Simultaneous inference for Berkson errors-in-variables regression under fixed design

Katharina Proksch[1] · Nicolai Bissantz[2] · Hajo Holzmann[3]

Abstract

In various applications of regression analysis, in addition to errors in the dependent observations also errors in the predictor variables play a substantial role and need to be incorporated in the statistical modeling process. In this paper we consider a nonparametric measurement error model of Berkson type with fixed design regressors and centered random errors, which is in contrast to much existing work in which the predictors are taken as random observations with random noise. Based on an estimator that takes the error in the predictor into account and on a suitable Gaussian approximation, we derive finite sample bounds on the coverage error of uniform confidence bands, where we circumvent the use of extreme-value theory and rather rely on recent results on anti-concentration of Gaussian processes. In a simulation study we investigate the performance of the uniform confidence sets for finite samples.

Keywords Berkson errors-in-variables · Deconvolution · Gaussian approximation · Uniform confidence bands

1 Introduction

In mean regression problems a predictor variable X, either a fixed design point or a random observation, is used to explain a response variable Y in terms of the conditional mean regression function $g(x) = \mathbb{E}[Y|X = x]$. The case of a random covariate

---

✉ Katharina Proksch
 k.proksch@utwente.nl

1 Department of Applied Mathematics, University of Twente, Drienerlolaan 5,
 7522 NB Enschede, The Netherlands

2 Fakultät für Mathematik, Ruhr-Universität Bochum, Universitätsstraße 150, 44801 Bochum,
 Germany

3 Fachbereich Mathematik und Informatik, Philipps-Universität Marburg, Hans-Meerwein-Straße
 6, 35032 Marburg, Germany

occurs when both X and Y are measured during an experiment and the case of fixed design corresponds to situations in which covariates can be set by the experimenter such as a machine setting, say, in a physical or engineering experiment. Writing $\epsilon = Y - \mathbb{E}[Y|X]$ gives the standard form of the nonparametric regression model $Y = g(X) + \epsilon$, that is, the response is observed with an additional error but the predictor can be set or measured error-free. In many experimental settings this is not a suitable model assumption since either the predictor can also not be measured precisely, or since the presumed setting of the predictor does not correspond exactly to its actual value. There are subtle differences between these two cases, which we illustrate by the example of drill core measurements of the content of climate gases in the polar ice. Assume that the content of climate gas Y at the bottom of a drill hole is quantified. The depth of the drill hole X is measured independently with error Δ giving the observation W. A corresponding regression model is of the following form

$$Y = g(X) + \epsilon, \quad W = X + \Delta, \tag{1}$$

where W, Δ and ϵ are independent, Δ and ϵ are centered, and observations of (Y, W) are available. This model is often referred to as *classical errors-in-variables model*. A change in the experimental set-up might require a change in the model that is imposed. Assume that in our drill core experiment we fix specific depths w at which the drill core is to be analyzed. However, due to imprecisions of the instrument we cannot accurately fix the desired value of w, rather the true (but unknown) depth where the measurement acquired is $w + \Delta$. In this case a corresponding model, referred to as *Berkson errors-in-variables model* (Berkson 1950), is of the form

$$Y = g(w + \Delta) + \epsilon, \tag{2}$$

where Δ and ϵ are independent and centered, w is set by the experimenter and Y is observed. In this paper we construct uniform confidence bands in the nonparametric Berkson errors-in-variables model with fixed design (2).

In particular, we provide finite sample bounds on the coverage error of these bands. We also address the question how to choose the grid when approximating the supremum of a Gaussian process on [0, 1]. For Berkson-type measurement errors, a fixed design as considered in the present paper seems to be of particular relevance in experimentation in physics and engineering. Instead of using the classical approach based on results from extreme-value theory (Bickel and Rosenblatt 1973), we propose a simulation based procedure and construct asymptotic uniform confidence regions by using anti-concentration properties of Gaussian processes which were recently derived by Chernozhukov et al. (2014). For an early, related contribution see Neumann and Polzehl (1998), who develop the wild bootstrap originally proposed by Wu (1986) to construct confidence bands in a nonparametric heteroscedastic regression model with irregular design. While their method could potentially also be adopted in our setting, we preferred to work with the simulation based method which allows for a more transparent analysis.

There is a vast literature on errors-in-variables models, where most of the earlier work is focused on parametric models (Berkson 1950; Anderson 1984; Stefanski

1985; Fuller 1987). A more recent overview of different models and methods can be found in the monograph by Carroll et al. (2006). In a nonparametric regression context, Fan and Truong (1993) consider the classical errors-in-variables setting (1), construct a kernel-type deconvolution estimator and investigate its asymptotic performance with respect to weighted L_p-losses and L_∞-loss and show rate-optimality for both ordinary smooth and super smooth known distributions of errors Δ. The case of Berkson errors-in-variables with random design is treated, e.g., in Delaigle et al. (2006), who also assume a known error distribution, Wang (2004), who assumes a parametric form of the error density, and Schennach (2013), whose method relies on the availability of an instrumental variable instead of the full knowledge of the error distribution. Furthermore, Delaigle et al. (2008) consider the case in which the error-distribution is unknown but repeated measurements are available to estimate the error distribution. A mixture of both types of errors-in-variables is considered in Carroll et al. (2007) and the estimation of the observation-error variance is studied in Delaigle and Hall (2011).

However, in the aforementioned papers the focus is on estimation techniques and the investigation of theoretical as well as numerical performance of the estimators under consideration. In the nonparametric setting only very little can be found about the construction of statistical tests or confidence statements. Model checks in the Berkson measurement error model are developed in Koul and Song (2008, 2009), who construct goodness-of-fit tests for a parametric point hypothesis based on an empirical process approach and on a minimum-distance principle for estimating the regression function, respectively. The construction of confidence statements seems to be discussed only for classical errors in variables models with random design in Delaigle et al. (2015), who focus on pointwise confidence bands based on bootstrap methods and in Kato and Sasaki (2019), who provide uniform confidence bands.

This paper is organized as follows. In Sect. 2 we discuss the mathematical details of our model and describe nonparametric methods for estimating the regression function in the fixed design Berkson model. In Sect. 3 we state the main theoretical results and in particular discuss the construction of confidence bands in Sect. 3.2, where we also discuss the choice of the bandwidth. The numerical performance of the proposed confidence bands is investigated in Sect. 4. Section 5 outlines an extension to error densities for which the Fourier transform is allowed to oscillate. Some auxiliary lemmas are stated in Sect. 6. Technical proofs of the main results from Sect. 3 are provided in Sect. 7, while details and proofs for the extension in Sect. 5 along with some additional technical details are given in an online supplement. In the following, for a function f, which is bounded on some given interval $[a, b]$, we denote by $\|f\| = \|f\|_{[a,b]} = \sup_{x \in [a,b]} |f(x)|$ its supremum norm. The L^p-norm of f over all of $\mathbb{R}$ is denoted by $\|f\|_p$. Further, for $w \in \mathbb{R}$ we set $\langle w \rangle := (1 + w^2)^{\frac{1}{2}}$.

2 The Berkson errors-in-variables model with fixed design

The Berkson errors-in-variables model with fixed design that we shall consider is given by

$$Y_j = g(w_j + \Delta_j) + \epsilon_j, \tag{3}$$

where $w_j = j/(n \, a_n)$, $j = -n, \ldots, n$, are the design points on a regular grid, a_n is a design parameter that satisfies $a_n \to 0$, $n a_n \to \infty$, and Δ_j and ϵ_j are unobserved, centered, independent and identically distributed errors for which $\mathrm{Var}[\epsilon_1] = \sigma^2 > 0$ and $\mathbb{E}|\epsilon_1|^M < \infty$ for some $M > 2$. The density f_Δ of the errors Δ_j is assumed to be known. For ease of notation, we consider an equally spaced grid of design points here. However, this somewhat restrictive assumption can be relaxed to more general designs with a mild technical effort, as we elaborate in Section II of the supplementary material. For random design Berkson errors-in-variables models, Delaigle et al. (2006) point out that identification of g on a given interval requires an infinitely supported design density if the error density is of infinite support. This corresponds to our assumption that asymptotically, the fixed design exhausts the whole real line, which is assured by the requirements on the design parameter a_n. Meister (2010) considers the particular case of normally distributed errors Δ and bounded design density, where a reconstruction of g is possible by using an analytic extension. If we define γ as the convolution of g and $f_\Delta(-\cdot)$, that is, $\gamma(w) = \int_{\mathbb{R}} g(z) f_\Delta(z - w) \, \mathrm{d}z$, then $\mathbb{E}[Y_j] = \gamma(w_j)$, and the calibrated regression model (Carroll et al. 2006) associated with (3) is given by

$$Y_j = \gamma(w_j) + \eta_j, \quad \eta_j = g(w_j + \Delta_j) - \gamma(w_j) + \epsilon_j. \tag{4}$$

Here the errors η_j are independent and centered as well but no longer identically distributed since their variances $v^2(w_j) = \mathbb{E}[\eta_j^2]$ depend on the design points. To be precise, we have that

$$v^2(w_j) = \int \left(g(w_j + \delta) - \gamma(w_j) \right)^2 f_\Delta(\delta) \, \mathrm{d}\delta + \sigma^2 \geq \sigma^2 > 0. \tag{5}$$

This reveals the increased variability due to the errors in the predictors.

The following considerations show that ignoring the errors in variables can lead to misinterpretations of the data at hand. To illustrate, in the setting of simulation Sect. 4, scenario 2, Fig. 1 (upper left panel) shows the alleged data points (w_j, Y_j), that is, the observations at the incorrect, presumed positions, for a sample of size $n = 100$. In addition to the usual variation introduced by the errors ϵ_i in y-direction this display shows a variation in x-direction introduced by the errors in the w_j. The upper right panel shows the actual but unobserved data points $(w_j + \Delta_j, Y_j)$ that only contain the variation in y-direction. Ignoring the errors-in-variables leads to estimating γ instead of g, which introduces a systematic error. The functions γ (solid line) and g (dashed line) are both shown in the lower right panel of Fig. 1. The corresponding variance function is shown in the lower left panel of Fig. 1 (solid line) in comparison with the constant variance σ^2 (dotted line). Apparently, there is a close connection between the calibrated model (4) and the classical deconvolution regression model as considered in Birke et al. (2010) and Proksch et al. (2015) in univariate and multivariate settings, respectively. In contrast to the calibrated regression model (4), in both works an i.i.d. error structure is assumed. Also, our theory provides finite sample bounds and is derived under weaker assumptions, requiring different techniques of proof. In particular, the previous,
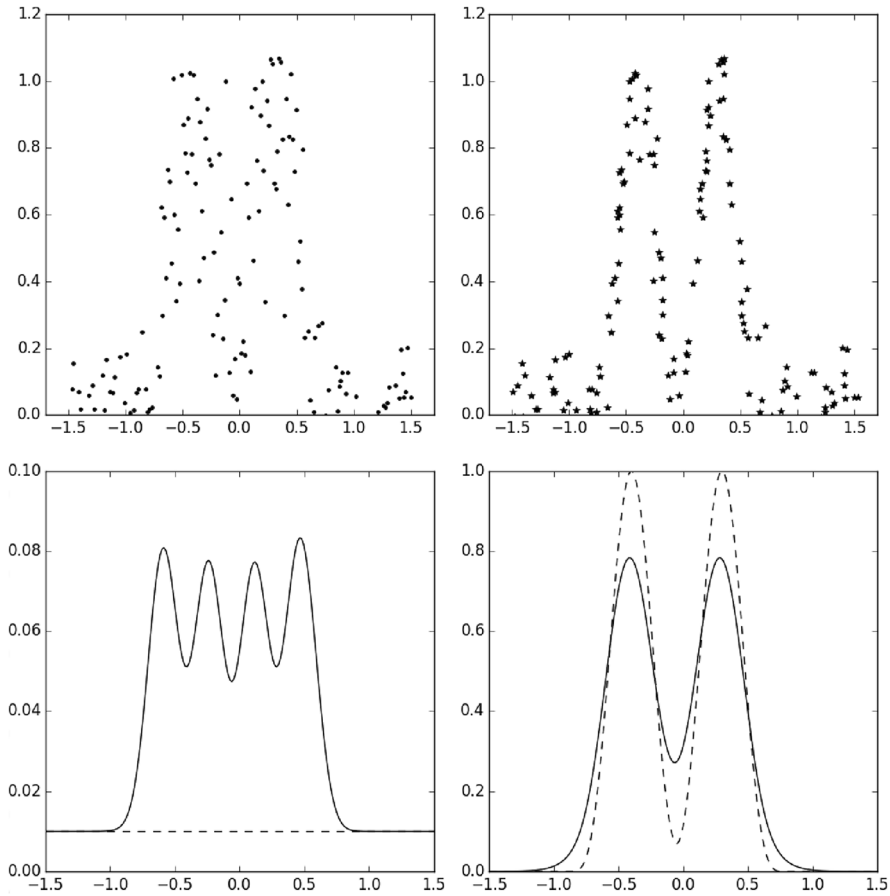
Fig. 1 Alleged data points (w_j, Y_j), actual data points $(w_j + \Delta_j, Y_j)$, a comparison between g (dashed line) and γ (solid line) and a comparison between σ^2 (dashed line) and v^2 (solid line) (clockwise from upper left to lower left)

asymptotic, results are derived under a stronger assumption on the convolution function f_Δ. To estimate g, we estimate the Fourier transform of γ,

$$\Phi_\gamma(t) = \int e^{itw} \gamma(w) \, dw, \quad \text{by} \quad \hat{\Phi}_\gamma(t) = \frac{1}{na_n} \sum_{j=-n}^{n} Y_j e^{itw_j}.$$

An estimator for g is then given by

$$\hat{g}_n(x;h) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \Phi_k(ht) \frac{\hat{\Phi}_\gamma(t)}{\Phi_{f_\Delta}(-t)} \, dt. \tag{6}$$

Here $h > 0$ is a smoothing parameter called the bandwidth, and Φ_k is the Fourier transform of a bandlimited kernel function k that satisfies Assumption 2 below. Notice that both Φ_γ and Φ_{f_Δ} tend to zero as $|t| \to \infty$ such that estimation of Φ_γ in (6) introduces instabilities for large values of $|t|$. Since the kernel k is bandlimited, the function Φ_k is compactly supported and the factor $\Phi_k(ht)$ discards large values of t, therefore serving as regularization. The estimator can be rewritten in kernel form as follows:

$$\hat{g}_n(x;h) = \frac{1}{na_n h} \sum_{j=-n}^{n} Y_j K\left(\frac{w_j - x}{h};h\right),$$

where the deconvolution kernel $K(\cdot;h)$ is given by

$$K(w;h) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itw} \frac{\Phi_k(t)}{\Phi_{f_\Delta}(-t/h)} dt. \tag{7}$$

3 Theory

By $\mathcal{W}^m(\mathbb{R})$ we denote the Sobolev spaces $\mathcal{W}^m(\mathbb{R}) = \{g \mid \|\Phi_g(\cdot)\langle\,\cdot\,\rangle^m\|_2 < \infty\}$, $m > 0$, where we recall that $\langle w \rangle := (1 + w^2)^{\frac{1}{2}}$ for $w \in \mathbb{R}$. We shall require the following assumptions.

Assumption 1 The functions g and f_Δ satisfy

(i) $g \in \mathcal{W}^m(\mathbb{R}) \cap L^r(\mathbb{R})$ for all $r \leq M$ and for some $m > 5/2$,
(ii) f_Δ is a bounded, continuous, square-integrable density,
(iii)   The function γ decays sufficiently fast in the following sense:

$$\int_{|z|>1/a_n} \langle z \rangle^s |\gamma(z)|^2 \, dz < \infty,$$

for some $s > 1/2$, where s may depend on n.

Assumption 1 (i) stated above is a smoothness assumption on the function g. In Lemma 1 in Sect. 6.1 we list the properties of g that are frequently used throughout this paper and that are implied by this assumption. In particular, by Sobolev embedding, $m > 5/2$ implies that the function g is twice continuously differentiable, which is used in the proof of Lemma 5. Assumption 1 (iii) will be discussed in more detail in Section III of the supplementary material.

Assumption 2 Let $\Phi_k \in C^2(\mathbb{R})$ be symmetric, $\Phi_k(t) \equiv 1$ for all $t \in [-D, D]$, $0 < D < 1, |\Phi_k(t)| \leq 1$ for $D < |t| \leq 1$, and $\Phi_k(t) = 0, |t| > 1$.

In contrast to kernel-estimators in a classical nonparametric regression context, the kernel $K(\cdot;h)$, defined in (7), depends on the bandwidth h and hence on the sample size via the factor $1/\Phi_{f_\Delta}(-t/h)$. For this reason, the asymptotic behavior of $K(\cdot;h)$ is determined by the properties of the Fourier transform of the error-density f_Δ. The following assumption on Φ_{f_Δ} is standard in the nonparametric deconvolution context (see, e.g., Kato and Sasaki 2019; Schmidt-Hieber et al. 2013) and will be relaxed in Sect. 5 below.

Assumption 3 Assume that $\Phi_{f_\Delta}(t) \neq 0$ for all $t \in \mathbb{R}$ and that there exist constants $\beta > 0$ and $0 < c < C, 0 < C_S$ such that

$$c\langle t\rangle^{-\beta} \leq |\Phi_{f_\Delta}(t)| \leq C\langle t\rangle^{-\beta} \quad \text{and} \quad \left|\Phi_{f_\Delta}^{(1)}(t)\right| \leq C_S\langle t\rangle^{-\beta-1}. \tag{S}$$

A standard example of a density that satisfies Assumption 3 is the Laplace density with parameter $a > 0$,

$$f_{\Delta,0}(a;x) = \frac{a}{2}\mathrm{e}^{-a|x|} \quad \text{with} \quad \Phi_{f_{\Delta,0}}(a;t) = \langle t/a\rangle^{-2}. \tag{8}$$

In this case we find $\beta = 2, C = a^2 \vee 1, c = a^2 \wedge 1$ and $C_S = 2/a^2 \vee 2a^2$.

Remark 1 Our asymptotic theory cannot accommodate the case of exponential decay of the Fourier transform of the density f_Δ, as the asymptotic behavior of the estimators in the supersmooth case and the ordinary smooth case differs drastically. While for the ordinary smooth case considered here $\hat{g}(x)$ and $\hat{g}(y)$ are asymptotically independent if $x \neq y$, convolution with a supersmooth distribution is no longer local and causes dependencies throughout the domain. This leads to different properties of the suprema $\sup_{x\in[0,1]} |\hat{g}(x;h) - \mathbb{E}[\hat{g}(x;h)]|$, which play a crucial role in the construction of our confidence bands. In particular, the asymptotics strongly depend on the exact decay of the characteristic function Φ_{f_Δ} and needs a treatment on a case to case basis (more details on the latter issue can be found in van Es and Gugushvili 2008).

3.1 Simultaneous inference

Our main goal is to derive a method to conduct uniform inference on the regression function g, which is based on a Gaussian approximation to the maximal deviation of $\hat{g}_n$ from g. We consider the usual decomposition of the difference $g(x) - \hat{g}_n(x;h)$ into deterministic and stochastic parts, that is

$$g(x) - \hat{g}_n(x;h) = g(x) - \mathbb{E}[\hat{g}_n(x;h)] + \mathbb{E}[\hat{g}_n(x;h)] - \hat{g}_n(x;h),$$

where

$$\hat{g}_n(x;h) - \mathbb{E}[\hat{g}_n(x;h)] = \frac{1}{na_nh} \sum_{j=-n}^{n} \eta_j K\left(\frac{w_j - x}{h};h\right). \tag{9}$$

If the bias, the rate of convergence of which is given in Lemma 5 in Sect. 6, is taken care of by choosing an undersmoothing bandwidth h, the stochastic term (9) in the above decomposition dominates.

Theorem 1 below is the basic ingredient for the construction of the confidence statements under Assumption 3. It guarantees that the random sum (9) can be approximated by a distribution free Gaussian version, uniformly with respect to $x \in [0, 1]$, that is, a weighted sum of independent, normally distributed random variables such that the required quantiles can be estimated from this approximation. In the following assumption conditions on the bandwidth and the design parameter are listed which will be needed for the theoretical results.

Assumption 4

(i) $\ln(n)n^{\frac{2}{M}-1}/(a_n h) + \frac{h}{a_n} + \ln(n)^2 h + \ln(n)a_n + \frac{1}{na_n h^{1+2\beta}} = o(1)$,

(ii) $\sqrt{na_n h^{2m+2\beta}} + \sqrt{na_n^{2s+1}h^2} + 1/\sqrt{na_n h^2} = o(1/\sqrt{\ln(n)})$.

The following example is a short version of a lengthy discussion given in Section III of the supplementary material. More details can be found there.

Example 1 When estimating a function g nonparametrically, the bandwidth is typically chosen such that bias and standard deviation of the estimator are balanced. It is well known from the literature that such a choice will not result in asymptotically correct coverage of the true function by our confidence bands. Therefore, we aim at choosing the bandwidth h slightly smaller than this, which is called *undersmoothing* (see, e.g., Giné and Nickl 2010; Chernozhukov et al. 2014; Kato and Sasaki 2018 or Yano and Kato 2020 for recent theoretical studies that employ undersmoothing and see also Sect. 3.3 for a more detailed discussion of the issue of bandwidth selection). The conditions listed in Assumption 4 are satisfied if h is such an undersmoothing bandwidth. To see this, consider the case of a function $g \in \mathcal{W}^m(\mathbb{R}), m > 5/2$, of bounded support, f_{Δ} as in (8) and $\mathbb{E}[\epsilon_1^4] < \infty$. Then $\beta = 2$ and Assumption 1 (iii) holds for $s = s_n = O(1/(\ln \ln(n)a_n))$ such that a_n can be chosen to be of logarithmic order in n (which is shown in Section III of the supplementary material). In the current setting, applying Theorem 1 and Lemma 5, a bias variance trade-off yields

$$h^* \sim \left(\frac{\ln(n)}{na_n} \right)^{\frac{1}{2(\beta+m)}}. \tag{10}$$

If we choose an undersmoothing bandwidth of order $h \sim h^*/\ln(n), a_n \sim 1/\ln(n)$ and $s_n \sim \ln(n)/\ln \ln(n)$, Assumption 4 (ii) becomes $\ln(n)^{1-2m-2\beta}(1 + o(1)) = o(1)$, which is satisfied since $m > 5/2$ and $\beta > 0$.

The first term in Assumption 4 (i) stems from the Gaussian approximation and becomes less restrictive if the number of existing moments, M, of the errors ϵ_i increases. The last term in (i) guarantees that the variance of the estimator tends to zero. The terms in between are only weak requirements and are needed for the

estimation of certain integrals. Assumption 4 (ii) guarantees that the bias is negligible under Assumption 3. The first term guarantees undersmoothing, the second term stems from the fact that only observations from the finite grid $[-1/a_n, 1/a_n]$ are available, while the third term accounts for the discretization bias. It is no additional restriction if $\beta > 1/2$. For a given interval $[a, b]$, recall that $\|f\| = \|f\|_{[a,b]}$ denotes the supremum norm of a bounded function on $[a, b]$.

Theorem 1 *Let Assumptions 2–3 and 4 (i) be satisfied. For some given interval $[a, b]$ of interest, let $\hat{v}_n$ be a nonparametric estimator of the standard deviation in model (4) such that $\hat{v} > \sigma/2$ and*

$$\mathbb{P}\left(\left\|\frac{1}{\hat{v}} - \frac{1}{v}\right\|_{[a,b]} > \frac{n^{2/M}}{\sqrt{na_nh}}\right) = o(1). \tag{11}$$

1. *There exists a sequence of independent standard normally distributed random variables $(Z_n)_{n\in\mathbb{Z}}$ such that for*

$$
\begin{aligned}
\mathbb{D}_n(x) &:= \frac{\sqrt{na_n h}h^\beta}{\hat{v}(x)}\left(\hat{g}_n(x;h) - \mathbb{E}[\hat{g}_n(x;h)]\right), \\
\mathbb{G}_n(x) &:= \frac{h^\beta}{\sqrt{na_n h}}\sum_{j=-n}^{n} Z_j K\left(\frac{w_j-x}{h};h\right),
\end{aligned}
\tag{12}
$$

we have that for all $\alpha \in (0, 1)$

$$\left|\mathbb{P}\left(\|\mathbb{D}_n\| \le q_{\|\mathbb{G}_n\|}(\alpha)\right) - \alpha\right| \le r_{n,1}, \tag{13}$$

where $q_{\|\mathbb{G}_n\|}(\alpha)$ is the α-quantile of $\|\mathbb{G}_n\|$ and for some constant $C > 0$

$$r_{n,1} = \mathbb{P}\left(\left\|\frac{1}{\hat{v}} - \frac{1}{v}\right\| > \frac{n^{2/M}}{\sqrt{na_nh}}\right) + C\left(\frac{1}{n} + \frac{n^{2/M}\sqrt{\ln(n)^3}}{\sqrt{na_nh}}\right).$$

2. *If, in addition, Assumption 4 (ii) and Assumption 1 are satisfied, $\mathbb{E}[\hat{g}_n(x;h)]$ in (12) can be replaced by $g(x)$ with an additional error term of order $r_{n,2} = \sqrt{na_n h^{2m+2\beta}} + \sqrt{na_n^{2s+1}h^2} + 1/\sqrt{na_n h^2}$.*

In particular, Theorem 1 implies that $\lim_{n\to\infty}\mathbb{P}\left(\|\mathbb{D}_n\| \le q_{\|\mathbb{G}_n\|}(\alpha)\right) = \alpha$ for all $\alpha \in (0, 1)$. Regarding assumption (11), properties of variance estimators in a heteroscedastic nonparametric regression model are discussed in Wang et al. (2008).

The following theorem is concerned with suitable grid widths of discrete grids $\mathcal{X}_{n,m} \subset [a, b]$ such that the maximum over $[a, b]$ and the maximum over $\mathcal{X}_{n,m}$ behave asymptotically equivalently.

Theorem 2 *For some given interval $[a, b]$ of interest, let $\mathcal{X}_{n,m} \subset [a, b]$ a grid of points $a = x_{0,n} \le x_{1,n} \le \cdots \le x_{m,n} = b$. Let $\|f\|_{\mathcal{X}_{n,m}} := \max_{x\in\mathcal{X}_{n,m}} |f(x)|$. If*

$|\mathcal{X}_{n,m}| := \max_{1 \le i \le m} |x_{i,m} - x_{i-1,m}| \le h^{1/2}/(na_n)^{1/2}$, *i.e., the grid is sufficiently fine, under the assumptions of Theorem* 1, *the following holds.*

1. *For all $\alpha \in (0,1)$*

$$\left| \mathbb{P}\left(\|\mathbb{D}_n\| \le q_{\|\mathbb{G}_n\|_{\mathcal{X}_{n,m}}}(\alpha) \right) - \alpha \right| \le r_{n,1}(1 + o(1)). \tag{14}$$

2. *If, in addition, Assumption* 4 *(ii) and Assumption* 1 *are satisfied, $\mathbb{E}[\hat{g}_n(x;h)]$ in* (14) *can be replaced by $g(x)$ with an additional error term of order $r_{n,2} = \sqrt{na_n h^{2m+2\beta}} + \sqrt{na_n^{2s+1}h^2} + 1/\sqrt{na_n h^2}$.*

3.2 Construction of the confidence sets

In this section we present an algorithm which can be used to construct uniform confidence sets based on Theorem 1. Let $\mathbb{G}_n(x)$ be the statistic defined in (12). In order to obtain quantiles that guarantee uniform coverage of a confidence band, generate B times $\|\mathbb{G}_n\|_{\mathcal{X}_{n,m}}$, where $|\mathcal{X}_{n,m}| = o(h^{3/2}a_n^{1/2}/\ln(n))$ (see Theorem 2), that is, calculate $\hat{v}_n(x)$ for $x \in \mathcal{X}_{n,m}$, generate B times $2n+1$ realizations of independent, standard normally distributed random variables $Z_{1,j}, \ldots, Z_{2n+1,j}$, $j = 1, \ldots, B$. Calculate $\mathbf{M}_{n,j} := \max_{x \in \mathcal{X}_{n,m}} |\mathbb{G}_{n,j}(x)|$. Estimate the $(1 - \alpha)$-quantile of $\|\mathbb{G}_n\|$ from $\mathbf{M}_{n,1}, \ldots, \mathbf{M}_{n,B}$ and denote the estimated quantile by $\hat{q}_{\|\mathbb{G}_n\|_{\mathcal{X}_{n,m}}}(1 - \alpha)$. From Theorem 1 we obtain the confidence band

$$\hat{g}_n(x;h) \pm \hat{q}_{\|\mathbb{G}_n\|_{\mathcal{X}_{n,m}}}(1 - \alpha) \frac{\hat{v}_n(x)}{\sqrt{na_n}h^{1/2+\beta}}, \quad x \in [a,b]. \tag{15}$$

Remark 2

(i) Given a suitable estimator for the variance v^2, Theorem 1 and Example 1 imply that the coverage error of the above bands will be of order

$$n^{2/M}\sqrt{\ln(n)}/\sqrt{na_n h} + \sqrt{na_n h^{2m+2\beta}}, \tag{16}$$

provided that both functions g and f_Δ decay sufficiently fast. The first term in (16) is determined by the accuracy of the Gaussian approximation and is negligible if the distribution of the errors ϵ_i possesses sufficiently many moments, while the second term is of order $\ln(n)^{1-2m-2\beta}$ if a bandwidth of order $h \sim h^*/\ln(n)$ is chosen, where h^* is given in (10).

(ii) If the surrogate function $\mathbb{E}[\hat{g}_n]$ is considered instead of g, the coverage error of the confidence bands decays polynomially fast.

(iii) If the function γ decays only polynomially fast, the design parameter a_n impacts the convergence rates, as it can no longer be chosen to be of logarithmic order

only. This is due to the fact that, in this case, sufficient amount of information on the function g is only gathered slowly as n increases. In this case, the term $\sqrt{na_n^{2s+1}h^2}$ in the coverage error is no longer negligible, which seems to be characteristic for Berkson EIV models.

Remark 3 In nonparametric regression without errors-in-variables the widths of uniform confidence bands are of order $\sqrt{\ln(n)}/\sqrt{nh}$ (see, e.g., Neumann and Polzehl 1998). Our bands (15) are wider by the factor $1/(a_n h^\beta)$ which is due to the ill-posedness (β) and the, possibly slow, decay of γ (expressed in terms of a_n).

3.3 Choice of the bandwidth

Selecting the rate optimal bandwidth does not lead to asymptotically correct coverage for the true function, as is well known in the literature (see, e.g., Bickel and Rosenblatt 1973; Hall 1992). One way of overcoming this difficulty would be to introduce a bias correction, i.e., subtracting a suitably estimated bias from $\hat{g}$ (see, e.g., Hall 1992; Eubank and Speckman 1993). This is, however, quite intricate in practice as it requires the estimation of derivatives of the function g. In this paper, we follow the very common strategy of undersmoothing. In the remaining part of this section, we sketch a way of choosing the bandwidth accordingly. For the choice of the bandwidth, Giné and Nickl (2010) (see also Chernozhukov et al. 2014) convincingly demonstrated how to use Lepski's method to adapt to unknown smoothness when constructing confidence bands. In our framework, choose an exponential grid of bandwidths $h_k = 2^{-k}$ for $k \in \{k_l, \ldots, k_u\}$, with $k_l, k_u \in \mathbb{N}$ being such that $2^{-k_u} \simeq 1/n$ and $2^{-k_l} \simeq \big((\ln n)/(na_n)\big)^{1/(\beta+\bar{m})}$ and where $\bar{m}$ corresponds to the maximal degree of smoothness to which one intends to adapt. Then for a sufficiently large constant $C_L > 0$ choose the index k according to

$$\hat{k} = \min\left\{k \in \{k_l, \ldots, k_u\} \mid \|\hat{g}(\cdot;h_k) - \hat{g}(\cdot;h_l)\| \le C_L\left(\frac{\ln n}{n\,a_n\,h_l^{1+2\beta}}\right)^{1/2}\right.$$

$$\left. \forall\, k \le l \le k_u\right\},$$

and choose an undersmoothing bandwidth according as $\hat{h} = h_{\hat{k}}/\ln n$. A result in analogy to Giné and Nickl (2010) would imply that under an additional *self-similarity* condition on the regression function g, using $\hat{h}$ in (15) produces confidence bands of width $(\ln n/(n\,a_n))^{(m-1/2)/(\beta+m)}(\ln n)^{\beta+1/2}$ if g has smoothness m. Technicalities in our setting would be even more involved due to the truncated exhaustive design involving the parameter a_n. Therefore, we refrain from going into the technical details. In the subsequent simulations we use a simplified bandwidth selection rule which, however, resembles the Lepski method.

Recently, in the context of Gaussian process regression, Yang et al. (2017) propose a fixed target smoothness approach in their frequentist analysis of Bayesian credible bands. The idea is to fix the target smoothness and to display the asymptotic behavior of the confidence bands in over-smooth, smoothness matching and undersmooth cases. While this approach would shed more light on the precise behavior of

our confidence bands in the over-smooth and smoothness matching case, we focus on the under-smooth case, as only in this case the confidence bands are asymptotically valid, guaranteeing the nominal coverage.

4 Simulations

In this section we investigate the numerical performance of our proposed methods in finite samples. We consider several different computational scenarios. As regression functions we consider $g_a(x) = (1 - 4(x - 0.1)^2)^5 I_{[0,1]}(2|x - 0.1|)$, and $g_b(x) = (1 - 4(x + 0.4)^2)^5 I_{[0,1]}(2|x + 0.4|) + (1 - 4(x - 0.3)^2)^5 I_{[0,1]}(2|x - 0.3|)$.

For the error distribution f_Δ we chose two densities of a Laplace distribution as defined in (8) with $a = \frac{0.1}{\sqrt{2}}$ and $a = \frac{0.05}{\sqrt{2}}$, i.e., standard deviations $\sigma_\delta = 0.1$ and $\sigma_\delta = 0.05$, respectively. Finally, $a_n = 2/3$ in all simulations discussed below. Our estimation is based on an application of the Fast Fourier transform implemented in python/scipy. The integration used a damped version of a spectral cut off with cut-off function $I(\omega) = 1 - \exp(-\frac{1}{(\omega \cdot h)^2})$ in spectral space.

Construction of the confidence bands requires the selection of a regularization parameter for the estimator $\hat{g}$. In our simulations, we have chosen this parameter by a visual inspection of a sequence of estimates for the regularization parameter, covering a range from over- to under-smoothing, see Fig. 2. We chose the minimal regularization parameter for which the estimates do not change systematically in overall amplitude, but appear to only exhibit additional random fluctuations at smaller values of the parameter. In the case shown here, we chose a regularization parameter of 0.27. The same procedure was followed for other combinations of n, σ, σ_δ and signal g_a resp. g_b) and the results can be found in Table 1. This regularization parameter was then kept fixed for each combination of n, σ, σ_δ and signal $g \in \{g_a, g_b\}$. Figures 3 and 4 show four random examples each for estimates of g_a and g_b, respectively, together with the associated confidence bands from 250 bootstrap simulations. Solid lines represent the true signal g_a and g_b and dashed lines the estimates $\hat{g}_n$ together with their associated confidence bands. Again, in both cases, $n = 100$, $\sigma = 0.1$ and $\sigma_\delta = 0.1$. Next, we discuss the practical performance of the bootstrap confidence bands in more detail for the first scenario, where the model is correctly specified and the errors in the predictors are taken into account as well. The results are shown in Tables 2 and 3 for the simulated rejection probabilities (one minus the coverage probability) at a nominal value of 5% and for the (average) width of the confidence bands. In

	$n = 100$ $\sigma = \sigma_\delta = 0.1$	$n = 100$ $\sigma = \sigma_\delta = 0.05$	$n = 750$ $\sigma = \sigma_\delta = 0.1$	$n = 750$ $\sigma = \sigma_\delta = 0.05$
g_a	0.25	0.24	0.21	0.12
g_b	0.20	0.22	0.22	0.11

Table 1 Regularization parameter used in the subsequent simulations

See text for details on its selection

Table 2 Simulated rejection probabilities for bootstrap confidence bands

	$n = 100$ (%) $\sigma = \sigma_\delta = 0.1$	$n = 100$ (%) $\sigma = \sigma_\delta = 0.05$	$n = 750$ (%) $\sigma = \sigma_\delta = 0.1$	$n = 750$ (%) $\sigma = \sigma_\delta = 0.05$
g_a	5.8	7.2	5.1	5.6
g_b	1.8	5.3	5.0	5.0

Table 3 Average width of bootstrap confidence bands

	$n = 100$ $\sigma = \sigma_\delta = 0.1$	$n = 100$ $\sigma = \sigma_\delta = 0.05$	$n = 750$ $\sigma = \sigma_\delta = 0.1$	$n = 750$ $\sigma = \sigma_\delta = 0.05$
g_a	0.44	0.16	0.21	0.14
g_b	0.86	0.28	0.24	0.22
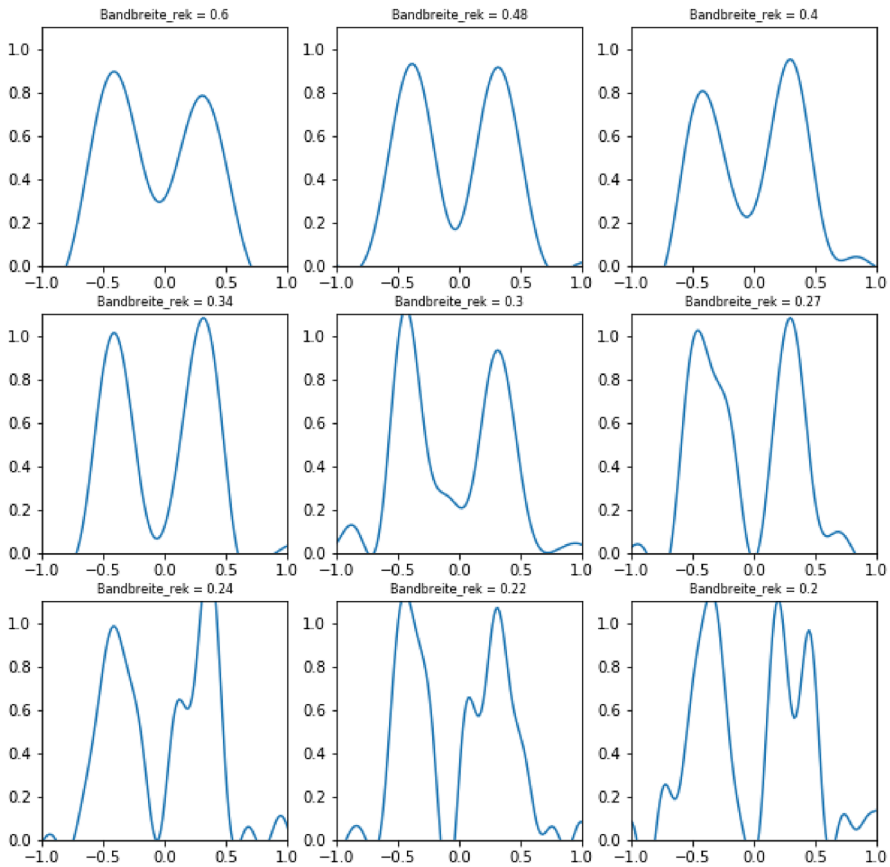
Fig. 2 Sequence of estimates for increasing regularization parameter from a random sample of observations of signal g_b, with $n = 100$ and $\sigma = \sigma_\delta = 0.1$
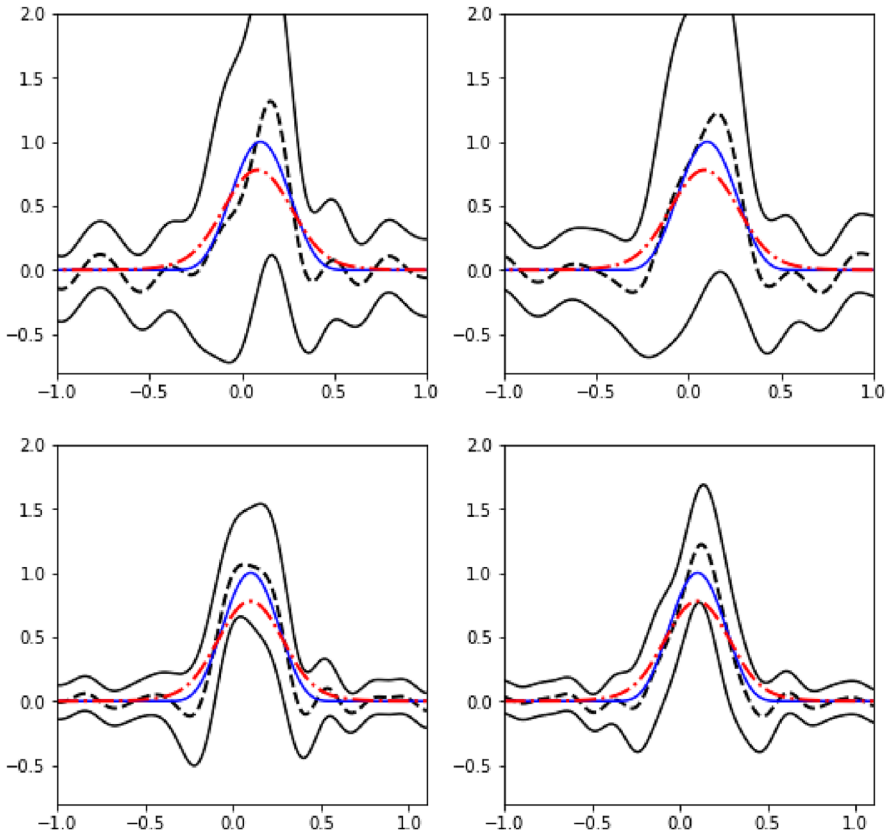
Fig. 3 True signal (solid line), observable signal (dash-dotted line) and estimates and associated confidence bands (dashed lines) from four random samples for g_a for $n = 100$ (top) and $n = 750$ (bottom) and $\sigma = \sigma_\delta = 0.1$

all cases, we performed simulations based on 500 random samples of data and nominal rejection probability 5% (i.e., confidence bands with nominal coverage probability of 95%). For each of these data samples, we repeated 250 times the following scenario.

First, we determined the width of the confidence bands from 250 bootstrap simulations and second, we evaluated whether the confidence bands cover the true signal everywhere in an interval of interest. The numbers shown in the table give the percentage of rejections, i.e., of where the confidence bands do not overlap the true signal everywhere in such an interval. Here, the intervals of interest are chosen as an interval where the respective signal is significantly different from 0. The intention of this is that in many practical applications the data analyst is particularly interested in those parts of the signal. Here, we chose the interval $[-0.7, 0.6]$ as 'interval of interest' for g_a and g_b. From the tables we conclude that the method performs well, particularly for $n = 750$, where the confidence bands are substantially less wide.
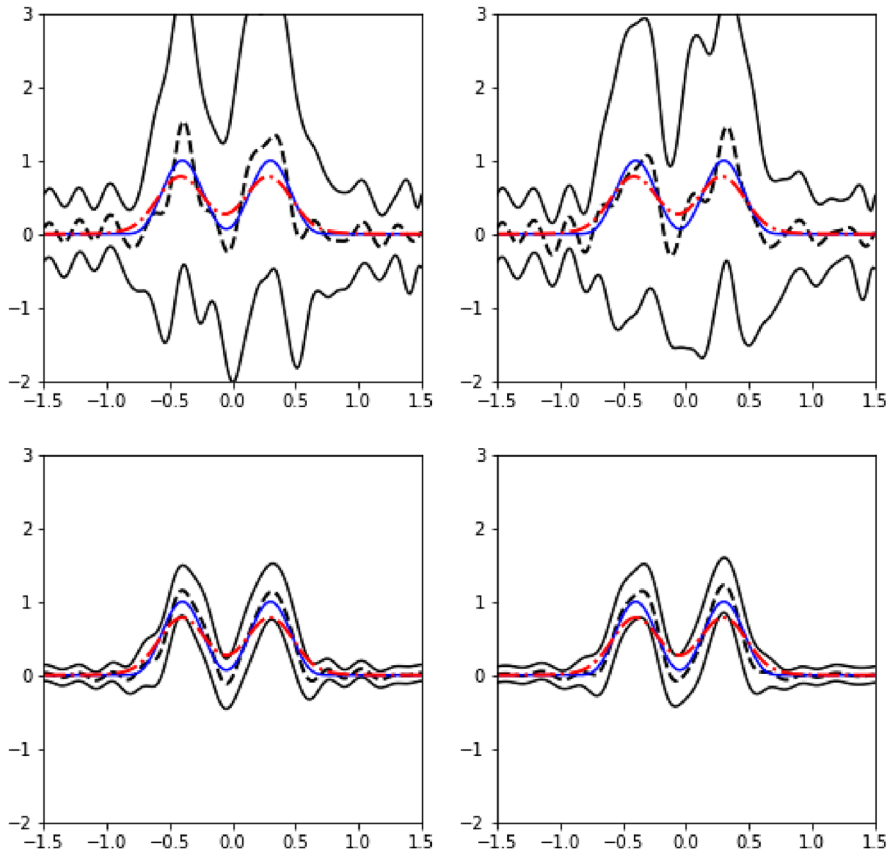
Fig. 4 True signal (solid line), observable signal (dash-dotted line) and estimates and associated confidence bands (dashed lines) from four random samples for g_b for $n = 100$ (top) and $n = 750$ (bottom) and $\sigma = \sigma_\delta = 0.1$

5 Extensions

The following assumption is less restrictive than Assumption 3, (S).

Assumption 5 Assume that $\Phi_{f_\Delta}(t) \neq 0$ for all $t \in \mathbb{R}$ and that there exist constants $\beta > 0$ and $0 < c < C, 0 < C_W$ such that

$$c\langle t\rangle^{-\beta} \leq |\Phi_{f_\Delta}(t)| \leq C\langle t\rangle^{-\beta} \quad \text{and} \quad |\Phi_{f_\Delta}^{(1)}(t)| \leq C_W\langle t\rangle^{-\beta}. \tag{W}$$

An example for a density that satisfies Assumption 5 but not Assumption 3 is given by the mixture

$$f_{\Delta,1}(1;x) = \frac{\lambda}{2}f_{\Delta,0}(1;x - \mu) + \frac{\lambda}{2}f_{\Delta,0}(1;x + \mu) + (1 - \lambda)f_{\Delta,0}(1;x), \tag{17}$$

where $\lambda \in (0, 1/2)$ and $\mu \neq 0$, and $f_{\Delta,0}$ is the Laplace density defined in (8). We find $\Phi_{f_{\Delta,1}}(t) = (1 - \lambda + \lambda \cos(\mu t))\langle t \rangle^{-2}$, which yields $\beta = 2$, $c = 1 - 2\lambda$ and $C_W = \lambda\mu + 4$. Technically, Assumption 3, (S) allows for sharper estimates of the tails of the deconvolution kernel (7) than does Assumption 4, (W), see Lemma 4 in Sect. 6. In this case we have to proceed differently as the approximation via a distribution free process such as $\mathbb{G}_n$ can no longer be guaranteed and we can only find a suitable Gaussian approximation depending on the standard deviation ν.

Roughly speaking, we approximate $\mathbb{D}_n(x)$ in (12) by the process

$$\widetilde{\mathbb{G}}_n(x) = \frac{\sqrt{n a_n h^{1+2\beta}}}{h \, \tilde{\nu}_n(x)} \sum_j \tilde{\nu}_n(\omega_j) Z_j K\left(\frac{w_j - x}{h}; h\right), \tag{18}$$

for a variance estimator $\tilde{\nu}_n$ on growing intervals $|x| \leq n a_n (1 - \delta)$ for some $\delta > 0$. We then replace the quantiles involving $\mathbb{G}_n$ in (13), (14) and in (15) by (the conditional quantiles given the sample) of $\widetilde{\mathbb{G}}_n$. Our theoretical developments involve a sample splitting, hence are somewhat cumbersome so that details are deferred to the supplementary material. We have also simulated a version of the bootstrap for the extended model. However, as simulations show, the results are clearly not as good as for the more restrictive assumptions on f_Δ. We have used $f_{\Delta,1}(x) = \frac{\lambda}{2} \cdot f_{\Delta,0}(a; x - 0.3) + (1 - \lambda) \cdot f_{\Delta,0}(a; x) + \frac{\lambda}{2} \cdot f_{\Delta,0}(a; x + 0.3)$, with $f_{\Delta,0}$ again the Laplace density defined in (8), $a = 0.05/\sqrt{2}$ and $\lambda = 0.2$. For the signal g_a in Sect. 4 we find confidence band widths of 0.686 and 0.462 for $n = 100$ and $n = 750$, respectively, at simulated coverage probabilities of 6.3% and 4.5% and bandwidths of 0.59 and 0.32, for $\sigma = \sigma_\delta = 0.1$.

6 Auxiliary lemmas

The proofs for the results in this section are given in Sect. 8.

6.1 Properties of g, ν and K

Assumption 1 stated above is basically a smoothness assumption on the function g. In the following lemma we list the properties of g that are frequently used throughout this paper and that are implied by Assumption 1.

Lemma 1 *Let Assumption 1 hold.*

(i)   *The function g is twice continuously differentiable.*
(ii) *The function g has uniformly bounded derivatives: $\|g^{(j)}\|_\infty < \infty$, $j \leq 2$.*

Given Assumption 1 (ii), the properties of the function g given in Lemma 1 are transferred to $\gamma = g * f_\Delta$. This is made precise in the following lemma.

Lemma 2 *Let Assumption 1 hold.*

(i) *The function $\gamma = g * f_\Delta$ is twice continuously differentiable with derivatives $\gamma^{(j)} = g^{(j)} * f_\Delta$.*

(ii) *$\gamma \in \mathcal{W}^m(\mathbb{R})$.*

(iii) *The function γ has uniformly bounded derivatives: $\|\gamma^{(j)}\|_\infty < \infty$, $j \leq 2$.*

Furthermore, the variance function v^2, defined in (5), is a function that depends on f_Δ, γ and g. The following lemma lists the properties of v^2, which are implied by the previous Lemmas 1 and 2, and that are frequently used throughout this paper.

Lemma 3 *Let Assumption 1 hold.*

(i) *The variance function v^2 is uniformly bounded and bounded away from zero.*

(ii) *The variance function v^2 is twice continuously differentiable with uniformly bounded derivatives.*

For the tails of the kernel, we have the following estimate.

Lemma 4 *For any $a > 1$ and $x \in [0, 1]$ we have*

$$\int_{\{|z|>a\}} \left(K\left(\frac{z-x}{h};h\right) \right)^2 \mathrm{d}z \leq \mathcal{C}\frac{2a}{a^2 - x^2} \cdot \begin{cases} h^{-2\beta}, & \text{if Ass. 4, (W) holds,} \\ h^{-2\beta+2}, & \text{if Ass. 3, (S) holds.} \end{cases}$$

Lemma 5 *Let Assumptions 1 and 2 be satisfied. Further assume that $h/a_n \to 0$ as $n \to \infty$.*

(i) *Then for the bias, we have that*

$$\sup_{x \in [0,1]} \left| \mathbb{E}[\hat{g}_n(x;h)] - g(x) \right| = O\left(h^{m-\frac{1}{2}} + \frac{1}{na_n h^{\beta+\frac{3}{2}}} \right) + \begin{cases} O\left(a_n^{s+1/2}h^{1-\beta}\right), & \text{Ass. 3, (S),} \\ O\left(a_n^{s+1/2}h^{-\beta}\right), & \text{Ass. 5, (W).} \end{cases}$$

(ii) a) *For the variance if Assumption 5, (W) holds and $na_n h^{1+\beta} \to \infty$, then we have that*

$$\frac{\sigma^2}{2C\pi}(1 + O(a_n)) \leq na_n h^{1+2\beta}\mathrm{Var}[\hat{g}_n(x;h)] \leq \frac{2^\beta \sup_{x \in \mathbb{R}} v^2(x)}{c\pi}.$$

(ii) b) *If actually Assumption 3, (S) holds and $na_n h^{1+\beta} \to \infty$, then*

$$\frac{v^2(x)}{C\pi}(1 + O(a_n)) \leq na_n h^{1+2\beta}\mathrm{Var}[\hat{g}_n(x;h)] \leq \frac{v^2(x)}{c\pi}(1 + O(h/a_n)).$$

Here c, C and β are the constants from Assumption 5 respectively (3).

6.2 Maxima of Gaussian processes

Let $\{\mathbb{X}_t \mid t \in T\}$ be a Gaussian process and ρ be a semi-metric on T. The packing number $D(T, \delta, \rho)$ is the maximum number of points in T with distance ρ strictly larger than $\delta > 0$. Similarly to the packing numbers, the covering numbers $N(T, \delta, \rho)$ are defined as the number of closed ρ-balls of radius δ, needed to cover T. Let further $d_{\mathbb{X}}$ denote the standard deviation semi-metric on T, that is, $d_{\mathbb{X}}(s, t) = \left(\mathbb{E}\left[|\mathbb{X}_t - \mathbb{X}_s|^2\right]\right)^{1/2}$ for $s, t \in T$. In the following, we drop the subscript if it is clear which process induces the pseudo-metric d.

Lemma 6 *There exist constants $C_E, C_{\widehat{E}} \in (0, \infty)$ such that*

(i) $N(T, \delta, d_{\mathbb{G}_n}) \leq D(T, \delta, d_{\mathbb{G}_n}) \leq C_E / (h^{3/2} a_n^{1/2} \delta)$.

(ii) $N(T, \delta, d_{\mathbb{G}_n^{\widehat{K}}}) \leq D(T, \delta, d_{\mathbb{G}_n^{\widehat{K}}}) \leq C_{\widehat{E}} / (h^{3/2} a_n^{1/2} \delta)$, where $\mathbb{G}_n^{\widehat{K}}$ is defined as $\mathbb{G}_n$ with K replaced by $\widehat{K}$, where $\widehat{K}(z; h) = zK(z; h)$.

Lemma 7 *Let $(\mathbb{X}_{n,1}(t), t \in T)$ and $(\mathbb{X}_{n,2}(t), t \in T)$ be almost surely bounded, centered Gaussian processes on a compact index set T and suppose that for any fixed $n \in \mathbb{N}$ $\text{diam}_{d_{\mathbb{X}_{n,1}}}(T) > D_n > 0$. If*

$$d_{\mathbb{X}_{n,1}}(s, t) \leq d_{\mathbb{X}_{n,2}}(s, t) \; \forall s, t \in T \quad \text{and} \quad \mathbb{E}[\|\mathbb{X}_{n,2}\|] = o(1/\sqrt{\ln(n)}),$$

we have that

$$\mathbb{E}\left[\|\mathbb{X}_{n,1}\|\right] \leq 2\mathbb{E}\left[\|\mathbb{X}_{n,2}\|\right] \quad \text{and hence} \quad \|\mathbb{X}_{n,1}\| = o_{\mathbb{P}}(1/\sqrt{\ln(n)}).$$

7 Proofs of Theorems 1 and 2

In the following, the letter $\mathcal{C}$ denotes a generic, positive constant, whose value may vary form line to line. The abbreviations R_n and $\widetilde{R}_n$, possibly with additional subscripts, are used to denote remainder terms and their definition may vary from proof to proof.

Proof of Theorem 1 We first prove assertion (i). Let $\rho_n := n^{2/M} \ln(n) / \sqrt{n a_n h}$ and notice that

$$\mathbb{P}\left(\|\mathbb{D}_n\| \leq q_{\|\mathbb{G}_n\|}(\alpha)\right) \leq \mathbb{P}\left(\|\mathbb{G}_n\| \leq q_{\|\mathbb{G}_n\|}(\alpha) + \rho_n\right) + \mathbb{P}\left(\left|\|\mathbb{D}_n\| - \|\mathbb{G}_n\|\right| > \rho_n\right)$$

$$\leq \alpha + \mathbb{P}\left(q_{\|\mathbb{G}_n\|}(\alpha) \leq \|\mathbb{G}_n\| \leq q_{\|\mathbb{G}_n\|}(\alpha) + \rho_n\right) + \mathbb{P}\left(\left|\|\mathbb{D}_n\| - \|\mathbb{G}_n\|\right| > \rho_n\right),$$

since the distribution of $\|\mathbb{G}_n\|$ is absolutely continuous. Analogously, it holds

$$\mathbb{P}\big(\|\mathbb{D}_n\| \leq q_{\|\mathbb{G}_n\|}(\alpha)\big) \geq \alpha - \mathbb{P}\big(q_{\|\mathbb{G}_n\|}(\alpha) - \rho_n \leq \|\mathbb{G}_n\| \leq q_{\|\mathbb{G}_n\|}(\alpha)\big)$$
$$- \mathbb{P}\big(\big|\|\mathbb{D}_n\| - \|\mathbb{G}_n\|\big| > \rho_n\big),$$

and therefore

$$\left|\mathbb{P}\big(\|\mathbb{D}_n\| \leq q_{\|\mathbb{G}_n\|}(\alpha)\big) - \alpha\right| \leq \sup_{x \in \mathbb{R}} \mathbb{P}\big(\big|\|\mathbb{G}_n\| - x\big| \leq \rho_n\big) + \mathbb{P}\big(\big|\|\mathbb{G}_n\| - \|\mathbb{D}_n\|\big| > \rho_n\big).$$

The first term on the right hand side of the inequality is the concentration function of the random variable $\|\mathbb{G}_n\|$, which can be estimated by Theorem 2.1 of Chernozhukov et al. (2014). This gives

$$\left|\mathbb{P}\big(\|\mathbb{D}_n\| \leq q_{\|\mathbb{G}_n\|}(\alpha)\big) - \alpha\right| \leq 4\rho_n\big(\mathbb{E}[\|\mathbb{G}_n\|] + 1\big) + \mathbb{P}\big(\big|\|\mathbb{G}_n\| - \|\mathbb{D}_n\|\big| > \rho_n\big).$$

By Lemma 6 we have $N([0,1], \delta, d_{\mathbb{G}_n}) \leq C_E/(h^{3/2}a_n^{1/2}\delta)$, which allows to estimate the expectation $\mathbb{E}[\|\mathbb{G}_n\|]$ as follows.

$$\mathbb{E}[\|\mathbb{G}_n\|] \leq C \int_0^{\mathrm{diam}_{d_{\mathbb{G}_n}}([0,1))} \sqrt{\ln\left(\frac{C_E}{h^{3/2}a_n^{1/2}\delta}\right)}\, d\delta \leq C\sqrt{\ln(n)}.$$

This yields

$$\left|\mathbb{P}\big(\|\mathbb{D}_n\| \leq q_{\|\mathbb{G}_n\|}(\alpha)\big) - \alpha\right| \leq C\sqrt{\ln(n)}\rho_n + \mathbb{P}\big(\big|\|\mathbb{G}_n\| - \|\mathbb{D}_n\|\big| > \rho_n\big).$$

We now estimate the term $\mathbb{P}\big(\big|\|\mathbb{G}_n\| - \|\mathbb{D}_n\|\big| > \rho_n\big)$ in several steps. With the definition

$$\mathbb{G}_{n,0}(x) := \frac{h^\beta}{v(x)\sqrt{na_nh}} \sum_{j=-n}^{n} v(w_j)Z_j K\left(\frac{w_j - x}{h}; h\right), \tag{19}$$

we find

$$\mathbb{P}\big(\big|\|\mathbb{G}_n\| - \|\mathbb{D}_n\|\big| > \rho_n\big) \leq \mathbb{P}\big(\|\mathbb{G}_n - \mathbb{D}_n\| > \rho_n\big)$$
$$\leq \mathbb{P}\Big(\|\mathbb{G}_{n,0} - \mathbb{D}_n\| > \frac{\rho_n}{2}\Big) + \mathbb{P}\Big(\|\mathbb{G}_n - \mathbb{G}_{n,0}\| > \frac{\rho_n}{2}\Big),$$

and thus

$$\mathbb{P}\big(\big|\|\mathbb{G}_n\| - \|\mathbb{D}_n\|\big| > \rho_n\big) \leq \mathbb{P}\Big(\|v\mathbb{G}_{n,0} - v_n\mathbb{D}_n\| > \frac{\sigma\rho_n}{8}\Big)$$
$$+ \mathbb{P}\Big(\Big\|\frac{1}{v} - \frac{1}{\hat{v}}\Big\|\|v\mathbb{G}_{n,0}\| > \frac{\rho_n}{4}\Big) + \mathbb{P}\Big(\|\mathbb{G}_n - \mathbb{G}_{n,0}\| > \frac{\rho_n}{2}\Big) =: R_{n,1} + R_{n,2} + R_{n,3}.$$

Consider first term $R_{n,2}$. Let $\kappa > 0$ be a constant and n sufficiently large such that $\kappa/\sqrt{\ln(n)} < 1$. Then

$$R_{n,2} \leq \mathbb{P}\Big(\Big\|\frac{1}{v} - \frac{1}{\hat{v}}\Big\| > \kappa\frac{\rho_n}{4\sqrt{\ln(n)}}\Big) + \mathbb{P}\Big(\|v\mathbb{G}_{n,0}\| > \frac{\sqrt{\ln n}}{\kappa}\Big) =: R_{n,2,1} + R_{n,2,2}.$$

The term $R_{n,2,1}$ is controlled by assumption and the term $R_{n,2,2}$ can be estimated by Borell's inequality. To this end, denote by d the pseudo distance induced by the process $v\mathbb{G}_{n,0}$. It holds that

$$\mathbb{E}\left[\sup_{x\in[0,1]} v(x)\mathbb{G}_{n,0}(x)\right] \leq \mathbb{E}[\|v\mathbb{G}_{n,0}\|] \leq C \int_0^{\mathrm{diam}([0,1])} \sqrt{\ln\left(N(\delta,[0,1],d)\right)}\,d\delta$$

$$\leq C \int_0^{\mathrm{diam}([0,1])} \sqrt{\ln\left(\frac{C}{h^{\frac{3}{2}} a_n^{\frac{1}{2}}\delta}\right)}\,d\delta,$$

where the last estimate follows by an application of Lemma 6. By a change of variables, using that for any $a \leq 1$

$$\frac{1}{a}\int_0^a \sqrt{-\ln(x)}\,dx \leq \sqrt{-\ln(a)} + 1/\sqrt{-2\ln(a)} \leq C\sqrt{-\ln(a)},$$

we obtain

$$\mathbb{E}\left[\sup_{x\in[0,1]} v(x)\mathbb{G}_{n,0}(x)\right] \leq \mathbb{E}[\|v\mathbb{G}_{n,0}\|] \leq C\sqrt{\ln(n)}. \tag{20}$$

Next,

$$R_{n,2,2} \leq 2\mathbb{P}\left(\sup_{x\in[0,1]}(v\mathbb{G}_{n,0})(x) > \frac{\sqrt{\ln n}}{2\kappa}\right)$$

$$= \mathbb{P}\left(\sup_{x\in[0,1]}(v\mathbb{G}_{n,0})(x) - \mathbb{E}\left[\sup_{x\in[0,1]}(v\mathbb{G}_{n,0})(x)\right] > \frac{\sqrt{\ln n}}{2\kappa} - \mathbb{E}\left[\sup_{x\in[0,1]}(v\mathbb{G}_{n,0})(x)\right]\right)$$

$$\leq \mathbb{P}\left(\sup_{x\in[0,1]}(v\mathbb{G}_{n,0})(x) - \mathbb{E}\left[\sup_{x\in[0,1]}(v\mathbb{G}_{n,0})(x)\right] > \frac{\sqrt{\ln n}}{4\kappa}\right),$$

for sufficiently small κ such that $\mathbb{E}\left[\sup_{x\in[0,1]} v(x)\mathbb{G}_{n,0}(x)\right] < \frac{\sqrt{\ln n}}{4\kappa}$. An application of Borell's inequality yields

$$R_{n,2,2} \leq \exp\left(-\frac{\ln(n)}{32\kappa^2\sigma_{[0,1]}^2}\right),$$

where $\sigma_{[0,1]}^2 := \sup_{x\in[0,1]} \mathrm{Var}[v(x)\mathbb{G}_{n,0}(x)]$ is a bounded quantity by Lemma 6. For sufficiently small κ, this yields the estimate

$$R_{n,2} \leq \mathbb{P}\left(\left\|\frac{1}{v} - \frac{1}{\hat{v}}\right\| > \kappa\frac{\rho_n}{4\sqrt{\ln(n)}}\right) + \frac{C}{n}.$$

Next, we estimate the term $R_{n,1}$, i.e., we consider the approximation of $\mathbb{D}_n$ by a suitable Gaussian process. To this end, consider the standardized random variables $\xi_j := \eta_j/v(w_j)$ and write

$$\hat{v}_n(x)\mathbb{D}_n(x) = \frac{h^\beta}{\sqrt{na_n h}} \sum_{j=-n}^{n} \xi_j v(w_j) K\left(\frac{w_j - x}{h};h\right) = \frac{h^\beta}{\sqrt{na_n h}} \left[\xi_0 v(w_0) K\left(-\frac{x}{h};h\right) \right.$$

$$\left. + \sum_{j=1}^{n} \xi_j v(w_j) K\left(\frac{w_j - x}{h};h\right) + \sum_{j=-n}^{-1} \xi_j v(w_j) K\left(\frac{w_j - x}{h};h\right) \right]$$

$$=: \mathbb{D}_n^0(x) + \mathbb{D}_n^+(x) + \mathbb{D}_n^-(x),$$

(21)

where the processes $\mathbb{D}_n^+(x), \mathbb{D}_n^-(x)$ and $\mathbb{D}_n^0(x)$ are defined in an obvious manner. Define the j-th partial sum $S_j := \sum_{v=1}^{j} \xi_v$, set $S_0 \equiv 0$ and write

$$\sum_{j=1}^{n} \xi_j v(w_j) K\left(\frac{w_j - x}{h};h\right) = S_n v(w_n) K\left(\frac{w_n - x}{h};h\right)$$

$$- \sum_{j=0}^{n-1} S_j \left[v(w_{j+1}) K\left(\frac{w_{j+1} - x}{h};h\right) - v(w_j) K\left(\frac{w_j - x}{h};h\right) \right]$$

$$. = S_n v(w_n) K\left(\frac{w_n - x}{h};h\right) - \sum_{j=1}^{n-1} S_j \int_{[w_j, w_{j+1}]} \frac{\mathrm{d}}{\mathrm{d}z}\left(v(z) K\left(\frac{z - x}{h};h\right)\right) \mathrm{d}z.$$

By assumption, there exists a constant M > 2 such that $\mathbb{E}[|\epsilon_1|^M] < \infty$. By Lemma 2, γ is uniformly bounded, which implies $\mathbb{E}[|\eta_j|^M] \leq M$ for some $M > 0$ and all j. By Corollary 4, §5 in Sakhanenko (1991) there exist iid standard normally distributed random variables $Z_1, \ldots, Z_n$ such that, for $W(j) := \sum_{j=1}^{n} Z_j$ the following estimate holds for any positive constant $\mathcal{C}$:

$$\mathbb{P}\left(\max_{1 \leq j \leq n} |S_j - W(j)| > \frac{n^{2/M}}{2\mathcal{C}} \right) \leq \sum_{j=1}^{n} \mathbb{E}[|\xi_j|^M]\left(\frac{\mathcal{C}}{n^{2/M}}\right)^M + \mathbb{P}\left(\max_{1 \leq j \leq n} \xi_j > \frac{n^{2/M}}{\mathcal{C}} \right).$$

(22)

Therefore,

$$\|\hat{v}_n \mathbb{D}_n^+ - v\mathbb{G}_{n,0}^+\| \leq \frac{\max_{1 \leq j \leq n} |S_j - W(j)|}{\sigma \sqrt{na_n h}} \left[\sup_{x \in [0,1]} v(w_n) h^\beta \left| K\left(\frac{w_n - x}{h};h\right) \right| \right.$$

$$\left. + \sup_{x \in [0,1]} \int_{[w_1, w_n]} \left| \frac{\mathrm{d}}{\mathrm{d}z}\left(h^\beta v(z) K\left(\frac{z - x}{h};h\right)\right) \right| \mathrm{d}z \right],$$

where $\mathbb{G}_{n,0}^+$ is defined in analogy to $\mathbb{D}_n^+$ in (21), with ξ_j replaced by Z_j. For n sufficiently large, we have $a_n < 1/2$ and thus, for $x \in [0,1]$ we have that $(w_n - x)/h \in [1/(2a_n h), 1/a_n h]$ and thus

$$v(w_n) h^\beta \left| K\left(\frac{w_n - x}{h};h\right) \right| \leq v(w_n) h^\beta \sup_{u > 1/(2a_n h)} |K(u;h)|$$

$$\leq v(w_n) h^\beta \sup_{u > 1/(2a_n h)} |2a_n h u K(u;h)| \leq 2a_n h^{\beta+1} v(w_n) \| \cdot K(\cdot;h)\|_\infty \leq \mathcal{C} a_n h,$$

by (23). Next,

$$\sup_{x\in[0,1]} \int_{[w_1,w_n]} \left| \frac{\mathrm{d}}{\mathrm{d}z}\left(h^\beta v(z) K\left(\frac{z-x}{h};h\right) \right) \right| \mathrm{d}z$$

$$\leq Ch^\beta \sup_{x\in[0,1]} \int_{[w_1,w_n]} \left| K\left(\frac{z-x}{h};h\right) \right| + \frac{1}{h}\left| K'\left(\frac{z-x}{h};h\right) \right| \mathrm{d}z$$

$$\leq Ch^\beta \sup_{x\in[0,1]} \int_{-\frac{1}{h}}^{\frac{1}{a_n h}} h|K(u;h)| + \left| K'(u;h) \right| \mathrm{d}z \leq C \ln(n)$$

by equation (S3) in the supplementary material. This yields

$$\|\hat{v}_n \mathbb{D}_n^+ - v\mathbb{G}_n^+\| \leq \|\hat{v}_n \mathbb{D}_n - v\mathbb{G}_n\| \leq C \ln(n) \frac{\max_{1\leq j\leq n} |S_j - W(j)|}{\sqrt{na_n h}}.$$

Hence,

$$R_{n,1} \leq \mathbb{P}\left(C \ln(n) \frac{\max_{1\leq j\leq n} |S_j - W(j)|}{\sqrt{na_n h}} > \frac{\rho_n}{2} \right)$$

$$\leq \mathbb{P}\left(\max_{1\leq j\leq n} |S_j - W(j)| > \frac{n^{2/M}}{2C} \right).$$

Since $0 < \sigma < v(w_j)$, $\mathbb{E}[|\xi_j|^M] \leq M/\sigma^M$ for all $1 \leq j \leq n$, we have $R_{n,2} \leq C/n$ by (22). Last, we need to estimate the term $R_{n,3}$. We have

$$R_{n,3} = \mathbb{P}\left(\|\mathbb{G}_n - \mathbb{G}_{n,0}\| > \frac{\rho_n}{4} \right) \leq \mathbb{P}\left(\|\widetilde{R}_n\| > \frac{\rho_n}{4} \right),$$

where

$$\widetilde{R}_n(x) := \frac{h^\beta}{\sqrt{na_n h}} \sum_{j=-n}^{n} Z_j(v(w_j) - v(x)) K\left(\frac{w_j - x}{h};h\right).$$

Using that by Lemma 3 $|v(w_j) - v(x)| \leq C|w_j - x| = hC|w_j - x|/h$, we find that $N\left([0,1], \delta, d_{\widetilde{R}_n}\right) \leq C/(\sqrt{a_n h}\delta)$. Furthermore, there exist positive constants $\hat{c}$ and $\hat{C}$ such that

$$\hat{c}\frac{h^2}{na_n h} \leq \sup_{x\in[0,1]} \mathrm{Var}\left[\widetilde{R}_n(x)\right] \leq \hat{C}\frac{h^2}{na_n h}.$$

By Theorem 4.1.2 in Adler and Taylor (2007), there exists a universal constant K such that, for all $u > 2\sqrt{\hat{C}\frac{h^2}{na_n h}}$,

$$\mathbb{P}\left(\sup_{x\in[0,1]} \widetilde{R}(x) \geq u \right) \leq \frac{Kun\sqrt{a_n h}}{\hat{c}h^2} \cdot \Psi\left(\frac{u}{\sqrt{\hat{C}\frac{h^2}{na_n h}}} \right),$$

where Ψ denotes the tail function of the standard normal distribution. Setting $u = \rho_n/8$ yields, for sufficiently large n,

$$\mathbb{P}\left(\sup_{x \in [0,1]} \widetilde{R}(x) \geq \frac{\rho_n}{8}\right) \leq \frac{K n^{\frac{1}{2} + \frac{2}{M}} \ln(n)}{8 \hat{c} h^2} \cdot \Psi\left(\frac{\ln(n) n^{\frac{2}{M}}}{8 h \sqrt{\hat{c}}}\right) \leq \mathcal{C} n \exp\left(-\frac{n^{\frac{4}{M}}}{h}\right) \leq \frac{\mathcal{C}}{n}.$$

Therefore, $R_{n,3} \leq \mathcal{C}/n$, which concludes the proof of assertion (i). Assertion (ii) is again an immediate consequence of Lemma 5. $\qquad\square$

Proof of Theorem 2 On the one hand,

$$\mathbb{P}\left(\|\mathbb{D}_n\| \leq q_{\|\mathbb{G}_n\|_{\mathcal{X}_{n,m}}}(\alpha)\right) \leq \mathbb{P}\left(\|\mathbb{D}_n\| \leq q_{\|\mathbb{G}_n\|}(\alpha)\right) \leq \alpha + r_{n,1},$$

by Theorem 1. On the other hand,

$$\mathbb{P}\left(\|\mathbb{D}_n\| \leq q_{\|\mathbb{G}_n\|_{\mathcal{X}_{n,m}}}(\alpha)\right)$$
$$\geq \alpha - \frac{\mathcal{C}}{n} - \mathbb{P}\left(q_{\|\mathbb{G}_n\|_{\mathcal{X}_{n,m}}}(\alpha) - \rho_n \leq \|\mathbb{G}_n\| \leq q_{\|\mathbb{G}_n\|_{\mathcal{X}_{n,m}}}(\alpha)\right).$$

Note that (25) implies

$$|s - t| \leq |\mathcal{X}_{n,m}| \quad \Rightarrow \quad d_{\mathbb{G}_n}(s,t) \leq \frac{|\mathcal{X}_{n,m}| h^\beta \|K^{(1)}(\cdot;h)\|_\infty}{h^{3/2} a_n^{1/2}} \leq \frac{\mathcal{C}|\mathcal{X}_{n,m}|}{h^{3/2} a_n^{1/2}}.$$

Hence,

$$\|\mathbb{G}_n\|_{\mathcal{X}_{n,m}} \leq \|\mathbb{G}_n\| \leq \|\mathbb{G}_n\|_{\mathcal{X}_{n,m}} + \sup_{s,t : |s-t| \leq |\mathcal{X}_{n,m}|} |\mathbb{G}_n(s) - \mathbb{G}_n(t)|$$
$$\leq \|\mathbb{G}_n\|_{\mathcal{X}_{n,m}} + \sup_{s,t : d_{\mathbb{G}_n}(s,t) \leq \mathcal{C} \frac{|\mathcal{X}_{n,m}|}{h^{3/2} a_n^{1/2}}} |\mathbb{G}_n(s) - \mathbb{G}_n(t)| =: \|\mathbb{G}_n\|_{\mathcal{X}_{n,m}} + \tau_n.$$

This yields

$$\mathbb{P}\left(\|\mathbb{D}_n\| \leq q_{\|\mathbb{G}_n\|_{\mathcal{X}_{n,m}}}(\alpha)\right)$$
$$\geq \alpha - \frac{\mathcal{C}}{n} - \mathbb{P}\left(q_{\|\mathbb{G}_n\|_{\mathcal{X}_{n,m}}}(\alpha) - \rho_n - \tau_n \leq \|\mathbb{G}_n\|_{\mathcal{X}_{n,m}} \leq q_{\|\mathbb{G}_n\|_{\mathcal{X}_{n,m}}}(\alpha)\right).$$

By Corollary 2.2.8 in van der Vaart and Wellner (1996) and Lemma 6, we find

$$\mathbb{E}[\tau_n] \leq \mathcal{C} \int_0^{\mathcal{C} \frac{|\mathcal{X}_{n,m}|}{h^{3/2} a_n^{1/2}}} \sqrt{\ln\left(N([0,1], d_{\mathbb{G}_n}, \eta)\right)} \, d\eta \leq \mathcal{C} \frac{|\mathcal{X}_{n,m}| \sqrt{-\ln(|\mathcal{X}_{n,m}|)}}{h^{3/2} a_n^{1/2}}.$$

Since $|\mathcal{X}_{n,m}| \leq h^{1/2}/n a_n^{1/2}$, we have that $\frac{|\mathcal{X}_{n,m}| \sqrt{-\ln(|\mathcal{X}_{n,m}|)}}{h^{3/2} a_n^{1/2}} = o(\rho_n^2)$ and therefore, by Markov's inequality,

$$\mathbb{P}\big(\tau_n \geq \rho_n\big) \leq C \frac{|\mathcal{X}_{n,m}|\sqrt{-\ln(|\mathcal{X}_{n,m}|)}}{h^{3/2}a_n^{1/2}} \frac{1}{\rho_n} = o(\rho_n).$$

This yields

$$
\begin{aligned}
\mathbb{P}\Big(& \|\mathbb{D}_n\| \leq q_{\|\mathbb{G}_n\|_{\mathcal{X}_{n,m}}}(\alpha)\Big) \\
&\geq \alpha - \frac{C}{n} - \mathbb{P}\Big(q_{\|\mathbb{G}_n\|_{\mathcal{X}_{n,m}}}(\alpha) - 2\rho_n \leq \|\mathbb{G}_n\|_{\mathcal{X}_{n,m}} \leq q_{\|\mathbb{G}_n\|_{\mathcal{X}_{n,m}}}(\alpha)\Big) - o(\rho_n) \\
&\geq \alpha - \frac{C}{n} - \sup_{x \in \mathbb{R}} \mathbb{P}\Big(\|\mathbb{G}_n\|_{\mathcal{X}_{n,m}} \in [x - 2\rho_n, x + 2\rho_n]\Big) - o(\rho_n) \geq \alpha - C\Big(\frac{1}{n} - \rho_n\Big),
\end{aligned}
$$

where we applied Theorem 2.1 in Chernozhukov et al. (2014). Claim 1 of this theorem now follows. Claim 2 is an immediate consequence of Lemma 5. □

8 Proofs of the auxiliary lemmas

Proof of Lemma 1 Assertion (i) is a direct consequence of Sobolev's Lemma.
(ii) By an application of the Hausdorff-Young inequality we obtain

$$\Big\|\frac{d^j}{dx^j}g\Big\|_\infty \leq \frac{1}{2\pi}\Big\|\Phi_{\frac{d^j}{dx^j}g}\Big\|_1, \quad j = 0, 1, 2.$$

Fourier transformation converts differentiation into multiplication, that is,

$$\Big\|\Phi_{\frac{d^j}{dx^j}g}\Big\|_1 = \|(\cdot)^j \Phi_g\|_1, \quad j = 0, 1, 2.$$

Since $g \in \mathcal{W}^m(\mathbb{R})$ for $m > 5/2$ by Assumption 1 it follows by an application of the Cauchy-Schwarz inequality that $\|(\cdot)^j \Phi_g\|_1 < \infty$ for $j = 0, 1, 2$ and the assertion follows. □

Proof of Lemma 2 Assertion (i) follows from Proposition 8.10 in Folland (1984) since f_Δ is a density and is hence integrable.
Assertion (ii) is a direct consequence of Assumption 1 and the convolution theorem: $\Phi_\gamma = \Phi_{g*f_\Delta(-\cdot)} = \Phi_g \cdot \overline{\Phi}_{f_\Delta}$, since Φ_{f_Δ} is bounded.
Assertion (iii) follows in the same manner as the second claim of Lemma 1. □

Proof of Lemma 3 (i) Recall from definition (5) that

$$v^2(z) = \int \big(g(z + \delta) - \gamma(z)\big)^2 f_\Delta(\delta) \, d\delta + \sigma^2, \quad \sigma^2 > 0.$$

Hence, it follows from Lemma 1 (ii) and Lemma 2 (iii) that

$$0 < \sigma^2 \leq v^2(z) \leq \sigma^2 + 2(\|g\|_\infty^2 + \|\gamma\|_\infty^2) < \infty.$$

(ii) By the first assertions of Lemma 1 and Lemma 2, the functions g and γ are twice continuously differentiable and f_Δ is continuous. This yields for $j = 1, 2$

$$\frac{d^j}{dz^j} v^2(z) = \int \frac{\partial^j}{\partial z^j} \left(\left(g(z + \delta) - \gamma(z) \right)^2 f_\Delta(\delta) \right) d\delta.$$

Since by Lemmas 1 and 2 the derivatives of g and γ are uniformly bounded and f_Δ is a probability density, we find for $j = 1, 2$

$$\left| \frac{d^j}{dz^j} v^2(z) \right| \leq \sup_{z \in \mathbb{R}} \left| \frac{\partial^j}{\partial z^j} \left(g(z + \delta) - \gamma(z) \right)^2 \right|.$$

$\square$

Proof of Lemma 4 From (7), we deduce for $w \in \mathbb{R}$

$$iwK(w;h) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itw} \frac{d}{dt} \left(\frac{\Phi_k(t)}{\Phi_{f_\Delta}(-t/h)} \right) dt$$

$$= \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itw} \left(\frac{\Phi_k^{(1)}(t)}{\Phi_{f_\Delta}(-t/h)} + \frac{\Phi_k(t) \cdot \Phi_{f_\Delta}^{(1)}(-t/h)}{h(\Phi_{f_\Delta}(-t/h))^2} \right) dt.$$

Hence,

$$\sup_{w \in \mathbb{R}} |wK(w;h)| \leq \frac{1}{2\pi c} \left\| \left| \Phi_k^{(1)} \right| \left\langle \frac{\cdot}{h} \right\rangle^\beta + \frac{\left| \Phi_k \cdot \Phi_{f_\Delta}^{(1)}(-\cdot/h) \right|}{ch} \left\langle \frac{\cdot}{h} \right\rangle^{2\beta} \right\|_1 \tag{23}$$

$$= \begin{cases} O(h^{-\beta-1}), & (W), \\ O(h^{-\beta}), & (S). \end{cases}$$

In particular, for all $w \in \mathbb{R} \backslash \{0\}$,

$$|K(w;h)| \leq \frac{\mathcal{C}}{|w|} \cdot \begin{cases} h^{-\beta-1}, & (W), \\ h^{-\beta}, & (S). \end{cases} \tag{24}$$

Now, let $a > 1$. Then

$$\int_{\{|z|>a\}} \left(K\left(\frac{z-x}{h};h \right) \right)^2 dz \leq \mathcal{C} \int_{\{|z|>a\}} \left(\frac{h}{z-x} \right)^2 dz \cdot \begin{cases} h^{-2\beta-2}, & (W) \\ h^{-2\beta}, & (s) \end{cases}$$

$$\leq \mathcal{C} \frac{2a}{a^2 - x^2} \begin{cases} h^{-2\beta}, & (W) \\ h^{-2\beta+2}, & (S) \end{cases}.$$

$\square$

Proof of Lemma 5 The proof of Lemma 5 is straightforward but tedious. We therefore omit the proof here and defer it to an online supplement. $\square$

$$d_{\mathbb{G}_n}(s,t)^2 = \mathbb{E}|\mathbb{G}_n(s) - \mathbb{G}_n(t)|^2 = \frac{h^{2\beta}}{na_n h} \sum_{j=-n}^{n} \left| K\left(\frac{w_j - s}{h};h\right) - K\left(\frac{w_j - t}{h};h\right) \right|^2$$

$$\leq \frac{h^{2\beta}\|K^{(1)}(\cdot;h)\|_\infty^2}{a_n h}\left(\frac{s-t}{h}\right)^2 \leq C\frac{h^{2\beta}}{a_n h}\left\| t\frac{\Phi_k(t)}{\Phi(-t/h)} \right\|_1^2 \left(\frac{s-t}{h}\right)^2,$$

Proof of Lemma 6

(25)

where the last estimate follows by the Hausdorff-Young inequality and definition (7).

Therefore, by Assumption 3, there exists a constant C_E such that $d_{\mathbb{G}_n}(s,t) \leq C_E|s-t|/(a_n^{\frac{2}{2}}h^{\frac{3}{2}})$. Now, consider the equidistant grid

$$\mathcal{G}_{n,\delta} := \left\{ t_j = j\frac{a_n^{\frac{1}{2}}}{C_E}\delta, \, j = 1, \ldots, \left\lfloor \frac{C_E}{a_n^{\frac{1}{2}}h^{\frac{3}{2}}\delta} \right\rfloor \right\} \subset [0,1]$$

and note that for each $s \in [0,1]$ there exists a $t_j \in \mathcal{G}_{n,\delta}$ such that $|s - t_j| \leq a_n^{1/2}h^{3/2}\delta/(2C_E)$, which implies $d_{\mathcal{G}_n}(s,t_j) \leq \delta/2$. Therefore, the closed $d_{\mathbb{G}_n}$-balls with centers $t_j \in \mathcal{G}_{n,\delta}$ and radius $\delta/2$ cover the space $[0,1]$, i.e., $N([0,1],\delta/2,d_{\mathbb{G}_n}) \leq \frac{C_E}{h^{\frac{3}{2}}a_n^{\frac{1}{2}}\delta}$.

The relationship $N([0,1],\delta,d_{\mathbb{G}_n}) \leq D([0,1],\delta,d_{\mathbb{G}_n}) \leq N([0,1],\delta/2,d_{\mathbb{G}_n})$ now yields the first claim of the lemma. Using that, by Assumption 3,

$$\|\widehat{K}^{(1)}(\cdot,h)\|_\infty \leq \|\cdot K^{(1)}(\cdot,h)\|_\infty + \|K(\cdot,h)\|_\infty$$

$$\leq C\left(\left\| \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{t\Phi_k(t)}{\Phi_{f_\Delta}(-t/h)}\right) \right\|_1 + \left\| \left(\frac{t\Phi_k(t)}{\Phi_{f_\Delta}(-t/h)}\right) \right\|_1 \right) \leq Ch^{-\beta},$$

the second claim follows along the lines of the first claim. $\square$

References

Adler, R. J., Taylor, J. E. (2007). *Random Fields and Geometry*. New York: Springer.

Anderson, T. W. (1984). Estimating linear statistical relationships. *The Annals of Statistics, 12*(1), 1–45.

Berkson, J. (1950). Are there two regressions? *Journal of the American Statistical Association, 45*, 164–180.

Bickel, P. J., Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *The Annals of Statistics, 1,* 1071–1095.

Birke, M., Bissantz, N., Holzmann, H. (2010). Confidence bands for inverse regression models. *Inverse Problems, 26,* 115020.

Carroll, R. J., Ruppert, D., Stefanski, L. A., Crainiceanu, C. M. (2006). *Measurement error in nonlinear models. A modern perspective.*, volume 105 of *Monographs on Statistics and Applied Probability*, 2nd ed. Boca Raton, FL: Chapman & Hall/CRC.

Carroll, R. J., Delaigle, A., Hall, P. (2007). Non-parametric regression estimation from data contaminated by a mixture of Berkson and classical errors. *Journal of the Royal Statistical Society, Series B. Statistical Methodology, 69,* 859–878.

Chernozhukov, V., Chetverikov, D., Kato, K. (2014). Anti-concentration and honest adaptive confidence bands. *The Annals of Statistics, 42,* 1564–1597.

Delaigle, A., Hall, P. (2011). Estimation of observation-error variance in errors-in-variables regression. *Statistica Sinica, 21,* 103–1063.

Delaigle, A., Hall, P., Qiu, P. (2006). Nonparametric methods for solving the Berkson errors-in-variables proble. *Journal of the Royal Statistical Society, Series B. Statistical Methodology, 68,* 201–220.

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

Delaigle, A., Hall, P., Jamshidi, F. (2015). Confidence bands in non-parametric errors-in-variables regression. *Journal of the Royal Statistical Society, Series B. Statistical Methodology, 77,* 149–169.

Eubank, R. L., Speckman, P. L. (1993). Confidence bands in nonparametric regression. *Journal of the American Statistical Association, 88,* 1287–1301.

Fan, J., Truong, Y. K. (1993). Nonparametric regression with errors in variables. *The Annals of Statistics, 21,* 1900–1925.

Folland, G. B. (1984). *Real analysis—Modern techniques and their applications*. New York: Wiley.

Fuller, W. A. (1987). *Measurement error models. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*. New York: John Wiley & Sons, Inc.

Giné, E., Nickl, R. (2010). Confidence bands in density estimation. *The Annals of Statistics, 38*(2), 1122–1170.

Hall, P. (1992). Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. *The Annals of Statistics, 20,* 675–694.

Kato, K., Sasaki, Y. (2018). Uniform confidence bands in deconvolution with unknown error distribution. *Journal of Econometrics, 207*(1), 129–161.

Kato, K., Sasaki, Y. (2019). Uniform confidence bands for nonparametric errors-in-variables regression. *Journal of Econometrics, 213*(2), 516–555.

Koul, H. L., Song, W. (2008). Regression model checking with Berkson measurement errors. *Journal of Statistical Planning and Inference, 138,* 1615–1628.

Koul, H. L., Song, W. (2009). Minimum distance regression model checking with Berkson measurement errors. *The Annals of Statistics, 37,* 132–156.

Meister, A. (2010). Nonparametric Berkson regression under normal measurement error and bounded design. *Journal of Multivariate Analysis, 101,* 1179–1189.

Neumann, M. H., Polzehl, J. (1998). Simultaneous bootstrap confidence bands in nonparametric regression. *Journal of Nonparametric Statistics, 9,* 307–333.

Proksch, K., Bissantz, N., Dette, H. (2015). Confidence bands for multivariate and time dependent inverse regression models. *Bernoulli, 21,* 144–175.

Sakhanenko, A. I. (1991). On the accuracy of normal approximation in the invariance principle. *Siberian Advances in Mathematics, 1,* 58–91.

Schennach, S. M. (2013). Regressions with Berkson errors in covariates—A nonparametric approach. *The Annals of Statistics, 41,* 1642–1668.

Schmidt-Hieber, J., Munk, A., Dümbgen, L., et al. (2013). Multiscale methods for shape constraints in deconvolution: Confidence statements for qualitative features. *The Annals of Statistics, 41*(3), 1299–1328.

Stefanski, L. A. (1985). The effects of measurement error on parameter estimation. *Biometrika, 72*(3), 583–592.

van der Vaart, A., Wellner, J. (1996). *Weak convergence and empirical processes. With applications to statistics*. New York: Springer.

van Es, B., Gugushvili, S. (2008). Weak convergence of the supremum distance for supersmooth kernel deconvolution. *Statistics and Probability Letters, 78*(17), 2932–2938.

Wang, L. (2004). Estimation of nonlinear models with Berkson measurement errors. *The Annals of Statistics, 32,* 2559–2579.

Wang, L., Brown, L. D., Cai, T. T., Levine, M. (2008). Effect of mean on variance function estimation in nonparametric regression. *The Annals of Statistics, 36*(2), 646–664.

Wu, C.-F.J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *The Annals of Statistics, 14*(4), 1261–1295.

Yang, Y., Bhattacharya, A., Pati, D. (2017). Frequentist coverage and sup-norm convergence rate in gaussian process regression. arXiv:1708.04753.

Yano, K., Kato, K. (2020). On frequentist coverage errors of Bayesian credible sets in moderately high dimensions. *Bernoulli, 26*(1), 616–641.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.