Delta}_1 = 0$. Then

$$\widetilde{\Delta}_1 = \mathbb{E}\big[\|\hat{f}_{W,\frac{1}{n}} - f_W\|^2 I(w \in \Omega_{\sigma,n})\big]\, I\big(\sigma > n^{-\frac{1}{2a+2k+1}} = h_{opt}\big)$$
$$+\mathbb{E}\big[\|\hat{f}_{W,\frac{1}{n}} - f_W\|^2 I(w \in \Omega_{\sigma,n})\big]\, I\big(n^{-\frac{1}{2a+1}} < \sigma \le n^{-\frac{1}{2a+2k+1}}\big) = \widetilde{\Delta}_{11} + \widetilde{\Delta}_{12}.$$

Here

$$\begin{aligned}\widetilde{\Delta}_{11} &\le \mathbb{E}\|\hat{f}_{W,\frac{1}{n}} - f_W\|^2\, I(\sigma > h_{opt})\\ &\le C\big[\sigma^{-2\vartheta} n^{-(2\vartheta+2k)} + n^{-1}\sigma^{-(2a+1)}\big]\\ &\le Cn^{-1}\sigma^{-(2a+1)} = C\Delta_{opt} \equiv C\Delta(n,\sigma,h_{opt}).\end{aligned}$$

For $\widetilde{\Delta}_{12}$, one has

$$\widetilde{\Delta}_{12} \le \sqrt{\mathbb{E}\|\hat{f}_{W,\frac{1}{n}} - f_W\|^4}\, \sqrt{\mathbb{P}\big[(w \in \Omega_{\sigma,n})\, I\big(\sigma \le n^{-\frac{1}{2a+2k+1}}\big)\big]}.$$

By Lemma 4, in cases I and IV, $\mathbb{E}\|\hat{f}_{W,\frac{1}{n}} - f_W\|^4 \le Cn^2$ and $\widetilde{\Delta}_{12} \le C[n\, n^{-\frac{\tau}{2}}] \le Cn^{-\frac{2k}{2k+2a+1}}$ provided $\tau \ge 4$. Therefore

$$\widetilde{\Delta}_1 \le C\Delta(n,\sigma,h_{opt}). \tag{6.15}$$

Now we find an upper bound for $\widetilde{\Delta}_2$. For $\widetilde{\Delta}_2$, $\hat{h} \ge h_{opt}$. Recall that, by definition of $\hat{h}$, if $\hat{h} = h \ge h_{opt}$, then

$$\|\hat{f}_{W,h} - \hat{f}_{W,h_{opt}}\|^2 \le 16\, C^2(\tau,q)\, I(\sigma,h_{opt})\, n^{-1}\, \log n.$$

Therefore

$$\begin{aligned}\widetilde{\Delta}_2 &\le \mathbb{E}\big[\|\hat{f}_{W,\hat{h}} - f_W\|^2 I(\hat{h} \ge h_{opt})\big]\\ &\le \mathbb{E}\big[\|\hat{f}_{W,\hat{h}} - \hat{f}_{W,h_{opt}}\|^2\, I(\hat{h} \ge h_{opt})\big] + 2\mathbb{E}\|\hat{f}_{W,h_{opt}} - f_W\|^2\\ &\le 32C^2(\tau,q)n^{-1}\, I(\sigma,h_{opt}) \log n + \Delta(n,\sigma,h_{opt}) \le C\Delta(n,\sigma,h_{opt})\, \log n,\end{aligned}$$

where $\Delta(n,\sigma,h)$ is defined in (3.5). Hence

$$\widetilde{\Delta}_2 \le C\Delta(n,\sigma,h_{opt})\, \log n. \tag{6.16}$$

Now we find an upper bound for $\widetilde{\Delta}_3$. Note that

$$\widetilde{\Delta}_3 \le \sum_{j=j_{opt}+1}^{j_{\max}} \mathbb{E}\big[\|\hat{f}_{W,h} - f_W\|^2 I(\hat{h} = h = 2^{-j})\big].$$

If $\hat{h} = h = 2^{-j}$ for $j \ge j_{opt}+1$, then $\hat{h} < h_{opt}$ and, by the definition of $\hat{h}$, there exist $\tilde{j}$ and $\tilde{h} = 2^{-\tilde{j}} < h_{opt}$ such that

$$\|\hat{f}_{W,h_{opt}} - \hat{f}_{W,\tilde{h}}\|^2 \ge 16\, C^2(\tau,q)\, I(\sigma,\tilde{h})\, n^{-1}\, \log n. \tag{6.17}$$

Since for any $h \le h_{opt}$,

$$\|f_{W,h} - f_W\|^2 \le C_0\, n^{-1} I(\sigma,h),$$

where C_0 is an absolute constant, one has

$$\begin{aligned}\|\hat{f}_{W,h_{opt}} - \hat{f}_{W,\tilde{h}}\| &\le \|\hat{f}_{W,h_{opt}} - f_{W,h_{opt}}\| + \|\hat{f}_{W,\tilde{h}} - f_{W,\tilde{h}}\| + \|f_{W,h_{opt}} - f_W\| + \|f_{W,\tilde{h}} - f_W\|\\ &\le C_0\sqrt{n^{-1} I(\sigma,h_{opt})} + C_0\sqrt{n^{-1} I(\sigma,\tilde{h})} + \|\hat{f}_{W,h_{opt}} - f_{W,h_{opt}}\| + \|\hat{f}_{W,\tilde{h}} - f_{W,\tilde{h}}\|.\end{aligned}$$

Hence, by Lemma 3, if n is large enough,

$$\begin{aligned}
&\mathbb{P}\Big\{\|\hat f_{W,h_{opt}} - f_{W,\tilde h}\| \ge 4C(\tau,q)\sqrt{\tfrac{\log n}{n} I(\sigma,\tilde h)}\Big\} \\
&\quad \le \mathbb{P}\Big\{\|\hat f_{W,h_{opt}} - f_{W,h_{opt}}\| \ge 2C(\tau,q)\sqrt{\tfrac{\log n}{n} I(\sigma,\tilde h)} - C_0\sqrt{\tfrac{I(\sigma,h_{opt})}{n}}\Big\} \\
&\qquad + \mathbb{P}\Big\{\|\hat f_{W,\tilde h} - f_{W,\tilde h}\| \ge 2C(\tau,q)\sqrt{\tfrac{\log n}{n} I(\sigma,\tilde h)} - C_0\sqrt{\tfrac{I(\sigma,\tilde h)}{n}}\Big\} \\
&\quad \le \mathbb{P}\Big\{\|\hat f_{W,h_{opt}} - f_{W,h_{opt}}\| \ge C(\tau,q)\sqrt{\tfrac{\log n}{n} I(\sigma,h_{opt})}\Big\} \\
&\qquad + \mathbb{P}\Big\{\|\hat f_{W,\tilde h} - f_{W,\tilde h}\| \ge C(\tau,q)\sqrt{\tfrac{\log n}{n} I(\sigma,\tilde h)}\Big\} \le 2(2+C_q)\, n^{-\tau}.
\end{aligned}$$

Therefore

$$\widetilde\Delta_3 \le \sum_{j=j_{opt}+1}^{j_{\max}} \sum_{\tilde j=j_{opt}+1}^{j_{\max}} \mathbb{E}\big[\|\hat f_{W,\hat h} - f_W\|^2 I(\hat h = 2^{-j}) I(\tilde h = 2^{-\tilde j})\big],$$

where $\tilde h$ is such that the inequality (6.17) holds. Let $\Omega_{\tilde h}$ be a set on which (6.17) is true. Then $\mathbb{P}(\Omega_{\tilde h}) \le 2(2+C_q)\, n^{-\tau}$ and

$$\begin{aligned}
\widetilde\Delta_3 &\le \sum_{j=j_{opt}+1}^{j_{\max}} \sum_{\tilde j=j_{opt}+1}^{j_{\max}} \sqrt{\mathbb{E}\|\hat f_{W,\hat h} - f_W\|^4}\sqrt{\mathbb{P}(\Omega_{\tilde h})} I(\tilde h = 2^{-\tilde j})\, I(\hat h = 2^{-j}) \\
&\lesssim \sum_{j=j_{opt}+1}^{j_{\max}} \sum_{\tilde j=j_{opt}+1}^{j_{\max}} n^{1-\frac{\tau}{2}} \le C(\log n)^2 n^{1-\frac{\tau}{2}} \le C\Delta(n,\sigma,h_{opt})
\end{aligned}$$

if $\tau/2 - 1 \ge 1$, which is true iff $\tau \ge 4$. Combination of the last inequality with (6.14), (6.15) and (6.16) complete the proof.

6.2. Supplementary Statements and Their Proofs

Lemma 3. *Consider $Y_1, Y_2, \cdots, Y_n$ i.i.d. such that $\mathbb{E}(|Y_1|^q) \le C_q$ with $q > 0$. Let $\tau \ge 1$, $C(\tau,q)$ satisfy assumption* (4.4)*, and $I(\sigma,h)$ be defined by* (3.4)*. Then there exists a set Ω such that for $w \in \Omega$ and all $h \ge 1/n$ simultaneously*

$$\|\hat f_{W,h}(x) - f_{W,h}(x)\|^2 \le C(\tau,q)^2\, I(\sigma,h)\, n^{-1} \log n \tag{6.18}$$

and

$$\mathbb{P}(\Omega) \ge 1 - (2+C_q) n^{-\tau}. \tag{6.19}$$

Proof. Let $f^*_Y(w) = \mathbb{E}(\hat f^*_Y(w))$, where $\hat f^*_Y(w) = \frac{1}{n}\sum_{k=1}^n \exp(iY_k w) = \frac{1}{n}\sum_{k=1}^n [\cos(Y_k w) + i\sin(Y_k w)]$. First we show that there exists a set Ω such that for $w \in \Omega$

$$\mathbb{P}\Big(\sup_{|w|\le n} |\hat f^*_Y(w) - f^*_Y(w)| > C(\tau,q)\sqrt{\log n/n}\ \Big) \le (2+C_q) n^{-\tau} \tag{6.20}$$

provided $C(\tau,q)$ satisfies assumption (4.4). Then it is sufficient to prove that

$$\mathbb{P}\left(\sup_{|w|\le n}\left|\frac{1}{n}\sum_{k=1}^n [\cos(Y_k w) - \mathbb{E}(\cos(Y_k w))]\right| > \frac{C(\tau,q)}{2}\sqrt{\frac{\log n}{n}}\ \right) \le \frac{2+C_q}{n^\tau}. \tag{6.21}$$

Let $\mathcal{B}$ be the set, where the inequality (6.21) holds. For any $\gamma > 0$,

$$\mathbb{P}(\mathcal{B}) \le \mathbb{P}\big(\mathcal{B} \cap \{\max_{1\le k\le n} |Y_k| \le n^\gamma\}\big) + \mathbb{P}\big(\max_{1\le k\le n} |Y_k| > n^\gamma\big). \tag{6.22}$$

By Markov's inequality,

$$\mathbb{P}\big(\max_{1\le k\le n}|Y_k|\ge n^{\gamma}\big)\le n^{-\gamma q}\,\mathbb{E}\big(\max_{1\le k\le n}|Y_k|^q\big)\le n^{-\gamma q}\sum_{k=1}^{n}\mathbb{E}|Y_k|^q\le n^{-\gamma q+1}\,\mathbb{E}|Y_1|^q. \tag{6.23}$$

Set $\gamma=(\tau+1)/q$, hence $\gamma q-1=\tau$. Then

$$\mathbb{P}(\mathcal{B})\le\mathbb{P}\big(\mathcal{B}\cap\{\max_{1\le k\le n}|Y_k|\le n^{\gamma}\}\big)+n^{-\tau}\,\mathbb{E}|Y_1|^q. \tag{6.24}$$

Partition the interval $[-n,n]$ into M sub-intervals by points w_j, $j=0,1,2,3,\dots,M$, such that $w_0=-n$, $w_j-w_{j-1}=n^{-(\gamma+1)}$, so that $M=2n^{\gamma+2}$. Consider a random function $Z_k(w)=\big[\cos(Y_k w)-\mathbb{E}(\cos(Y_k w))\big]\mathbb{I}(|Y_k|\le n^{\gamma})$. Since $|Y_k|\le n^{\gamma}$ and $|\partial(\cos(Yw))/\partial w|\le|Y|\le n^{\gamma}$, we obtain

$$|Z_k(w)-Z_k(w')|\le 2n^{\gamma}|w-w'|.$$

Therefore $Z_k(w)$ satisfies the Lipschitz condition and, for any $w\in[-n,n]$, there exists w_j such that

$$\Big|\frac{1}{n}\sum_{k=1}^{n}Z_k(w)\Big|\le\Big|\frac{1}{n}\sum_{k=1}^{n}Z_k(w_j)\Big|+2n^{\gamma}.\frac{1}{n^{\gamma+1}}.$$

Hence

$$\begin{aligned}\mathbb{P}\big(\mathcal{B}\cap\{\max_{1\le k\le n}|Y_k|\le n^{\gamma}\}\big)&\le\mathbb{P}\Big(\max_{1\le j\le M}\Big|\frac{1}{n}\sum_{k=1}^{n}Z_k(w_j)\Big|+\frac{2}{n}>\frac{C(\tau,q)}{2}\sqrt{\frac{\log n}{n}}\ \Big)\\&\le\mathbb{P}\Big(\max_{1\le j\le M}\Big|\frac{1}{n}\sum_{k=1}^{n}Z_k(w_j)\Big|>\frac{C(\tau,q)}{4}\sqrt{\frac{\log n}{n}}\ \Big)\\&\le\sum_{j=1}^{M}\mathbb{P}\Big(\Big|\frac{1}{n}\sum_{k=1}^{n}Z_k(w_j)\Big|>\frac{C(\tau,q)}{4}\sqrt{\frac{\log n}{n}}\ \Big)\end{aligned}$$

provided

$$\frac{C(\tau,q)}{4}\sqrt{\frac{\log n}{n}}\ge\frac{2}{n},$$

which is guaranteed by condition (4.4).

Using Hoeffding's inequality with $\xi_k=Z_k(w_j)$ where $|\xi_k|\le 2$ and $t=\frac{C(\tau,q)}{4}\sqrt{\frac{\log n}{n}}$, we obtain that

$$\mathbb{P}\Big(\Big|\frac{1}{n}\sum_{k=1}^{n}Z_k(w_j)\Big|>\frac{C(\tau,q)}{4}\sqrt{\frac{\log n}{n}}\ \Big)\le 2\exp\Big(-\frac{(C(\tau,q))^2\log n}{128}\Big)$$

and

$$\mathbb{P}\big(\mathcal{B}\cap\{\max_{1\le k\le n}|Y_k|\le n^{\gamma}\}\big)\le 2n^{-\tau} \tag{6.25}$$

is guaranteed by condition (4.4). Validity of (6.20) follows from inequalities (6.24) and (6.25).

In order to prove (6.18), note that $1/h\le n$ and

$$\begin{aligned}\|\hat f_{W,h}-f_{W,h}\|^2&=\tfrac{1}{2\pi}\,\|\hat f^*_{W,h}-f^*_{W,h}\|^2\\&=\tfrac{1}{2\pi}\int_{-1/h}^{1/h}\frac{|g^*(\sigma s)|^2}{|f^*_\xi(s)|^2}\,|\hat f^*_Y(s)-f^*_Y(s)|^2\,ds\\&\le\sup_{|s|\le n}|\hat f^*_Y(s)-f^*_Y(s)|^2\,I(\sigma,h),\end{aligned}$$

which completes the proof. ☐

Lemma 4. *Let* $h_{\min} = \max\{\sigma, n^{-\frac{1}{2a+1}}\}$. *Then, for any* $h \in [h_{\min}, 1/2]$ *, one has*

$$\mathbb{E}\|\hat{f}_{W,h} - f_W\|^4 \le \begin{cases} C\sigma^{-(4a+3)}n^{-1}, & \textit{cases I, IV}, \\ C\sigma^{-(4a+3)}n^{-1}\log n, & \textit{case II}, \\ Cn^2, & \textit{case III}. \end{cases} \tag{6.26}$$

In particular, if $\sigma \ge n^{-\frac{1}{2a+1}}$, *then* $\mathbb{E}\|\hat{f}_{W,h} - f_W\|^4 \le Cn^2$.

Proof. Note that

$$\mathbb{E}\|\hat{f}_{W,h} - f_W\|^4 = \mathbb{E}\|\hat{f}_{W,h} - \mathbb{E}\hat{f}_{W,h} + \mathbb{E}\hat{f}_{W,h} - f_W\|^4 \le 8\mathbb{E}\|\hat{f}_{W,h} - \mathbb{E}\hat{f}_{W,h}\|^4 + 8\|\mathbb{E}\hat{f}_{W,h} - f_W\|^4. \tag{6.27}$$

Then

$$\|\mathbb{E}\hat{f}_{W,h} - f_W\|^4 = [R_1(\hat{f}_{W,h}, f_W)]^2 \le \Delta_1^2 \le 1,$$

where Δ_1 is defined in (3.7). To find an upper bound for the first term, note that for any x

$$\begin{aligned} |\hat{f}_{W,h}(x)| &\le \frac{1}{2\pi}\int_{-1/h}^{1/h} \frac{|g^*(\sigma s)|}{|f_\xi^*(s)|}\,ds \\ &\le \frac{1}{2\pi}\int_{-1/h}^{1/h} \frac{C_g(\sigma^2 s^2+1)^{-\frac{\vartheta}{2}}\exp(-\gamma|\sigma s|^\beta)}{c_\xi(s^2+1)^{-\frac{a}{2}}}\,ds \\ &\le C\min\left(h^{-(a+1)}, \sigma^{-(a+1)}\right)\tilde{\Delta}_{h,\sigma}, \end{aligned}$$

where

$$\tilde{\Delta}_{h,\sigma} = \begin{cases} 1 & \text{in cases I and IV}, \\ \max\left\{\log\left(\frac{\sigma}{h}\right), 1\right\} & \text{in case II}, \\ \max\left\{1, \left(\frac{\sigma}{h}\right)^{a-\vartheta+1}\right\} & \text{in case III}. \end{cases}$$

The same upper bound holds for $f_{W,h} = \mathbb{E}\hat{f}_{W,h}$. Hence

$$\|\hat{f}_{W,h} - \mathbb{E}\hat{f}_{W,h}\|_\infty^2 \le C\min(h^{-2(a+1)}, \sigma^{-2(a+1)})\tilde{\Delta}_{h,\sigma}^2. \tag{6.28}$$

Therefore

$$\begin{aligned} \mathbb{E}\|\hat{f}_{W,h} - f_{W,h}\|^4 &\le \mathbb{E}[\|\hat{f}_{W,h} - f_{W,h}\|^2]\|\hat{f}_{W,h} - f_{W,h}\|_\infty^2 \\ &\le C\,n^{-1}\min(h^{-(4a+3)}, \sigma^{-(4a+3)})\tilde{\Delta}_{h,\sigma}^2\Delta_{h,\sigma}^2, \end{aligned}$$

where, according to Lemma 2, $\Delta_{h,\sigma}$ is of the form (6.7). Observe that an upper bound for the first term in (6.27) is larger than the second term and that

$$\mathbb{E}\|\hat{f}_{W,h} - f_W\|^4 \le \begin{cases} C\sigma^{-(4a+3)}n^{-1} & \text{in cases I, IV}, \\ C\sigma^{-(4a+3)}n^{-1}\log\left(\frac{1}{h_{\min}}\right) & \text{in case II}, \\ h^{-(4a+3)}n^{-1} & \text{in case III}. \end{cases}$$

Since $h_{\min} \ge n^{-\frac{1}{2a+1}}$, we finally obtain (6.26).

Now, let $\sigma \ge n^{-\frac{1}{2a+1}}$, then $\sigma^{-(4a+3)}n^{-1} \le n^{-1}n^{\frac{4a+3}{2a+1}} \le n^{\frac{2a+2}{2a+1}} \le n^2$, which completes the proof. □

Lemma 5. *Let* $\sigma \le n^{-\frac{1}{2a+2k+1}}$ *and* $\Omega_{\sigma,n}$ *be defined by formula* (4.5). *Then in the cases* I *and* IV, *if* n *is large enough,*

$$\mathbb{P}(\Omega_{\sigma,n}) \le (2 + C_q)\,n^{-\tau}. \tag{6.29}$$

Proof. Note that

$$\|\hat f_{W,\sigma} - \hat f_{W,\frac{1}{n}}\| \le \|\hat f_{W,\sigma} - f_{W,\sigma}\| + \|\hat f_{W,\frac{1}{n}} - f_{W,\frac{1}{n}}\| + \|f_{W,\sigma} - f_W\| + \|f_{W,\frac{1}{n}} - f_W\|. \tag{6.30}$$

Then by Lemma 1, for some absolute constant $\tilde C$, $\|f_{W,\sigma} - f_W\| \le \tilde C\sigma^{-k}$; $\|f_{W,\frac{1}{n}} - f_W\| \le \tilde C n^{-k} < \tilde C\sigma^k$. Also, by Lemma 3, for $w \in \Omega$

$$\|\hat f_{W,\sigma} - f_{W,\sigma}\| \le C(\tau,q)\sqrt{\frac{I(\sigma,\sigma)\log n}{n}} \le C(\tau,q)\sqrt{\frac{I(\sigma,1/n)\log n}{n}}$$

and

$$\|\hat f_{W,\frac{1}{n}} - f_{W,\frac{1}{n}}\| \le C(\tau,q)\sqrt{\frac{I(\sigma,1/n)\log n}{n}}.$$

Hence it follows from (6.30) that for $w \in \Omega$

$$\|\hat f_{W,\sigma} - \hat f_{W,\frac{1}{n}}\| \le 2C(\tau,q)\sqrt{\frac{I(\sigma,1/n)\log n}{n}} + 2\tilde C\sigma^k.$$

Therefore, for $w \in \Omega$,

$$\|\hat f_{W,\sigma} - \hat f_{W,\frac{1}{n}}\| \ge 4C(\tau,q)\sqrt{\frac{I(\sigma,1/n)\log n}{n}}$$

cannot be true, unless

$$C(\tau,q)\sqrt{\frac{I(\sigma,1/n)\log n}{n}} < \tilde C\sigma^k. \tag{6.31}$$

By Lemma 2, in cases I and IV, one has $I(\sigma,1/n) \le \tilde{\tilde C}\sigma^{-(2a+1)}$. So, inequality (6.31) holds only if $C(\tau,q)\ (\tilde{\tilde C})^2\sigma^{-(a+\frac{1}{2})}\sqrt{\log n/n} < \tilde C\sigma^k$, which is equivalent to $\sigma > \bar C(n^{-1}\log n)^{\frac{1}{2k+2a+1}}$, where $\bar C = \tilde{\tilde C}C(\tau,q)/\tilde C$. Therefore, if $w \in \Omega$ and $\sigma \le n^{-\frac{1}{2k+2a+1}}$, where n is such that $\bar C(\log n)^{\frac{1}{2k+2a+1}} \ge 1$, then (6.31) is not true. Hence $w \notin \Omega_{\sigma,n}$, so that $\Omega_{\sigma,n} \subset \Omega^c$ and (6.29) holds. □

Lemma 6. *Consider an integral of the form*

$$I(\lambda) = \int_{m_1}^{m_2} P_\lambda(z)\exp(Q_\lambda(z))\,dz, \tag{6.32}$$

where $0 \le m_1 < m_2 < \infty$ and $P_\lambda(z)$ and $Q_\lambda(z)$ are real-valued differentiable functions of z and $\lambda \to \infty$ is a large parameter. Let $z_0 \equiv z_{0,\lambda} = \underset{z\in[m_1,m_2]}{\text{argmax}}\, Q_\lambda(z)$ be a unique global maximum of $Q_\lambda(z)$ on the interval $[m_1,m_2]$. Assume that the following conditions hold:

(1) *A function P is a positive slowly varying function, i.e., for any $t > 0$ one has*

$$\lim_{x\to\infty} P(tx)/P(x) = 1.$$

(2) *$Q_\lambda(z_0) - Q_\lambda(z)$ increases monotonically for $\lambda \ge \lambda_0$ as $\lambda \to \infty$.*

(3) *If $Q'_\lambda(z_0) = 0$, then for every $\lambda \ge \lambda_0$*

$$\lim_{x\to 0}\frac{Q_\lambda(z_0+x) - Q_\lambda(z_0)}{x^2} = \frac{Q''_\lambda(z_0)}{2} < 0. \tag{6.33}$$

(4) *If $Q'_\lambda(z_0) \ne 0$, then for every $\lambda \ge \lambda_0$*

$$\lim_{x\to 0}\frac{Q_\lambda(z_0+x) - Q_\lambda(z_0)}{x} = Q'_\lambda(z_0) \ne 0. \tag{6.34}$$

Then, as $\lambda \to \infty$,

$$I(\lambda) \asymp \begin{cases} \frac{\exp\{Q_\lambda(z_0)\}\, P_\lambda(z_0)}{\sqrt{|Q''_\lambda(z_0)|}} & \text{if (6.33) holds,} \\ \frac{\exp\{Q_\lambda(z_0)\}\, P_\lambda(z_0)}{Q'_\lambda(z_0)} & \text{if (6.34) holds.} \end{cases} \tag{6.35}$$

Proof. Comparing (6.32) with the integral

$$I(\lambda) = \int G(z) \exp(-F(z)) dz \tag{6.36}$$

we obtain $F(z) = -Q_\lambda(z), G(z) = P_\lambda(z)$. Then, following the calculations in [8] with $F(z_0) = -Q_\lambda(z_0)$, $F_1(z_0) = -Q'_\lambda(z_0)$, we obtain from formulas (3) and (4), p. 111,

$$I(\lambda) = [-Q'_\lambda(z_0)]^{-1} \exp\{Q_\lambda(z_0)\} \sum_0^\infty L_r,$$

where L_r is given by

$$L_r = -Q'_\lambda(z_0) \Big(\frac{d}{Q'_\lambda(z)\, dz}\Big)^r \frac{P_\lambda(z)}{Q'_\lambda(z)}\Big|_{z=z_0}.$$

Hence, taking the term with $r = 0$, we obtain, when (6.33) holds:

$$I(\lambda) \approx \frac{\exp\{Q_\lambda(z_0)\}\, P_\lambda(z_0)}{Q'_\lambda(z_0)}.$$

Now, consider the case when (6.34) holds.

Then following the calculations in [8], p. 118, we obtain

$$I(\lambda) = \exp\{Q_\lambda(z_0)\} \int \exp\{-f^2\} P_\lambda(z)\, dz,$$

where

$$f = \sqrt{Q_\lambda(z_0) - Q_\lambda(z)} \sim \sqrt{F_2/2}\, z \quad \text{as} \quad z \to z_0$$

with $F_2(z_0) = -Q''_\lambda(z_0)$. Therefore, from formulas (16) and (17), p. 119, we obtain

$$I(\lambda) = \Big[\frac{\pi}{-2Q''_\lambda(z_0)}\Big]^{\frac{1}{2}} \exp\{Q_\lambda(z_0)\} \sum_0^\infty L_r,$$

where L_r is given by

$$L_r = Q'_\lambda(z_0) \Big(\frac{d}{2\, f'\, dz}\Big)^r \frac{P_\lambda(z)}{f'}\Big|_{z=z_0}.$$

Hence, taking the term with $r = 0$, we obtain, when (6.34) holds:

$$I(\lambda) \approx \frac{\sqrt{\pi} \exp\{Q_\lambda(z_0)\}\, P_\lambda(z_0)}{\sqrt{-2Q''_\lambda(z_0)}},$$

which is equivalent to the second expression of (6.35). □

Lemma 7. *Let* n *be large and* $z \in \mathbb{R}$ *be a fixed quantity. Then, as* $n \to \infty$, *the solution of the equation*

$$e^m m^z = n \tag{6.37}$$

is given by

$$m = (\log n - z \log\log n)(1 + o(1)), \quad n \to \infty. \tag{6.38}$$

Proof. Since $e^m m^z = n$, then $m + z\log m = \log n$ and $m = \log n - z\log m$. Plugging this m back into (6.37), we obtain $e^{\log n - z\log m}(\log n - z\log m)^z = n$. Since for large values of n, one has $(\log n - z\log m)^z \approx (\log n)^z$, the previous equation becomes $(\log n)^z n e^{-z\log m} \approx n$, so that $z\log\log n \approx z\log m$, which yields (6.38). □

ACKNOWLEDGMENTS

Marianna Pensky and Ramchandra Rimal were partially supported by National Science Foundation (NSF), grants DMS-1407475 and DMS-1712977. The authors also thank Alexander Tsybakov for the help with the proof of Lemma 3.

REFERENCES

1. J. Berkson, "Are There Two Regressions?", J. Amer. Statist. Assoc. **45**, 164–180 (1950).
2. J. Bovy, J. F. Hennawi, D. W. Hogg, A. D. Myers, J. A. Kirkpatrick, D. J. Schlegel, N. P. Ross, E. S. Sheldon, I. D. McGreer, D. P. Schneider, and B. A. Weaver, "Think outside the color box: Probabilistic target selection and the SDSS-XDQSO Quasar targeting catalog", The Astrophysical Journal **729(2)**, #141 (2011).
3. R. J. Carroll, D. Rupport, L. A. Stefanski, and C. M. Crainiceanu, *Measurement Error in Nonlinear Models, A Modern Perspective* (Chapman and Hall, New York, 2006), 2nd ed..
4. R. J.Carroll, A. Delaigle, and P. Hall, "Nonparametric Prediction in Measurement Error Models", J. Amer. Statist. Assoc. **104**, 993–1003 (2009).
5. F. Comte and J. Kappus, "Density Deconvolution from Repeated Measurements without Symmetry Assumption on the Errors", J. Multivar. Anal. **140**, 31–46 (2015).
6. A. Delaigle, "Nonparametric Density Estimation from Data with a Mixture of Berkson and Classical Errors", Can. J. Statist. **35**, 89–104 (2007).
7. A. Delaigle, "An Alternative View of the Deconvolution Problem", Statist. Sinica **18**, 1025–1045 (2008).
8. R. B. Dingle, *Asymptotic Expansions: Their Derivation and Interpretation* (Academic Press Inc., London–New York, 1973).
9. L. Du, C. Zou, and Z. Wang, "Nonparametric Regression Function Estimation for Errors-in-variables Models with Validation Data", Statist. Sinica **21**, 1093–1113 (2011).
10. P. Geng and H. L. Koul, "Minimum Distance Model Checking in Berkson Measurement Error Models with Validation Data", Test **28** (3), 879–899 (2019).
11. A. Goldenshluger, "On Pointwise Adaptive Nonparametric Deconvolution", Bernoulli **5**, 907–925 (1999).
12. A. Goldenshluger and O. Lepski, "Bandwidth Selection in Kernel Density Estimation: Oracle Inequality and Adaptive Minimax Optimality", Ann. Statist. **39**, 1608–1632 (2011).
13. K. H. Kim, W. K. Härdle, and S.-K. Chao, "Simultaneous Inference for the Partially Linear Model with a Multivariate Unknown Function when the Covariates are Measured with Errors", *SFB 649 Discussion Papers SFB649DP2016-024* (Humboldt Univ., Berlin, Germany, 2016).
14. C. Lacour and F. Comte, "Data-Driven Density Estimation in the Presence of Additive Noise with Unknown Distribution", J. Roy. Statist. Soc., Ser. B, **73**, 601–627 (2011).
15. O. V. Lepski and V. G. Spokoiny, "Optimal Pointwise Adaptive Methods in Nonparametric Estimation", Ann. Statist. **25**, 2512–2546 (1997).
16. J. P. Long, N. E. Karoui, and J. A. Rice, "Kernel Density Estimation with Berkson Error", Canad. J. Statist. **44**, 142–160 (2016).
17. A. Meister, *Deconvolution Problems in Nonparametric Statistics* (Springer, New York, 2009).
18. E. A. Robinson, *Seismic Inversion and Deconvolution: Part B: Dual-Sensor Technology* (Elsevier, Oxford, 1999).
19. A. B. Tsybakov, *Introduction to Nonparametric Estimation* (Springer, New York, 2009).
20. L. Wang, "Estimation of Nonlinear Berkson-type Measurement Error Models", Statist. Sinica **13**, 1201–1210 (2003).
21. L. Wang, "Estimation of Nonlinear Models with Berkson Measurement Errors", Ann. Statist. **32**, 2559–2579 (2004).
22. C. B. Wason, J. L. Black, and G. A. King, "Seismic Modeling and Inversion", Proc. IEEE. **72**, 1385–1393 (1984).
23. R. Wong, *Asymptotic Approximations of Integrals* (SIAM, Philadelphia, 2001).