Journal of Multivariate Analysis 101 (2010) 1179–1189

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Nonparametric Berkson regression under normal measurement error and bounded design

Alexander Meister *

Graduiertenkolleg 1100, Universität Ulm, Helmholtzstrasse 22, D-89081 Ulm, Germany

ARTICLE INFO

Article history:
Received 30 January 2009
Available online 5 November 2009

AMS 2000 subject classifications:
62G08

Keywords:
Berkson error
Deconvolution
Errors-in-variables regression
Inverse problems
Orthogonal polynomials

ABSTRACT

Regression data often suffer from the so-called Berkson measurement error which contaminates the design variables. Conventional nonparametric approaches to this errors-in-variables problem usually require rather strong conditions on the support of the design density and that of the contaminated regression function, which seem unrealistic in many cases. In the current note, we introduce a novel nonparametric regression estimator, which is able to identify the regression function on the whole real line under normal Berkson error although the location of the design variables is restricted to some bounded interval. The asymptotic properties of this estimator are investigated and some numerical simulations are provided.

1. Introduction

Nonparametric regression estimation has become an important tool to investigate the link between two random quantities. Recently, there has been increasing interest in the so-called errors-in-variables problems where the design variables are corrupted by measurement errors. In the classical measurement error model where the design variables are contaminated before the dependent variables are observed, Fan and Truong [1] introduce a deconvolution procedure to estimate the regression function despite the contamination of the data and study the optimal convergence rates.

On the contrary, the current paper deals with the Berkson regression model where one observes the data (X_j, Y_j), $j = 1, \ldots, n$, with

$$Y_j = g(X_j + \delta_j) + \varepsilon_j, \tag{1}$$

where the random variables $X_1, \delta_1, \varepsilon_1, \ldots, X_n, \delta_n, \varepsilon_n$ are independent; the δ_j are assumed to have the known density f_δ while we only assume that $E\varepsilon_j = 0$ and $E\varepsilon_j^2 \le C_\varepsilon < \infty$ with respect to the ε_j. Assuming exact knowledge of the error density is the standard approach in nonparametric errors-in-variables problems. In practice, this is a limitation, of course. Sometimes, additional direct observations of the distribution of δ_j may be available so that f_δ is accessible; see e.g. [2]. The goal is to estimate the unknown regression function g. Therefore, the Berkson model differs from the standard nonparametric regression model by the occurrence of the δ_j which are degenerated to zero otherwise; it is also different from the problem studied in [1] as, in the Berkson setting, the design variables are contaminated after the X_j have been recorded.

The Berkson model, which has been introduced by Berkson [3], has mainly been studied in a parametric or semiparametric framework; see for example [4] or [5]. To our knowledge, the first nonparametric approach to the Berkson problem is given in [2]. Meister [6] studies that setting in the case of Fourier-oscillating error densities. Carroll et al. [7] consider a model which combines the classical and the Berkson measurement error.

* Corresponding address: Universitat Rostock, Institut fuer Mathematik, Ulmenstr. 69, D-18057 Rostock, Germany.
E-mail addresses: alexander.meister@uni-ulm.de, alexander.meister@uni-rostock.de.

0047-259X/$ – see front matter
doi:10.1016/j.jmva.2009.10.010

Many practical examples of errors-in-variables problems are given in the book of Carroll et al. [8]. In Delaigle et al. [2], a real data example for the Berkson problem is described: The data originate from a survey conducted by the US Department of Agriculture. Its goal is to compare three methods of estimating areas growing specific crops. These methods are aerial photography, satellite imagery and personal statements of the farmers. The authors' goal is to detect the link between the data gained by aerial photography and data acquired by satellite imagery. However, in absence of photography data, the authors consider the data from personal interviews as the predictor variables X_j and the satellite imagery data as the response variables Y_j—transferred to our notation. Then, the unobserved random variables $X_j+\delta_j$ represent the photography data. Finally the function g describes the desired link. Thus those data can be studied via the nonparametric Berkson model. Another real data set to which the Berkson problem is applicable is given in the Nevada Test Side Thyroid Disease Study. The goal of that study is the investigation of the link between radiation exposure and thyroid disease outcomes. The predictor data X_j denote some observations of dosimetry, which obviously represent only a part of (the logarithm of) the true radiation exposure of an individual. This latter quantity is unobserved and a contaminated version of X_j. Hence, it may be interpreted as $X_j + \delta_j$. The response data indicate absence or presence of thyroid disease. That example is given in [7].

The estimator proposed by Delaigle et al. [2] leans on the fact that the convolved regression function

$$p(x) := \left[g * f_\delta^-\right](x) = E\left(Y_j \mid X_j = x\right), \tag{2}$$

translated to our notation, is directly accessible by the empirical data set. Throughout this work, $*$ denotes convolution; and $E\left(Z \mid X_j = x\right)$ denotes the conditional expectation of a random variable Z given $X_j = x$. Then the authors propose a deconvolution procedure based on Fourier methods to estimate the function g. The authors consider a random design model; and they argue that, if $\operatorname{supp} f_\delta = \mathbb{R}$, then the design density must also be infinitely supported in order to be able to identify g on a given interval where, throughout, we write $\operatorname{supp} f$ for the support of a function f, i.e.

$$\operatorname{supp} f = \overline{\left\{x \in \mathbb{R} : \limsup_{h\downarrow 0} \frac{1}{2h}\int_{x-h}^{x+h} f(y)\mathrm{d}y \neq 0\right\}},$$

and $\overline{A}$ denotes the closure of a set A. That requires some strong conditions on the relation between the support of f_δ, g and f_X, i.e. the design density. More concretely, if the support of f_δ, g and f_X equals $[a_\delta, b_\delta]$, $[a_g, b_g]$ and $[a_X, b_X]$, respectively, the inequalities $a_X \leq a_\delta + a_g$ and $b_X \geq b_\delta + b_g$ must be satisfied; however, they seem to be rather unrealistic in many real data applications. In [7], the case of infinitely supported g is discussed; however, the authors consider estimation of g only on a domain where f_X is bounded away from zero.

In the current note, we show that such strong assumptions on the support can be omitted when the error density f_δ is normal. Although normal measurement errors occur rather frequently in practice due to the central limit theorem, Gaussian models are not very popular in deconvolution problems. Models involving ordinary smooth error densities – in the notation of Fan [9] – with polynomial Fourier tails are usually favoured as the attainable convergence rates are faster; they are algebraic for ordinary smooth error densities while supersmooth error densities such as the normal density only admit logarithmic rates. However, in the framework of this work, we show that just the supersmoothness of the normal density turns out to be an advantage. We introduce a novel procedure, which estimates g consistently on the whole real line, in the setting where all the design variables $X_1, \ldots, X_n$ are located in a fixed compact interval J. Even though the support of g may be bounded in most practical applications that support could exceed the interval J. Then, our estimator is able to estimate the function g as a whole while the approach of Delaigle et al. [2] is not and the estimator of Carroll et al. [7] is consistent only for $g(x)$, $x \in J$. It may be seen as surprising that g is identifiable at all under the given conditions. To provide some better understanding of this fact, we mention that the function p in (2) is analytic if f_δ is normal even if only finitely many derivatives of g are supposed to exist. That latter condition is the standard assumption in nonparametric curve estimation. That is also why, unfortunately, our procedure cannot be extended to ordinary smooth error densities. This means that $p(x)$, for all $x \in \mathbb{R}$, is uniquely determined by its restriction to an open non-void interval, take $(\min\{X_1, \ldots, X_n\}, \max\{X_1, \ldots, X_n\})$ as that interval. Therefore, p as a whole is empirically accessible by the data set drawn from model (1); and then g is reconstructable by a deconvolution procedure. Motivated by that result, we derive a nonparametric estimator for g under model (1) in Section 2. In Section 3, the convergence rates of this estimator are investigated. The graphical plots of some numerical simulations are shown in Section 4. The proofs are deferred to Section 5.

2. Methodology

We assume throughout that f_δ is the normal density with the mean 0 and the variance σ^2, denoted by $N(0, \sigma^2)$. Also, we assume that g is Lebesgue measurable and satisfies

$$|g(x)| \leq C\exp(D|x|), \tag{3}$$

for all $x \in \mathbb{R}$ and some constants $C, D > 0$. This condition is significantly weaker than the assumptions imposed in [7] where a polynomial bound on $|g|$ is stipulated. As the normal density is supersmooth, i.e. its Fourier transform has exponential tails, the function p in (2) is differentiable infinitely often; its derivatives are

$$p^{(k)}(x) = \frac{1}{\sqrt{2\pi}\sigma}\int g(y)\frac{d^k}{dx^k}\exp\left(-(x-y)^2/(2\sigma^2)\right)\mathrm{d}y$$

$$= \frac{(-1)^k}{\sqrt{2\pi}\sigma^k} \int g(x-\sigma y) \exp\left(-y^2/2\right) H_k(y) dy,$$

where H_k denotes the kth Hermite polynomial. Applying Taylor's expansion, we obtain that

$$p(x) = \sum_{k=0}^{K-1} \frac{(-1)^k}{\sqrt{2\pi} k! \sigma^k} \int g(-\sigma y) \exp\left(-y^2/2\right) H_k(y) dy \cdot x^k + R_K(x). \tag{4}$$

By Lagrange's representation and (3), the residual term R_K satisfies the inequality

$$\begin{aligned} |R_K(x)| &\leq \frac{1}{K!\sigma^K} C \exp(D|x|)|x|^K \frac{1}{\sqrt{2\pi}} \int \exp\left(-y^2/2\right) \exp(D\sigma|y|)|H_K(y)| dy \\ &\leq \frac{1}{K!\sigma^K} C \exp(D|x|)|x|^K \frac{1}{\sqrt{2\pi}} \left(\int \exp\left(-y^2/2 + 2D\sigma|y|\right) dy\right)^{1/2} \cdot \left(\int \exp\left(-y^2/2\right) |H_K(y)|^2 dy\right)^{1/2} \\ &\leq \text{const} \cdot (K!)^{-1/2} \exp(D|x|)|x/\sigma|^K. \end{aligned} \tag{5}$$

Here and elsewhere const denotes a generic positive constant.

The representation (4) motivates us to apply a polynomial approach to estimate the function p. Therefore we introduce

$$L_k(x) = \left((X_{(n)} - X_{(1)})/2\right)^{-1/2} \sqrt{k+1/2} \cdot \tilde{L}_k\left(2[x - (X_{(n)} + X_{(1)})/2]/(X_{(n)} - X_{(1)})\right),$$

where $\tilde{L}_k$ denotes the kth Legendre polynomial, which can be represented by

$$\tilde{L}_k(x) = \frac{1}{2^k k!} \frac{d^k}{dx^k}\left[(x^2-1)^k\right]. \tag{6}$$

That famous equality is also known as the Rodrigues formula. For details about Legendre polynomials, see e.g. the book of El Attar [10]. Here and elsewhere, $X_{(1)} < \cdots < X_{(n)}$ denotes all elements of the set $\{X_1, \ldots, X_n\}$ with increasing order. Thus, L_k is the normalization of $\tilde{L}_k$ so that

$$\int_{X_{(1)}}^{X_{(n)}} |L_k(x)|^2 dx = 1.$$

Furthermore, the sequence $(L_k)_{\{k \geq 0\}}$ is an orthonormal basis of the Hilbert space consisting of all squared-integrable functions on the domain $[X_{(1)}, X_{(n)}]$. That inspires us to consider

$$\hat{p}_K(x) = \sum_{k=0}^{K-1} \hat{c}_k \cdot L_k(x),$$

with

$$\hat{c}_k = \sum_{j=1}^{n-1} Y_{(j)} \cdot \int_{X_{(j)}}^{X_{(j+1)}} L_k(z) dz, \tag{7}$$

as an estimator of $p(x)$ where $Y_{(j)}$ denotes that Y_l which correlates with the design variable $X_{(j)}$. Therein, note that the polynomials L_k can uniquely be continued from the domain $[X_{(1)}, X_{(n)}]$ to the whole of $\mathbb{R}$ in a natural way. The parameter K is still to be selected.

We denote the coefficients of L_k by $\lambda_{k,j}$, i.e.

$$L_k(x) = \sum_{j=0}^{k} \lambda_{k,j} x^j.$$

Hence, according to (4), it is reasonable to consider

$$\hat{d}_j = \sum_{k=j}^{K-1} \hat{c}_k \lambda_{k,j},$$

as an empirical version of

$$\frac{(-1)^j}{\sqrt{2\pi} j! \sigma^j} \int g(-\sigma y) \exp\left(-y^2/2\right) H_j(y) dy.$$

Moreover, it is well known that the Hermite polynomials form an orthogonal basis of the Hilbert space of all functions f satisfying $\|f\|_w^2 = \langle f, f\rangle_w < \infty$ with respect to the inner product

$$\langle f, g\rangle_w := \int f(x)g(x)\exp(-x^2/2)\mathrm{d}x.$$

In particular, we have the representation

$$g_\sigma(x) = \sum_{k=0}^{\infty} \frac{1}{\sqrt{2\pi}k!}\langle g_\sigma, H_k\rangle_w H_k,$$

where the infinite sum must be understood as a $\langle\cdot,\cdot\rangle_w$-limit; and $g_\sigma(x) = g(-\sigma x)$.

Therefore, we define

$$\hat{g}(x) := \sum_{k=0}^{\lfloor c_K(K-1)\rfloor} (-\sigma)^k \hat{d}_k \cdot H_k(-x/\sigma), \tag{8}$$

as our nonparametric linear estimator of g with some parameter $c_K \in (0, 1)$. Note that (8) is a spectral cut-off estimator and $\lfloor c_K(K-1)\rfloor$ denotes that integer where the cut-off takes place. It corresponds to one by the bandwidth in kernel regularization. In the related field of density deconvolution, a similar regularization techniques involving Hermite polynomials is introduced in [11]. In the context of the current work, where we aim at estimating a regression function rather than a density, the polynomial approach seems to be even more appropriate.

As a great advantage of our method, we point out that only one smoothing parameter K remains to be selected while the choice of c_K will not be critical as shown in the next section; unlike the approach of Delaigle et al. [2] where a local smoother, including an additional smoothing parameter, is used to estimate the convolved regression function.

3. Asymptotics

In this section, we will study the convergence rates attained by the estimator $\hat{g}$ as defined in (8). With respect to the design variables $X_1, \ldots, X_n$, we give two alternative conditions.

Condition D: We consider either the fixed design model and assume that

$$\text{const} \cdot 1/n \le X_{(j+1)} - X_{(j)} \le \text{const} \cdot \log n/n,$$

holds true for all integer $n > 0$, $j = 1, \ldots, n-1$ and uniform constants, and that all $X_1, \ldots, X_n$ are located in some fixed compact interval J; or consider the random design model and assume that all $X_1, \ldots, X_n$ lie in some fixed compact interval J almost surely and have the design density f_X which is bounded away from zero by a constant $c_X > 0$ and bounded above by C_X on J.

The following lemma is valid under both alternatives.

Lemma 1. *Under Condition D, we have*

$$E\left(\max\left\{X_{(j+1)} - X_{(j)} : j = 1, \ldots, n\right\}\right)^4 = O\left(n^{-4}\log^4 n\right),$$

and

$$E\left(X_{(n)} - X_{(1)}\right)^{-\alpha} \le const^\alpha,$$

for all $\alpha \in (0, n/2)$ and some positive constant which is independent of n.

With respect to the target regression function g, we impose some smoothness conditions by stipulating that g is β-fold differentiable ($\beta \in \mathbb{N}$) on the whole real line and satisfies

$$\int |x^k g^{(j)}(x)|^2 \exp(-x^2/(2\sigma^2))\mathrm{d}x/\sigma \le C_G, \tag{9}$$

for some positive constant C_G and all integers $k + j \le \beta$, $k, j \ge 0$. All functions g which satisfy (9) as well as (3) are collected in the function class $\mathcal{G} = \mathcal{G}_{\beta,C_G,C,D}$. As an important property, the functions contained in $\mathcal{G}$ are well approximable by the basis of the Hermite polynomials, which is shown in the following lemma.

Lemma 2. *For any $g \in \mathcal{G}$, we have*

$$\sum_{k=0}^{\infty} k^\beta \frac{1}{\sqrt{2\pi}k!}|\langle g_\sigma, H_k\rangle_w|^2 \le C_{\mathcal{G}},$$

for a constant $C_{\mathcal{G}} > 0$ which is dependent of only $\mathcal{G}$ but not of g explicitly.

The following theorem gives us an upper bound on the convergence rates of the MISE (mean integrated squared error) of estimator (8) with the $N(0,\sigma^2)$-density as the weight function. Thus, we consider estimation of g on the whole real line and not only on the support of f_X as in [7].

Theorem 1. *Assume that Condition D holds. Choose $K = K_n \asymp \lfloor \log n/\log\log n\rfloor$, i.e. $K_n = const \cdot \lfloor \log n/\log\log n\rfloor$, and $c_K \in (0,1)$ such that*

$$\lim_{n\to\infty} c_K K/\frac{\log n}{\log\log n} \in (0,1).$$

Then, the estimator (8) satisfies

$$\sup_{g\in\mathcal{G}} E\int |\hat{g}(x)-g(x)|^2 \exp\left(-x^2/(2\sigma^2)\right) \mathrm{d}x/\sigma = O\left((\log n/\log\log n)^{-\beta}\right).$$

Apparently, the estimator (8) achieves the common rates of convergence for deconvolution with normal error densities under smoothness constraints (see e.g. [9]) up to the deterioration caused by the iterated logarithm. Moreover, the choice of parameter K is auto-adaptive, i.e. it is independent of the unknown smoothness degree β of g. Obviously, no data-driven selection procedure for K such as cross-validation or plug-in is required to attain those rates.

As a corollary of Theorem 1, we may conclude that the same convergence rates can be established when we consider the uniform MISE

$$\sup_{g\in\mathcal{G}} E\int_a^b |\hat{g}(x)-g(x)|^2 \mathrm{d}x,$$

for all fixed real numbers $a < b$ since the normal density chosen as the weight function is continuous and non-vanishing. Note that $S = [a,b]$ may, by far, exceed the interval J, which contains all the observations $X_1,\ldots,X_n$.

In the following theorem, we give a lower bound on the attainable convergence rates with respect to any estimator in the given statistical experiment, provided the ε_j are normally distributed.

Theorem 2. *Assume that Condition D holds. Let $\hat{g}$ be an arbitrary estimator of g based on the data $(X_1,Y_1),\ldots,(X_n,Y_n)$ drawn from model (1). Assume that $\beta > 2$ and allow for C_G sufficiently large; and suppose that the ε_j have the standard normal distribution. Then, for any $a < b$, we have*

$$\sup_{g\in\mathcal{G}} E\int_a^b |\hat{g}(x)-g(x)|^2 \mathrm{d}x \geq const \cdot (\log n)^{-\beta},$$

for n sufficiently large.

Hence, combining Theorems 1 and 2, we have shown that estimator (8) achieves nearly optimal convergence rates up to an iterated logarithmic factor under smoothness constraints.

4. Simulations

Now the finite sample performance of the estimator (8) is considered based on numerical simulations. As we face a difficult nonparametric problem with slow convergence rates, our simulation study is restricted to the case of a large sample size; we choose $n = 1000$. The design variables $X_1,\ldots,X_n$ are drawn from the uniform density on the interval $[-1,1]$. The Berkson errors δ_j have a standard normal density; and the regression errors ε_j are $N(0,0.2^2)$-distributed and independent of the Berkson errors. We have studied two different target regression functions; they are

$$g(x) = g_1(x) = 2 - 0.8x,$$
$$g(x) = g_2(x) = -0.5 + 0.3x - 1.5x^2.$$

With respect to the selection of the parameters K and c_K, one must avoid choosing K too large, in particular, as the variance term may explode. In both of our considered examples, we found that $K = 4$ and $c_K = 0.7$ is an appropriate choice. Obviously, this selection also leads to satisfying results in a broader class of regression functions. One could also think of data-driven selectors such as cross-validation methods; however, the slow rates and the restriction of K to the set of all integers seem to limit those interests.

Fig. 1 shows five independent replicates of the estimator (8) under the linear regression function g_1; while, in Fig. 2, g_2 is the underlying regression function. The true curves are plotted with dotted linestyle. We realize that our estimator is able to detect the basic structure of the curves in both cases. In particular, we highlight the fact that our procedure estimates the regression curves well on the domain $[-2,2]$ which is larger than the support $[-1,1]$ of the design density. That has been the main motivation for the construction of the estimator (8).

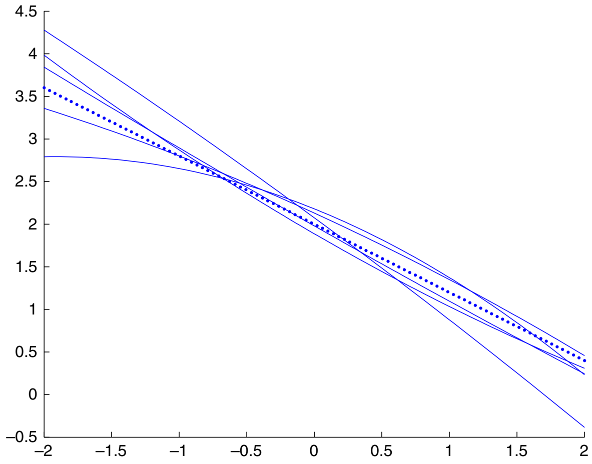

Fig. 1.

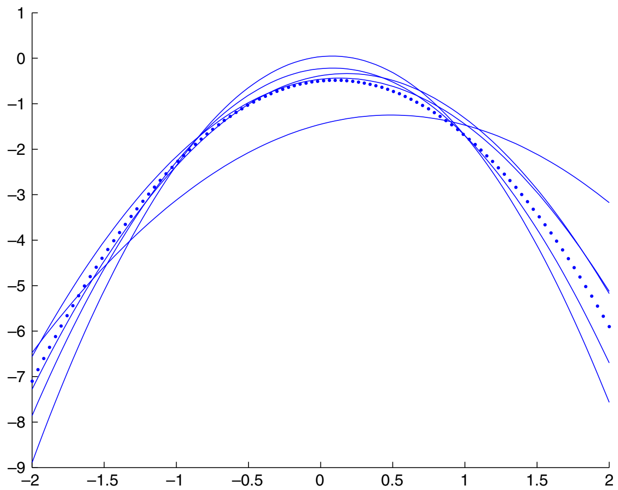

Fig. 2.

5. Proofs

Proof of Lemma 1. Under the first alternative of Condition D, both assertions are trivial. Under the second alternative, we consider that

$$E\left[\max\left\{X_{(j+1)}-X_{(j)}:j=1,\ldots,n\right\}\right]^4=\int_0^\infty P\left[\max\left\{X_{(j+1)}-X_{(j)}:j=1,\ldots,n\right\}\geq x^{1/4}\right]\mathrm{d}x$$
$$\leq M^{-4}+(\lambda^4-M^{-4})\cdot P\left[\max\left\{X_{(j+1)}-X_{(j)}:j=1,\ldots,n\right\}\geq 1/M\right],$$

for some integer $M>0$ where λ denotes the Lebesgue measure of J. Thus, the assertion

$$\max\left\{X_{(j+1)}-X_{(j)}:j=1,\ldots,n\right\}\geq 1/M$$

implies the existence of at least one $k=1,\ldots,2M$ such that none of the X_j is located in T_k where T_k, $k=1,\ldots,2M$ denotes the partition of the interval J consisting of $2M$ equidistant intervals with the length $\lambda/(2M)$. That gives us

$$E\left[\max\left\{X_{(j+1)}-X_{(j)}:j=1,\ldots,n\right\}\right]^4\leq M^{-4}+\lambda^4\sum_{k=1}^{2M}P\left[X_j\not\in T_k,\forall j=1,\ldots,n\right]$$
$$\leq M^{-4}+2\lambda^4M\left(1-c_X\lambda/(2M)\right)^n.$$

Then, selecting $M=c_Mn/\log n$ with a constant $c_M>0$ sufficiently small proves the first assertion.

With respect to the second assertion, we consider that

$$\begin{aligned} E(X_{(n)}-X_{(1)})^{-\alpha} &= \int_0^\infty P\left[X_{(n)}-X_{(1)} < s^{-1/\alpha}\right] \mathrm{d}s \\ &\le E\int_0^\infty P\left[|X_j-X_1| < s^{-1/\alpha}, \quad \forall j\in\{1,\ldots,n\} \mid X_1\right] \mathrm{d}s \\ &\le \int_0^\infty \min\left\{1, (2C_X)^{n-1}s^{-(n-1)/\alpha}\right\} \mathrm{d}s \\ &\le C^\alpha + \frac{\alpha}{n-\alpha-1}C^\alpha\,(2C_X/C)^{n-1} \\ &\le (C')^\alpha, \end{aligned}$$

for some constants C, C' large enough so that the lemma has been proved. ■

Proof of Lemma 2. First, we mention that (9) implies that

$$\|g_{k,j,\sigma}\|_w^2 \le \sigma^{2k} C_G =: C_G',$$

by an elementary substitution for all nonnegative integers $j+k \le \beta$ where $g_{k,j,\sigma}(x) = x^k g^{(j)}(-\sigma x)$. We define the normalized Hermite polynomials $\tilde{H}_k = H_k/\left((2\pi)^{1/4}\sqrt{k!}\right)$, which form an orthonormal basis with respect to the inner product $\langle x, y\rangle_w$ for two functions x, y. Utilizing the equality

$$H_{k+1}'(x) = (k+1)H_k(x), \quad \forall x \in \mathbb{R},$$

for all integer k, we derive by integration by parts that

$$\begin{aligned} \langle g_\sigma, \tilde{H}_k\rangle_w &= \frac{1}{\sqrt{k+1}}\int g(-\sigma x)\tilde{H}_{k+1}'(x)\exp(-x^2/2)\mathrm{d}x \\ &= \frac{\sigma}{\sqrt{k+1}}\int g'(-\sigma x)\tilde{H}_{k+1}(x)\exp(-x^2/2)\mathrm{d}x + \frac{1}{\sqrt{k+1}}\int g(-\sigma x)x\tilde{H}_{k+1}(x)\exp(-x^2/2)\mathrm{d}x, \end{aligned}$$

where the boundary terms vanish as the function $g(-\sigma x)\tilde{H}_{k+1}(x)\exp(-x^2/2)$ is integrable so that there exist two sequences $(R_n)_n \uparrow \infty$ and $(S_n)_n \downarrow -\infty$ such that

$$\lim_{n\to\infty} g(-\sigma R_n)\tilde{H}_{k+1}(R_n)\exp(-R_n^2/2) = \lim_{n\to\infty} g(-\sigma S_n)\tilde{H}_{k+1}(S_n)\exp(-S_n^2/2) = 0.$$

Repeating that procedure β-times where in the next step the function $g(-\sigma x)$ must be replaced by $xg(-\sigma x)$ and $g'(-\sigma x)$, we obtain that

$$|\langle g, \tilde{H}_k\rangle_w| \le \text{const}\cdot k^{-\beta/2}\cdot\max\left\{\left|\int g^{(j)}(-\sigma x)x^l\exp(-x^2/2)\tilde{H}_{k+\beta}(x)\mathrm{d}x\right| : j+l\le\beta, \text{integer } j, l \ge 0\right\}.$$

Therefore, we may conclude that

$$\begin{aligned} \sum_{k=0}^\infty k^\beta |\langle g_\sigma, \tilde{H}_k\rangle_w|^2 &\le \text{const}\cdot\sum_{j,l\ge 0, j+l\le\beta}\left(\sum_{k=0}^\infty |\langle g_{l,j,\sigma}, \tilde{H}_{k+\beta}\rangle|^2\right) \\ &\le \text{const}\cdot\sum_{j,l\ge 0, j+l\le\beta}\|g_{l,j,\sigma}\|_w^2 \le \text{const}\cdot C_G', \end{aligned}$$

by Parseval's identity. Rescaling $\tilde{H}_k$ to H_k gives us the desired inequality. ■

Proof of Theorem 1. Utilizing Parseval's identity for general orthogonal bases, we obtain that

$$\begin{aligned} &E\int |\hat{g}(x)-g(x)|^2\exp\left(-x^2/(2\sigma^2)\right)\mathrm{d}x/\sigma = E\int |\hat{g}(-\sigma x)-g(-\sigma x)|^2\exp\left(-x^2/2\right)\mathrm{d}x \\ &= \sum_{k=0}^{\lfloor c_K(K-1)\rfloor}\sqrt{2\pi}k!EE\left(\left|(-\sigma)^k\hat{d}_k - \frac{1}{\sqrt{2\pi}k!}\langle g_\sigma, H_k\rangle_w\right|^2 \mid \sigma_X\right) + \sum_{k=\lfloor c_K(K-1)\rfloor+1}^\infty \frac{1}{\sqrt{2\pi}k!}|\langle g_\sigma, H_k\rangle_w|^2. \end{aligned} \tag{10}$$

Therein, σ_X denotes the σ-algebra generated by the $X_1,\ldots,X_n$ and we write $E(\cdots\mid\sigma_X)$ and $\mathrm{var}(\cdots\mid\sigma_X)$ for the conditional expectation and variance, respectively, given σ_X.

We show that the first addend in (10) is asymptotically negligible compared to the second one. As $p - R_K$ as in (4) is a polynomial with the degree $\leq K-1$ we have

$$\begin{aligned} p(x) - R_K(x) &= \sum_{m=0}^{K-1} L_m(x)\langle L_m, p - R_K\rangle_I = \sum_{m=0}^{K-1}\sum_{j=0}^{m} \lambda_{m,j} x^j \langle L_m, p - R_K\rangle_I \\ &= \sum_{j=0}^{K-1}\left(\sum_{m=j}^{K-1} \lambda_{m,j}\langle L_m, p - R_K\rangle_I\right) x^j, \end{aligned}$$

for all $x \in \mathbb{R}$ almost surely. Note that two polynomials which coincide on an open non-void interval coincide with each other on the whole real line. Therein, we write

$$\langle f, g\rangle_I = \int_I f(x)\overline{g(x)}\mathrm{d}x,$$

and put $I = [X_{(1)}, X_{(n)}]$. Comparison of the coefficients gives us

$$\frac{(-1)^j}{\sqrt{2\pi}j!\sigma^j}\langle g_\sigma, H_j\rangle_w = \sum_{m=j}^{K-1}\lambda_{m,j}\langle L_m, p - R_K\rangle_I,$$

for all $j = 0, \ldots, K-1$. With respect to the first term in (10), we obtain that

$$\begin{aligned} &EE\left(\left|(-\sigma)^k\hat{d}_k - \frac{1}{\sqrt{2\pi}k!}\langle g_\sigma, H_k\rangle_w\right|^2 \,\middle|\, \sigma_X\right) = \sigma^{2k} EE\left(\left|\sum_{j=k}^{K-1}\lambda_{j,k}\left(\hat{c}_j - \langle L_j, p - R_K\rangle_I\right)\right|^2 \,\middle|\, \sigma_X\right) \\ &\quad = \sigma^{2k} E\left|\sum_{j=k}^{K-1}\lambda_{j,k}\left(E(\hat{c}_j \mid \sigma_X) - \langle L_j, p - R_K\rangle_I\right)\right|^2 + \sigma^{2k} E\mathrm{var}\left(\sum_{j=k}^{K-1}\lambda_{j,k}\left(\hat{c}_j - \langle L_j, p - R_K\rangle_I\right)\,\middle|\, \sigma_X\right). \qquad (11) \end{aligned}$$

Let us calculate the following conditional expectation

$$\begin{aligned} E(\hat{c}_k \mid \sigma_X) &= \sum_{j=1}^{n-1} p(X_{(j)}) \cdot \int_{X_{(j)}}^{X_{(j+1)}} L_k(z)\mathrm{d}z \\ &= \int_{X_{(1)}}^{X_{(n)}} p(z)L_k(z)\mathrm{d}z + \sum_{j=1}^{n-1}\int_{X_{(j)}}^{X_{(j+1)}} \big(p(X_{(j)}) - p(z)\big)\cdot L_k(z)\mathrm{d}z. \end{aligned}$$

Hence, we have

$$\begin{aligned} \left|E(\hat{c}_k \mid \sigma_X) - \langle p, L_k\rangle_I\right| &\leq \|p'\|_{J,\infty}\cdot \max\left\{|X_{(j+1)} - X_{(j)}| : j = 1, \ldots, n-1\right\}\cdot \int_{X_{(1)}}^{X_{(n)}} |L_k(z)|\mathrm{d}z \\ &\leq \|p'\|_{J,\infty}\cdot \max\left\{|X_{(j+1)} - X_{(j)}| : j = 1, \ldots, n-1\right\}\cdot (X_{(n)} - X_{(1)})^{1/2}, \end{aligned}$$

by the Cauchy–Schwarz inequality and the orthonormality of the L_k, where $\|p'\|_{J,\infty}$ denotes the constraint supremum of $|p'(x)|$ on the compact set $x \in J$ with J as in Condition D. By (3), $\|p'\|_{J,\infty}$ is uniformly bounded above. Moreover, $(X_{(n)} - X_{(1)})^{1/2}$ is uniformly bounded above by the square root of the Lebesgue measure of J. Thus the bias term contained in (11) is bounded above by a uniform constant times

$$\begin{aligned} &\sigma^{2k} E\left(\sum_{j=k}^{K-1}|\lambda_{j,k}|\left(\max\left\{|X_{(l+1)} - X_{(l)}| : l = 1, \ldots, n-1\right\} + |\langle L_j, R_K\rangle_I|\right)\right)^2 \\ &\quad \leq \mathrm{const}\cdot\sigma^{2k}\cdot\left[E\left(\sum_{j=k}^{K-1}|\lambda_{j,k}|\right)^4\right]^{1/2}\cdot\left(\left[E\left(\max\left\{|X_{(l+1)} - X_{(l)}| : l = 1, \ldots, n-1\right\}\right)^4\right]^{1/2} + \mathrm{const}^K/K!\right), \end{aligned}$$

due to (5). In order to derive an upper bound on the coefficients $|\lambda_{j,k}|$, we calculate the derivatives in (6), hence we have for all $k = 0, \ldots, j$,

$$\begin{aligned} |\lambda_{j,k}| &\leq \sqrt{2j+1}\cdot 2^{-j}\left|\sum_{m=0}^{\lfloor (j-k)/2\rfloor}\binom{j}{m}\binom{2j-2m}{j}\binom{j-2m}{k}\frac{[(X_{(1)} + X_{(n)})/2]^{j-2m-k}}{[(X_{(n)} - X_{(1)})/2]^{j-2m+1/2}}\right| \\ &\leq \mathrm{const}^K\cdot\left[(X_{(n)} - X_{(1)})^{-K+1/2} + \mathrm{const}\right]. \qquad (12) \end{aligned}$$

Therefore we conclude by Lemma 1 that for all $k = 0, \dots, j$,

$$E\left(\sum_{j=k}^{K-1} |\lambda_{j,k}|\right)^4 \le \text{const}^K,$$

for some sufficiently large constant. Using Lemma 1 again, the bias term in (11) has the upper bound

$$\text{const}^K \cdot \left[n^{-2}\log^2 n + 1/K!\right].$$

With respect to the variance term in (11), we derive that

$$\sigma^{2k} E\text{var}\left(\sum_{j=k}^{K-1} \lambda_{j,k}\left(\hat{c}_j - \langle L_j, p - R_K\rangle_I\right) \,\Big|\, \sigma_X\right) = \sigma^{2k} E\text{var}\left(\sum_{j=k}^{K-1} \lambda_{j,k}\hat{c}_j \,\Big|\, \sigma_X\right)$$
$$\le \sigma^{2k} E \sum_{l=1}^{n-1}\left[\sum_{j=k}^{K-1} \lambda_{j,k}\int_{X_{(l)}}^{X_{(l+1)}} L_j(z)dz\right]^2 E\left(Y_{(l)}^2 \mid \sigma_X\right),$$

as the random variable $\langle L_j, p - R_K\rangle_I \lambda_{j,k}$ is measurable in σ_X where $E\left(Y_{(l)}^2 \mid \sigma_X\right) \le$ const follows from (3) and $f_\delta = N(0, \sigma^2)$. Applying (12), $\sup_{x\in[-1,1]} |\tilde{L}_k(x)| \le 1, \forall k$, and Lemma 1 again, we obtain that the variance term is bounded above by

$$\text{const} \cdot \sigma^{2k} \sum_{l=1}^{n-1}\left(E\left[\sum_{j=k}^{K-1} |\lambda_{j,k}| \sup_{z\in I} |L_j(z)|\right]^4\right)^{1/2} \cdot \left(E(X_{(l+1)} - X_{(l)})^4\right)^{1/2}$$
$$\le \text{const} \cdot n^{-1}\log^2 n\sigma^{2k}\left(E\left[\sum_{j=k}^{K-1} |\lambda_{j,k}|\right]^4 \frac{4}{(X_{(n)} - X_{(1)})^2}\right)^{1/2} (K + 1/2)$$
$$\le \text{const}^K \cdot n^{-1}\log^2 n.$$

Inserting those upper bounds into the first addend in (11) and then in (10) gives us the following upper bound on this term,

$$\text{const}^K \cdot (\lfloor c_K(K-1)\rfloor)! \cdot \left(n^{-1}\log^2 n + 1/K!\right) = o\left(\log^{-\gamma} n\right),$$

for all $\gamma > 0$ by the specific selection of K as proposed in the theorem and Stirling's formula. Hence, the first addend in (10) is asymptotically negligible and we face a dominating bias phenomenon as usual in supersmooth deconvolution.

Now we focus on the second term in (10). Lemma 2 gives us $O\left(K^{-\beta}\right)$ as an upper bound on this term. Inserting K as suggested in the theorem completes the proof. ■

Proof of Theorem 2. We define the function

$$g_n(x) = a_n \frac{\sin(x/b_n^2)}{\pi x}\cos(2b_n x),$$

for all $x \in \mathbb{R}$ where the sequences $(a_n)_n \downarrow 0$ and $(b_n)_n \uparrow \infty$ remain to be chosen; also we introduce the function $\tilde{g}_n = -g_n$. We can guarantee that $g_n, \tilde{g}_n \in \mathcal{G}$ holds true when we assume that

$$a_n \asymp b_n^{2-\beta},$$

with appropriate constants. Then, g_n and $\tilde{g}_n$ compete to be the true regression function. First we consider the fixed design model. Then, in the fixed design alternative of Condition D, the joint density of the data $Y_1, \dots, Y_n$ is given by

$$f_\xi(y_1, \dots, y_n) = \prod_{j=1}^n \frac{1}{\sqrt{2\pi}}\int \exp\left(-\frac{1}{2}\left(y_j - \xi(X_j + s)\right)^2\right) f_\delta(s)ds,$$

when $\xi = g_n, \tilde{g}_n$, respectively, is the true regression function where f_δ denotes the density of the δ_j, that is $N(0, \sigma^2)$. We derive an upper bound on the $L_1(\mathbb{R}^n)$-distance between f_{g_n} and $f_{\tilde{g}_n}$. Note that each of the factors in the definition of f_ξ is a univariate density, which integrates to one. By the telescopic sum, we obtain that

$$\|f_{g_n} - f_{\tilde{g}_n}\|_1 \le \frac{1}{\sqrt{2\pi}}\sum_{j=1}^n \int\left|\int\left[\exp\left(-\frac{1}{2}\left(y + g_n(X_j + s)\right)^2\right) - \exp\left(-\frac{1}{2}\left(y - g_n(X_j + s)\right)^2\right)\right] \cdot f_\delta(s)ds\right| dy$$
$$= \frac{1}{\sqrt{2\pi}}\sum_{j=1}^n \int\left|\sum_{k=0}^\infty \frac{1}{k!}\left(\frac{d^k}{dy^k}\exp(-y^2/2)\right)\left(1 - (-1)^k\right)\int g_n^k(X_j + s)f_\delta(s)ds\right| dy$$

$$\leq \frac{2}{\sqrt{2\pi}} \sum_{j=1}^{n} \sum_{k=0}^{\infty} \frac{1}{(1+2k)!} \cdot \int |H_{1+2k}(y)| \exp(-y^2/2)\mathrm{d}y \cdot |g_n^{2k+1} * f_\delta|(X_j),$$

by Taylor's expansion where $g_n^{2k+1} * f_\delta$ denotes the convolution of the functions g_n^{2k+1} and f_δ. Therein, note that $g_n \in L_2(\mathbb{R})$. The fact that the even terms in Taylor's expansion vanish is essential. Applying the Cauchy–Schwarz inequality, we obtain that

$$\begin{aligned}\int |H_{1+2k}(y)| \exp(-y^2/2)\mathrm{d}y &\leq \left(\int |H_{1+2k}(y)|^2 \exp(-y^2/2)\mathrm{d}y\right)^{1/2} \left(\int \exp(-y^2/2)\mathrm{d}y\right)^{1/2} \\ &\leq \text{const} \cdot \sqrt{(1+2k)!}.\end{aligned}$$

From there we may conclude by Fourier inversion that

$$\begin{aligned}\|f_{g_n} - f_{\tilde{g}_n}\|_1 &\leq \text{const} \cdot \sum_{k=0}^{\infty} \frac{1}{\sqrt{(1+2k)!}} \cdot \sum_{j=1}^{n} |g_n^{2k+1} * f_\delta|(X_j) \\ &\leq \text{const} \cdot \sum_{k=0}^{\infty} \frac{1}{\sqrt{(1+2k)!}} \sum_{j=1}^{n} \left| \frac{1}{2\pi} \int \exp(-itX_j)[g_n^{2k+1}]^{ft}(t) f_\delta^{ft}(t)\mathrm{d}t \right| \\ &\leq \text{const} \cdot \sum_{k=0}^{\infty} \frac{n}{\sqrt{(1+2k)!}} \int |[g_n^{2k+1}]^{ft}(t) f_\delta^{ft}(t)|\mathrm{d}t. \end{aligned} \tag{13}$$

where f^{ft} denotes the Fourier transform of a function f.

Under the random design alternative of Condition D, the density f_ξ must be seen as the conditional density of the $Y_1, \ldots, Y_n$ given the $X_1, \ldots, X_n$. All the inequalities derived above follow analogously as almost sure versions.

We calculate

$$\begin{aligned}[g_n^{2k+1}]^{ft}(t) &= 2^{-2k-1} a_n^{2k+1} \int \exp(itx) \left(\frac{\sin(x/b_n^2)}{\pi x}\right)^{2k+1} (\exp(2b_n xi) + \exp(-2b_n xi))^{2k+1} \,\mathrm{d}x \\ &= 2^{-2k-1} a_n^{2k+1} \sum_{j=0}^{2k+1} \binom{2k+1}{j} \eta_k^{ft}(t - 2(2j-2k-1)b_n),\end{aligned}$$

where $\eta_k(x) = \left(\sin(x/b_n^2)/(\pi x)\right)^{2k+1}$ so that $\eta_0^{ft} = \chi_{[-1/b_n^2, 1/b_n^2]}$, i.e. the indicator function of the interval $[-1/b_n^2, 1/b_n^2]$. For $k \geq 1$ we have that

$$\|\eta_k^{ft}\|_\infty \leq \text{const} \cdot \int \left|\frac{\sin x}{\pi x}\right|^3 \mathrm{d}x < \infty,$$

via estimation by the $L_1(\mathbb{R})$-norm of η_k. Summarizing, we have established that $\|\eta_k\|_\infty$ is bounded above by a constant uniformly with respect to all $k \geq 0$. Furthermore, by Fourier inversion, we have

$$\eta_k^{ft} = (2\pi)^{-2k} \left[\underbrace{\eta_0^{ft} * \cdots * \eta_0^{ft}}_{(2k+1)\text{-times}}\right],$$

where η_0^{ft} equals $2/b_n^2$ times the uniform density on the interval $[-1/b_n^2, 1/b_n^2]$ so that we are guaranteed that η_k^{ft} is supported on $[-1, 1]$ whenever $k \leq c_b \cdot b_n^2$ for some positive constant c_b. By the binomial formula, we conclude that

$$\|[g_n^{2k+1}]^{ft}\|_\infty \leq \text{const} \cdot a_n^{2k+1}, \quad \forall n, k \in \mathbb{N},$$

and that the support of $\left[g_n^{2k+1}\right]^{ft}$ is included in $\mathbb{R} \setminus (-b_n, b_n)$ if $k \leq c_b \cdot b_n^2$. Inserting those results into (13), we obtain that

$$\|f_{g_n} - f_{\tilde{g}_n}\|_1 \leq \text{const} \cdot \sum_{k=0}^{\lfloor c_b b_n^2 \rfloor} \frac{n}{\sqrt{(1+2k)!}} a_n^{2k+1} \exp\left(-b_n^2\sigma^2/2\right) + \text{const} \cdot \sum_{k=\lfloor c_b b_n^2 \rfloor}^{\infty} \frac{n}{\sqrt{(1+2k)!}} a_n^{2k+1}.$$

We choose

$$b_n = c_B (\log n)^{1/2},$$

with a constant C_B sufficiently large. Then, $\|f_{g_n} - f_{\tilde{g}_n}\|_1$ converges to zero by the normality of f_δ and Stirling's formula. In the random design case, this convergence occurs almost surely so that $E\|f_{g_n} - f_{\tilde{g}_n}\|_1 \to 0$ as $n \to \infty$ follows.

In the sequel E_ξ denotes the expectation with respect to the data $X_1, \ldots, X_n, Y_1, \ldots, Y_n$ when ξ is the true regression function. We conclude that

$$\begin{aligned}
\sup_{g \in \mathcal{G}} E \int_a^b |\hat{g}(x) - g(x)|^2 \mathrm{d}x &\geq \frac{1}{2}\left(E_{g_n} \int_a^b |\hat{g}(x) - g_n(x)|^2 \mathrm{d}x + E_{\tilde{g}_n} \int_a^b |\hat{g}(x) - \tilde{g}_n(x)|^2 \mathrm{d}x \right) \\
&\geq \frac{1}{2} E \int_a^b \int \cdots \int \left(|\hat{g}(x, y_1, \ldots, y_n) - g_n(x)|^2 + |\hat{g}(x, y_1, \ldots, y_n) + g_n(x)|^2 \right) \\
&\quad \times \min\left\{ f_{g_n}(y_1, \ldots, y_n), f_{\tilde{g}_n}(y_1, \ldots, y_n) \right\} dy_1 \cdots dy_n \\
&\geq \int_a^b g_n^2(x) \mathrm{d}x \cdot \left(1 - \frac{1}{2} E \|f_{g_n} - f_{\tilde{g}_n}\|_1 \right),
\end{aligned}$$

where the expectation can be removed in the last three lines in the fixed design case. Hence, we have

$$\begin{aligned}
\sup_{g \in \mathcal{G}} E \int_a^b |\hat{g}(x) - g(x)|^2 \mathrm{d}x &\geq \text{const} \cdot \int_a^b g_n^2(x) \mathrm{d}x \\
&\geq \text{const} \cdot a_n^2 b_n^{-2} \int_{a/b_n^2}^{b/b_n^2} \left| \frac{\sin x}{\pi x} \right|^2 \cos^2(2 b_n^3 x) \mathrm{d}x \\
&\geq \text{const} \cdot a_n^2 b_n^{-4} \asymp b_n^{-2\beta} \asymp (\log n)^{-\beta},
\end{aligned}$$

where the periodicity of the function $\cos^2(x)$ must be taken into account. That completes the proof of the theorem. ■

References

[1] J. Fan, Y. Truong, Nonparametric regression with errors in variables, Ann. Statist. 21 (1993) 1900–1925.
[2] A. Delaigle, P. Hall, P. Qiu, Nonparametric methods for solving the Berkson errors-in-variables problem, J. Roy. Statist. Soc. Ser. B 68 (2006) 201–220.
[3] J. Berkson, Are there two regression problems? J. Amer. Statist. Assoc. 45 (1950) 164–180.
[4] L. Huwang, Y.H.S. Huang, On errors-in-variables in polynomial regression – Berkson case, Statist. Sinica 10 (2000) 923–936.
[5] J.P. Buonaccorsi, C.-D. Lin, Berkson measurement error in designed repeated measures studies with random coefficients, J. Statist. Plann. Inference 104 (2002) 53–72.
[6] A. Meister, Deconvolution from Fourier-oscillating error densities under decay and smoothness restrictions, Inverse Problems 24 (2008) 015003 14 pages.
[7] R.J. Carroll, A. Delaigle, P. Hall, Nonparametric regression estimation from data contaminated by a mixture of Berkson and classical errors, J. Roy. Statist. Soc. Ser. B 69 (2007) 859–878.
[8] R.J. Carroll, D. Ruppert, L.A. Stefanski, C.M. Crainiceanu, Measurement Error in Nonlinear Models, 2nd ed., Chapman and Hall CRC Press, Boca Raton, 2006.
[9] J. Fan, On the optimal rates of convergence for non-parametric deconvolution problems, Ann. Statist. 19 (1991) 1257–1272.
[10] R. El Attar, Legendre polynomials and functions, CreateSpace. ISBN 978-1441490124, 2009.
[11] E. Masry, J.A. Rice, Gaussian deconvolution via differentiation, Canad. J. Statist. 20 (1992) 9–21.