UNIFORM CONFIDENCE BANDS IN DECONVOLUTION WITH UNKNOWN ERROR DISTRIBUTION

KENGO KATO AND YUYA SASAKI

Abstract. This paper develops a method to construct uniform confidence bands in deconvolution when the error distribution is unknown. We work with the setting where an auxiliary sample from the error distribution is available and the error density is ordinary smooth. The construction is based upon the "intermediate" Gaussian approximation and the Gaussian multiplier bootstrap, but not on explicit limit distributions such as Gumbel distributions, which enables us to prove validity of the multiplier bootstrap confidence band under weak regularity conditions. Importantly, the previous literature on this topic has focused on the case where the error distribution is known, but has not formally covered the case with unknown error distribution. However, in many applications, the assumption that the error density is known is unrealistic, and the present paper fills this important void. In addition, we also discuss extensions of the results on confidence bands to super-smooth error densities. We conduct simulation studies to verify the performance of the multiplier bootstrap confidence band in the finite sample. Finally, we apply our method to the Outer Continental Shelf (OCS) Auction Data and draw confidence bands for the density of common values of mineral rights on oil and gas tracts.

1. Introduction

Consider the deconvolution model

$$Y = X + \varepsilon, \tag{1}$$

where X and ε are independent, real-valued random variables with unknown densities f_X and f_ε, respectively. We observe Y, but do not observe X nor ε. The model (1) implies that the density f_Y of Y exists and is given by the convolution of f_X and f_ε, namely,

$$f_Y(y) = (f_X * f_\varepsilon)(y) = \int_{\mathbb{R}} f_X(x) f_\varepsilon(y-x)dx. \tag{2}$$

Date: First version: August 4, 2016. This version: September 6, 2016.

2000 *Mathematics Subject Classification.* 62G07; 62G15.

Key words and phrases. bootstrap, confidence band, deconvolution, empirical process theory.

K. Kato is supported by Grant-in-Aid for Scientific Research (C) (15K03392) from the JSPS.

The deconvolution problem refers to the problem of estimating and making inference on f_X from available data.

The goal of this paper is to develop a method to construct uniform confidence bands for f_X when the error density f_ε is unknown. In addition to the availability of an independent sample from f_Y, denoted by $Y_1, \dots, Y_n$, we assume that an auxiliary independent sample from f_ε, denoted by $\eta_1, \dots, \eta_m$, is available to pre-estimate the unknown density f_ε. This assumption is satisfied in various ways depending on an application of interest. A simple case is where auxiliary experiments provide a sample $\eta_1, \dots, \eta_m$. Another case, which accommodates a wider spectrum of relevant applications, is where panel data or repeated measurements $(Y^{(1)}, Y^{(2)})$ for a common signal X with symmetrically and identically distributed errors $(\varepsilon^{(1)}, \varepsilon^{(2)})$ are available (see Example 1 ahead). Under these two representative situations with unknown f_ε, we develop a deconvolution estimator and confidence bands for f_X based on a multiplier bootstrap. We use a standard deconvolution kernel estimator that goes back to Carroll and Hall (1988) and Stefanski and Carroll (1990), but with the error characteristic function replaced by the empirical characteristic function based on the auxiliary sample $\eta_1, \dots, \eta_m$ from f_ε.

It is well understood that the difficulty of estimating f_X critically depends on how fast the modulus of the error characteristic function $\varphi_\varepsilon(t) = \mathrm{E}[e^{it\varepsilon}]$ with $i = \sqrt{-1}$ decays as $|t| \to \infty$, in addition to the smoothness of f_X (the faster $|\varphi_\varepsilon(t)|$ decays as $|t| \to \infty$, the more difficult estimation of f_X becomes). Informally, the error density f_ε is said to be *ordinary smooth* if $|\varphi_\varepsilon(t)|$ decays at most polynomially fast as $|t| \to \infty$, while f_ε is said to be *super-smooth* if $|\varphi_\varepsilon(t)|$ decays exponentially fast as $|t| \to \infty$ (cf. Fan, 1991a). We first consider ordinary smooth error densities and prove asymptotic validity of the multiplier bootstrap confidence band under weak regularity conditions. In this ordinary smooth case, the sample size m for f_ε need not be large in comparison with n. As a corollary, we derive uniform convergence rates of the deconvolution kernel density estimator with the estimated error characteristic function. Furthermore, we extend the results on confidence bands to super-smooth error densities; in the super-smooth case, however, we require $m/n \to \infty$ for a technical reason.

We conduct simulation studies to verify the performance of the multiplier bootstrap confidence band in the finite sample. The simulation studies show that the simulated coverage probabilities are very close to nominal coverage probabilities even with sample sizes as small as 250 and 500, suggesting practical benefits of our confidence band. Finally, following Li et al. (2002), we apply our method to the Outer Continental Shelf (OCS)

Auction Data (Hendricks et al., 1987) and draw confidence bands for the density of common values of mineral rights on oil and gas tracts in the Gulf of Mexico. Our result provides statistical support for some qualitative features of the common value density that Li et al. (2002) find visually in their estimate. In order to illustrate the advantage of our method over existing alternatives which assume to know the error density, we also construct a uniform confidence band for the common value density by assuming that the error density is known to take a particular form. We find qualitative differences between the two confidence bands, highlighting practical values of our approach which requires no prior knowledge of the error distribution.

The literature related to this paper is broad. We refer to Carroll et al. (2006) and Meister (2009) as general references on measurement error models and deconvolution. In the present paper, among other approaches, we build upon the deconvolution kernel density estimation method, which is pioneered by Carroll and Hall (1988); Stefanski and Carroll (1990); Fan (1991a,b). These earlier studies, as well as many of more recent papers, focus on the case where the error density f_ε is assumed to be known, even though it is often unknown in applications.

The deconvolution problem with unknown f_ε is studied by Diggle and Hall (1993); Efromovich (1997); Neumann (1997); Johannes (2009); Comte and Lacour (2011); Dattner et al. (2016), where the unknown density is estimated from auxiliary measurements of ε. Horowitz and Markatou (1996) and Li and Vuong (1998) consider to estimate f_ε from repeated measurements (panel data) of Y, instead of assuming measurements from the error distribution *per se*; see also Neumann (2007) and Delaigle et al. (2008) for further developments. In addition, a recent work by Delaigle and Hall (2016) relaxes the requirement of repeated measurements under the assumption of a symmetric error distribution. Despite the richness of this literature, however, uniform confidence bands for f_X have not been developed in any of these papers allowing for unknown f_ε.

The deconvolution problem is an example of statistical ill-posed inverse problems, and developing formal theories for inference in ill-posed inverse problems is considered to be challenging. In fact, it is only (relatively) recently that studies on uniform confidence bands in deconvolution have appeared, and those previous studies focus on the case where f_ε is known. To the best of our knowledge, Bissantz et al. (2007) is the first paper that formally studies uniform confidence bands in deconvolution. They assume that f_ε is known and ordinary smooth, and prove, under a number of technical conditions, a Smirnov-Bickel-Rosenblatt type limit theorem (cf. Smirnov, 1950; Bickel and Rosenblatt, 1973) for the deconvolution kernel density estimator based on the Komlós-Major-Tusnády (KMT) strong approximation (Komlós et al., 1975) and the extreme value theory (cf. Leadbetter et al., 1983), namely, they prove that the supremum deviation of the deconvolution kernel density estimator, suitably normalized, converges in distribution to a Gumbel distribution. Further, they prove consistency of the nonparametric bootstrap. See also Bissantz and Holtzmann (2008). Interestingly, in the case where the error density is super-smooth, van Es and Gugushvili (2008) show that the limit distribution of the supremum deviation of the deconvolution kernel density estimator in general differs from Gumbel distributions. We also refer to Lounici and Nickl (2011); Schmidt-Hieber et al. (2013); Delaigle et al. (2015).

Importantly, none of these papers formally considers the case where f_ε is unknown. It should be noted that Delaigle et al. (2015, Section 4.2) discuss how to possibly accommodate the case of unknown error density, but a theory to support this argument is not provided.[1] While the effect of pre-estimating the unknown error density for the purpose of estimating f_X is modest, its effect for the purpose of inference on f_X is non-trivial. We attempt to contribute to the literature on deconvolution by formally establishing a method to construct uniform confidence bands for f_X under the realistic situation where the error density f_ε is unknown. It is worthwhile to point out that the multiplier bootstrap confidence band proposed in the present paper is valid for both cases where the error density is ordinary- and super-smooth (although in the latter case we require $m/n \to \infty$), despite the fact that, as already mentioned, the limit distributions of the supremum deviation of the deconvolution kernel density estimator (with known f_ε) in general differ.

From a technical point of view, the present paper builds upon the "intermediate" Gaussian and multiplier bootstrap approximation theorems developed in Chernozhukov et al. (2014a,b, 2016). These approximation theorems are applicable to the general empirical process under weaker regularity conditions than those for the KMT and Gumbel approximations. In particular, they enable us to prove validity of the multiplier bootstrap confidence band without relying on explicit limit distributions such as Gumbel distributions. See the discussion in Remark 4 in Section 3. However, it should be stressed that those theorems are not directly applicable to our problems, primarily because: 1) the "deconvoluting" kernel K_n (see Section 2 ahead) is implicitly defined via the Fourier inversion, and verifying conditions in those approximation theorems with this K_n is not trivial; 2)

[1] The focus in Delaigle et al. (2015) is on pointwise confidence intervals for nonparametric regression functions, and differs from our objective to conduct uniform inference on nonparametric density functions.

f_ε is unknown and we have to work with the estimated deconvoluting kernel $\widehat{K}_n$, and so the estimation error has to be taken into account, which requires delicate cares since $\widehat{K}_n$ depends on the inverse of the estimated error characteristic function, and the error characteristic function is, although it is assumed to be non-vanishing on the entire real line, approaching zero for the tail; see the discussion after Theorem 1.

The rest of the paper is organized as follows. Section 2 presents our methodology of constructing confidence bands for f_X. Section 3 presents the main theoretical results of this paper where we consider the ordinary smooth case. Section 4 presents the results on numerical simulations. Section 5 presents an empirical application of our method. Section 6 presents extensions of the results on confidence bands to super-smooth error densities. Section 7 contains the proofs of all the results in the paper. Appendix A contains an auxiliary lemma on uniform convergence rates of the empirical characteristic function.

1.1. Notation.

For $a, b \in \mathbb{R}$, let $a \wedge b = \min\{a, b\}$. For a non-empty set T and a (complex-valued) function f on T, we use the notation $\|f\|_T = \sup_{t \in T} |f(t)|$. Let $\ell^\infty(T)$ denote the Banach space of all bounded real-valued functions on T with norm $\|\cdot\|_T$. The inverse Fourier transform of an integrable function f on $\mathbb{R}$ is defined by

$$\varphi_f(t) = \int_{\mathbb{R}} e^{itx} f(x) dx, \ t \in \mathbb{R},$$

where $i = \sqrt{-1}$ denotes the imaginary unit throughout the paper. We refer to Folland (1999) as a basic reference on Fourier analysis.

2. Methodology

In this section, we informally present our methodology to construct confidence bands for f_X. The formal analysis of our methodology will be carried out in the following sections.

Let φ_Y, φ_X, and φ_ε denote the inverse Fourier transforms (the characteristic functions) of f_Y, f_X, and f_ε, respectively. The model (1) implies that these characteristic functions satisfy the relation

$$\varphi_Y(t) = \varphi_X(t)\varphi_\varepsilon(t), \ t \in \mathbb{R}.$$

Under the assumptions that φ_ε does not vanish on $\mathbb{R}$ and $|\varphi_X|$ is integrable on $\mathbb{R}$, the Fourier inversion formula yields that f_X is recovered from f_Y and f_ε as

$$f_X(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \frac{\varphi_Y(t)}{\varphi_\varepsilon(t)} dt. \tag{3}$$

Precisely speaking, if $|\varphi_X|$ is integrable on $\mathbb{R}$, then f_X has a version that is bounded and continuous, and satisfies (3) for each $x \in \mathbb{R}$; we take this version in what follows. Suppose that independent copies $Y_1, \dots, Y_n$ of Y are observed. To describe a basic approach to constructing confidence bands for f_X, suppose first that f_ε were known. In that case, a standard kernel estimator of f is given by

$$\widehat{f}_X^*(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \widehat{\varphi}_Y(t) \frac{\varphi_K(th_n)}{\varphi_\varepsilon(t)} dt = \frac{1}{nh_n} \sum_{j=1}^n K_n((x - Y_j)/h_n),$$

where

$$\widehat{\varphi}_Y(t) = \frac{1}{n} \sum_{j=1}^n e^{itY_j}, \quad K_n(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \frac{\varphi_K(t)}{\varphi_\varepsilon(t/h_n)} dt,$$

and where $K : \mathbb{R} \to \mathbb{R}$ is a real-valued kernel function such that its inverse Fourier transform φ_K is symmetric and supported in $[-1, 1]$, and h_n is a sequence of positive numbers (bandwidths) such that $h_n \to 0$ (cf. Carroll and Hall, 1988; Stefanski and Carroll, 1990). The symmetry and the compactness of the support of φ_K ensure that K_n is real-valued and bounded. The function K_n is called a deconvoluting kernel.

Suppose that we are interested in constructing a confidence band for f_X on a compact interval $I \subset \mathbb{R}$. A confidence band $\mathcal{C}_n$ at level $(1 - \tau)$, where $\tau \in (0, 1)$ is given, is a family of random intervals $\mathcal{C}_n = \{\mathcal{C}_n(x) = [c_L(x), c_U(x)] : x \in I\}$ such that

$$\mathrm{P}\{f_X(x) \in [c_L(x), c_U(x)] \ \forall x \in I\} \ge 1 - \tau.$$

A standard approach to constructing such a confidence band is based on approximating the distribution of the supremum in absolute value of the following stochastic process:

$$Z_n^*(x) = \frac{\widehat{f}_X^*(x) - \mathrm{E}[\widehat{f}_X^*(x)]}{\sqrt{\mathrm{Var}(\widehat{f}_X^*(x))}} = \frac{\sqrt{n} h_n \{\widehat{f}_X^*(x) - \mathrm{E}[\widehat{f}_X^*(x)]\}}{\sigma_n(x)}, \ x \in I,$$

where we assume that $\sigma_n^2(x) = \mathrm{Var}(K_n((x - Y)/h_n)) > 0$ for all $x \in I$, and $\sigma_n(x)$ is defined by $\sigma_n(x) = \sqrt{\sigma_n^2(x)}$. Letting

$$c_n^*(1 - \tau) = (1 - \tau)\text{-quantile of } \|Z_n^*\|_I,$$

a band of the form

$$\mathcal{C}_n^*(x) = \left[\widehat{f}_X^*(x) - \frac{\sigma_n(x)}{\sqrt{n} h_n} c_n^*(1 - \tau), \widehat{f}_X^*(x) + \frac{\sigma_n(x)}{\sqrt{n} h_n} c_n^*(1 - \tau)\right], \ x \in I \tag{4}$$

contains $\{\mathrm{E}[\widehat{f}_X^*(x)] : x \in I\}$ with probability at least $(1 - \tau)$, since

$$\mathrm{P}\{\mathrm{E}[\widehat{f}_X^*(x)] \in \mathcal{C}_n^*(x) \ \forall x \in I\} = \mathrm{P}\{\|Z_n^*\|_I \le c_n^*(1 - \tau)\} \ge 1 - \tau,$$

and provided that the bias $\|f_X - \mathrm{E}[\widehat{f}_X^*(\cdot)]\|_I$ is sufficiently small (which is typically achieved by using undersmoothing), the band of the form (4) is a valid confidence band for f_X on I at level approximately $(1-\tau)$.

Constructing the band of the form (4) is, however, infeasible because both the distribution of $\|Z_n^*\|_I$ and the variance function $\sigma_n^2(\cdot)$ are unknown. More fundamentally, in many applications, the error density f_ε is unknown, and so the kernel estimator $\widehat{f}_X^*$ using f_ε is simply infeasible to compute. In this paper, we assume that f_ε is unknown, but independent observations from f_ε, denoted by $\eta_1, \dots, \eta_m$, are available:

$$\eta_1, \dots, \eta_m \sim f_\varepsilon,$$

where $m = m_n \to \infty$ as $n \to \infty$. We do not assume that $(\eta_1, \dots, \eta_m)$ are independent from $(Y_1, \dots, Y_n)$. A simple case is where auxiliary experiments provide a sample $\eta_1, \dots, \eta_m$. Another typical example where such observations are available is the case where we observe repeated measurements on X with errors obeying common symmetric distribution, described as follows.

Example 1 (Carroll et al. (2006), p.298)**.** Suppose that we observe repeated measurements on X with errors as

$$\begin{cases} Y^{(1)} = X + \varepsilon^{(1)}, \\ Y^{(2)} = X + \varepsilon^{(2)}, \end{cases}$$

where $X, \varepsilon^{(1)}$, and $\varepsilon^{(2)}$ are independent, and $\varepsilon^{(1)}$ and $\varepsilon^{(2)}$ are identically distributed with *symmetric* common distribution. Then we have

$$\frac{1}{2}(Y^{(1)} + Y^{(2)}) = X + \frac{1}{2}(\varepsilon^{(1)} + \varepsilon^{(2)}),$$

where $(\varepsilon^{(1)} + \varepsilon^{(2)})/2$ has the same distribution as $(\varepsilon^{(1)} - \varepsilon^{(2)})/2 = (Y^{(1)} - Y^{(2)})/2$. So if we define $Y = (Y^{(1)} + Y^{(2)})/2, \varepsilon = (\varepsilon^{(1)} + \varepsilon^{(2)})/2$, and $\eta = (Y^{(1)} - Y^{(2)})/2$, then we have $Y = X + \varepsilon$, where η is observed with the same distribution as ε. In this example, $m = n$.

In such a case, a natural estimator of φ_ε is the empirical characteristic function based on $\eta_1 \dots, \eta_m$, namely,

$$\widehat{\varphi}_\varepsilon(t) = \frac{1}{m} \sum_{j=1}^m e^{it\eta_j}.$$

Suppose for a while that $\inf_{|t| \le h_n^{-1}} |\widehat{\varphi}_\varepsilon(t)| > 0$ with probability approaching one (which is indeed guaranteed under the assumptions stated below); then we may estimate K_n as

$$\widehat{K}_n(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \frac{\varphi_K(t)}{\widehat{\varphi}_\varepsilon(t/h_n)} dt, \tag{5}$$

and define the feasible version of $\widehat{f}_X^*$ as

$$\widehat{f}_X(x) = \frac{1}{nh_n} \sum_{j=1}^n \widehat{K}_n((x - Y_j)/h_n).$$

(Formally think of $\widehat{K}_n(x) = 0$ if the integral in (5) is not well-defined.) This estimator was first considered by Diggle and Hall (1993). Note that $\widehat{K}_n$ is also real-valued by symmetry of φ_K. In addition, we may estimate $\sigma_n^2(x)$ by

$$\widehat{\sigma}_n^2(x) = \frac{1}{n} \sum_{j=1}^n \widehat{K}_n^2((x - Y_j)/h_n) - \left(\frac{1}{n} \sum_{j=1}^n \widehat{K}_n((x - Y_j)/h_n) \right)^2 .$$

Now, consider the stochastic process

$$\widehat{Z}_n(x) = \frac{\sqrt{n}h_n \{\widehat{f}_X(x) - \mathrm{E}[\widehat{f}_X^*(x)]\}}{\widehat{\sigma}_n(x)}, \ x \in I,$$

where $\widehat{\sigma}_n(x) = \sqrt{\widehat{\sigma}_n^2(x)}$, and let

$$c_n(1 - \tau) = (1 - \tau)\text{-quantile of } \|\widehat{Z}_n\|_I.$$

(Formally think of $\widehat{Z}_n(x)$ as 0 if $\widehat{\sigma}_n(x) = 0$; the probability that $\widehat{\sigma}_n(x) = 0$ for some $x \in I$ is approaching zero under the assumptions stated below. The same rule applies to $\widehat{Z}_n^\xi$ defined below.) Then a band of the form

$$\left[\widehat{f}_X(x) - \frac{\widehat{\sigma}_n(x)}{\sqrt{n}h_n} c_n(1 - \tau), \widehat{f}_X(x) + \frac{\widehat{\sigma}_n(x)}{\sqrt{n}h_n} c_n(1 - \tau) \right], \ x \in I$$

will be a valid confidence band for f_X on I at level approximately $(1 - \tau)$ provided that the bias $\|f_X - \mathrm{E}[\widehat{f}_X^*(\cdot)]\|_I$ is sufficiently small.

The quantiles of $\|\widehat{Z}_n\|_I$ are still unknown, but it will be shown below that under suitable regularity conditions the distribution of $\|\widehat{Z}_n\|_I$ can be approximated by that of the supremum in absolute value of a tight Gaussian random variable Z_n^G in $\ell^\infty(I)$ with mean zero and the same covariance function as Z_n^*, So we propose to estimate the quantiles of $\|\widehat{Z}_n\|_I$ via the *Gaussian multiplier bootstrap* as in Chernozhukov et al. (2014b), described as follows.

Generate independent standard normal random variables $\xi_1, \ldots, \xi_n$, independent of the data $\mathcal{D}_n = \{Y_1, \ldots, Y_n, \eta_1, \ldots, \eta_m\}$, and consider the stochastic process

$$\widehat{Z}_n^\xi(x) = \frac{1}{\widehat{\sigma}_n(x)\sqrt{n}} \sum_{j=1}^n \xi_j \left\{ \widehat{K}_n((x - Y_j)/h_n) - n^{-1} \textstyle\sum_{j'=1}^n \widehat{K}_n((x - Y_{j'})/h_n) \right\}$$

for $x \in I$. Conditionally on the data $\mathcal{D}_n$, $\widehat{Z}_n^{\xi}(x), x \in I$ is a Gaussian process with mean zero and the covariance function "close" to that of Z_n^G. Hence we propose to estimate $c_n(1-\tau)$ by

$$\widehat{c}_n(1-\tau) = \text{conditional } (1-\tau)\text{-quantile of } \|\widehat{Z}_n^{\xi}\|_I \text{ given } \mathcal{D}_n,$$

and the resulting confidence band is

$$\widehat{\mathcal{C}}_n(x) = \left[\widehat{f}_X(x) - \frac{\widehat{\sigma}_n(x)}{\sqrt{n}h_n}\widehat{c}_n(1-\tau), \widehat{f}_X(x) + \frac{\widehat{\sigma}_n(x)}{\sqrt{n}h_n}\widehat{c}_n(1-\tau) \right], \ x \in I. \tag{6}$$

Note that the analysis below covers the case where I is singleton, i.e., $I = \{x_0\}$, and in that case, the above confidence band gives a confidence interval for $f_X(x_0)$.

In practice, the presence of $\widehat{\varphi}_\varepsilon$ in the denominator in the integrand in (5) could make the estimate $\widehat{f}_X$ numerically unstable; an approach to cope with this problem is to restrict the integral in (5) to the set $\{t \in \mathbb{R} : |\widehat{\varphi}_\varepsilon(t)| \geq m^{-1/2}\}$ (cf. Neumann, 1997). Likewise, in practice, replacing $\widehat{\sigma}_n(x)$ by $\max\{\widehat{\sigma}_n(x), \sqrt{h_n}\}$ in the definition of $\widehat{Z}_n$ and $\widehat{Z}_n^{\xi}$ would make resulting confidence bands numerically more stable. These modifications do not alter the asymptotic results presented below, and we will work with the original definitions of $\widehat{K}_n$ and $\widehat{\sigma}_n$.

3. Main results

In this section, we present the theorems that provide asymptotic validity of our confidence band described in the previous section. We first consider the case where the error density f_ε is ordinary smooth. We begin with stating the assumptions that guarantee our asymptotic results.

Assumption 1. *$|\varphi_X|$ is integrable on $\mathbb{R}$.*

Assumption 2. *Let K be a real-valued integrable function (kernel) on $\mathbb{R}$, not necessarily non-negative, such that $\int_{\mathbb{R}} K(x)dx = 1$, and its inverse Fourier transform φ_K is symmetric, continuously differentiable, and supported in $[-1, 1]$.*

Both of these assumptions are standard in the literature on deconvolution. Note that Assumption 1 implies that f_X is bounded and continuous, which in turn implies that f_Y is bounded and continuous via the relation $f_Y(y) = \int_{\mathbb{R}} f_X(y-x)f_\varepsilon(x)dx$.

The next assumption is concerned with the tail behavior of the error characteristic function φ_ε, which is an important factor that determines the difficulty of estimating f_X

(another factor is the smoothness of f_X). We assume here that the error density f_ε is ordinary smooth, i.e., $|\varphi_\varepsilon(t)|$ decays at most polynomially fast as $|t| \to \infty$.

Assumption 3. *The error characteristic function φ_ε is continuously differentiable and does not vanish on $\mathbb{R}$, and there exist constants $C_1 > 1$ and $\alpha > 0$ such that*

$$C_1^{-1}|t|^{-\alpha} \le |\varphi_\varepsilon(t)| \le C_1|t|^{-\alpha}, \quad |\varphi_\varepsilon'(t)| \le C_1|t|^{-\alpha-1}, \ \forall |t| \ge 1.$$

It is not difficult to see that Assumption 3 implies that

$$\inf_{|t|\le h_n^{-1}} |\varphi_\varepsilon(t)| \ge C_1^{-1}(1-o(1))h_n^\alpha$$

as $n \to \infty$.

Assumption 4. *Let $I \subset \mathbb{R}$ be a compact interval such that $f_Y(y) > 0$ for all $y \in I$.*

Now, recall that

$$\sigma_n^2(x) = \mathrm{Var}(K_n((x-Y)/h_n)).$$

In development of our theory, we will need that $\sigma_n^2(x)/h_n^{-2\alpha+1}$ is bounded away from zero on I, on which we would like to make inference on f_X. It will be shown in Lemma 3 that Assumptions 1–4 guarantee that $\sigma_n^2(x)/h_n^{-2\alpha+1}$ is bounded away from zero on I for sufficiently large n.

The next assumption is a mild moment condition on the error distribution, which is used in establishing uniform convergence rates of the empirical characteristic function (see Lemma 4).

Assumption 5. $\mathrm{E}[|\varepsilon|^p] < \infty$ *for some $p > 0$.*

The next assumption (mildly) restricts the bandwidth h_n and the sample size $m = m_n$ for f_ε.

Assumption 6. *(a) $n^{-1/2}h_n^{-1} = o\{(\log(1/h_n))^{-1}\}$. (b) $m = m_n \to \infty$ as $n \to \infty$, and*

$$m^{-1} = o\left[\{(nh_n \log(1/h_n)\}^{-1} \wedge \left\{h_n^{2\alpha+2}(\log(1/h_n))^{-2}\right\}\right].$$

Remark 1. For an illustrative purpose, consider the canonical case where $m = n$. Then Assumption 6 reduces to the following simple condition:

$$nh_n^{2+2\alpha}/(\log(1/h_n))^2 \to \infty. \tag{7}$$

The conventional "optimal" bandwidth that minimizes the MISE of the kernel estimator (when f_ε is known) is proportional to $n^{-1/(2\alpha+2\beta+1)}$ where β is the "smoothness" of f_X

(cf. Fan, 1991a), and so condition (7) is satisfied with this bandwidth if $\beta > 1/2$. See also Corollary 2 below.

The following theorem establishes that the distribution of the supremum in absolute value of the stochastic process $\widehat{Z}_n(x), x \in I$ can be approximated by that of a tight Gaussian random variable Z_n^G in $\ell^\infty(I)$ with mean zero and the same covariance function as Z_n^*. This theorem is a building block for establishing validity of the Gaussian multiplier bootstrap described in the previous section. Recall that a Gaussian process $Z = \{Z(x) : x \in I\}$ indexed by I is a tight random variable in $\ell^\infty(I)$ if and only if I is totally bounded for the intrinsic pseudo-metric $\rho_2(x, y) = \sqrt{\mathrm{E}[\{Z(x) - Z(y)\}^2]}, x, y \in I$, and Z has sample paths almost surely uniformly ρ_2-continuous; see van der Vaart and Wellner (1996, p.41).

Theorem 1. *Under Assumptions 1–6, for each sufficiently large n, there exists a tight Gaussian random variable Z_n^G in $\ell^\infty(I)$ with mean zero and the same covariance function as Z_n^*, and such that as $n \to \infty$,*

$$\sup_{z \in \mathbb{R}} |\mathrm{P}\{\|\widehat{Z}_n\|_I \le z\} - \mathrm{P}\{\|Z_n^G\|_I \le z\}| \to 0.$$

When I is not singleton, it is possible to further show that $\|Z_n^G\|_I$ (and hence $\|\widehat{Z}_n\|_I$) properly normalized converges in distribution to a Gumbel distribution (i.e., a Smirnov-Bickel-Rosenblatt type limit theorem) under additional (substantial) conditions, as in Bissantz et al. (2007). However, we intentionally stop at the "intermediate" Gaussian approximation and do not derive the Gumbel approximation, because of the following two reasons.

- The Gumbel approximation is poor. The coverage error of the resulting confidence band is of order $1/\log n$ (Hall, 1991).
- Deriving the Gumbel approximation requires additional substantial conditions.

Because of the slow rate of the Gumbel approximation, it is often preferred to use versions of bootstraps to construct confidence bands for nonparametric density and regression functions (see, e.g., Claeskens and van Keilegom, 2003; Bissantz et al., 2007), but the Gumbel approximation was considered to be a building block for validity of the bootstraps. It was, however, pointed out in Chernozhukov et al. (2014b) that the intermediate Gaussian approximation (such as that in Theorem 1) is sufficient for showing validity of bootstraps. We defer the discussion on the regularity conditions to the end of this section.

Another technicality in the proof of Theorem 1 is to bound the estimation error of $\widehat{\varphi}_\varepsilon$. Dattner et al. (2016, p.172) derive a bound on $\|\widehat{f}_X - \widehat{f}_X^*\|_{\mathbb{R}}$ which is of order $O_{\mathrm{P}}\{h_n^{-\alpha}(mh_n)^{-1/2}\}$, but this rate is not sufficient for our purpose and in particular excludes the case with $m = n$ in Theorem 1; see Step 2 in the proof of Theorem 1.

As a byproduct of the techniques used to prove Theorem 1, we can derive uniform convergence rates of $\widehat{f}_X$ on $\mathbb{R}$. In the next corollary, Assumption 4 is not needed.

Corollary 1. *Suppose that Assumptions 1–3, 5, and 6 are satisfied. Then as $n \to \infty$,*

$$\|\widehat{f}_X - \mathrm{E}[\widehat{f}_X^*(\cdot)]\|_{\mathbb{R}} = O_{\mathrm{P}}\left\{h_n^{-\alpha}(nh_n)^{-1/2}\sqrt{\log(1/h_n)}\right\}.$$

Corollary 1 does not take into account the bias $\|\mathrm{E}[\widehat{f}_X^*(\cdot)] - f_X\|_{\mathbb{R}}$, but the above rate is the correct one for the "variance part" when f_ε is known. The situation here is bit complicated since f_ε is unknown and because of the role of m. So we only present the following corollary, which deals with the canonical case where $m = n$. In the following, for $\beta > 0$ and $B > 0$, let $\Sigma(\beta, B)$ denote a Hölder ball of functions on $\mathbb{R}$ with smoothness β and radius B, namely,

$$\begin{aligned}\Sigma(\beta, B) = \{f : \mathbb{R} \to \mathbb{R} : f \text{ is } k\text{-times differentiable and such that} \\ |f^{(k)}(x) - f^{(k)}(y)| \le B|x-y|^{\beta-k}\ \forall x, y \in \mathbb{R}\},\end{aligned}$$

where k is the integer such that $k < \beta \le k+1$.

Corollary 2. *Suppose that Assumptions 1–3 and 5 are satisfied, and consider the case where $m = n$. Further, suppose that $f_X \in \Sigma(\beta, B)$ for some $\beta > 1/2, B > 0$, and that φ_K is $(k+3)$-times continuously differentiable with $\varphi_K^{(\ell)}(0) = 0$ for all $\ell = 1, \dots, k$ where k is the integer such that $k < \beta \le k+1$. Take $h_n = C(n/\log n)^{-1/(2\alpha+2\beta+1)}$ for any constant $C > 0$; then*

$$\|\widehat{f}_X - f_X\|_{\mathbb{R}} = O_{\mathrm{P}}\{(n/\log n)^{-\beta/(2\alpha+2\beta+1)}\}.$$

Remark 2. Informally, $(n/\log n)^{-\beta/(2\alpha+2\beta+1)}$ is the minimax rate of convergence for estimating f_X under the sup-norm loss when f_ε is known and when $f_X \in \Sigma(\beta, B)$ and $|\varphi_\varepsilon(t)|$ decays like $|t|^{-\alpha}$ as $|t| \to \infty$. See Theorem 1 in Lounici and Nickl (2011) for the precise formulation. In fact, the proof of Theorem 1 in Lounici and Nickl (2011) continues to hold when f_ε is unknown and $(\eta_1, \dots, \eta_n)$ are independent from $(Y_1, \dots, Y_n)$ – in their proof modify P_k^n to be the distribution admitting the joint density $\prod_{j=1}^n (f_X^k * f_\varepsilon)(y_j) f_\varepsilon(\eta_j)$ with f_X^k denoting their f_k. So $(n/\log n)^{-\beta/(2\alpha+2\beta+1)}$ gives a lower bound on the minimax rate under the sup-norm loss for the class of joint distributions of (Y, η) where the marginal densities of Y and η are given by $f_X * f_\varepsilon$ and f_ε, respectively, and where $f_X \in \Sigma(\beta, B)$

with $\max_{0\le \ell\le k}\|f_X^{(\ell)}\|_{\mathbb{R}} \le B$ and f_ε verifies Assumption 3. Corollary 2 then shows that $\widehat{f}_X$ attains the minimax rate under the sup-norm loss.

Remark 3. The literature on uniform convergence rates in deconvolution is limited. Lounici and Nickl (2011) and Giné and Nickl (2016, Section 5.3.2) derive uniform convergence rates for deconvolution wavelet and kernel density estimators on the entire real line assuming that the error density is known; Dattner et al. (2016, Proposition 2.6) derive uniform convergence rates for the deconvolution kernel density estimator with the estimated error characteristic function, but on a bounded interval. So their results do not cover the above corollaries, although combination of a bound on $\|\widehat{f}_X - \widehat{f}_X^*\|_{\mathbb{R}}$ in the proof of Proposition 2.6 in Dattner et al. (2016) with the results in Giné and Nickl (2016, Section 5.3.2) would yield Corollary 2.

Now, we present validity of the multiplier bootstrap confidence band described in the previous section.

Theorem 2. *Under Assumptions 1–6, as $n\to\infty$,*

$$\sup_{z\in\mathbb{R}} |\mathrm{P}\{\|\widehat{Z}_n^\xi\| \le z \mid \mathcal{D}_n\} - \mathrm{P}\{\|Z_n^G\| \le z\}| \xrightarrow{\mathrm{P}} 0, \tag{8}$$

where $\mathcal{D}_n = \{Y_1,\dots,Y_n,\eta_1,\dots,\eta_m\}$ and Z_n^G is given in Theorem 1. Therefore, if we denote by $\widehat{c}_n(1-\tau)$ the $(1-\tau)$-quantile of the conditional distribution of $\|\widehat{Z}_n^\xi\|_I$ given $\mathcal{D}_n$ where $\tau\in(0,1)$ is given, then we have

$$\mathrm{P}\{\|\widehat{Z}_n\| \le \widehat{c}_n(1-\tau)\} \to 1-\tau$$

as $n\to\infty$.

Theorem 2 shows that the multiplier bootstrap confidence band $\widehat{\mathcal{C}}_n$ defined in (6) contains the surrogate function $\mathrm{E}[\widehat{f}_X^*(\cdot)]$ on I with probability at least $1-\tau+o(1)$ as $n\to\infty$. If f_X belongs to a Hölder ball $\Sigma(\beta,B)$, then $\widehat{\mathcal{C}}_n$ will be a valid confidence band for f_X provided that h_n is chosen in such a way that $h_n^{\alpha+\beta}\sqrt{nh_n\log(1/h_n)} \to 0$, which corresponds to undersmoothing.

Corollary 3. *Suppose that Assumptions 1–6 are satisfied. Further, suppose that $f_X\in\Sigma(\beta,B)$ for some $\beta>0, B>0$, and that φ_K is $(k+3)$-times continuously differentiable with $\varphi^{(\ell)}(0)=0$ for all $\ell=1,\dots,k$ where k is the integer such that $k<\beta\le k+1$. Consider the multiplier bootstrap confidence band $\widehat{\mathcal{C}}_n$ defined in (6). Then as $n\to\infty$,*

$$\mathrm{P}\{f_X(x)\in\widehat{\mathcal{C}}_n(x)\ \forall x\in I\}\to 1-\tau,$$

provided that

$$h_n^{\alpha+\beta}\sqrt{nh_n\log(1/h_n)} \to 0. \tag{9}$$

Consider the canonical case where $m = n$. Then the conditions on the bandwidth h_n in Corollary 3 reduce to

$$nh_n^{2\alpha+2}/(\log(1/h_n))^2 \to \infty, \ h_n^{\alpha+\beta}\sqrt{nh_n\log(1/h_n)} \to 0,$$

and so we need $\beta > 1/2$ in order to ensure the existence of the bandwidth h_n satisfying these conditions.

Remark 4 (Comparison with the conditions in Bissantz et al. (2007))**.** Bissantz et al. (2007) is an important pioneering work on confidence bands in deconvolution. They assume that the error density is known and ordinary smooth, and show that

$$\sqrt{2\log(1/h_n)}(\|\sqrt{n}h_n^{\alpha+1/2}(\widehat{f}_X^* - \mathrm{E}[\widehat{f}_X^*(\cdot)])/\sqrt{f_Y(\cdot)}\|_{[0,1]}/C_{K,1}^{1/2} - d_n)$$

converges in distribution to a Gumbel distribution, where

$$d_n = \sqrt{2\log(1/h_n)} + \frac{\log(C_{K,2}^{1/2}/2\pi)}{\sqrt{2\log(1/h_n)}},$$

and where $C_{K,1}$ and $C_{K,2}$ are some constants; see Bissantz et al. (2007) for their explicit values. Further, they show validity of the nonparametric bootstrap for approximating the distribution of $\|(\widehat{f}_X^* - \mathrm{E}[\widehat{f}_X^*(\cdot)])/\sqrt{f_Y(\cdot)}\|_{[0,1]}$; see their Theorem 2.

Since we work with a different setting than Bissantz et al. (2007), namely we assume that f_ε is unknown and an auxiliary sample from f_ε is available, the regularity conditions in the present paper are not directly comparable to those of Bissantz et al. (2007). However, it is worthwhile to point out that conditions on the error characteristic function are significantly relaxed in the present paper. Indeed, their Assumption 2 is substantially more restrictive than our Assumption 3. The reasons that they require their Assumption 2 are: 1) they use the KMT strong approximation (Komlós et al., 1975) to the empirical process $[0,1] \ni x \mapsto \sqrt{n}h_n^{\alpha+1/2}(\widehat{f}_X^*(x) - \mathrm{E}[\widehat{f}_X^*(x)])/\sqrt{f_Y(x)}$, to which end a bound on the total variation of K_n is needed, and their Assumption 2 (a) serves that role; 2) their analysis relies on the Gumbel approximation, to which end, they require further approximations beyond the KMT approximation based on the extreme value theory (cf Leadbetter et al., 1983) and consequently require some extra assumptions, namely, their Assumption 2 (b).

In the present paper, we build upon the intermediate Gaussian and multiplier bootstrap approximation theorems developed in Chernozhukov et al. (2014a,b, 2016), and regularity conditions needed to apply those techniques are typically much weaker than those for the KMT and Gumbel approximations. In particular, we do not need a bound on the total variation of K_n; instead, we need that the class of functions $\{y \mapsto K_n((x-y)/h_n) : x \in I\}$ is of Vapnik-Chervonenkis type, and to that end, thanks to Lemma 1 in Giné and Nickl (2009), it is enough to prove that K_n has bounded quadratic variation of order $h_n^{-\alpha}$, which is ensured by our Assumptions 2 and 3 (see Lemmas 1 and 2 ahead). In addition, in contrast to Bissantz et al. (2007), we do not need that $\sigma_n^2(x)/h_n^{-2\alpha+1}$ has a fixed limit; we only need that $\sigma_n^2(x)/h_n^{-2\alpha+1}$ is bounded away from zero uniformly in $x \in I$.

Furthermore, the intermediate Gaussian and multiplier bootstrap approximations apply not only to the ordinary smooth case, but also to the super-smooth case, as discussed in Section 6, and so they enable us to study confidence bands for f_X in a unified way (although in the super-smooth case we require $m/n \to \infty$). On the other hand, as shown in van Es and Gugushvili (2008), the Gumbel approximation does not hold for the super-smooth case in general (see also Remark 5 ahead).

4. Simulation studies

4.1. **Simulation framework.** In this section, we present simulation studies to evaluate finite-sample performance of the inference method developed in the previous two sections. We generate data from the model introduced in Example 1. For distributions of the primitive latent variables $(X, \varepsilon^{(1)}, \varepsilon^{(2)})$, we consider two alternative models described below.

In the first model, X is drawn from the centered normal distribution $N(0, \sigma_X^2)$, and $\varepsilon^{(1)}$ and $\varepsilon^{(2)}$ are drawn from the Laplace distribution with $(0, 1)$ as the location and scale parameters. This Laplace distribution is symmetric around zero, and therefore the premise of Example 1 regarding the error variables is satisfied. The characteristic function for the error variables is given by $\varphi_{\varepsilon^{(1)}}(t) = \varphi_{\varepsilon^{(2)}}(t) = 1/(1+t^2)$. Consequently, the distribution of $\varepsilon = (\varepsilon^{(1)} + \varepsilon^{(2)})/2$ has its characteristic function not vanishing on $\mathbb{R}$ and is ordinary smooth with $\alpha = 4$. On the other hand, the signal X has a super-smooth distribution. This setting conveniently yields the signal-to-noise ratio given by

$$\sqrt{\frac{\operatorname{Var}(X)}{\operatorname{Var}(\varepsilon)}} = \sqrt{\frac{\sigma_X^2}{\operatorname{Var}(\varepsilon^{(1)})/4 + \operatorname{Var}(\varepsilon^{(2)})/4}} = \sigma_X.$$

In the second model, X is drawn from the chi-squared distribution $\chi^2(df)$, and $\varepsilon^{(1)}$ and $\varepsilon^{(2)}$ are drawn from the Laplace distribution with $(0, \sqrt{2})$ as the location and scale parameters. The characteristic function for the error variables is given by $\varphi_{\varepsilon^{(1)}}(t) =$

	Model 1	Model 2
Distribution of X	$N(0,\sigma_X^2)$	$\chi^2(df)$
Distribution of $(\varepsilon^{(1)},\varepsilon^{(2)})$	Laplace $(0,1)$	Laplace $(0,\sqrt{2})$
Smoothness of X	Super	Ordinary
Smoothness of ε	Ordinary	Ordinary
Signal-to-noise ratio	σ_X	$\sqrt{df}$

Table 1. A summary of the two models considered for simulation studies.

$\varphi_{\varepsilon^{(2)}}(t) = 1/(1+2t^2)$. The distribution of $\varepsilon = (\varepsilon^{(1)}+\varepsilon^{(2)})/2$ has its characteristic function not vanishing on $\mathbb{R}$ and is ordinary smooth with $\alpha = 4$. The signal X also has an ordinary-smooth distribution. In this setting, the signal-to-noise ratio is given by

$$\sqrt{\frac{\mathrm{Var}(X)}{\mathrm{Var}(\varepsilon)}} = \sqrt{\frac{2df}{\mathrm{Var}(\varepsilon^{(1)})/4 + \mathrm{Var}(\varepsilon^{(2)})/4}} = \sqrt{df}.$$

Table 1 summarizes the two models and their relevant properties.

The observed potion of data, $\mathcal{D}_n = \{Y_1,\dots,Y_n,\eta_1,\dots,\eta_n\}$, is constructed by $Y_j = (Y_j^{(1)}+Y_j^{(2)})/2$ and $\eta_j = (Y_j^{(1)}-Y_j^{(2)})/2$, where $Y_j^{(1)} = X_j + \varepsilon_j^{(1)}$ and $Y_j^{(2)} = X_j + \varepsilon_j^{(2)}$, for each $j = 1,\dots,n$. The three primitive latent variables, X, $\varepsilon^{(1)}$, and $\varepsilon^{(2)}$ are independently generated. We use Monte Carlo simulations to compute the coverage probabilities of our multiplier bootstrap confidence bands for f_X on the interval $I = [-2\sigma_X, 2\sigma_X]$ for Model 1 and on the interval $I = [\mu_X/2, \mu_X + 2\sigma_X]$ where $(\mu_X,\sigma_X) = (df, \sqrt{2df})$ for Model 2. We use the kernel function K defined by its inverse Fourier transform φ_K as follows:

$$\varphi_K(t) = \begin{cases} 1 & \text{if } |t| \le c \\ \exp\left\{\frac{-b\exp(-b/(|t|-c)^2)}{(|t|-1)^2}\right\} & \text{if } c < |t| < 1 \\ 0 & \text{if } |t| \ge 1 \end{cases},$$

where $b = 1$ and $c = 0.05$ (cf. Politis and Romano, 1999; Bissantz et al., 2007). Note that φ_K is symmetric and supported in $[-1,1]$, and its Fourier transform K is integrable with $\int_{\mathbb{R}} K(x)dx = 1$. While we use this "flat top kernel" following earlier work in the literature, we emphasize that this flat-top feature is not required in our theory. Given this kernel function and the data $\mathcal{D}_n$, we have all the ingredients, namely, $\widehat{K}_n$, $\widehat{f}_X$, and $\widehat{\sigma}_n$, to compute confidence bands, provided an appropriate choice of the bandwidth h_n.

Our theory prescribes admissible asymptotic rates for the bandwidth h_n that require undersmoothing. The literature provides data-driven approaches to bandwidth selection, which are usually based on minimizing the MISE. These data-driven approaches tend to yield non-under-smoothing bandwidths, and do not conform with our requirements. We adopt the two-step selection method developed in Bissantz et al. (2007, Section 5.2) that aims to select undersmoothing bandwidths. The first step selects a pilot bandwidth h_n^P based on a data-driven approach. We simply use a normal reference bandwidth (Delaigle and Gijbels, 2004, Section 3.1) for h_n^P. Once the pilot bandwidth h_n^P is obtained, we next make a list of candidate bandwidths $h_{n,j} = (j/J)h_n^P$ for $j = 1, \ldots, J$. The deconvolution estimate based on the j-th candidate bandwidth is denoted by $\widehat{f}_{X,j}$. The second step in the two step approach chooses the largest bandwidth $h_{n,j}$ such that the adjacent uniform distance $\|\widehat{f}_{n,j-1} - \widehat{f}_{n,j}\|_I$ is larger than $\rho\|\widehat{f}_{n,J-1} - \widehat{f}_{n,J}\|_I$ in the pilot case for some $\rho > 1$. Similarly to the values recommended by Bissantz et al. (2007), we find that $J \approx 20$ and $\rho \approx 3$ work well in our simulation studies.

4.2. **Simulation results.** Simulated uniform coverage probabilities are computed for each of the three nominal coverage probabilities, 80%, 90%, and 95%, based on 2,000 Monte Carlo iterations. In each run of the simulation, we generate 2,500 multiplier bootstrap replications given the observed data $\mathcal{D}_n$ to compute the estimated critical values, $\widehat{c}_n(1-\tau)$. Results under Model 1 are summarized in Table 2 for each of the three different cases of the signal-to-noise ratio: $\sigma_X \in \{2.0, 4.0, 8.0\}$, and for each of the three sample sizes $n = m \in \{250, 500, 1,000\}$. Observe that the simulated probabilities are close to the respective nominal probabilities. Not surprisingly, the size tends to be more accurate for the results based on larger sample sizes. There is some difference in the simulated coverage probabilities across the signal-to-noise ratios of the data generating model, but the variation is not notable.

Results under Model 2 are summarized in Table 3 for each of the three different cases of the signal-to-noise ratio: $\sqrt{\mu_X} \in \{2.0, 4.0, 8.0\}$, and for each of the three sample sizes $n = m \in \{250, 500, 1,000\}$. Compared to the results in Table 2, the simulated probabilities are less close to the respective nominal probabilities in Table 3. The simulated probabilities perform better for the results based on larger sample sizes. Unlike the relatively homogeneous results in Table 2, we do observe notable differences in the simulated coverage probabilities across the signal-to-noise ratios in Table 3.

We next use Monte Carlo simulations to show the power of uniform specification tests based on our uniform confidence band. Thus far, simulation results for coverage probabilities are reported for the true density function f_X. We now consider a list of alternative specifications of f_X, given by $f_{X,\mu_X}(x) = (2\pi)^{-1/2}e^{-(x-\mu_X)^2/2}$ for $\mu_X \in \{0.0, 0.1, 0.2, 0.3, 0.4, 0.5\}$,

Nominal Coverage Probability $(1-\tau)$	Sample Size (n)	Signal-to-Noise Ratio 2.0	4.0	8.0
0.800	250	0.762	0.755	0.728
	500	0.786	0.746	0.763
	1,000	0.784	0.750	0.754
0.900	250	0.870	0.866	0.843
	500	0.897	0.862	0.863
	1,000	0.897	0.862	0.863
0.950	250	0.930	0.936	0.907
	500	0.951	0.929	0.923
	1,000	0.942	0.933	0.933

TABLE 2. Simulated uniform coverage probabilities of f_X by estimated confidence bands in $I = [-2\sigma_X, 2\sigma_X]$ under a normal random variable X and an independent Laplace random vector $(\varepsilon^{(1)}, \varepsilon^{(2)})$. The simulated probabilities are computed for each of the three nominal coverage probabilities, 80%, 90%, and 95%, based on 2,000 Monte Carlo iterations.

and $f_{X,\sigma_X}(x) = (2\pi\sigma_X^2)^{-1/2}e^{-x^2/(2\sigma_X^2)}$ for $\sigma_X \in \{1.0, 1.1, 1.2, 1.3, 1.4, 1.5\}$. For the errors, we again consider the independent Laplace random vector $(\varepsilon^{(1)}, \varepsilon^{(2)})$ as in Model 1. Figure 1 plots simulated coverage probabilities for the list of the alternative specifications of f_{X,μ_X} (top) and for the list of the alternative specifications of f_{X,σ_X} (bottom) for the nominal coverage probability of $(1-\tau) = 0.90$. The three curves are drawn for each of the three sample sizes $n \in \{250, 500, 1,000\}$. Observe that, under the true specification (i.e., $\mu_X = 0.0$ in the top graph and $\sigma_X = 1.0$ in the bottom graph), the simulated coverage probabilities are close to the nominal coverage probability of 0.90, with the case of $n = 1,000$ being the closest and the case of $n = 250$ being the farthest. On the other hand, as the specification deviates away from the truth (i.e., as μ_X or σ_X increases), the nominal coverage probabilities decrease, with the case of $n = 1,000$ being the fastest and the case of $= 250$ being the slowest. These results evidence the power of the uniform specification tests.

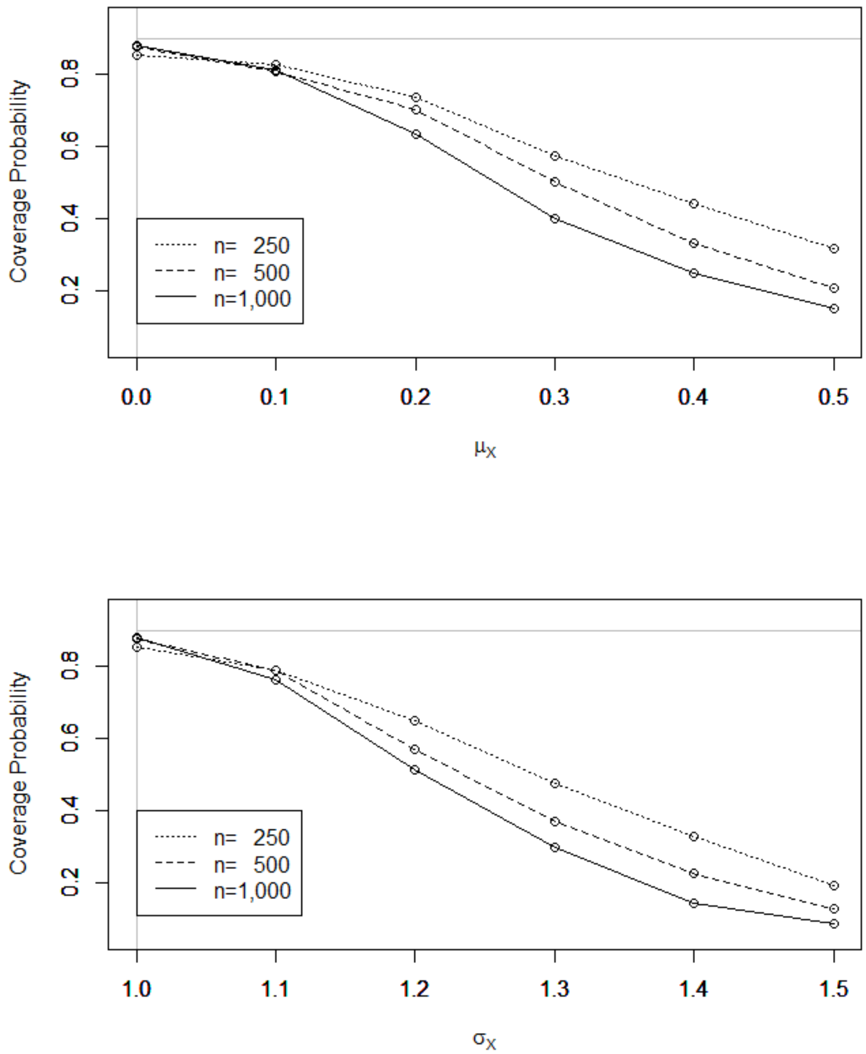

FIGURE 1. Simulated uniform coverage probabilities for alternative specifications of f_X by estimated confidence bands in $I = [-2, 2]$ under the standard normal random variable X and an independent Laplace random vector $(\varepsilon^{(1)}, \varepsilon^{(2)})$. For the top graph, the list of alternative specifications are given by $f_{X,\mu_X}(x) = (2\pi)^{-1/2} e^{-(x-\mu_X)^2/2}$ for $\mu_X \in \{0.0, 0.1, 0.2, 0.3, 0.4, 0.5\}$. For the bottom graph, the list of alternative specifications are given by $f_{X,\sigma_X}(x) = (2\pi\sigma_X^2)^{-1/2} e^{-x^2/(2\sigma_X^2)}$ for $\sigma_X \in \{1.0, 1.1, 1.2, 1.3, 1.4, 1.5\}$. The simulated probabilities are computed for the nominal coverage probability of 90% based on 2,000 Monte Carlo iterations.

Nominal Coverage Probability $(1-\tau)$	Sample Size (n)	Signal-to-Noise Ratio 2.0	4.0	8.0
0.800	250	0.636	0.740	0.666
	500	0.698	0.751	0.685
	1,000	0.732	0.749	0.702
0.900	250	0.760	0.853	0.799
	500	0.822	0.867	0.822
	1,000	0.848	0.878	0.836
0.950	250	0.834	0.927	0.871
	500	0.834	0.928	0.897
	1,000	0.912	0.940	0.906

Table 3. Simulated uniform coverage probabilities of f_X by estimated confidence bands in $I = [\mu_X/2, \mu_X + 2\sigma_X]$ under a chi-squared variable X and an independent Laplace random vector $(\varepsilon^{(1)}, \varepsilon^{(2)})$. The simulated probabilities are computed for each of the three nominal coverage probabilities, 80%, 90%, and 95%, based on 2,000 Monte Carlo iterations.

5. Application to OCS Wildcat auctions

In this section, we apply our method to the Outer Continental Shelf (OCS) Auction Data (see Hendricks et al. (1987) for details), and construct a confidence band for the density of mineral rights on oil and gas on offshore lands off the coasts of Texas and Louisiana in the gulf of Mexico. We focus on "wildcat sales," referring to sales of those oil and gas tracts whose geological or seismic characteristics are unknown to participating firms. The sales rule follows the first-price sealed-bid auction mechanism, where participating firms simultaneously submit sealed bids, and the highest bidder pays the price they submitted to receive the right for the tract. Firms who are willing to participate in sales can carry out a seismic investigation before the sales date in order to estimate the value of mineral rights. The private value $Y^{(1)}$ (in the logarithm of US dollars per acre) obtained by firm 1 through its investigation of the tract is treated as a measure of the ex post value X (also known as the common value, in the logarithm of US dollars per acre) with an error $\varepsilon^{(1)}$, i.e., $Y^{(1)} = X + \varepsilon^{(1)}$. Collecting the private values $(Y^{(1)}, Y^{(2)})$ for pairs of firms across various wildcat auctions, we can obtain data necessary to construct a uniform confidence band for the density f_X of ex post mineral right values under our assumptions.

For this setup and for this data set, Li et al. (2002) apply the method of Li and Vuong (1998) to nonparametrically estimate f_X, but they do not obtain a confidence band. In their analysis, firms' private values $(Y^{(1)}, Y^{(2)})$ are first recovered from bid data through a widely used method in economics which is based on an equilibrium restriction (Bayesian Nash equilibrium) for the first-price sealed-bid auction mechanism – see Li et al. (2002) for details. In this paper, we directly take these private values $(Y^{(1)}, Y^{(2)})$ as the data to be used as an input for our analysis. The sample size is 434, consisting of 217 tracts with 2 firms in each tract. Figure 2 depicts a simple kernel density estimate of the private values $(Y^{(1)}, Y^{(2)})$ in the logarithm of US dollars per acre. This figure essentially reproduces Figure 3 of Li et al. (2002) using the same data set of $(Y^{(1)}, Y^{(2)})$ values. Note that the private value distribution is bimodal. We next construct our auxiliary data (Y, η) following Example 1 under the assumption that the errors $(\varepsilon^{(1)}, \varepsilon^{(2)})$ in the firms' private values are symmetrically and identically distributed. It is supposed that, "because bidders have equal access to the same information about the tract, they can be considered as identical ex ante" (Li et al., 2002). This supports at least the identical distribution assumption that we make.

We continue to use the same kernel function and the same bandwidth selection rule as those ones used for simulation studies in Section 4. Confidence bands for f_X are constructed using 25,000 multiplier bootstrap replications. Figure 3 shows the obtained 90% and 95% confidence bands in light gray and dark gray, respectively. The black curve draws our nonparametric estimate of f_X, and runs at the center of the bands. Not surprisingly, our nonparametric estimate, our confidence bands, and the nonparametric estimate obtained by Li et al. (2002) – shown in their Figure 4 – are all similar to each other, and share the same qualitative characteristics. First, unlike the bimodal density of the private values $(Y^{(1)}, Y^{(2)})$, the density of the common value X is suggested to be single-peaked in all the results of our analysis and that of Li et al. (2002). Second, as Li et al. (2002) also emphasized, the small bump in the estimated density f_X around $x = 6$ is common in all the results. Furthermore, our confidence bands do not include zero at this locality, $x = 6$. This result can be viewed as a statistical evidence in support of a significant presence of such a bump pointed out by Li et al. (2002). A more global look into the graph suggests that the 95% confidence band is bounded away from zero on the interval $[3.90, 6.32]$. In other words, we can conclude that the ex post value of mineral rights in the US dollars per acre is supported on a superset of the interval $[\exp(3.90), \exp(6.32)] \approx [49, 555]$, if we consider a set of density functions contained in the 95% confidence band.

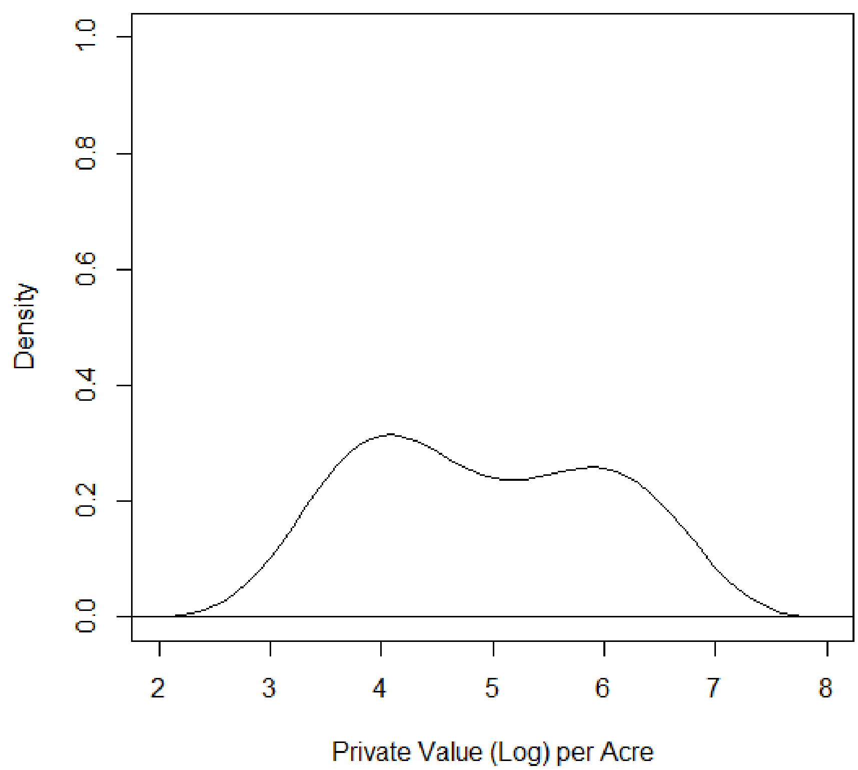

FIGURE 2. A kernel density estimate of the private values $(Y^{(1)}, Y^{(2)})$ in the logarithm of US dollars per acre.

Finally, for the purpose of illustrating the advantage of our method over the existing others, we show an empirical result that we would obtain under the assumption that the distribution of ε were known as in the previous literature. Specifically, for this exercise, we assume that ε follows Laplace $(0, \sigma_\varepsilon^2/2)$, where $\sigma_\varepsilon^2 = 0.168$. Figure 4 shows the result. Some qualitative features still remain the same as those of the previous result based on unknown error distribution. On the other hand, there are some notable differences too. Most importantly, the 95% confidence band now contains zero around the locality, $x = 6$, of the aforementioned bump in the estimated density of X. The contrast between Figures 3 and 4 show that there can be non-trivial differences in statistical implications of confidence bands between the case of assuming known error distribution and the case of assuming unknown error distribution.

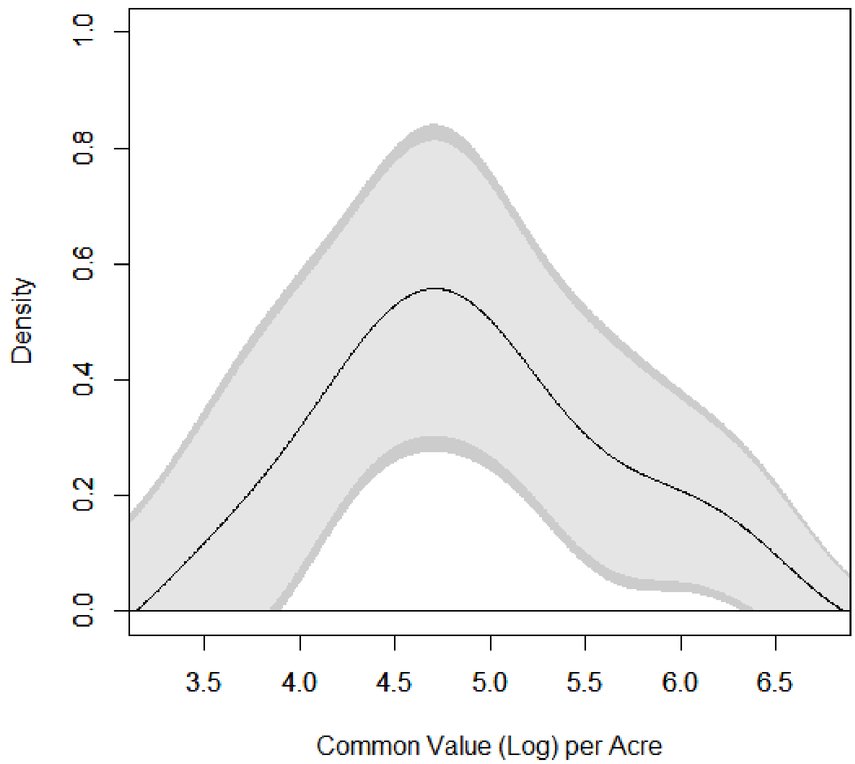

Figure 3. 90% and 95% confidence bands for f_X marked in light gray and dark gray, respectively. The black curve in the middle indicates our deconvolution kernel density estimate of f_X.

6. Extensions to super-smooth case

In this section, we consider extensions of the results on confidence bands to the case where the error density f_ε is super-smooth. We still keep Assumptions 1 and 5, but require a different set of assumptions on the kernel K, the bandwidth h_n, and the sample size m for f_ε. It turns out that from a technical reason, we require $m/n \to \infty$ in the super-smooth case, and so Example 1 is formally not covered in the super-smooth case. Still, we believe that the extensions to the super-smooth case are of some interest.

We modify Assumptions 2, 3, and 6 as follows. First, for the kernel and error characteristic functions, we assume the following conditions.

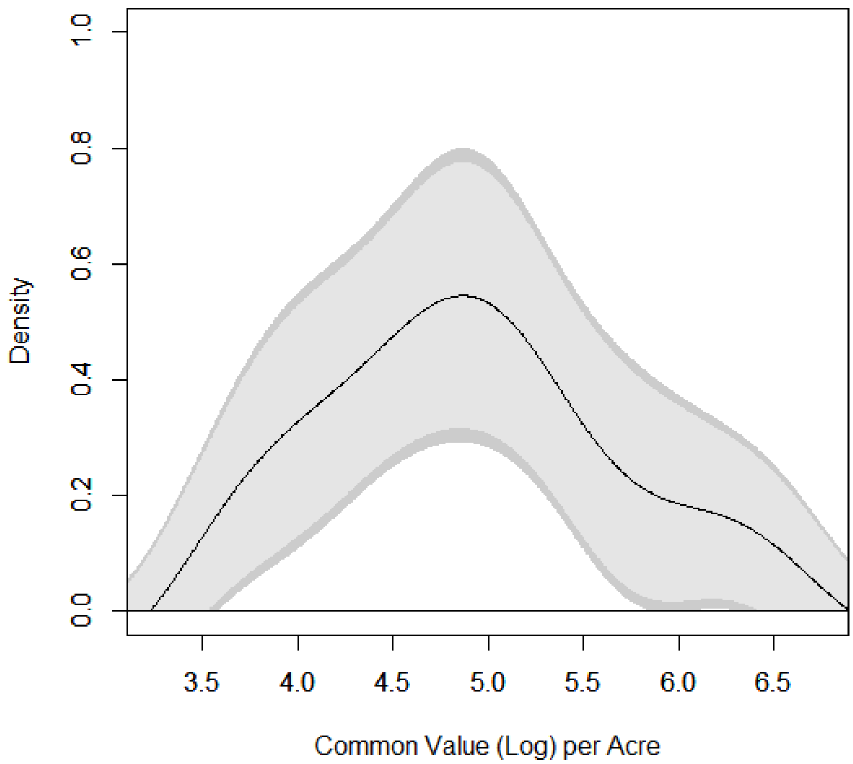

FIGURE 4. 90% and 95% confidence bands for f_X marked in light gray and dark gray, respectively, under the assumption that the distribution of η is Laplace. The black curve in the middle indicates our deconvolution kernel density estimate of f_X.

Assumption 7. *Let K be a kernel on $\mathbb{R}$ such that its inverse Fourier transform φ_K is symmetric and supported in $[-1,1]$. Further, there exist constants $C_2 \neq 0$ and $\lambda \geq 0$ such that*

$$\varphi_K(1-t) = C_2 t^\lambda + o(t^\lambda), \quad t \downarrow 0.$$

Assumption 8. *The error characteristic function φ_ε does not vanish on $\mathbb{R}$, and there exist constants $C_3 \in \mathbb{R} \setminus \{0\}, \gamma > 1, \gamma_0 \in \mathbb{R}, \nu > 0$ such that*

$$\varphi_\varepsilon(t) = C_3(1 + o(1))|t|^{\gamma_0} e^{-\nu|t|^\gamma}, \quad |t| \to \infty.$$

These assumptions are adapted from van Es and Uh (2005). Assumption 8 covers cases where the error characteristic function decays exponentially fast as $|t| \to \infty$, thereby covering cases where the error density is super-smooth. However, Assumption 8 is more restrictive than standard super-smoothness conditions; e.g., it excludes the Cauchy error. This assumption is needed to derive a lower bound on $\sigma_n^2(x)$; see the following discussion.

Assumptions 7 and 8, together with the assumption that $\mathrm{E}[Y^2] < \infty$, ensure that $\sigma_n^2(x) = \mathrm{Var}(K_n(x-Y)/h_n)$ is expanded as

$$\sigma_n^2(x) = (1+o(1))\frac{C_2^2}{2C_3^2\pi^2}(\nu\gamma)^{-2\lambda-2}(\Gamma(\lambda+1))^2 h_n^{2\gamma(1+\lambda)+2\gamma_0}e^{2\nu h_n^{-\gamma}} \tag{10}$$

as $n \to \infty$; see the proof of Theorem 1.5 in van Es and Uh (2005). It is not difficult to verify from their proof that $o(1)$ in (10) is uniform in $x \in I$ for any compact interval $I \subset \mathbb{R}$. It is worthwhile to point out that, in contrast to the ordinary smooth case, the lower bound on $\sigma_n^2(x)$ in (10) does not explicitly depend on x nor f_Y. Further, Assumption 8 implies that

$$\inf_{|t|\le h_n^{-1}} |\varphi_\varepsilon(t)| \ge |C_3|(1-o(1))h_n^{-\gamma_0}e^{-\nu h_n^{-\gamma}} \tag{11}$$

as $n \to \infty$. It turns out that (10) and (11) are the only differences to take care of when proving the analogues of Theorems 1 and 2 in the super-smooth case. Finally, we modify Assumption 6 as follows.

Assumption 9. *(a)* $n^{-1/2}h_n^{-2\gamma(1+\lambda)} = o\{(\log(1/h_n))^{-1}\}$. *(b)* $m = m_n \to \infty$ *as* $n \to \infty$, *and*

$$m^{-1} = o\Big[\Big\{n^{-1}h_n^{2\gamma(1+\lambda)-2}(\log(1/h_n))^{-1}\Big\} \wedge \Big\{h_n^{4\gamma(1+\lambda)-2\gamma_0}e^{-2\nu h_n^{-\gamma}}(\log(1/h_n))^{-2}\Big\}\Big].$$

The requirement that $m^{-1} = o\{n^{-1}h_n^{2\gamma(1+\lambda)-2}(\log(1/h_n))^{-1}\}$ implies that we at least need $m/n \to \infty$. This condition is used to ensure that the effect of estimating φ_ε is negligible. To be precise, in our proof, a bound on $\|\widehat{f}_X - \widehat{f}_X^*\|_{\mathbb{R}}$ involves a term of order $m^{-1/2}h_n^{\gamma_0}e^{\nu h_n^{-\gamma}}$, which has to be of smaller order than $n^{-1/2}h_n^{\gamma(1+\lambda)+\gamma_0-1}e^{\nu h_n^{-\gamma}}(\log(1/h_n))^{-1/2}$. Technically, this problem happens because the ratio of $1/\inf_{|t|\le h_n^{-1}}|\varphi_\varepsilon(t)|$ over $\inf_{x\in I}\sigma_n(x)$ is larger in the super-smooth case than that in the ordinary smooth case; the ratio is $O(h_n^{-\gamma(1+\lambda)})$ in the super-smooth case, while it is $O(h_n^{-1/2})$ in the ordinary smooth case. It is not known at the current moment whether we could relax this condition on m in the super-smooth case.

In any case, these assumptions guarantee that the conclusions of Theorems 1 and 2 hold true in the super-smooth case.

Theorem 3. *Suppose that Assumptions 1, 5, and 7–9 are satisfied. Let $I \subset \mathbb{R}$ be any compact interval, and suppose in addition that $\mathrm{E}[Y^2] < \infty$. Then the conclusions of Theorems 1 and 2 hold true.*

Remark 5 (Comparisons with van Es and Gugushvili (2008))**.** van Es and Gugushvili (2008) prove that, under the assumptions that f_ε is known and verifies Assumption 8 (of the present paper) with $\gamma = 2$,

$$\frac{\sqrt{n}}{h_n^{2(1+\lambda)+\gamma_0-1} e^{\nu h_n^{-2}}} \| \widehat{f}_X^* - \mathrm{E}[\widehat{f}_X^*(\cdot)] \|_{[0,1]} \stackrel{d}{\to} \frac{|C_2|}{\sqrt{2}|C_3|\pi} (2\nu)^{-\lambda-1}\Gamma(\lambda+1)V,$$

where V follows the Rayleigh distribution, i.e., V is a random variable having density $f_V(v) = ve^{-v^2/2}1_{[0,\infty)}(v)$ (see van Es and Gugushvili, 2008, for the precise regularity conditions). Interestingly, the limit distribution differs from Gumbel distributions.

Despite this non-standard feature, Theorem 3 shows that the multiplier bootstrap "works", i.e., the conditional distribution of $\|\widehat{Z}_n^\xi\|_I$ can consistently estimate the distribution of $\|\widehat{Z}_n\|_I$ in the sense that $\sup_{z\in\mathbb{R}} |\mathrm{P}\{\|\widehat{Z}_n^\xi\|_I \le z \mid \mathcal{D}_n\} - \mathrm{P}\{\|\widehat{Z}_n\|_I \le z\}| \stackrel{\mathrm{P}}{\to} 0$. Further, Theorem 3 extends the admissible range of γ compared with the result of van Es and Gugushvili (2008).

If f_X belongs to a Hölder ball $\Sigma(\beta, B)$, then we have the following corollary.

Corollary 4. *Assume all the conditions of Theorem 3. Further, suppose that $f_X \in \Sigma(\beta, B)$ for some $\beta > 0, B > 0$, and that φ_K is $(k+3)$-times continuously differentiable with $\varphi_K^{(\ell)}(0) = 0$ for all $\ell = 1, \dots, k$ where k is the integer such that $\beta < k \le \beta + 1$. Consider the multiplier bootstrap confidence band $\widehat{\mathcal{C}}_n$ defined in (6). Then as $n \to \infty$,*

$$\mathrm{P}\{f_X(x) \in \widehat{\mathcal{C}}_n(x) \ \forall x \in I\} \to 1 - \tau,$$

provided that

$$h_n^{\beta+1-\gamma(1+\lambda)-\gamma_0} e^{-\nu h_n^{-\gamma}} \sqrt{n \log(1/h_n)} \to 0.$$

7. Proofs

In what follows, the notation $\lesssim$ signifies that the left hand side is bounded by the right hand side up to some constant independent of n and x.

7.1. **Proof of Theorem 1.** Before proving Theorem 1, we shall prove the following lemmas. For a class of measurable functions $\mathcal{F}$ on a measurable space $(S, \mathcal{S})$ and a probability measure Q on $\mathcal{S}$, let $N(\mathcal{F}, \|\cdot\|_{Q,2}, \delta)$ denote the δ-covering number for $\mathcal{F}$ with respect to the $L^2(Q)$-seminorm $\|\cdot\|_{Q,2}$; see Section 2.1 in van der Vaart and Wellner (1996) for details.

Lemma 1. *Let K be a kernel on $\mathbb{R}$ such that φ_K is symmetric and supported in $[-1,1]$, and suppose that φ_ε does not vanish on $\mathbb{R}$. Let $r_n = 1/\inf_{|t|\le h_n^{-1}} |\varphi_\varepsilon(t)|$. Consider the class of functions $\mathcal{K}_n = \{y \mapsto K_n((x-y)/h_n) : x \in \mathbb{R}\}$, where K_n denotes the corresponding deconvoluting kernel. Then there exist constants $A, v > 0$ independent of n such that for all $n \ge 1$,*

$$\sup_Q N(\mathcal{K}_n, \|\cdot\|_{Q,2}, r_n\delta) \le (A/\delta)^v, \ 0 < \delta \le 1,$$

where $\sup_Q$ is taken over all Borel probability measures Q on $\mathbb{R}$.

In view of Lemma 1 in Giné and Nickl (2009) (see also Proposition 3.6.12 in Giné and Nickl (2016)), Lemma 1 readily follows if K_n has quadratic variation $\lesssim r_n^2$. Recall that a complex-valued function f on $\mathbb{R}$ is said to be of bounded p-variation for $p \in [1,\infty)$ if

$$V_p(f) := \sup\left\{ \sum_{\ell=1}^N |f(x_\ell) - f(x_{\ell-1})|^p : -\infty < x_0 < \cdots < x_N < \infty, N = 1, 2, \dots \right\}$$

is finite. A function of bounded 2-variation is said to be of bounded quadratic variation. Now, Lemma 1 follows from the next lemma.

Lemma 2. *Assume the same conditions as in Lemma 1. Then the deconvoluting kernel K_n is of bounded quadratic variation with $V_2(K_n) \lesssim r_n^2$.*

Proof. The basic idea of the proof is due to the proof of Lemma 1 in Lounici and Nickl (2011). In view of the continuous embedding of the homogeneous Besov space $\dot{B}^{1/2}_{2,1}(\mathbb{R})$ into $BV_2(\mathbb{R})$, the space of functions of bounded quadratic variation (Bourdaud et al., 2006, Theorem 5), it is enough to show that $\|K_n\|_{\dot{B}^{1/2}_{2,1}} \lesssim r_n$, where

$$\|f\|_{\dot{B}^{1/2}_{2,1}} = \int_{\mathbb{R}} \frac{1}{|u|^{3/2}} \left(\int_{\mathbb{R}} |f(x+u) - f(x)|^2 dx \right)^{1/2} du.$$

Let $\psi_n(t) = \varphi_K(t)/\varphi_\varepsilon(t/h_n)$, and observe that, using Plancherel's theorem,

$$\begin{aligned}\int_{\mathbb{R}} |K_n(x+u) - K_n(x)|^2 dx &= \frac{1}{2\pi} \int_{\mathbb{R}} |e^{-itu} - 1|^2 |\psi_n(t)|^2 dt \\ &= \frac{1}{\pi} \int_{\mathbb{R}} (1 - \cos(tu)) |\psi_n(t)|^2 dt.\end{aligned}$$

Using the inequality $1-\cos(tu) \le \min\{2,(tu)^2/2\}$, we conclude that

$$\begin{aligned}\|K_n\|_{\dot{B}^{1/2}_{2,1}} &\lesssim \left(\int_{\mathbb{R}} t^2|\psi_n(t)|^2 dt\right)^{1/2}\int_{[-1,1]}\frac{1}{|u|^{1/2}}du + \left(\int_{\mathbb{R}}|\psi_n(t)|^2dt\right)^{1/2}\int_{[-1,1]^c}\frac{1}{|u|^{3/2}}du \\ &\lesssim r_n.\end{aligned}$$

This completes the proof. □

Remark 6. The authors noticed that Lemma 1 is essentially contained in Lemma 5.3.5 in Giné and Nickl (2016), but we still keep presenting its proof since restricting to the convolution kernel case leads to a simpler proof.

Lemma 3. *Assumptions 1–4 imply that*

$$\inf_{x\in I}\sigma_n^2(x) \gtrsim h_n^{-2\alpha+1}.$$

Proof. The proof is essentially due to Fan (1991b). Observe first that $\|K_n\|_{\mathbb{R}} \lesssim h_n^{-\alpha}$. Second, integration by parts yields that

$$\begin{aligned}K_n(x) &= \frac{1}{2\pi i x}\int_{\mathbb{R}} e^{-itx}\left\{\frac{\varphi_K(t)}{\varphi_\varepsilon(t/h_n)}\right\}' dt \\ &= \frac{1}{2\pi i x}\int_{\mathbb{R}} e^{-itx}\left\{\frac{\varphi_K'(t)}{\varphi_\varepsilon(t/h_n)} - \frac{\varphi_K(t)\varphi_\varepsilon'(t/h_n)}{h_n\varphi_\varepsilon^2(t/h_n)}\right\}dt.\end{aligned}$$

It is not difficult to verify that

$$\int_{\mathbb{R}}\left|\frac{\varphi_K'(t)}{\varphi_\varepsilon(t/h_n)}\right|dt \lesssim h_n^{-\alpha}.$$

Splitting the integral into $|t/h_n|\le 1$ and $|t/h_n|>1$, we also see that

$$\left\{\int_{|t/h_n|\le 1} + \int_{|t/h_n|>1}\right\}\left|\frac{\varphi_K(t)\varphi_\varepsilon'(t/h_n)}{h_n\varphi_\varepsilon^2(t/h_n)}\right|dt \lesssim 1 + h_n^{-\alpha}\int_{\mathbb{R}}|t|^{\alpha-1}|\varphi_K(t)|dt,$$

which is $\lesssim h_n^{-\alpha}$. This yields that $h_n^{2\alpha}K_n^2(x) \lesssim 1/x^2$, and so $h_n^{2\alpha}K_n^2(x) \lesssim \min\{1, 1/x^2\}$.

Now, observe that

$$\begin{aligned}|\mathrm{E}[K_n((x-Y)/h_n)]| &= h_n\left|\int_{\mathbb{R}} K(y)f_X(x-h_ny)dy\right| \\ &\le h_n\|f_X\|_{\mathbb{R}}\int_{\mathbb{R}}|K(y)|dy = O(h_n),\end{aligned}$$

and

$$\mathrm{E}[K_n^2((x-Y)/h_n)] = h_n\int_{\mathbb{R}} K_n^2(y)f_Y(x-h_ny)dy.$$

So it is enough to prove that

$$\inf_{x \in I} \int_{\mathbb{R}} K_n^2(y) f_Y(x - h_n y) dy \gtrsim h_n^{-2\alpha}.$$

To this end, since, by Plancherel's theorem,

$$\int_{\mathbb{R}} K_n^2(y) dy = \frac{1}{2\pi} \int_{\mathbb{R}} \frac{|\varphi_K(t)|^2}{|\varphi_\varepsilon(t/h_n)|^2} dt \gtrsim h_n^{-2\alpha},$$

it is enough to prove that

$$\sup_{x \in I} \left| h_n^{2\alpha} \int_{\mathbb{R}} K_n^2(y) \{ f_Y(x - h_n y) - f_Y(x) \} dy \right| \to 0.$$

Since f_Y is continuous and I is compact, for any $\rho > 0$, there exists $\delta > 0$ such that $\sup_{x \in I} |f_Y(x - y) - f_Y(x)| \le \rho$ whenever $|y| \le \delta$. So

$$\begin{aligned}
&\sup_{x \in I} h_n^{2\alpha} \int_{\mathbb{R}} K_n^2(y) |f_Y(x - h_n y) - f_Y(x)| dy \\
&\lesssim \rho \int_{|y| \le \delta/h_n} \min\{1, 1/y^2\} dy + 2\|f_Y\|_{\mathbb{R}} \int_{|y| > \delta/h_n} y^{-2} dy \\
&\lesssim \rho,
\end{aligned}$$

which yields the desired conclusion. □

Proof of Theorem 1. We divide the proof into three steps.

Step 1. (Gaussian approximation to Z_n^*). Recall the empirical process $Z_n^*(x), x \in I$ defined as

$$\begin{aligned}
Z_n^*(x) &= \frac{\sqrt{n} h_n \{ \widehat{f}_X^*(x) - \mathrm{E}[\widehat{f}_X^*(x)] \}}{\sigma_n(x)} \\
&= \frac{1}{\sigma_n(x)\sqrt{n}} \sum_{j=1}^n \{ K_n((x - Y_j)/h_n) - \mathrm{E}[K_n((x - Y)/h_n) \}, \ x \in I.
\end{aligned}$$

Consider the class of functions

$$\mathcal{G}_n = \left\{ \frac{1}{\sigma_n(x)} K_n((x - \cdot)/h_n) : x \in I \right\},$$

together with the empirical process indexed by $\mathcal{G}_n$ defined as

$$\mathcal{G}_n \ni g \mapsto \nu_n(g) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \{ g(Y_j) - \mathrm{E}[g(Y)] \}.$$

Observe that $\|Z_n^*\|_I = \|\nu_n\|_{\mathcal{G}_n}$. We apply Corollary 2.2 in Chernozhukov et al. (2014a) to $\|\nu_n\|_{\mathcal{G}_n}$. First, since the set $\{1/\sigma_n(x) : x \in I\}$ is bounded with $\|1/\sigma_n(\cdot)\|_I \lesssim h_n^{\alpha - 1/2}$, in view of Lemma 1 of the present paper and Corollary A.1 in Chernozhukov et al. (2014a), there exist constants $A', v' > 0$ independent of n such that

$$\sup_Q N(\mathcal{G}_n, \|\cdot\|_{Q,2}, \delta/\sqrt{h_n}) \le (A'/\delta)^{v'},\ 0 < \delta \le 1, \tag{12}$$

which ensures that there exists a tight Gaussian random variable G_n in $\ell^\infty(\mathcal{G}_n)$ with mean zero and the same covariance function as ν_n. Since $\|K_n(x-\cdot)/h_n)/\sigma_n(x)\|_{\mathbb{R}} \lesssim 1/\sqrt{h_n}$ and $\mathrm{Var}(K_n(x-Y)/h_n)/\sigma_n(x)) = 1$, application of Corollary 2.1 in Chernozhukov et al. (2014a) with $q = \infty, b \lesssim 1/\sqrt{h_n}, \sigma = 1, K_n \lesssim \log n$, and $\gamma = 1/\log n$, yields that there exists a sequence of random variables W_n with $W_n \stackrel{d}{=} \|G_n\|_{\mathcal{G}_n}$ (where the notation $\stackrel{d}{=}$ signifies equality in distribution) and such that

$$|\|\nu_n\|_{\mathcal{G}_n} - W_n| = O_{\mathrm{P}}\{(\log n)/(nh_n)^{1/6}\}, \tag{13}$$

where the left hand side is equal to $|\|Z_n^*\|_I - W_n|$.

Next, let

$$Z_n^G(x) = G_n\left(K_n((x-\cdot)/h_n)/\sigma_n(x)\right),\ x \in I,$$

and observe that Z_n^G is a tight Gaussian random variable in $\ell^\infty(I)$ with mean zero and the same covariance function as Z_n^*, and such that $\|Z_n^G\|_I = \|G_n\|_{\mathcal{G}_n} \stackrel{d}{=} W_n$. It should be noted that deducing from (13) a bound on

$$\sup_{z\in\mathbb{R}} \left|\mathrm{P}\{\|Z_n^*\|_I \le z\} - \mathrm{P}\{\|Z_n^G\|_I \le z\}\right|$$

is a non-trivial step, because the distribution of the approximating Gaussian process Z_n^G changes with n. To this end, we rely on the anti-concentration inequality for the supremum of a Gaussian process, which yields that

$$\sup_{z\in\mathbb{R};\Delta>0} \frac{\mathrm{P}\{z-\Delta \le \|Z_n^G\|_I \le z+\Delta\}}{\Delta} \lesssim \mathrm{E}[\|Z_n^G\|_I]. \tag{14}$$

See Lemma 6.1 in Chernozhukov et al. (2014a) (see also Theorem 3 in Chernozhukov et al. (2015)). To apply this inequality, we shall bound $\mathrm{E}[\|Z_n^G\|_I] = \mathrm{E}[\|G_n\|_{\mathcal{G}_n}]$, but given (12) and $\mathrm{Var}(K_n((x-Y)/h_n)/\sigma_n(x)) = 1$, Dudley's entropy integral bound (cf. van der Vaart and Wellner, 1996, Corollary 2.2.8) yields that

$$\mathrm{E}[\|G_n\|_{\mathcal{G}_n}] \lesssim \int_0^1 \sqrt{1+\log(1/(\delta\sqrt{h_n}))}d\delta \lesssim \sqrt{\log(1/h_n)}.$$

Now, combining (13) with the anti-concentration inequality (14), we conclude that

$$\sup_{z\in\mathbb{R}} \left|\mathrm{P}\{\|Z_n^*\|_I \le z\} - \mathrm{P}\{\|Z_n^G\|_I \le z\}\right| \to 0,$$

provided that

$$\frac{(\log n)\sqrt{\log(1/h_n)}}{(nh_n)^{1/6}} \to 0,$$

which is satisfied under our assumption.

Step 2. (Gaussian approximation to the intermediate process). Define the intermediate process

$$\widetilde{Z}_n(x) = \frac{\sqrt{n}h_n\{\widehat{f}_X(x) - \mathrm{E}[\widehat{f}^*_X(x)]\}}{\sigma_n(x)},\ x \in I,$$

where the difference from $\widehat{Z}_n(x)$ is that $\widehat{\sigma}_n(x)$ is replaced by $\sigma_n(x)$. In this step, we wish to prove that

$$\sup_{z\in\mathbb{R}} |\mathrm{P}\{\|\widetilde{Z}_n\|_I \le z\} - \mathrm{P}\{\|Z^G_n\|_I \le z\}| \to 0, \tag{15}$$

where Z^G_n is given in the previous step.

Since φ_ε does not vanish on $\mathbb{R}$, $\varphi_Y(t) = 0$ if and only if $\varphi_X(t) = 0$, from which we have, setting $T = \{t \in \mathbb{R} : \varphi_X(t) \ne 0\}$,

$$\widehat{f}^*_X(x) = \frac{1}{2\pi}\left\{\int_T e^{-itx}\varphi_K(th_n)\frac{\widehat{\varphi}_Y(t)}{\varphi_Y(t)}\varphi_X(t)dt + \int_{T^c} e^{-itx}\frac{\varphi_K(th_n)}{\varphi_\varepsilon(t)}\widehat{\varphi}_Y(t)dt\right\},$$

and

$$\widehat{f}_X(x) = \frac{1}{2\pi}\left\{\int_T e^{-itx}\varphi_K(th_n)\frac{\widehat{\varphi}_Y(t)}{\varphi_Y(t)}\frac{\varphi_\varepsilon(t)}{\widehat{\varphi}_\varepsilon(t)}\varphi_X(t)dt + \int_{T^c} e^{-itx}\frac{\varphi_K(th_n)}{\varphi_\varepsilon(t)}\widehat{\varphi}_Y(t)\frac{\varphi_\varepsilon(t)}{\widehat{\varphi}_\varepsilon(t)}dt\right\}.$$

So letting

$$\widetilde{f}_X(x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}\varphi_K(th_n)\frac{\varphi_\varepsilon(t)}{\widehat{\varphi}_\varepsilon(t)}\varphi_X(t)dt,$$

we obtain the following decomposition:

$$\begin{aligned}
&\widehat{f}_X(x) - \widehat{f}^*_X(x) \\
&= \left[\{\widehat{f}^*_X(x) - \widetilde{f}_X(x)\} - \{\widehat{f}^*_X(x) - \mathrm{E}[\widehat{f}^*_X(x)]\}\right] + \{\widetilde{f}_X(x) - \mathrm{E}[\widehat{f}^*_X(x)]\} \\
&= \frac{1}{2\pi}\int_T e^{-itx}\varphi_K(th_n)\left\{\frac{\widehat{\varphi}_Y(t)}{\varphi_Y(t)} - 1\right\}\left\{\frac{\varphi_\varepsilon(t)}{\widehat{\varphi}_\varepsilon(t)} - 1\right\}\varphi_X(t)dt \\
&\quad + \frac{1}{2\pi}\int_{T^c} e^{-itx}\frac{\varphi_K(th_n)}{\varphi_\varepsilon(t)}\widehat{\varphi}_Y(t)\left\{\frac{\varphi_\varepsilon(t)}{\widehat{\varphi}_\varepsilon(t)} - 1\right\}dt \\
&\quad + \frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}\varphi_K(th_n)\left\{\frac{\varphi_\varepsilon(t)}{\widehat{\varphi}_\varepsilon(t)} - 1\right\}\varphi_X(t)dt.
\end{aligned}$$

Hence the Cauchy-Schwarz inequality yields that

$$
\begin{aligned}
&|\widehat{f}_X(x) - \widehat{f}_X^*(x)|^2 \\
&\lesssim \left\{ \int_{T \cap [-h_n^{-1}, h_n^{-1}]} \left| \frac{\widehat{\varphi}_Y(t)}{\varphi_Y(t)} - 1 \right|^2 |\varphi_X(t)|^2 dt \right\} \left\{ \int_{-h_n^{-1}}^{h_n^{-1}} \left| \frac{\varphi_\varepsilon(t)}{\widehat{\varphi}_\varepsilon(t)} - 1 \right|^2 dt \right\} \\
&\quad + h_n^{-2\alpha} \left\{ \int_{T^c \cap [-h_n^{-1}, h_n^{-1}]} |\widehat{\varphi}_Y(t)|^2 dt \right\} \left\{ \int_{-h_n^{-1}}^{h_n^{-1}} \left| \frac{\varphi_\varepsilon(t)}{\widehat{\varphi}_\varepsilon(t)} - 1 \right|^2 dt \right\} \\
&\quad + \int_{-h_n^{-1}}^{h_n^{-1}} \left| \frac{\varphi_\varepsilon(t)}{\widehat{\varphi}_\varepsilon(t)} - 1 \right|^2 |\varphi_X(t)| dt,
\end{aligned}
$$

where we have used the fact that $|\varphi_X|$ is integrable on $\mathbb{R}$. We will bound the following four terms:

$$
\begin{aligned}
&\int_{T \cap [-h_n^{-1}, h_n^{-1}]} \left| \frac{\widehat{\varphi}_Y(t)}{\varphi_Y(t)} - 1 \right|^2 |\varphi_X(t)|^2 dt, \quad \int_{T^c \cap [-h_n^{-1}, h_n^{-1}]} |\widehat{\varphi}_Y(t)|^2 dt, \\
&\int_{-h_n^{-1}}^{h_n^{-1}} \left| \frac{\varphi_\varepsilon(t)}{\widehat{\varphi}_\varepsilon(t)} - 1 \right|^2 dt, \quad \int_{-h_n^{-1}}^{h_n^{-1}} \left| \frac{\varphi_\varepsilon(t)}{\widehat{\varphi}_\varepsilon(t)} - 1 \right|^2 |\varphi_X(t)| dt.
\end{aligned}
$$

To bound the first term, pick and fix any $t \in \mathbb{R}$ such that $\varphi_Y(t) \neq 0$; let

$$
\zeta_j = e^{itY_j} / \varphi_Y(t), \ j = 1, \dots, n.
$$

Then we have

$$
\mathrm{E}\left[\left| \frac{\widehat{\varphi}_Y(t)}{\varphi_Y(t)} - 1 \right|^2 \right] = \mathrm{E}\left[\left| \frac{1}{n} \sum_{j=1}^n \{\zeta_j - \mathrm{E}[\zeta_j]\} \right|^2 \right] \le \frac{1}{n|\varphi_Y(t)|^2}.
$$

Therefore,

$$
\begin{aligned}
\mathrm{E}\left[\int_{T \cap [-h_n^{-1}, h_n^{-1}]} \left| \frac{\widehat{\varphi}_Y(t)}{\varphi_Y(t)} - 1 \right|^2 |\varphi_X(t)|^2 dt \right] &\le n^{-1} \int_{-h_n^{-1}}^{h_n^{-1}} \frac{1}{|\varphi_\varepsilon(t)|^2} dt \\
&\lesssim h_n^{-2\alpha} (nh_n)^{-1}.
\end{aligned}
$$

On the other hand, since $\varphi_Y(t) = 0$ for any $t \in T^c$, we have

$$
\mathrm{E}\left[\int_{T^c \cap [-h_n^{-1}, h_n^{-1}]} |\widehat{\varphi}_Y(t)|^2 dt \right] \le 2(nh_n)^{-1}.
$$

To bound the third and fourth terms, we first note that, from Lemma 4 in Appendix A together with the fact that $\mathrm{E}[|\varepsilon|^p] < \infty$ for some $p > 0$,

$$
\|\widehat{\varphi}_\varepsilon - \varphi_\varepsilon\|_{[-h_n^{-1}, h_n^{-1}]} = O_{\mathrm{P}}\{m^{-1/2} \log(1/h_n)\}, \tag{16}
$$

which is $o_{\mathrm{P}}(h_n^{\alpha})$ by Assumption 6 (b). Hence

$$\inf_{t\in[-h_n^{-1},h_n^{-1}]}|\widehat{\varphi}_{\varepsilon}(t)| \geq \inf_{t\in[-h_n^{-1},h_n^{-1}]}|\varphi_{\varepsilon}(t)| - o_{\mathrm{P}}(h_n^{\alpha}) \gtrsim (1-o_{\mathrm{P}}(1))h_n^{\alpha},$$

from which we have

$$\begin{aligned}
&\int_{-h_n^{-1}}^{h_n^{-1}}\left|\frac{\varphi_{\varepsilon}(t)}{\widehat{\varphi}_{\varepsilon}(t)}-1\right|^2 dt \lesssim h_n^{-2\alpha}\int_{-h_n^{-1}}^{h_n^{-1}}|\varphi_{\varepsilon}(t)-\widehat{\varphi}_{\varepsilon}(t)|^2 dt = O_{\mathrm{P}}(h_n^{-2\alpha-1}m^{-1}),\\
&\int_{-h_n^{-1}}^{h_n^{-1}}\left|\frac{\varphi_{\varepsilon}(t)}{\widehat{\varphi}_{\varepsilon}(t)}-1\right|^2 |\varphi_X(t)|dt = O_{\mathrm{P}}(h_n^{-2\alpha}m^{-1}),
\end{aligned}$$

where we have used the fact that $|\varphi_X|$ is integrable on $\mathbb{R}$.

Taking these together, we have

$$\|\widehat{f}_X - \widehat{f}_X^*\|_{\mathbb{R}} = o_{\mathrm{P}}\left[h_n^{-\alpha}\{(nh_n)\log(1/h_n)\}^{-1/2}\right]$$

by Assumption 6 (b), from which we conclude that

$$\|\widetilde{Z}_n - Z_n^*\|_I = o_{\mathrm{P}}\{(\log(1/h_n))^{-1/2}\}.$$

This shows that there exists a sequence $\Delta_n \downarrow 0$ such that $\mathrm{P}\{\|Z_n - Z_n^*\|_I > \Delta_n(\log(1/h_n))^{-1/2}\} \leq \Delta_n$ (which follows from the fact that convergence in probability is metrized by the Ky Fan metric; see Theorem 9.2.2 in Dudley (2002)), and so

$$\begin{aligned}
\mathrm{P}\{\|\widetilde{Z}_n\|_I \leq z\} &\leq \mathrm{P}\{\|Z_n^*\|_I \leq z + \Delta_n(\log(1/h_n))^{-1/2}\} + \Delta_n\\
&\leq \mathrm{P}\{\|Z_n^G\|_I \leq z + \Delta_n(\log(1/h_n))^{-1/2}\} + o(1)\\
&\leq \mathrm{P}\{\|Z_n^G\|_I \leq z\} + o(1)
\end{aligned}$$

uniformly in $z \in \mathbb{R}$, where the second inequality follows from the previous step, and the last inequality follows from the anti-concentration inequality (14). Likewise, we have

$$\mathrm{P}\{\|\widetilde{Z}_n\|_I \leq z\} \geq \mathrm{P}\{\|Z_n^G\|_I \leq z\} - o(1)$$

uniformly in $z \in \mathbb{R}$, so that we obtain the conclusion of this step.

Step 3. (Proof of the theorem). Observe that

$$\begin{aligned}
\|\widehat{K}_n - K_n\|_{\mathbb{R}} &\lesssim \int_{\mathbb{R}}\left|\frac{1}{\widehat{\varphi}_{\varepsilon}(t/h_n)} - \frac{1}{\varphi_{\varepsilon}(t/h_n)}\right||\varphi_K(t)|dt\\
&\lesssim h_n^{-2\alpha}\int_{\mathbb{R}}|\widehat{\varphi}_{\varepsilon}(t/h_n) - \varphi_{\varepsilon}(t/h_n)|\,|\varphi_K(t)|dt,
\end{aligned}$$

which is $O_{\mathrm{P}}(h_n^{-2\alpha}m^{-1/2})$ because

$$\int_{\mathbb{R}}\mathrm{E}\left[|\widehat{\varphi}_{\varepsilon}(t/h_n) - \varphi_{\varepsilon}(t/h_n)|\right]|\varphi_K(t)|dt \lesssim m^{-1/2}.$$

Together with the fact that $\|K_n\|_{\mathbb{R}} = O(h_n^{-\alpha})$, we have

$$\|\widehat{K}_n^2 - K_n^2\|_{\mathbb{R}} \leq \|\widehat{K}_n - K_n\|_{\mathbb{R}} \|\widehat{K}_n + K_n\|_{\mathbb{R}} = O(h_n^{-3\alpha} m^{-1/2}),$$

which yields that

$$\widehat{\sigma}_n^2(x) = \frac{1}{n}\sum_{j=1}^n K_n^2((x - Y_j)/h_n) - \left(\frac{1}{n}\sum_{j=1}^n K_n((x-Y_j)/h_n)\right)^2 + O_{\mathrm{P}}(h_n^{-3\alpha} m^{-1/2})$$

uniformly in $x \in I$. It is not difficult to verify that

$$\begin{aligned}
\frac{1}{n}\sum_{j=1}^n K_n^2((x-Y_j)/h_n) &= \mathrm{E}[K_n^2((x-Y)/h_n)] + O_{\mathrm{P}}(h_n^{-2\alpha} n^{-1/2}), \\
\frac{1}{n}\sum_{j=1}^n K_n((x-Y_j)/h_n) &= \mathrm{E}[K_n((x-Y)/h_n)] + O_{\mathrm{P}}(h_n^{-\alpha} n^{-1/2})
\end{aligned}$$

uniformly in $x \in I$. Indeed, in view of Lemma 1 of the present paper and Corollary A.2 in Chernozhukov et al. (2014a), these estimates follow from Theorem 2.14.1 in van der Vaart and Wellner (1996). Therefore, we have

$$\widehat{\sigma}_n^2(x) = \sigma_n^2(x) + O_{\mathrm{P}}\{h_n^{-2\alpha}(h_n^{-\alpha} m^{-1/2} + n^{-1/2})\}$$

uniformly in $x \in I$, and so

$$\widehat{\sigma}_n^2(x)/\sigma_n^2(x) = 1 + O_{\mathrm{P}}\{h_n^{-1}(h_n^{-\alpha} m^{-1/2} + n^{-1/2})\}$$

uniformly in $x \in I$ by Assumption 4, where $h_n^{-1}(h_n^{-\alpha} m^{-1/2} + n^{-1/2}) = o\{(\log(1/h_n))^{-1}\}$ by Assumption 6. This yields that

$$\|\sigma_n(\cdot)/\widehat{\sigma}_n(\cdot) - 1\|_I = o_{\mathrm{P}}\{(\log(1/h_n))^{-1}\}.$$

By Steps 1 and 2 together with the fact that $\mathrm{E}[\|Z_n^G\|_I] = O(\sqrt{\log(1/h_n)})$, we see that $\|\widetilde{Z}_n\|_I = O_{\mathrm{P}}(\sqrt{\log(1/h_n)})$. So

$$\|\widehat{Z}_n - \widetilde{Z}_n\|_I \leq \|\sigma_n(\cdot)/\widehat{\sigma}_n(\cdot) - 1\|_I \, \|\widetilde{Z}_n\|_I = o_{\mathrm{P}}\{(\log(1/h_n))^{-1/2}\},$$

and arguing as in the last part of the proof of Step 2, we conclude that

$$\sup_{z\in\mathbb{R}} |\mathrm{P}\{\|\widehat{Z}_n\|_I \leq z\} - \mathrm{P}\{\|Z_n^G\|_I \leq z\}| \to 0.$$

This completes the proof of Theorem 1. $\square$

7.2. **Proofs of Corollaries 1 and 2.**

Proof of Corollary 1. Step 2 in the proof of Theorem 1 yields that $\|\widehat{f}_X - \widehat{f}_X^*\|_{\mathbb{R}} = o_{\mathrm{P}}\{h_n^{-\alpha}(nh_n)^{-1/2}\}$ (it is not difficult to verify that Assumption 4 was not used to derive this rate). Hence we have to show that $\|\widehat{f}_X^* - \mathrm{E}[\widehat{f}_X^*(\cdot)]\|_{\mathbb{R}} = O_{\mathrm{P}}\{h_n^{-\alpha}(nh_n)^{-1/2}\sqrt{\log(1/h_n)}\}$. To this end, we make use of Corollary 5.1 in Chernozhukov et al. (2014a). Invoke Lemma 1 and observe that $\|K_n\|_{\mathbb{R}} \lesssim h_n^{-\alpha}$ and $\sigma_n^2(x) \lesssim h_n^{-2\alpha+1}$. Application of Corollary 5.1 in Chernozhukov et al. (2014a) to $\mathcal{K}_n$ then yields that

$$
\begin{aligned}
&\mathrm{E}\left[\left\|\sum_{j=1}^n \{K((\cdot - Y_j)/h_n) - \mathrm{E}[K_n((\cdot - Y)/h_n)]\}\right\|_{\mathbb{R}}\right] \\
&\quad \lesssim h_n^{-\alpha}\sqrt{nh_n \log(1/h_n)} + h_n^{-\alpha}\log(1/h_n) \\
&\quad \lesssim h_n^{-\alpha}\sqrt{nh_n \log(1/h_n)},
\end{aligned}
$$

which in turn yields that

$$
\mathrm{E}[\|\widehat{f}_X^* - \mathrm{E}[\widehat{f}_X^*(\cdot)]\|_{\mathbb{R}}] \lesssim h_n^{-\alpha}(nh_n)^{-1/2}\sqrt{\log(1/h_n)}.
$$

This completes the proof. □

Proof of Corollary 2. Since φ_K is $(k+3)$-times continuously differentiable, the kernel $K(x) = (1/2\pi)\int_{\mathbb{R}} e^{-itx}\varphi_K(t)dt$ is such that $\int_{\mathbb{R}} |x|^{k+1}|K(x)|dx < \infty$, and in addition $\int_{\mathbb{R}} x^\ell K(x)dx = i^{-\ell}\varphi_K^{(\ell)}(0) = 0$ for all $\ell = 1, \ldots, k$. Hence

$$
\begin{aligned}
\|\mathrm{E}[\widehat{f}_X^*(\cdot)] - f_X(\cdot)\|_{\mathbb{R}} &= \left\|\int_{\mathbb{R}} \{f_X(\cdot - h_n y) - f_X(\cdot)\}K(y)dy\right\|_{\mathbb{R}} \\
&\le \frac{Bh_n^\beta}{k!}\int_{\mathbb{R}} |y|^\beta |K(y)|dy.
\end{aligned}
$$

Combining the result of Corollary 1, and in view of the remark after Assumption 6, we obtain the desired conclusion. □

7.3. **Proofs of Theorem 2 and Corollary 3.**

Proof of Theorem 2. We divide the proof into three steps.

Step 1. Define

$$
Z_n^\xi(x) = \frac{1}{\sigma_n(x)\sqrt{n}}\sum_{j=1}^n \xi_j \left\{K_n((x - Y_j)/h_n) - n^{-1}\textstyle\sum_{j'=1}^n K_n((x - Y_{j'})/h_n)\right\}
$$

for $x \in I$. We first prove that

$$\sup_{z\in\mathbb{R}} |\mathrm{P}\{\|Z_n^{\xi}\|_I \le z \mid \mathcal{D}_n\} - \mathrm{P}\{\|Z_n^G\|_I \le z\}| \xrightarrow{\mathrm{P}} 0.$$

To this end, we make use of Theorem 2.2 in Chernozhukov et al. (2016). Recall the class of functions $\mathcal{G}_n$ defined in the proof of Theorem 1, and let

$$\nu_n^{\xi}(g) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \xi_j \left\{ g(Y_j) - n^{-1}\textstyle\sum_{j'=1}^n g(Y_{j'}) \right\}, \ g \in \mathcal{G}_n.$$

Then application of Theorem 2.2 in Chernozhukov et al. (2016) to $\|\nu_n\|_{\mathcal{G}_n}$ with $B(f) \equiv 0, b \lesssim 1/\sqrt{h_n}, \sigma = 1, K_n \lesssim \log n, \gamma = 1/\log n$, and q sufficiently large, yields that there exists a random variable W_n^{ξ} of which the conditional distribution given $\mathcal{D}_n$ is the same as the distribution of $\|G_n\|_{\mathcal{G}_n} (= \|Z_n^G\|_I)$, and such that

$$|\|\nu_n^{\xi}\|_{\mathcal{G}_n} - W_n^{\xi}| = o_{\mathrm{P}}\{(\log(1/h_n))^{-1/2}\},$$

which shows that there exists a sequence $\Delta_n \downarrow 0$ such that

$$\mathrm{P}\{|\|\nu_n^{\xi}\|_{\mathcal{G}_n} - W_n^{\xi}| > \Delta_n (\log(1/h_n))^{-1/2} \mid \mathcal{D}_n\} \xrightarrow{\mathrm{P}} 0$$

by Markov's inequality. Since $\|\nu_n^{\xi}\|_{\mathcal{G}_n} = \|Z_n^{\xi}\|_I$, we have

$$\begin{aligned} \mathrm{P}\{\|Z_n^{\xi}\|_I \le z \mid \mathcal{D}_n\} &\le \mathrm{P}\{W_n^{\xi} \le z + \Delta_n(\log(1/h_n))^{-1/2} \mid \mathcal{D}_n\} + o_{\mathrm{P}}(1) \\ &= \mathrm{P}\{\|Z_n^G\|_I \le z + \Delta_n(\log(1/h_n))^{-1/2}\} + o_{\mathrm{P}}(1) \end{aligned}$$

uniformly in $z \in \mathbb{R}$, and the anti-concentration inequality (14) yields that

$$\mathrm{P}\{\|Z_n^G\|_I \le z + \Delta_n(\log(1/h_n))^{-1/2}\} \le \mathrm{P}\{\|Z_n^G\|_I \le z\} + o(1)$$

uniformly in $z \in \mathbb{R}$. Likewise, we have

$$\mathrm{P}\{\|Z_n^{\xi}\|_I \le z \mid \mathcal{D}_n\} \ge \mathrm{P}\{\|Z_n^G\|_I \le z\} - o_{\mathrm{P}}(1)$$

uniformly in $z \in \mathbb{R}$. Therefore, we obtain the conclusion of this step.

Step 2. In view of the proof of Step 1, in order to prove the result (8), it is enough to prove that

$$\|\widehat{Z}_n^{\xi} - Z_n^{\xi}\|_I = o_{\mathrm{P}}\{(\log(1/h_n))^{-1/2}\}.$$

Define

$$\widetilde{Z}_n^{\xi} = \frac{1}{\sigma_n(x)\sqrt{n}} \sum_{j=1}^n \xi_j \left\{ \widehat{K}_n((x-Y_j)/h_n) - n^{-1}\textstyle\sum_{j'=1}^n \widehat{K}_n((x-Y_{j'})/h_n) \right\}$$

for $x \in I$. We first prove that $\|\widetilde{Z}_n^\xi - Z_n^\xi\|_I = o_{\mathrm{P}}\{(\log(1/h_n))^{-1/2}\}$. Step 2 in the proof of Theorem 1 shows that

$$\left\| \frac{1}{\sigma_n(\cdot)\sqrt{n}} \sum_{j=1}^n \{\widehat{K}_n((\cdot - Y_j)/h_n) - K_n((\cdot - Y_j)/h_n)\} \right\|_I$$

is $o_{\mathrm{P}}\{(\log(1/h_n))^{-1/2}\}$ (which is more that what we need), and so it remains to prove that

$$\left\| \frac{1}{\sigma_n(\cdot)\sqrt{n}} \sum_{j=1}^n \xi_j \{\widehat{K}_n((\cdot - Y_j)/h_n) - K_n((\cdot - Y_j)/h_n)\} \right\|_I$$

is $o_{\mathrm{P}}\{(\log(1/h_n))^{-1/2}\}$. Since $\|1/\sigma_n(\cdot)\|_I \lesssim h_n^{\alpha-1/2}$, it is enough to prove that

$$\left\| \sum_{j=1}^n \xi_j \{\widehat{K}_n((\cdot - Y_j)/h_n) - K_n((\cdot - Y_j)/h_n)\} \right\|_I$$

is $o_{\mathrm{P}}\{h_n^{-\alpha}(nh_n)^{1/2}(\log(1/h_n))^{-1/2}\}$. Observe that

$$\begin{aligned}
&\left| \sum_{j=1}^n \xi_j \{\widehat{K}_n((x - Y_j)/h_n) - K_n((x - Y_j)/h_n)\} \right| \\
&\lesssim \int_{\mathbb{R}} \left| \sum_{j=1}^n \xi_j e^{itY_j/h_n} \right| \left| \frac{1}{\widehat{\varphi}_\varepsilon(t/h_n)} - \frac{1}{\varphi_\varepsilon(t/h_n)} \right| |\varphi_K(t)| dt \\
&\lesssim \left\{ \int_{\mathbb{R}} \left| \sum_{j=1}^n \xi_j e^{itY_j/h_n} \right|^2 |\varphi_K(t)| dt \right\}^{1/2} \left\{ \int_{\mathbb{R}} \left| \frac{1}{\widehat{\varphi}_\varepsilon(t/h_n)} - \frac{1}{\varphi_\varepsilon(t/h_n)} \right|^2 |\varphi_K(t)| dt \right\}^{1/2} \\
&= O_{\mathrm{P}}(n^{1/2} h_n^{-2\alpha} m^{-1/2}),
\end{aligned}$$

which is $o_{\mathrm{P}}\{h_n^{-\alpha}(nh_n)^{1/2}(\log(1/h_n))^{-1/2}\}$ by Assumption 6 (b). Therefore, we have $\|\widetilde{Z}_n^\xi - Z_n^\xi\|_I = o_{\mathrm{P}}\{(\log(1/h_n))^{-1/2}\}$.

By Step 1, the fact that $\mathrm{E}[\|Z_n^G\|_I] = O(\sqrt{\log(1/h_n)})$ (see the proof of Theorem 1), and the previous result, we see that $\|\widetilde{Z}_n^\xi\|_I = O_{\mathrm{P}}(\sqrt{\log(1/h_n)})$. Now, because $\|\sigma_n(\cdot)/\widehat{\sigma}_n(\cdot) - 1\|_I = o_{\mathrm{P}}\{(\log(1/h_n))^{-1}\}$ from Step 3 in the proof of Theorem 1, we conclude that

$$\|\widehat{Z}_n^\xi - \widetilde{Z}_n^\xi\|_I \le \|\sigma_n(\cdot)/\widehat{\sigma}_n(\cdot) - 1\|_I \|\widetilde{Z}_n^\xi\|_I = o_{\mathrm{P}}\{(\log(1/h_n))^{-1/2}\},$$

which leads to (8).

Step 3. In this step, we shall verify the last assertion of the theorem. The result (8) implies that there exists a sequence $\Delta_n \downarrow 0$ such that with probability greater than $1 - \Delta_n$,

$$\sup_{z \in \mathbb{R}} |\mathrm{P}\{\|\widehat{Z}_n^\xi\|_I \le z \mid \mathcal{D}_n\} - \mathrm{P}\{\|Z_n^G\|_I \le z\}| \le \Delta_n. \tag{17}$$

Further, taking $\Delta_n \downarrow 0$ more slowly if necessary, we have

$$\sup_{z\in\mathbb{R}} |\mathrm{P}\{\|\widehat{Z}_n\|_I \le z\} - \mathrm{P}\{\|Z_n^G\|_I \le z\}| \le \Delta_n.$$

Let $\mathcal{E}_n$ denote the event on which (17) happens, and let $c_n^G(u)$ denote the u-quantile of $\|Z_n^G\|_I$ for $u \in (0,1)$. Then on the event $\mathcal{E}_n$,

$$\mathrm{P}\{\|\widehat{Z}_n^{\xi}\|_I \le c_n^G(1-\tau+\Delta_n) \mid \mathcal{D}_n\} \ge \mathrm{P}\{\|Z_n^G\|_I \le c_n^G(1-\tau+\Delta_n)\} - \Delta_n = 1-\tau,$$

where the last equality holds since $\|Z_n^G\|$ has continuous distribution (recall the anti-concentration inequality (14)). This yields that the inequality

$$\widehat{c}_n(1-\tau) \le c_n^G(1-\tau+\Delta_n)$$

holds on $\mathcal{E}_n$, and so

$$\begin{aligned}
\mathrm{P}\{\|\widehat{Z}_n\|_I \le \widehat{c}_n(1-\tau)\} &\le \mathrm{P}\{\|\widehat{Z}_n\|_I \le c_n^G(1-\tau+\Delta_n)\} + \mathrm{P}(\mathcal{E}_n^c) \\
&\le \mathrm{P}\{\|Z_n^G\| \le c_n^G(1-\tau+\Delta_n)\} + 2\Delta_n \\
&= 1-\tau+3\Delta_n.
\end{aligned}$$

Likewise, we have

$$\mathrm{P}\{\|\widehat{Z}_n\|_I \le \widehat{c}_n^G(1-\tau)\} \ge 1-\tau-3\Delta_n.$$

This completes the proof. □

Proof of Corollary 3. Recall the stochastic process $\widetilde{Z}_n(x), x \in I$ defined in the proof of Theorem 1. Observe that $\|f_X - \mathrm{E}[\widehat{f}_X^*(\cdot)]\|_{\mathbb{R}} = O(h_n^\beta)$. Condition (9) then yields that

$$\begin{aligned}
\frac{\sqrt{n}h_n(\widehat{f}_X(x) - f_X(x))}{\widehat{\sigma}_n(x)} &= \{1 + o_{\mathrm{P}}\{(\log(1/h_n))^{-1}\}\}\frac{\sqrt{n}h_n(\widehat{f}_X(x) - f_X(x))}{\sigma_n(x)} \\
&= \{1 + o_{\mathrm{P}}\{(\log(1/h_n))^{-1}\}\}\widetilde{Z}_n(x) + o\{(\log(1/h_n))^{-1/2}\} \\
&= \widetilde{Z}_n(x) + o_{\mathrm{P}}\{(\log(1/h_n))^{-1/2}\}
\end{aligned}$$

uniformly in $x \in I$, where we have used the facts that $\|\sigma_n(\cdot)/\widehat{\sigma}_n(\cdot) - 1\|_I = o_{\mathrm{P}}\{(\log(1/h_n))^{-1}\}$ and $\|\widetilde{Z}_n\|_I = O_{\mathrm{P}}(\sqrt{\log(1/h_n)})$ (these estimates are derived in the proof of Theorem 1). Using the anti-concentration inequality (14) together with the result of Step 2 in the proof of Theorem 1, we have

$$\sup_{z\in\mathbb{R}} \left| \mathrm{P}\left\{ \left\| \frac{\sqrt{n}h_n(\widehat{f}_X(\cdot) - f_X(\cdot))}{\widehat{\sigma}_n(\cdot)} \right\|_I \le z \right\} - \mathrm{P}\{\|Z_n^G\|_I \le z\} \right| \to 0.$$

Now, arguing as in Step 3 in the proof of Theorem 2, we conclude that

$$\mathrm{P}\left\{\left\|\frac{\sqrt{n}h_n(\widehat{f}_X(\cdot)-f_X(\cdot))}{\widehat{\sigma}_n(\cdot)}\right\|_I \le \widehat{c}_n(1-\tau)\right\} \to 1-\tau,$$

which yields the desired result. □

7.4. **Proofs for Section 6.** The proof of Theorem 3 is almost identical to the proofs of Theorems 1 and 2 in the ordinary smooth case. The only changes that have to be taken into account are (10) and (11), which imply that $\|K((x-\cdot)/h_n)/\sigma_n(x)\|_{\mathbb{R}} \lesssim h_n^{-\gamma(1+\lambda)}$, for example. To avoid repetitions, we omit the details for brevity. In view of the proof of Corollary 3, Corollary 4 directly follows from Theorem 3. □

Appendix A. Uniform convergence rates of the empirical characteristic function

In this appendix, we establish rates of convergence of the empirical characteristic function on expanding sets. The proof of the following lemma is due essentially to Neumann and Reiß (2009, Theorem 4.1).

Let F be a distribution function on $\mathbb{R}$ with characteristic function $\varphi(t) = \int_{\mathbb{R}} e^{itx} dF(x)$, and let $X_1, \dots, X_n$ be an independent sample from F. Let $F_n(x) = n^{-1}\sum_{j=1}^n 1_{(-\infty,x]}(X_j)$ be the empirical distribution function, and let $\varphi_n(t) = \int_{\mathbb{R}} e^{itx} dF_n(x) = n^{-1}\sum_{j=1}^n e^{itX_j}$ be the empirical characteristic function.

Lemma 4. *Suppose that $\int_{\mathbb{R}} |x|^p dF(x) < \infty$ for some $p > 0$. Then for any $\delta > 0$ and any $T_n \to \infty$, we have*

$$\|\varphi_n - \varphi\|_{[-T_n,T_n]} = O_{\mathrm{P}}\{n^{-1/2}(\log T_n)^{1/2+\delta}\}.$$

Proof. Let $w(t) = (\log(e+|t|))^{-1/2-\delta}$. According to Theorem 4.1 in Neumann and Reiß (2009), it follows that

$$C := \sup_{n \ge 1} \mathrm{E}[\|\sqrt{n}(\varphi_n - \varphi)w\|_{\mathbb{R}}] < \infty.$$

Now, because

$$\|\sqrt{n}(\varphi_n - \varphi)w\|_{\mathbb{R}} \ge \sqrt{n}\|\varphi_n - \varphi\|_{[-T_n,T_n]} \inf_{|t|\le T_n} w(t),$$

we conclude that

$$\mathrm{E}[\|\varphi_n - \varphi\|_{[-T_n,T_n]}] \le \frac{C}{\sqrt{n}\inf_{|t|\le T_n} w(t)} = O\{n^{-1/2}(\log T_n)^{1/2+\delta}\},$$

which leads to the desired result by Markov's inequality. □

It is worthwhile to point out that the restriction to the set $|t| \le T_n$ in Lemma 4 is essential. Intuitively, this can be explained by the observation that φ_n is the characteristic function of the discrete distribution (i.e., the empirical distribution function), and so is periodic and approaches its maximum value (i.e., 1) arbitrarily often as $|t| \to \infty$, while if F is absolutely continuous, then $\varphi(t) \to 0$ as $|t| \to \infty$ by the Riemann-Lebesgue lemma. In fact, although the class of functions $\{x \mapsto e^{itx} : t \in \mathbb{R}\}$ is uniformly bounded, it is in general not Glivenko-Cantelli (nor Donsker, of course). See Feurerverger and Mureika (1977).

References

Bickel, P. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.* **1** 1071-1095. Correction (1975) **3** 1370.

Bissantz, N., Dümbgen, L., Holtzmann, H., and Munk, A. (2007). Non-parametric confidence bands in deconvolution density estimation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 483-506.

Bissantz, N. and Holzmann, H. (2008). Statistical inference for inverse problems. *Inverse Problems* 24:034009.

Bourdaud, G., Lanza de Cristoforis, M., and Sickel, W. (2006). Superposition operators and functions of bounded p-variation. *Rev. Mat. Iberoamericana* **22** 455-487.

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

Carroll, R.J., Ruppert, D., Stefanski, L.A., and Crainiceanu, C.M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective (2nd Edition).* Chapman & Hall/CRC.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2014a). Gaussian approximation of suprema of empirical processes. *Ann. Statist.* **42** 1564-1597.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2014b). Anti-concentration and honest, adaptive confidence bands. *Ann. Statist.* **42** 1787-1818.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2015). Comparison and anti-concentration bounds for maxima of Gaussian random vectors. *Probab. Theory Related Fields* **162** 47-70.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2016). Empirical and multiplier bootstraps for suprema of empirical processes of increasing complexity, and related Gaussian couplings. *Stochastic Process. Appl.*, to appear. arXiv:1502:00352.

Claeskens, G. and van Keilegom, I. (2003). Bootstrap confidence bands for regression curves and their derivatives. *Ann. Statist.* **31** 1852-1884.

Comte, F. and Lacour, C. (2011). Data-driven density estimation in the presence of additive noise with unknown distribution. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 601-627.

Dattner, I., Reiß, M., and Trabs, M. (2016). Adaptive quantile estimation in deconvolution with unknown error distribution. *Bernoulli* **22** 143192.

Delaigle, A. and Gijbels, I. (2004). Practical bandwidth selection in deconvolution kernel density estimation. *Comput. Statist. Data Anal.* **45** 249-267.

Delaigle, A. and Hall, P. (2016). Methodology for nonparametric deconvolution when the error distribution is unknown. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 231-252.

Delaigle, A., Hall, P., and Jamshidi, F. (2015). Confidence bands in nonparametric errors-in-variables regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 149-169.

Delaigle, A., Hall, P., and Meister, A. (2008). On deconvolution with repeated measurements. *Ann. Statist.* **36** 665-685.

Diggle, P.J. and Hall, P. (1993). A Fourier approach to nonparametric deconvolution of a density estimate. *J. Roy. Stat. Soc. Ser. B. Stat. Methodol.* **55** 523-531.

Dudley, R.M. (2002). *Real Analysis and Probability.* Cambridge University Press.

Efromovich, S. (1987). Density estimation for the case of supersmooth measurement error. *J. Amer. Stat. Assoc.* **92** 526-535.

van Es, B. and Gugushvili, S. (2008). Weak convergence of the supremum distance for supersmooth kernel deconvolution. *Statist. Probab. Lett.* **78** 2932-2938.

van Es, B. and Uh, H.-W. (2005). Asymptotic normality of kernel-type deconvolution estimators. *Scand. J. Statist.* **32** 467-483.

Fan, J. (1991a). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19** 1257-1272.

Fan, J. (1991b). Asymptotic normality for deconvolution kernel density estimators. *Sankhya A* **53** 97-110.

Feurerverger, A. and Mureika, R.A. (1977). Empirical characteristic function and its applications. *Ann. Statist.* **5** 88-97.

Folland, G.B. (1999). *Real Analysis (2nd Edition).* Wiley.

Giné, E. and Nickl, R. (2009). Uniform limit theorems for wavelet density estimators. *Ann. Probab.* **37** 1605-1646.

Giné, E. and Nickl, R. (2016). *Mathematical Foundations of Infinite-Dimensional Statistical Models.* Cambridge University Press.

Hall, P. (1991). On convergence rates of suprema. *Probab. Theory Related Fields* **89** 447-455.

Hendricks, K., Porter, R.H., and Boudreau, B. (1987). Information, returns, and bidding behavior in OCS auctions: 1954-1969. *J. Indust. Econom.* **35** 517-542

Horowitz, J.L. and Markatou, M. (1996). Semiparametric estimation of regression models for panel data. *Rev. Econom. Stud.* **63** 145-168.

Johannes, J. (2009). Deconvolution with unknown error distribution. *Ann. Statist.* **37** 2301-2323.

Komlós, J., Major, P., and Tusnády, G. (1975). An approximation for partial sums of independent rv's and the sample df I. *Z. Warhsch. Verw. Gabiete* **32** 111-131.

Leadbetter, M.R., Lindgren, G., and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer.

Li, T., Perrigne, I., and Vuong, Q. (2002). Conditionally independent private information in OCS wildcat auctions. *J. Econometrics* **98** 129-161.

Li, T. and Vuong, Q. (1998). Nonparametric estimation of the measurement error model using multiple indicators. *J. Multivariate Anal.* **65** 139-165.

Lounici, K. and Nickl, R. (2011). Global uniform risk bounds for wavelet deconvolution estimators. *Ann. Statist.* **39** 201-231.

Meister, A. (2009). *Deconvolution Problems in Nonparametric Statistics*. Springer.

Neumann, M.H. (1997). On the effect of estimating the error density in nonparametric deconvolution. *J. Nonparametric Statist.* **7** 307-330.

Neumann, M.H. (2007). Deconvolution from panel data with unknown error distribution. *J. Multivariate Anal.* **98** 1955-1968.

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

Politis, D.N. and Romano, J.P. (1999). Multivariate density estimation with general flat-top kernels of infinite order. *J. Multivariate Anal.* **68** 1-25.

Schmidt-Hieber, J., Munk, A., and Dümbgen, L. (2013). Multiscale methods for shape constraints in deconvolution: confidence statements for qualitative features. *Ann. Statist.* **41** 1299-1328.

Smirnov, N.V. (1950). On the construction of confidence regions for the density of distribution of random variables. *Doklady Akad. Nauk SSSR* **74** 189-191 (Russian).

Stefanski, L. and Carroll, R.J. (1990). Deconvoluting kernel density estimators. *Statistics* **21** 169-184.

van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer.

(K. Kato) Graduate School of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan.

E-mail address: kkato@e.u-tokyo.ac.jp

(Y. Sasaki) Department of Economics, Johns Hopkins University, Wyman Park Building 560, 3400 N. Charles St, Baltimore, MD 21218, USA.

E-mail address: sasaki@jhu.edu