# INFERENCE ON DISTRIBUTION FUNCTIONS UNDER MEASUREMENT ERROR

KARUN ADUSUMILLI, TAISUKE OTSU, AND YOON-JAE WHANG

Abstract. This paper is concerned with inference on the cumulative distribution function (cdf) $F_{X^*}$ in the classical measurement error model $X = X^* + \epsilon$. We show validity of asymptotic and bootstrap approximations for the distribution of the deviation in the sup-norm between the deconvolution cdf estimator of Hall and Lahiri (2008) and $F_{X^*}$. We allow the density of $\epsilon$ to be ordinary or super smooth, or to be estimated by repeated measurements. Our approximation results are applicable to various contexts, such as confidence bands for $F_{X^*}$ and its quantiles, and cdf-based tests for goodness-of-fit of parametric models of $F_{X^*}$, homogeneity of two samples, and stochastic dominance. Simulation and real data examples illustrate satisfactory performance of the proposed methods.

## 1. Introduction

This paper is concerned with inference on the cumulative distribution function (cdf) $F_{X^*}$ in the classical measurement error model $X = X^* + \epsilon$, where we observe $X$ instead of $X^*$ and $\epsilon$ is a measurement error. For estimation of the probability density function (pdf) $f_{X^*}$, there is rich literature on the density deconvolution (see, Meister, 2009, for a review). In contrast, literature of estimation and inference on the cdf $F_{X^*}$ is relatively thin. Fan (1991) proposed a cdf estimator by integrating the deconvolution density estimator with some truncation. This truncation for the integral is circumvented in Hall and Lahiri (2008) (for the case where the pdf $f_\epsilon$ of $\epsilon$ is symmetric) and Dattner, Goldenshluger and Juditsky (2011) (for the case where $f_\epsilon$ is possibly asymmetric). Hall and Lahiri (2008) studied the $L_2$-risk properties of the cdf estimator. Dattner, Goldenshluger and Juditsky (2011) considered minimax rate optimal estimation of $F_{X^*}$. Both Hall and Lahiri (2008) and Dattner, Goldenshluger and Juditsky (2011) focused on the risk properties of the estimator $\hat{F}_{X^*}(t_0)$ at a given $t_0$ and assumed ordinary smooth densities for $f_\epsilon$. These papers demonstrate that in contrast to the no measurement error case, the cdf estimator $\hat{F}_{X^*}(t_0)$ converges to $F_{X^*}(t_0)$ typically at a nonparametric rate. On the other hand, Söhl and Trabs (2012) established a uniform central limit theorem for linear functionals of the deconvolution estimator that can be applied to derive a Donsker-type theorem, i.e., the weak convergence of $\sqrt{n}\{\hat{F}_{X^*}(\cdot) - F_{X^*}(\cdot)\}$ to a Gaussian process. Söhl and Trabs (2012) considered the case of ordinary smooth $f_\epsilon$, and for the Donsker-type result obtained therein, it is demanded that the Fourier transform $f_\epsilon^{\text{ft}}$ satisfies $|f_\epsilon^{\text{ft}}(\cdot)| \leq C|\cdot|^{-\beta}$ for some $\beta < 1/2$ and $C > 0$, which excludes the Laplace distribution for example. It must be emphasized that (except for Fan, 1991, on the truncated estimator) all these papers concentrated on the case of ordinary smooth and known $f_\epsilon$, so the cases of super smooth and unknown $f_\epsilon$ (with repeated measurements) are not covered.

In this paper, we investigate validity of asymptotic and bootstrap approximations for the distribution of the maximal deviation $T_n = \sup_{t\in\mathcal{T}} |\hat{F}_{X^*}(t) - F_{X^*}(t)|$ in the sup-norm over some set $\mathcal{T}$ between the deconvolution cdf estimator $\hat{F}_{X^*}$ of Hall and Lahiri (2008) and $F_{X^*}$. Our analysis allows $f_\epsilon$ to be ordinary or super smooth, or to be unknown and estimated by repeated measurements. We also characterize the convergence rate of the bootstrap approximation error and find that it is of polynomial order under ordinary smooth errors and logarithmic order under super smooth errors. Our approximation results on the distribution of $T_n$ are applicable to various contexts, such as confidence bands for $F_{X^*}$ and its quantiles, and cdf-based tests for goodness-of-fit of parametric models of $F_{X^*}$, homogeneity of two samples, and stochastic dominance. We emphasize that some inference problems, such as testing for stochastic dominance, are cumbersome to be handled by density-based methods. Also, even for problems which can be dealt with density-based methods (e.g., goodness-of-fit testing), the cdf-based methods are expected to have desirable power properties.

In the context of the density deconvolution, Bissantz, Dümbgen, Holzmann and Munk (2007) extended Bickel and Rosenblatt's (1973) construction of uniform confidence bands for densities to the classical measurement error model with the ordinary smooth $f_\epsilon$. A recent paper by Kato and Sasaki (2016) considered confidence bands of the pdf $f_{X^*}$ with unknown $f_\epsilon$. In contrast to the above papers, this paper is concerned with inference on the cdf $F_{X^*}$. Dattner, Reiß and Trabs (2016) proposed a quantile estimator of $X^*$ and obtained the uniform convergence rate. This paper provides a confidence band for the quantile function of $X^*$.

This paper is organized as follows. In Section 2, we focus on the case of known $f_\epsilon$ and present the asymptotic and bootstrap approximations for $T_n$. Section 3 considers the case where $f_\epsilon$ is unknown but repeated measurements on $X^*$ are available, and studies validity of a bootstrap approximation for the distribution of $T_n$. Section 4 contains four applications of the main results: a confidence band for quantiles (Section 4.1), goodness-of-fit test for parametric models of $F_{X^*}$ (Section 4.2), homogeneity test for two samples (Section 4.3), and test for stochastic dominance (Section 4.4). Section 5 presents some simulation evidences. In Section 6, we provide a real data example, stochastic dominance tests for income data in Korea. All proofs are contained in the Supplement.

## 2. Case of known measurement error distribution

2.1. **Setup.** We first introduce our basic setup. Suppose we observe a random sample $\{X_i\}_{i=1}^n$ generated from

$$X = X^* + \epsilon, \tag{1}$$

where $X^*$ is an unobservable variable of interest and $\epsilon$ is its measurement error. Throughout the paper, $\epsilon$ is assumed to be independent of $X^*$ (i.e., $\epsilon$ is the classical measurement error). Let $\mathrm{i} = \sqrt{-1}$ and $f^{\mathrm{ft}}$ be the Fourier transform of a function $f$. If the pdf $f_\epsilon$ of $\epsilon$ is known, the pdf $f_{X^*}$ of $X^*$ can be estimated by the so-called deconvolution kernel density estimator (see, e.g., Stefanski and Carroll, 1990)

$$\hat{f}_{X^*}(t) = \frac{1}{nh}\sum_{i=1}^{n}\mathbb{K}\left(\frac{t - X_i}{h}\right), \quad \text{where } \mathbb{K}(u) = \frac{1}{2\pi}\int_{-1}^{1} e^{-\mathrm{i}\omega u}\frac{K^{\mathrm{ft}}(\omega)}{f_\epsilon^{\mathrm{ft}}(\omega/h)}d\omega, \tag{2}$$

where $h$ is a bandwidth and $K$ is a kernel function with $K^{\mathrm{ft}}$ supported on $[-1, 1]$. Furthermore, if $f_\epsilon$ is symmetric, integration of $\hat{f}_{X^*}$ yields the following estimator for the cdf $F_{X^*}$ of $X^*$ (see, Hall and Lahiri, 2008)

$$\hat{F}_{X^*}(t) = \frac{1}{2} + \frac{1}{n}\sum_{i=1}^{n}\mathbb{L}\left(\frac{t - X_i}{h}\right), \quad \text{where } \mathbb{L}(u) = \frac{1}{2\pi}\int_{-1}^{1}\frac{\sin(\omega u)}{\omega}\frac{K^{\mathrm{ft}}(\omega)}{f_\epsilon^{\mathrm{ft}}(\omega/h)}d\omega. \tag{3}$$

For the general case of possibly asymmetric $f_\epsilon$, an estimator for $F_{X^*}$ is obtained by replacing $\mathbb{L}(u)$ with $\mathbb{L}_a(u) = \frac{1}{\pi}\int_0^1 \frac{1}{\omega}\mathrm{Im}\left[e^{-\mathrm{i}\omega u}\frac{K^{\mathrm{ft}}(\omega)}{f_\epsilon^{\mathrm{ft}}(\omega/h)}\right]d\omega$ (Dattner, Goldenshluger and Juditsky, 2011), where $\mathrm{Im}[\cdot]$ stands for the imaginary part. Although we hereafter focus on the cdf estimator in (3), our results can be extended to the general asymmetric case.

This section is concerned with approximation for the distribution of the maximal deviation

$$T_n = \sup_{t\in\mathcal{T}}|\hat{F}_{X^*}(t) - F_{X^*}(t)|, \tag{4}$$

under the sup-norm, where $\mathcal{T}$ is a compact interval specified by the researcher. A direct use of such approximation is construction of the confidence band for $F_{X^*}$ over $\mathcal{T}$. Several other ways to use this approximation are presented in Section 4. In Section 2.2 below, we consider a bootstrap approximation for the distribution of $T_n$. In Section 2.3, we also present an asymptotic approximation based on the Gumbel distribution for ordinary smooth measurement error densities.

2.2. **Bootstrap approximation.** Consider a nonparametric bootstrap resample $\{X_i^\#\}_{i=1}^n$ from $\{X_i\}_{i=1}^n$ with equal weights. The bootstrap counterpart of $T_n$ is given by $T_n^\# = \sup_{t\in\mathcal{T}}|\hat{F}_{X^*}^\#(t) - \hat{F}_{X^*}(t)|$, where $\hat{F}_{X^*}^\#$ is defined as in (3) using $X_i^\#$. To establish validity of the bootstrap approximation, we impose the following assumptions.

**Assumption C.** *(i) $\{X_i\}_{i=1}^n$ is an i.i.d. sample from $X = X^* + \epsilon$. $X^*$ and $\epsilon$ are independent. (ii) The densities $f_X$, $f_{X^*}$, and $f_\epsilon$ are bounded and continuous on $\mathbb{R}$, and $\inf_{t\in\mathcal{T}^\delta} f_X(t) > c$ for some $c > 0$ and $\delta$-expansion $\mathcal{T}^\delta$ of $\mathcal{T}$. Also, $E|X^*| < \infty$ and $E|\epsilon| < \infty$. (iii) $\sup_{\omega\in\mathbb{R}}\{(1 + |\omega|)^\gamma|f_{X^*}^{\mathrm{ft}}(\omega)|\} < C$ for some $\gamma, C > 0$. (iv) $f_\epsilon^{\mathrm{ft}}(\omega) \neq 0$ for all $\omega \in \mathbb{R}$, $f_\epsilon^{\mathrm{ft}}(\omega)$ is differentiable at all $\omega \in \mathbb{R}$, and $f_\epsilon$ is an even function.*

Assumption C (i) is on the setup wherein we assume that $\epsilon$ is a classical measurement error.[1] Assumption C (ii) is mild but excludes the Cauchy measurement error. This assumption is required for characterizing the bias of the estimator (see, e.g., Hall and Lahiri, 2008). The Cauchy measurement error is also ruled out in van Es and Uh (2005) who show pointwise asymptotic normality of the deconvolution density estimator. Assumption C (iii), analogous to the so-called Sobolev condition, is used to characterize the rate for the bias term (cf. Hall and Lahiri, 2008).

[1]The independence assumption between $X^*$ and $\epsilon$ is standard but, if necessary, can be relaxed to the sub-independence assumption, see Schennach (2013).

Assumption C (iv) contains conditions on $f_\epsilon$. The first condition is common in the density deconvolution literature but may be relaxed by taking a ridge approach as in Hall and Meister (2007). The last condition is used to derive the cdf estimator in (3) as in Hall and Lahiri (2008). Also when we consider estimation of $f_\epsilon$ using repeated measurements, symmetry of $f_\epsilon$ gives us a simple estimator (Delaigle, Hall and Meister, 2008).

We now present two classes of assumptions on the tail behavior of $f_\epsilon$. The first class is called the ordinary smooth densities.

**Assumption OS.** *(i) There exist* $\beta > 1/2$ *and* $c, C, \omega_0 > 0$ *such that*

$$c|\omega|^{-\beta} \leq |f_\epsilon^{\mathrm{ft}}(\omega)| \leq C|\omega|^{-\beta},$$

*for all* $|\omega| \geq \omega_0$. *(ii)* $K$ *is an even function with* $K^{\mathrm{ft}}(\omega) = (1-\omega^q)^r \mathbb{I}\{|\omega| \leq 1\}$ *for some* $q, r \geq 2$. *There exist* $c_1, C_1 > 0$ *such that*

$$n^{-1/4} h^{\beta-1/2} \int |\mathbb{K}(u)| du < C_1 n^{-c_1}, \tag{5}$$

*for all* $n$ *large enough. Also, letting* $\bar{\mathbb{K}}(u) = \frac{1}{\pi}\int_0^1 \cos(\omega u) \frac{K^{\mathrm{ft}}(\omega)}{f_\epsilon^{\mathrm{ft}}(\omega/h)} \mathbb{I}\{|\omega| \geq h\omega_0\} d\omega$, *it holds that*

$$h^{\beta-1/2} \int |\mathbb{K}(u) - \bar{\mathbb{K}}(u)| du = O(h^s), \tag{6}$$

*for some* $s > 0$. *(iii) As* $n \to \infty$, *it holds* $h \to 0$, $\sqrt{n} h^{\beta-1/2} \to \infty$, $n^\nu h \to 0$ *for some* $\nu \in (0, 1/2)$, *and* $n^{1+2\xi} h^{2(\beta+\gamma)-1} \to 0$ *for some* $\xi > 0$.

Assumption OS (i) is a standard condition to characterize ordinary smooth densities. Note that we focus on the case of $\beta > 1/2$, where the cdf estimator $\hat{F}_{X^*}$ converges at a nonparametric rate (Dattner, Goldenshluger and Juditsky, 2011). For the case of $\beta < 1/2$, the estimator $\hat{F}_{X^*}$ typically converges at the $\sqrt{n}$-rate and the Donsker-type theorem applies (Söhl and Trabs, 2012). Assumption OS (ii) contains conditions for the kernel function. The first condition specifies a particular form for $K$ that is commonly used in the literature (e.g., Delaigle and Hall, 2006). The second condition ensures that the deconvolution kernel $\mathbb{K}$ is $L_1$-integrable. The term $n^{-1/4}$ in (5) is required to ensure that the bootstrap counterpart $T_n^{\#}$ convergences to a Gaussian process at a polynomial rate in $n$ (see, Lemma 2). If $f_\epsilon^{\mathrm{ft}}$ is twice differentiable, applying the integration by parts formula twice gives

$$\mathbb{K}(u) = \frac{1}{u^2} \int_0^1 \cos(\omega u) \left\{ \frac{K^{\mathrm{ft}}(\omega)}{f_\epsilon^{\mathrm{ft}}(\omega/h)} \right\}'' d\omega,$$

and a sufficient condition for (5) is

$$n^{-1/4} h^{\beta-1/2} \sup_{|\omega| \leq 1} \left| \left\{ \frac{K^{\mathrm{ft}}(\omega)}{f_\epsilon^{\mathrm{ft}}(\omega/h)} \right\}'' \right| = O(n^{-c_1}),$$

for some $c_1 > 0$. The third condition assures that $\mathbb{K}$ is well approximated by its trimmed version $\bar{\mathbb{K}}$. Since

$$\int |\mathbb{K}(u) - \bar{\mathbb{K}}(u)| du = \frac{1}{\pi} \int \left| \int_0^{h\omega_0} \cos(\omega u) \frac{K^{\mathrm{ft}}(\omega)}{f_\epsilon^{\mathrm{ft}}(\omega/h)} d\omega \right| du,$$

applying the integration by parts formula twice again implies that a sufficient condition for (6) is given by

$$h^{\beta+1/2} \sup_{|\omega|\leq h\omega_0} \max\left\{\left|\left(\frac{K^{\mathrm{ft}}(\omega)}{f_\epsilon^{\mathrm{ft}}(\omega/h)}\right)'\right|, \left|\left(\frac{K^{\mathrm{ft}}(\omega)}{f_\epsilon^{\mathrm{ft}}(\omega/h)}\right)''\right|\right\} = O(h^s),$$

for some $s > 0$. Based on the above sufficient conditions, it is possible to show that Assumption OS (ii) is satisfied by a large class of ordinary smooth error distributions including Laplace and its convolutions. Intuitively these conditions mean that $f_\epsilon^{\mathrm{ft}}$ should not oscillate too wildly around its trend implied by the ordinary smooth density. Finally, Assumption OS (iii) contains conditions for the bandwidth $h$.

The second class of measurement error densities, called the super smooth densities, is presented as follows.

**Assumption SS.** *(i) There exist $\mu, c, C, \omega_0, \lambda > 0$ and $\lambda_0 \in \mathbb{R}$ such that*

$$c|\omega|^{\lambda_0} \exp(-|\omega|^\lambda/\mu) \leq |f_\epsilon^{\mathrm{ft}}(\omega)| \leq C|\omega|^{\lambda_0} \exp(-|\omega|^\lambda/\mu),$$

*for all $|\omega| \geq \omega_0$. (ii) $K$ is an even function with $K^{\mathrm{ft}}(\omega) = (1-\omega^q)^r \mathbb{I}\{|\omega| \leq 1\}$ for some $q, r \geq 2$. There exist $\mu_1 > 2\mu$ and $c_1, C_1 < \infty$ such that*

$$\frac{1}{\varsigma(h)} \int |\mathbb{K}(u)| du < C_1 h^{-c_1} \exp\left(\frac{1}{\mu_1 h^\lambda}\right), \tag{7}$$

*for all $n$ large enough, where*

$$\varsigma(h) = h^\vartheta \exp\left(\frac{1}{\mu h^\lambda}\right) \tag{8}$$

*with $\vartheta = \lambda(s+1/2) + \lambda_0 + 1/2$. Also, letting $\bar{\mathbb{K}}(u) = \frac{1}{\pi}\int_0^1 \cos(\omega u) \frac{K^{\mathrm{ft}}(\omega)}{f_\epsilon^{\mathrm{ft}}(\omega/h)} \mathbb{I}\{|\omega| \geq h\omega_0\} d\omega$, it holds that*

$$\frac{1}{\varsigma(h)} \int |\mathbb{K}(u) - \bar{\mathbb{K}}(u)| du = O(n^{-s}), \tag{9}$$

*for some $s > 0$. (iii) $h = \left(\frac{\mu}{2}\log n + \mu\theta \log\log n\right)^{-1/\lambda}$ for some $\theta \in ((\vartheta-\gamma)/\lambda + 1, \vartheta/\lambda)$.*

Assumption SS (i) a standard condition to characterize super smooth densities. Assumption SS (ii) contains conditions for the kernel function, and similar comments apply as the ordinary smooth case. The condition $\mu_1 > 2\mu$ is required to guarantee that the bootstrap counterpart $T_n^{\#}$ convergences to a Gaussian process at a polynomial rate in $n$ (see, Lemma 5). If $f_\epsilon^{\mathrm{ft}}$ is twice differentiable, a sufficient condition for (7) is

$$\frac{1}{\varsigma(h)} \sup_{|\omega|\leq 1} \left|\left(\frac{K^{\mathrm{ft}}(\omega)}{f_\epsilon^{\mathrm{ft}}(\omega/h)}\right)''\right| d\omega = O\left(h^{-a} \exp\left(\frac{1}{\mu_1 h^\lambda}\right)\right),$$

for some $a > 0$. Also, a sufficient condition for (9) is

$$\exp\left(-\frac{1}{\mu h^\lambda}\right) \sup_{|\omega|\leq h\omega_0} \max\left\{\left|\left(\frac{K^{\mathrm{ft}}(\omega)}{f_\epsilon^{\mathrm{ft}}(\omega/h)}\right)'\right|, \left|\left(\frac{K^{\mathrm{ft}}(\omega)}{f_\epsilon^{\mathrm{ft}}(\omega/h)}\right)''\right|\right\} = O(n^{-a_1}),$$

for some $a_1 > 0$. For instance, these conditions are satisfied if

$$\sup_{|\omega|\leq 1} \max\{|A'(\omega/h)|, |A''(\omega/h)|\} = O\left(h^{-a_1} \exp\left(\frac{1}{\mu_1 h^\lambda}\right)\right), \tag{10}$$

for some $a_1 > 0$, where $A(\omega) = \frac{\exp(-|\omega|^\lambda/\mu)}{f_\epsilon^{\mathrm{ft}}(\omega)}$. Based on (10), we can see that Assumption SS (ii) is satisfied by a large class of super smooth error distributions including Gaussian and its convolutions. Since the function $A$ inherits the differentiability properties of $f_\epsilon^{\mathrm{ft}}$, the condition (10) intuitively means that $f_\epsilon^{\mathrm{ft}}$ should not oscillate too wildly around its trend implied by the super smooth density. Assumption SS (iii) is on the bandwidth $h$. Note that this condition implicitly requires $\gamma > \lambda$.

Let $\hat{c}_\alpha$ denote the $(1-\alpha)$-th quantile of the bootstrap statistic $T_n^{\#}$. Under these assumptions, validity of the bootstrap approximation is established as follows.[2]

**Theorem 1.** *Suppose that Assumption C holds true. Then*

$$P\{T_n \leq \hat{c}_\alpha\} \geq 1 - \alpha - \delta_n, \tag{11}$$

*for some positive sequence $\delta_n = O(n^{-c})$ (under Assumption OS) or $\delta_n = O((\log n)^{-c})$ (under Assumption SS) with $c > 0$.*

**Remark 1.** Based on this theorem, we can construct an asymptotic confidence band for $F_{X^*}$ over $\mathcal{T}$ with level $\alpha$ as $\mathcal{C}_n(t) = [\hat{F}_{X^*}(t) \pm \hat{c}_\alpha]$ for $t \in \mathcal{T}$ in the sense that

$$P\{F_{X^*}(t) \in \mathcal{C}_n(t) \text{ for all } t \in \mathcal{T}\} \geq 1 - \alpha - \delta_n,$$

for $\delta_n = O(n^{-c})$ (under Assumption OS) or $\delta_n = O((\log n)^{-c})$ (under Assumption SS) with some $c > 0$. Note that the approximation error $\delta_n$ is of polynomial order under Assumption OS (the ordinary smooth case) and of logarithmic order under Assumption SS (the super smooth case). We also note from the proof of the theorem that the slower approximation rate for the super-smooth case is solely due to the bias; if bias correction were possible, the bootstrap approximation error would be of polynomial order in both cases.

**Remark 2.** To implement the bootstrap approximation in Theorem 1, we need to choose the bandwidth $h$. For estimation of the cdf $F_{X^*}(t_0)$ at a given $t_0$, Hall and Lahiri (2008) suggested to choose $h$ to minimize the approximate integrated MSE based on the normal reference distribution on $X^*$. For estimation of the quantile function of $X^*$, Dattner, Reiß and Trabs (2016) developed an adaptive method to choose $h$ based on Lepski (1990). In Section 5 for simulations, we suggest a bandwidth selection rule based on Bissantz, Dümbgen, Holzmann and Munk (2007). The basic idea is to estimate the ideal bandwidth that minimizes the maximal deviation between $\hat{F}_{X^*}$ and $F_{X^*}$ under the sup-norm by utilizing a series of estimates $\hat{F}_{X^*}$ based on different values of $h$.

2.3. **Asymptotic Gumbel approximation for ordinary smooth case.** For the ordinary smooth case, it is also possible to characterize the asymptotic distribution of the standardized object

$$t_n = \sup_{t \in \mathcal{T}} |f_X(t)^{-1/2}\{\hat{F}_{X^*}(t) - F_{X^*}(t)\}|, \tag{12}$$

[2]Here we present the bootstrap approximation result for the statistic $T_n$ decaying to zero. The result may be presented for its normalized counterpart. This is analogous to whether we present the bootstrap approximation for the non-normalized object $\hat{\theta} - \theta$ or normalized one $\sqrt{n}(\hat{\theta} - \theta)$, where $\theta$ is some parameter and $\hat{\theta}$ is its estimator.

by the Gumbel distribution. Under additional assumptions, listed in Assumption G in the Supplement (Appendix C), we can follow similar steps in Bickel and Rosenblatt (1973) and Bissantz, Dümbgen, Holzmann and Munk (2007) to show the following result.

**Theorem 2.** *Suppose that Assumptions C,OS, and G hold, and $(nh)^{-1}(\log n)^3 \to 0$ as $n \to \infty$. Then*

$$P\left\{(-2\log h)^{1/2}(B^{-1/2}t_n - b_n) \le c\right\} \to \exp(-2\exp(-c)), \tag{13}$$

*for all $c \in \mathbb{R}$, where the constant $B$ and sequence $b_n$ are defined in the Supplement (eq. (25)).*

See the Supplement (Appendix C) for a detailed discussion on Assumption G and proof of this theorem.

**Remark 3.** As shown in (13), the limiting behavior of $t_n$ is characterized by the Gumbel distribution. Based on (13) and the conventional kernel density estimator $\hat{f}_X$ for $f_X$, we can also obtain an asymptotically valid critical value to conduct inference. For example, the asymptotic confidence band at level $\alpha$ for $F_{X^*}$ is given by

$$\mathcal{C}_n^G(t) = [\hat{F}_{X^*}(t) \pm B^{1/2}\hat{f}_X(t)^{1/2}\{c_\alpha^G(-2\log h)^{-1/2} + b_n\}],$$

for $t \in \mathcal{T}$, where $c_\alpha^G$ solves $\exp(-2\exp(-c_\alpha^G)) = \alpha$. However, as discussed in the next remark, the asymptotic Gumbel approximation requires additional assumptions and tends to be less accurate than the bootstrap approximation.

**Remark 4.** Compared to the bootstrap approximation, the asymptotic Gumbel approximation has two drawbacks. First, the Gumbel approximation requires an additional assumption (Assumption G). Second, as indicated by Bissantz, Dümbgen, Holzmann and Munk (2007), the approximation error (i.e., $\delta_n$ in (11) for the bootstrap approximation) by (13) is typically a logarithmic rate even under Assumption OS, and therefore tends to be less accurate than the bootstrap approximation in (11). This contrast between the asymptotic Gumbel and bootstrap approximations was first clarified by Chernozhukov, Chetverikov and Kato (2014) for construction of confidence bands on the density with no measurement error. Kato and Sasaki (2016) extended their results for confidence bands on the pdf $f_{X^*}$ with unknown $f_\epsilon$. We obtain analogous results for confidence bands on the cdf $F_{X^*}$. We also note that in contrast to Chernozhukov, Chetverikov and Kato (2014) and Kato and Sasaki (2016) who employed Gaussian multiplier bootstrap methods, Theorem 1 shows validity of the conventional nonparametric bootstrap. Accordingly the techniques used in the proof of Theorem 1 are quite different: in particular, we employ Komlós, Major and Tusnády's (1975) coupling along with anti-concentration inequalities for Gaussian processes (Chernozhukov, Chetverikov and Kato, 2015) while the latter employ the Slepian-Stein type coupling for suprema of empirical processes constructed in Chernozhukov, Chetverikov and Kato (2014). Finally, we also obtain deterministic bounds on the approximation error of the bootstrap; to the best of our knowledge this is new in the literature on deconvolution.

**Remark 5.** We note that the asymptotic Gumbel approximation in (13) is available only for the ordinary smooth case. It remains an open question whether we can establish such an asymptotic approximation for the super smooth case. As discussed in Bissantz, Dümbgen, Holzmann and

Munk (2007, p. 486) for the density deconvolution, the main difficulty is that the limiting form of the deconvolution kernel (eq. (24) in Appendix C of the Supplement) is not available for the super smooth case. On the other hand, as shown in Theorem 1, we emphasize that the bootstrap approximation is valid even for the super smooth case.

## 3. Case of unknown measurement error distribution

The assumption of known measurement error density $f_\epsilon$ is unrealistic in many applications. In this section, we consider the situation where $f_\epsilon$ is unknown and needs to be estimated. In general, $f_\epsilon$ cannot be identified by a single measurement. Identification of $f_\epsilon$ can be restored however if we have two or more independent noisy measurements of the variable $X^*$. More specifically, suppose that we observe

$$X_{i,j} = X_i^* + \epsilon_{i,j} \quad \text{for } j = 1, \dots, N_i \text{ and } i = 1, \dots, n,$$

where $X_i^*$ is the error-free variable and $\epsilon_{i,j}$'s are independently distributed measurement errors from the density $f_\epsilon$. We thus have $N_i$ repeated measurements of each variable $X_i^*$. We shall assume that the number of repeated observations is bounded above (i.e., $N_i \leq C < \infty$ for all $i$). This assumption is not critical for our theory but allows us to simplify the proofs considerably. Since in practice the number of repeated measurements is small anyway, we do not pursue the generalization to growing $C$. Under the assumption that $f_\epsilon$ is symmetric (Assumption C (iv)), its Fourier transform $f_\epsilon^{\text{ft}}$ can be estimated by (Delaigle, Hall and Meister, 2008)

$$\hat{f}_\epsilon^{\text{ft}}(\omega) = \left| \frac{1}{N} \sum_{i=1}^{n} \sum_{(j_1,j_2)\in\mathcal{J}_i}^{N_i} \cos\{\omega(X_{i,j_1} - X_{i,j_2})\} \right|^{1/2}, \tag{14}$$

where $N = \frac{1}{2}\sum_{i=1}^{n} N_i(N_i-1)$ , $\mathcal{J}_i$ is the set of $\frac{1}{2}N_i(N_i-1)$ distinct pairs $(j_1, j_2)$ with $1 \leq j_1 < j_2 \leq N$, and we ignore all the observations with $N_i = 1$. By plugging this estimator into (3), we can estimate the cdf $F_{X^*}$ by

$$\tilde{F}_{X^*}(t) = \frac{1}{2} + \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{N_i} \tilde{\mathbb{L}}\left(\frac{t - X_{i,j}}{h}\right), \quad \text{where } \tilde{\mathbb{L}}(u) = \frac{1}{2\pi}\int_{-1}^{1} \frac{\sin(\omega u)}{\omega}\frac{K^{\text{ft}}(\omega)}{\hat{f}_\epsilon^{\text{ft}}(\omega/h)} d\omega. \tag{15}$$

In this section, we consider bootstrap approximation of the distribution of the maximal deviation $\tilde{T}_n = \sup_{t\in\mathcal{T}} |\tilde{F}_{X^*}(t) - F_{X^*}(t)|$. To construct the bootstrap counterpart of $\tilde{T}_n$, we suggest resampling from the set of observed variables $\{X_{i,j}\}$ while keeping the estimated measurement error density $\hat{f}_\epsilon^{\text{ft}}$ the same. More precisely, the bootstrap version of $\tilde{F}_{X^*}$ is given by

$$\tilde{F}_{X^*}^{\#}(t) = \frac{1}{2} + \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{N_i} \tilde{\mathbb{L}}\left(\frac{t - X_{i,j}^{\#}}{h}\right),$$

where $X_{i,j}^{\#}$ is randomly drawn from the pooled observations $\{X_{i,j}\}$. The bootstrap counterpart of $\tilde{T}_n$ is obtained as $\tilde{T}_n^{\#} = \sup_{t\in\mathcal{T}} |\tilde{F}_{X^*}^{\#}(t) - \tilde{F}_{X^*}(t)|$.

To establish validity of the bootstrap approximation by $\tilde{T}_n^{\#}$, we first show that the cdf estimator $\tilde{F}_{X^*}$ under repeated measurements converges fast enough under the sup-norm to $\hat{F}_{X^*}$ so that the distributional properties of the latter would continue to hold. Previously, for the case of density deconvolution, Delaigle, Hall and Meister (2008) showed that under certain conditions, the deconvolution pdf estimator $\tilde{f}_{X^*}$ using $\hat{f}_\epsilon^{\text{ft}}$ enjoys the same first-order asymptotic properties as the estimator $\hat{f}_{X^*}$ in (2) for the case of known $f_\epsilon$. Also, this result was obtained in terms of the uniform MSE metric, $\sup_t E|\tilde{f}_{X^*}(t) - \hat{f}_{X^*}(t)|^2$. Since validity of the confidence bands rests on controlling the sup-norm, we derive a corresponding result for the cdf estimators under the sup-norm. To this end, we add the following conditions.

**Assumption B.** *(i) There exist $c \in (0,1)$ and $C > 0$ such that $P\{|\epsilon| \geq M\} \leq C(\log M)^{-1/c}$ for all $M > 0$. (ii) As $n \to \infty$, it holds $h \to 0$, $\sqrt{n}h^{\beta-1/2} \to \infty$, $\log n/(nh^{4\beta}) \to 0$, and $nh^{4\beta+1} \to \infty$.*

Based on these conditions, we are able to prove the following theorem.

**Theorem 3.** *Suppose that Assumptions C, OS, and B hold with $\gamma > \beta + 1$. Then for some $c > 0$,*

$$\sqrt{n}h^{\beta-1/2} \sup_{t\in\mathcal{T}} |\tilde{F}_{X^*}(t) - \hat{F}_{X^*}(t)| = o_p(n^{-c}).$$

Let $\tilde{c}_\alpha$ be the $(1-\alpha)$-th quantile of the bootstrap statistic $\tilde{T}_n^{\#}$. Based on the above theorem, validity of the bootstrap approximation is established as follows.

**Theorem 4.** *Suppose that Assumptions C, OS, and B hold with $\gamma > \beta + 1$. Then*

$$P\{\tilde{T}_n \leq \tilde{c}_\alpha\} \geq 1 - \alpha - o_p(1). \tag{16}$$

**Remark 6.** Based on this theorem, we can construct an asymptotic confidence band for $F_{X^*}$ over $\mathcal{T}$ with level $\alpha$ as $[\tilde{F}_{X^*}(t) \pm \tilde{c}_\alpha]$ for $t \in \mathcal{T}$. The key additional requirement $\gamma > \beta + 1$ says that $f_{X^*}$ is smoother than $f_\epsilon$ by up to a derivative. As shown in Theorem 3, this ensures that the error from estimating $f_\epsilon^{\text{ft}}$ is asymptotically negligible. Also, we note that the conditions $nh^{4\beta+1} \to \infty$ in Assumption B (ii) and $n^{1+2\xi}h^{2(\beta+\gamma)-1} \to 0$ for some $\xi > 0$ in Assumption OS (iii) hold simultaneously only if $\gamma > \beta + 1$.

**Remark 7.** Note that the above theorems are presented only for the ordinary smooth case. A similar result can be derived for the super smooth case under the assumption that $f_{X^*}$ is smoother than $f_\epsilon$, i.e. the former is also super smooth. One such sufficient condition in the super smooth case could be

$$\int \left| \frac{\omega^a f_{X^*}^{\text{ft}}(\omega)}{f_\epsilon^{\text{ft}}(\omega)} \right|^2 d\omega < \infty,$$

for all $a > 0$.

**Remark 8.** Here we focus on the case where repeated measurements on $X^*$ are available and $f_\epsilon^{\text{ft}}$ can be estimated by (14) under the symmetry assumption on $f_\epsilon$. If $f_\epsilon$ is not necessarily symmetric but repeated measurements are available, then we can employ the estimator by Li and Vuong (1998) or Comte and Kappus (2015) based on Kotlarski's identity. Also, in some applications, a separate independent experiment may give us observations from $f_\epsilon$ (see, e.g., Efromovich, 1997, and Neumann, 1997). See Meister (2009, Section 2.6) for an overview on estimation of $f_\epsilon$. We can expect that similar results hold true for other estimators of $f_\epsilon^{\text{ft}}$ under different setups as far as the estimator $\hat{f}_\epsilon^{\text{ft}}$ to construct $\tilde{F}_{X^*}$ converges sufficiently fast to $f_\epsilon^{\text{ft}}$ so that the estimation error $\tilde{F}_{X^*}(t) - \hat{F}_{X^*}(t)$ is negligible.

## 4. Applications

4.1. **Confidence band for quantile function.** In addition to the confidence band for $F_{X^*}$, the results in the previous sections can be utilized to obtain the confidence band for the quantile function of $X^*$. Hall and Lahiri (2008) proposed to estimate the $u$-th quantile $Q(u) = F_{X^*}^{-1}(u)$ by

$$\hat{Q}(u) = \sup\{t : \hat{F}_{X^*}^{\text{m}}(t) \leq u\},$$

where $\hat{F}_{X^*}^{\text{m}}(t) = \sup_{y \leq t} \hat{F}_{X^*}(y)$ is a monotone version of $\hat{F}_{X^*}(t)$. To obtain the confidence band for the quantile function $Q(u)$ over some interval $[u_1, u_2]$, we impose the following assumptions.

**Assumption Q.** *(i) $F_{X^*}^{-1}(u)$ exists and is unique for all $u \in [u_1, u_2]$ such that $0 < u_1 < u_2 < 1$. There exists an interval $\mathcal{H}$ satisfying $F_{X^*}^{-1}[u_1 - \varepsilon, u_2 + \varepsilon] \subset \mathcal{H}$ for some $\varepsilon > 0$, $\inf_{x \in \mathcal{H}} f_X(x) > 0$, and $0 < \inf_{x \in \mathcal{H}} f_{X^*}(x) \leq \sup_{x \in \mathcal{H}} f_{X^*}(x) < \infty$. (ii) $\sup_{x \in \mathcal{H}} |f_{X^*}(x + \delta) - f_{X^*}(x)| \leq M|\delta|^a$ for all $\delta$ sufficiently small, with $a > 0$ (under Assumption OS) and $a = 1$ (under Assumption SS).*

Based on these assumptions, we can obtain the asymptotic confidence bands for the quantile function as follows.

**Theorem 5.** *Suppose that Assumptions C, Q, and either OS or SS hold true. Then,*

$$P\left\{\hat{Q}(u) - \frac{\hat{c}_\alpha}{\hat{f}_{X^*}(\hat{Q}(u))} \leq Q(u) \leq \hat{Q}(u) + \frac{\hat{c}_\alpha}{\hat{f}_{X^*}(\hat{Q}(u))} \text{ for all } u \in [u_1, u_2]\right\} \geq 1 - \alpha - o(1).$$

**Remark 9.** Dattner, Reiß and Trabs (2016) have obtained the uniform convergence rate of their quantile estimator, say $\bar{Q}(u)$, based on the M-estimation method. In particular, Dattner, Reiß and Trabs (2016, Proposition 2.6) obtained that under an MSE optimal choice of the bandwidth,

$$\sup_{u \in [u_1, u_2]} |\bar{Q}(u) - Q(u)| = O_p\left(\left(\frac{\log n}{n}\right)^{\frac{\gamma}{2(\beta+\gamma)-1}}\right).$$

Thus, Theorem 5 is complementary in that it provides a confidence band for $Q(u)$ over $u \in [u_1, u_2]$. Note that as with the case of the cdf, we require under-smoothing to obtain the asymptotically valid confidence band, which excludes the MSE optimal bandwidth.

4.2. **Goodness-of-fit testing.** Another useful application of our results is goodness-of-fit testing on parametric models for $F_{X^*}$. Consider a parametric model $\{G_{X^*}(\cdot, \theta) : \theta \in \Theta\}$ for the distribution of the error-free variable $X^*$ of interest. For simplicity, suppose the measurement error density $f_\epsilon$ is known as in Section 2. Our method can be adapted to the case of unknown $f_\epsilon$. The goodness-of-fit testing problem of our interest is

$$H_0 : F_{X^*}(t) = G_{X^*}(t, \theta) \text{ over } t \in \mathcal{T} \text{ for some } \theta \in \Theta,$$

against negation of $H_0$. Let $\hat{\theta}$ be some $\sqrt{n}$-consistent estimator of the true parameter $\theta_0$ under $H_0$. A typical example of $\hat{\theta}$ is the maximum likelihood estimator using the density function $\int g_{X^*}(t - a, \theta) f_\epsilon(a) da$ on the observable $X$, where $g_{X^*}$ is the density of $G_{X^*}$.

To test $H_0$, we can employ the Kolmogorov-type statistic

$$K_n = \sup_{t \in \mathcal{T}} |\hat{F}_{X^*}(t) - G_{X^*}(t, \hat{\theta})|,$$

and its bootstrap counterpart is given by

$$K_n^\# = \sup_{t \in \mathcal{T}} |\hat{F}_{X^*}^\#(t) - G_{X^*}(t, \hat{\theta}^\#)|,$$

where $\hat{F}_{X^*}^\#$ and $\hat{\theta}^\#$ are computed by the (parametric) bootstrap resample $\{X_i^\#\}_{i=1}^n$ from $X^\# = X_*^\# + \epsilon^\#$ with $X_*^\# \sim G_{X^*}(\cdot, \hat{\theta})$ and $\epsilon^\# \sim f_\epsilon$. In contrast to the no measurement error case, the cdf estimator $\hat{F}_{X^*}$ converges at a slower rate than $\sqrt{n}$. Therefore, if $\hat{\theta}$ is $\sqrt{n}$-consistent, then the estimation error of $\hat{\theta}$ is negligible under $H_0$, and the validity of the bootstrap critical value follows by a modification of the proof of Theorem 1. The result is summarized in the following corollary. Let $\hat{c}_\alpha^K$ be the $(1-\alpha)$-th quantile of $K_n^\#$.

**Corollary 1.** *Suppose that Assumption C holds true, the null $H_0$ is satisfied at $\theta_0$, $\sqrt{n}(\hat{\theta} - \theta_0) = O_p(1)$, and the density of $G_{X^*}(\cdot, \theta)$ is bounded for all $\theta$ in a neighborhood of $\theta_0$. Then*

$$P\{K_n > \hat{c}_\alpha^K\} \leq \alpha + \delta_n,$$

*for some positive sequence $\delta_n = O(n^{-c})$ (under Assumption OS) or $\delta_n = O((\log n)^{-c})$ (under Assumption SS) with $c > 0$.*

Consistency of the test can be shown analogously. If $f_\epsilon$ is unknown but repeated measurements on $X^*$ are available, a similar result holds true by replacing $\hat{F}_{X^*}$ and $\hat{F}_{X^*}^\#$ with $\tilde{F}_{X^*}$ and $\tilde{F}_{X^*}^\#$, respectively.

4.3. **Homogeneity test.** Our bootstrap and asymptotic approximation results can be extended to two sample problems. Let $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^m$ be two independent samples of $X$ and $Y$. $X$ is generated as in (1). Also $Y$ is generated as

$$Y = Y^* + \delta,$$

where $Y^*$ is the unobservable error-free variable with the distribution function $F_{Y^*}$ and $\delta$ is its measurement error. We assume $\delta$ is independent of $Y^*$. Suppose we wish to test the homogeneity hypothesis

$$H_0 : F_{X^*}(t) = F_{Y^*}(t) \quad \text{for all } t \in \mathcal{T},$$

against the negation of $H_0$. The Kolmogorov-type statistic presented in the last subsection can be modified as follows

$$S_{n,m} = \sup_{t \in \mathcal{T}} |\hat{F}_{X^*}(t) - \hat{F}_{Y^*}(t)|,$$

where $\hat{F}_{Y^*}$ is the estimator for $F_{Y^*}$as in (3) using the sample $\{Y_i\}_{i=1}^m$. In this case, the bootstrap counterpart of $S_{n,m}$ is given by

$$S_{n,m}^\# = \sup_{t \in \mathcal{T}} \left| \hat{F}_{X^*}^\#(t) - \hat{F}_{Y^*}^\#(t) - \{\hat{F}_{X^*}(t) - \hat{F}_{Y^*}(t)\} \right|,$$

where $\hat{F}^{\#}_{Y^*}$ using the sample $\{Y_i\}_{i=1}^m$ is defined in the same manner as $\hat{F}^{\#}_{X^*}$. The $(1-\alpha)$-th quantile $\hat{c}^S_\alpha$ of $S^{\#}_{n,m}$ provides an asymptotically valid critical value as follows.

**Corollary 2.** *Suppose that Assumption C holds true for both $X = X^* + \epsilon$ and $Y = Y^* + \delta$, and that $n/(n+m) \to \tau \in (0,1)$ as $n, m \to \infty$. Then under $H_0$*

$$P\{S_{n,m} > \hat{c}^S_\alpha\} \le \alpha + \delta_{n,m},$$

*for some positive sequence $\delta_{n,m} = O(n^{-c})$ (under Assumption OS for both $\epsilon$ and $\delta$) or $\delta_{n,m} = O((\log n)^{-c})$ (under Assumption SS for both $\epsilon$ and $\delta$) with $c > 0$.*

An analogous result is available for the case of unknown $f_\epsilon$ by replacing $\hat{F}_{X^*}$ and $\hat{F}_{Y^*}$ with their repeated measurements versions. Also, if we wish to test the homogeneity hypothesis $H_0$ but $Y$ has no measurement error (i.e., $Y = Y^*$), we can replace $\hat{F}_{Y^*}$ with the empirical distribution function of the sample $\{Y_i\}_{i=1}^m$.

4.4. **Stochastic dominance test.** Another intriguing application of our main results is testing the hypothesis of the (first-order) stochastic dominance

$$H_0 : F_{X^*}(t) \le F_{Y^*}(t) \quad \text{for all } t \in \mathcal{T}, \tag{17}$$

against the negation of $H_0$. By modifying the Kolmogorov-type test in Section 4.3, the test statistic for (17) and its bootstrap counterpart are given by

$$\begin{aligned} D_{n,m} &= \sup_{t\in\mathcal{T}}\{\hat{F}_{X^*}(t) - \hat{F}_{Y^*}(t)\}, \\ D^{\#}_{n,m} &= \sup_{t\in\mathcal{T}}\left\{\hat{F}^{\#}_X(t) - \hat{F}^{\#}_Y(t) - \{\hat{F}_X(t) - \hat{F}_Y(t)\}\right\}, \end{aligned}$$

where $\hat{F}^{\#}_X$ and $\hat{F}^{\#}_Y$ are computed as in (3) using nonparametric bootstrap resamples $\{X^{\#}_i\}_{i=1}^n$ and $\{Y^{\#}_i\}_{i=1}^m$ from $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^m$, respectively.

Let $\hat{c}^D_\alpha$ denote the $(1-\alpha)$-th quantile of the bootstrap statistic $D^{\#}_{n,m}$. The bootstrap validity of our stochastic dominance test is established as follows.

**Theorem 6.** *Suppose that Assumption C holds true for both $X = X^* + \epsilon$ and $Y = Y^* + \delta$, and that $n/(n+m) \to \tau \in (0,1)$ as $n, m \to \infty$.*

**(i):** *Under $H_0$,*

$$P\{D_{n,m} > \hat{c}^D_\alpha\} \le \alpha + \varrho_{n,m},$$

*for some positive sequence $\varrho_{n,m} = O(n^{-c})$ (under Assumption OS for both $\epsilon$ and $\delta$) or $\varrho_{n,m} = O((\log n)^{-c})$ (under Assumption SS for both $\epsilon$ and $\delta$) with $c > 0$.*

**(ii):** *Let $\mathcal{P}_0$ be the set of probability measures of $(X,Y)$ satisfying $H_0$ (but $f_\delta$ and $f_\epsilon$ are fixed) and*

$$\begin{aligned}
&0 < c_X \le \inf_{t\in\mathcal{T}} f_X(t) \le \sup_{t\in\mathcal{T}} f_X(t) \le C_X < \infty,\\
&0 < c_Y \le \inf_{t\in\mathcal{T}} f_Y(t) \le \sup_{t\in\mathcal{T}} f_Y(t) \le C_Y < \infty,\\
&\sup_{\omega\in\mathbb{R}}\{(1+|\omega|)^{\gamma_X}|f^{\mathrm{ft}}_{X^*}(\omega)|\} \le M_X < \infty,\\
&\sup_{\omega\in\mathbb{R}}\{(1+|\omega|)^{\gamma_Y}|f^{\mathrm{ft}}_{Y^*}(\omega)|\} \le M_Y < \infty,
\end{aligned}$$

*for some $c_X, c_Y, \gamma_X, \gamma_Y, C_X, C_Y, M_X, M_Y > 0$ that are independent of $(f_X, f_Y)$. Then*

$$\sup_{P\in\mathcal{P}_0} P\{D_{n,m} > \hat{c}^D_\alpha\} \le \alpha + \varrho_{n,m},$$

*for some positive sequence $\varrho_{n,m} = O(n^{-c})$ (under Assumption OS) or $\varrho_{n,m} = O((\log n)^{-c})$ (under Assumption SS) with $c > 0$.*

**(iii):** *Under the alternative $H_1$ (i.e., $H_0$ is false) and Assumption OS or SS,*

$$P\{D_{n,m} > \hat{c}^D_\alpha\} \to 1.$$

**Remark 10.** Based on the proof of Theorem 6 (iii), we can characterize some local power properties. Suppose that both measurement errors are ordinary smooth. For any sequence $M_n \to \infty$, $D_{n,m}$ is consistent (i.e., $P\{D_{n,m} > \hat{c}^D_\alpha\} \to 1$) against local alternatives of the form

$$H_{1n} : F_{Y^*}(t) > F_{X^*}(t) + M_n\gamma_n \text{ for some } t \in \mathcal{T},$$

where

$$\gamma_n = n^{-1/2}\left(\frac{\sqrt{\log(1/h_X)}}{h_X^{\beta_X - 1/2}} \vee \frac{\sqrt{\log(1/h_Y)}}{h_X^{\beta_Y - 1/2}}\right),$$

and $h_X$ and $h_Y$ are (possibly different) bandwidths for the estimators $\hat{F}_{X^*}$ and $\hat{F}_{Y^*}$, respectively. A similar expression is available for $\gamma_n$ in the super smooth case with $h_X^{\beta_X-1/2}$, $h_Y^{\beta_Y-1/2}$ replaced by $\varsigma_X^{-1}(h_X)$, $\varsigma_Y^{-1}(h_Y)$ respectively. Finally in the mixed error case, i.e when one of the errors is ordinary smooth while the other is super-smooth, the value of $\gamma_n$ is determined by the super-smooth error (e.g $\gamma_n = n^{-1/2}\varsigma_X(h_X)\sqrt{\log(1/h_X)}$ if $\epsilon$ is super-smooth).

## 5. Simulation

In this section, we investigate the finite sample performance of the bootstrap uniform confidence band discussed in Theorem 1 using simulation experiments.

### 5.1. Simulation designs.

We generate data from the model (1), where the unobserved variable of interest $X^*$ is drawn from the normal distribution $N(0, \sigma^2_{X^*})$ and the measurement error $\varepsilon$ is drawn from the Laplace distribution $L(0, \sigma^2_\varepsilon)$ or the normal distribution $N(0, \sigma^2_\varepsilon)$. We fix $\sigma_{X^*} = 1$ and choose $\sigma_\varepsilon$ so that 'signal-to-noise ratio (SNR)' is given by $\sigma_{X^*}/\sigma_\varepsilon = 2, 3, 4$. We use the kernel function $K$ defined by

$$K(\omega) = \frac{48\cos\omega}{\pi\omega^4}\left(1 - \frac{15}{\omega^2}\right) - \frac{144\sin\omega}{\pi\omega^5}\left(2 - \frac{5}{\omega^2}\right),$$

whose Fourier transformation is given by $K^{\mathrm{ft}}(\omega) = (1-\omega^2)^3 \cdot \mathbb{I}\{|\omega| \le 1\}$. We consider four different sample sizes $n = 100, 250, 500, 1000$ and three different confidence levels $1-\alpha = 0.80, 0.90, 0.95$. The number of simulation and bootstrap repetitions are 2000 and 1000, respectively. We compute the coverage probabilities of our confidence bands for $F_{X^*}$ over the interval $[-2\sigma_{X^*},\ 2\sigma_{X^*}]$.

5.2. **Bandwidth choice.** We adapt the bandwidth selection method of Bissantz, Dümbgen, Holzmann and Munk (2007, Section 5.2) to the cdf estimation. First we consider $J$ different bandwidths: $h_j = h_0 j/J$ for $j = 1, 2, \ldots, J$, where $h_0$ is a pilot bandwidth. A pilot bandwidth is an over-smoothing bandwidth obtained by multiplying $\gamma > 1$ to the normal reference rule of Hall and Lahiri (2008, Section 4.2). The normal reference rule was originally suggested by Delaigle and Gijbels (2004) to estimate density functions and was modified by Hall and Lahiri (2008) to the setting of estimating distribution functions. For $j = 2, \ldots, J$, define the distances

$$L_\infty(\hat{F}_{X^*},\ F_{X^*}) = ||\hat{F}_{X^*} - F_{X^*}||_\infty, \quad d^{(\infty)}_{j-1,j} = ||\hat{F}_{X^*,j-1} - \hat{F}_{X^*,j}||_\infty,$$

where $\hat{F}_{X^*,j}$ denotes the deconvolution estimator (3) with bandwidth $h = h_j$ and $||\cdot||_\infty$ denotes the supremum norm. For over-smoothing bandwidths, $L_\infty(\hat{F}_{X^*},\ F_{X^*})$ changes only moderately with increasing bandwidth, while with undersmoothing bandwidth the distance suddenly increases with decreasing bandwidth. Based on this observation, Bissantz, Dümbgen, Holzmann and Munk (2007) suggest to choose the bandwidth to be the largest bandwidth at which $d^{(\infty)}_{j-1,j}$ is more than $\tau(\tau > 1)$ times greater than $d^{(\infty)}_{J-1,J}$. In our simulations, we choose $J = 20$ (number of bandwidths), $\tau = 3$ and $\gamma = 1.5$. (We find that the simulation results are insensitive to the precise choice of the parameters.

Figures 1 and 2 illustrate the distances over different bandwidths for 3 different random samples with the measurement error drawn from the Laplace and normal distributions, respectively. A comparison of two plots in a Figure indicates that the bandwidth at which $d^{(\infty)}_{j-1,j}$ changes suddenly (marked by a circle, a square, or star) is a good indicator of the bandwidth at which the true distance $L_\infty$ is about to stagnate.

5.3. **Simulation results.** Table 1 presents the empirical coverage probabilities of our bootstrap confidence bands. The simulated probabilities are generally close to the nominal confidence levels. As we expected, the coverage errors tend to be smaller when the sample size is larger or when the signal-to-ratio is larger.

Figures 3 and 4 depict some typical examples for the true cdf (CDF, $F_{X^*}$), deconvolution cdf estimate (ECDF, $\hat{F}_{X^*}$), and uniform confidence bands (CB), when the latent true distribution is standard normal and the measurement errors are drawn from Laplace and normal distributions. They show that the uniform confidence bands perform reasonably well even for small sample size $n = 100$ and the widths of the bands shrink substantially as the sample size increases from $n = 100$ to $n = 500$.

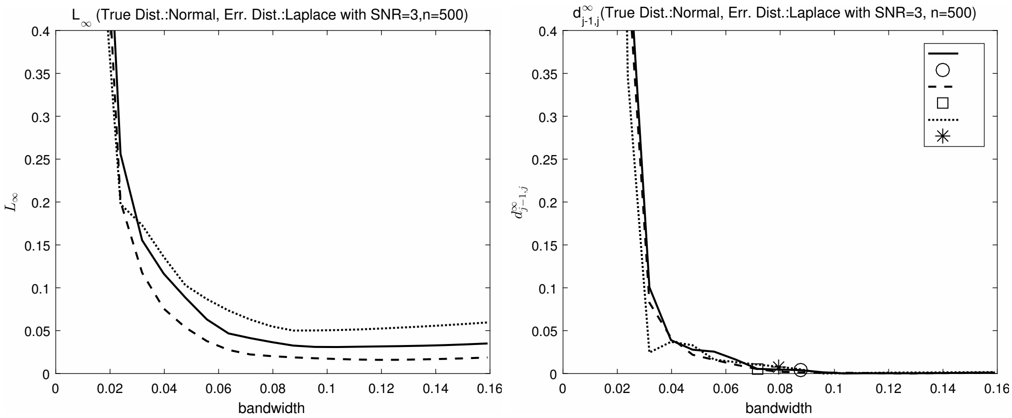


FIGURE 1. $L_\infty$ and $d^{\infty}_{j-1,j}$ distances under Laplace error

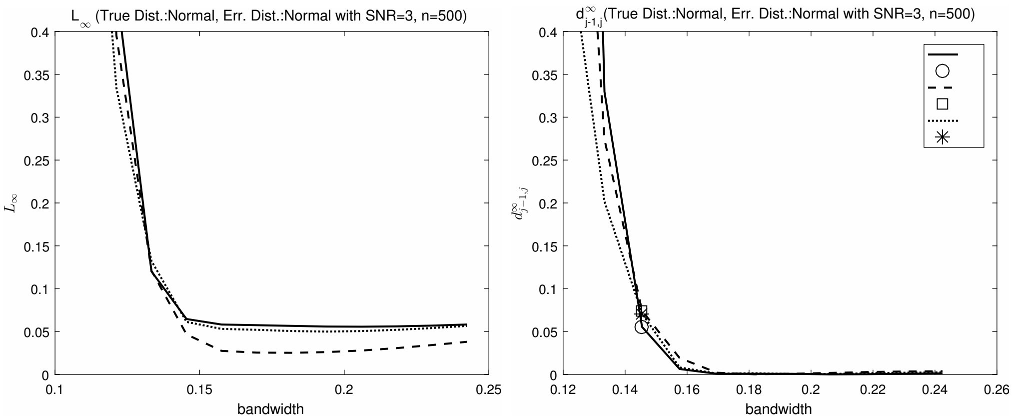


FIGURE 2. $L_\infty$ and $d^{\infty}_{j-1,j}$ distances under Normal error

| Level | $n$ | Laplace Error | | | Normal Error | | |
|---|---|---|---|---|---|---|---|
| | | SNR=2 | SNR=3 | SNR=4 | SNR=2 | SNR=3 | SNR=4 |
| 0.80 | 100 | 0.818 | 0.828 | 0.828 | 0.780 | 0.833 | 0.826 |
| | 250 | 0.811 | 0.818 | 0.823 | 0.790 | 0.803 | 0.810 |
| | 500 | 0.807 | 0.812 | 0.830 | 0.793 | 0.805 | 0.817 |
| | 1000 | 0.811 | 0.824 | 0.836 | 0.763 | 0.789 | 0.812 |
| 0.90 | 100 | 0.911 | 0.919 | 0.924 | 0.882 | 0.920 | 0.924 |
| | 250 | 0.897 | 0.913 | 0.916 | 0.888 | 0.899 | 0.903 |
| | 500 | 0.902 | 0.915 | 0.921 | 0.880 | 0.892 | 0.911 |
| | 1000 | 0.898 | 0.907 | 0.919 | 0.883 | 0.886 | 0.903 |
| 0.95 | 100 | 0.963 | 0.961 | 0.961 | 0.943 | 0.956 | 0.967 |
| | 250 | 0.957 | 0.958 | 0.963 | 0.938 | 0.947 | 0.956 |
| | 500 | 0.953 | 0.959 | 0.962 | 0.936 | 0.949 | 0.959 |
| | 1000 | 0.951 | 0.955 | 0.958 | 0.932 | 0.945 | 0.955 |

TABLE 1. Simulated uniform coverage probabilities for $F_{X^*}$ under Laplace and Normal errors.

## 6. Real data example

6.1. **Data description.** In this section, we apply the stochastic dominance test to the Korea Household Income and Expenditure Survey data to investigate welfare changes of different population sub-groups between 2006 and 2012. We use the data because the OECD report (2008)

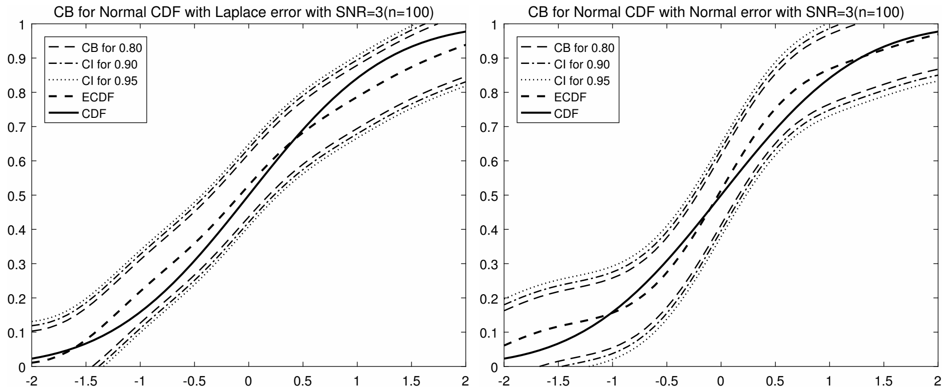


FIGURE 3. Uniform confidence bands under Laplace (left) and Normal (right) errors with $n = 100$

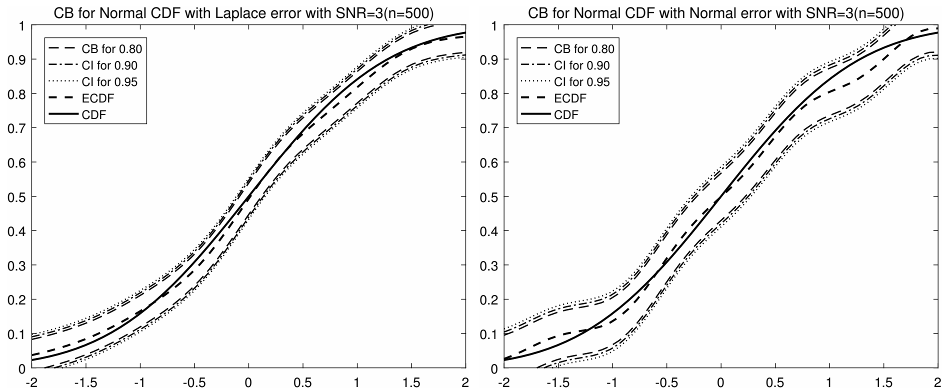


FIGURE 4. Uniform confidence bands under Laplace (left) and Normal (right) errors with $n = 500$

shows that, among OECD countries, Korea has the most significant variations in within-age group inequality and, compared to the inequality within the working age group, the relative inequality within the retirement age group is the worst. The data fit into our framework because it is well known that survey data are inherently affected by various sources of measurement errors, see Deaton (1997) and Bound, Brown and Mathiowetz (2000) for potential sources of measurement errors in household-based survey data. The survey reports incomes from various sources and consumption of goods and services for each household. We first compute the real household disposable income by adding all incomes, public pension, social benefits and transfers, minus tax, public pension premium and social security fees, after adjusting for inflation using the 2010 consumer price index. We then obtain the individualized data by adjusting the total household disposable income using the square-root equivalization scale, which is a common practice to approximate individual welfare.

Table 2 shows the descriptive statistics for the data. It shows that average real incomes of individuals in all age group except those over 70 have increased from 2006 to 2012. Standard deviations of all incomes have also slightly increased over the same period. The results are consistent with the finding of OECD (2008). However, unless the income distributions are normal, comparison of only the first two moments is not sufficient to draw a conclusion on the

| Year | Age Group | Sample Size | Mean | S.D. |
|---|---|---|---|---|
| 2006 | 25-45 | 12045 | 1,650 | 910 |
| | 45-65 | 8512 | 1,575 | 1,034 |
| | 60+ | 4605 | 1,047 | 862 |
| | 65+ | 3250 | 968 | 823 |
| | 70+ | 2050 | 944 | 823 |
| 2012 | 25-45 | 8722 | 1,800 | 910 |
| | 45-65 | 7653 | 1,814 | 1,106 |
| | 60+ | 5166 | 1,105 | 934 |
| | 65+ | 3700 | 974 | 879 |
| | 70+ | 2439 | 891 | 857 |

TABLE 2. Descriptive Statistics (Income unit: 1,000 won)

uniform ordering of nonparametric income distributions that does not depend on a specific social welfare function. This motivates us to consider a stochastic dominance criterion (see, e.g., Levy (2016)).

6.2. **Results.** We consider two different null hypotheses for each age group: (i) The 2006 income distribution first-order stochastically dominates that the 2012 income distribution (abbreviated to 06 FSD 12) (ii) The 2012 income distribution first-order stochastically dominates the 2006 income distribution (abbreviated to 12 FSD 06). As a benchmark test, we consider the Barrett and Donald (2003, BD)'s test based on the observed incomes, neglecting the presence of measurement errors. We choose the bandwidth as in our simulation experiments and assume Laplace and normal measurement errors. The variance of measurement errors is determined so that the signal-to-noise ratio (SNR) is 2,3, or 4.[3]

Table 3 reports the bootstrap p-values of the tests. The BD test implies that, for age groups 25-45 and 45-65, the 2012 income significantly dominates the 2006 income and, for age group 60+, there appears to be no dominance relationship (i.e. the two distributions cross), while for age group 70+ the 2006 income dominates the 2012 income. Similar results hold when we apply our test assuming Laplace measurement errors. However, when the measurement errors are normal, our test shows that, for age group 60+, there is a significant evidence that the 2012 income dominates the 2006 income. This implies that the ambiguous result (crossing of two distribution functions) for the age-group 60+ might be due to the presence of measurement errors in the observed data.

[3]In practice, as mentioned in Section 3, the error variance is generally not identified unless repeated measurements or extraneous information is available. However, in the case of the CPS income survey data, Bound and Krueger(1991) mentioned that "the error variance represents 27.6% of the total variance in CPS earnings for men and 8.9% for women." According to their remark, the signal-to-ratios are 1.9 for men and 3.35 for women, both of which lie in the range we considered.

| Age Group | Null Hypothesis | BD | Laplace Error | | | Normal Error | | |
|---|---|---|---|---|---|---|---|---|
| | | | SNR=2 | SNR=3 | SNR=4 | SNR=2 | SNR=3 | SNR=4 |
| 25-45 | 06 FSD 12 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 12 FSD 06 | 1.000 | 0.998 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 45-65 | 06 FSD 12 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 12 FSD 06 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 60+ | 06 FSD 12 | 0.000 | 0.037 | 0.023 | 0.013 | 0.000 | 0.000 | 0.000 |
| | 12 FSD 06 | 0.039 | 0.000 | 0.000 | 0.000 | 0.305 | 0.234 | 0.054 |
| 65+ | 06 FSD 12 | 0.353 | 0.400 | 0.652 | 0.704 | 0.143 | 0.240 | 0.189 |
| | 12 FSD 06 | 0.000 | 0.001 | 0.000 | 0.000 | 0.027 | 0.013 | 0.003 |
| 70+ | 06 FSD 12 | 0.928 | 0.501 | 0.934 | 0.988 | 0.664 | 0.698 | 0.715 |
| | 12 FSD 06 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.001 |

Table 3. Bootstrap P-values from BD and our tests

## References


[1] Barrett, G. F. and S. G. Donald (2003) Consistent tests for stochastic dominance, *Econometrica*, 71, 71-104.

[2] Bickel, P. J. and M. Rosenblatt (1973) On some global measures of the deviations of density function estimates, *Annals of Statistics*, 1, 1071-1095.

[3] Bissantz, N., Dümbgen, L., Holzmann, H. and A. Munk (2007) Non-parametric confidence bands in deconvolution density estimation, *Journal of the Royal Statistical Society*, B 69, 483-506.

[4] Bound, J., Brown, C. and N. Mathiowetz (2000) Measurement error in survey data, *Handbook of Econometrics*, vol. 5, chapter 59, Elsevier.

[5] Bound, J. and A. B. Krueger (1991) The extent of measurement error in longitudinal earnings data: Do two wrongs make a right?, *Journal of Labor Economics*, 9, 1-24.

[6] Chernozhukov, V., Chetverikov, D. and K. Kato (2014) Anti-concentration and honest, adaptive confidence bands, *Annals of Statistics*, 42, 1787-1818.

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

[8] Comte, F. and J. Kappus (2015) Density deconvolution from repeated measurements without symmetry assumption on the errors, *Journal of Multivariate Analysis*, 140, 31-46.

[9] Dattner, I., Goldenshluger, A. and A. Juditsky (2011) On deconvolution of distribution functions, *Annals of Statistics*, 39, 2477-2501.

[10] Dattner, I., Reiß, M. and M. Trabs (2016) Adaptive quantile estimation in deconvolution with unknown error distribution, *Bernoulli*, 22, 143-192.

[11] Deaton, A. (1997) *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*, Baltimore: Johns Hopkins University Press for the World Bank.

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

[13] Delaigle, A. and P. Hall (2006) On optimal kernel choice for deconvolution, *Statistics & Probability Letters*, 76, 1594-1602.

[14] Delaigle, A., Hall, P. and A. Meister (2008) On deconvolution with repeated measurements, *Annals of Statistics*, 36, 665-685.

[15] Efromovich, S. (1997) Density estimation for the case of supersmooth measurement error, *Journal of the American Statistical Association*, 92, 526-535.

[16] Fan, J. (1991) On the optimal rates of convergence for nonparametric deconvolution problems, *Annals of Statistics*, 19, 1257-1272.

[17] Hall, P. and S. N. Lahiri (2008) Estimation of distributions, moments and quantiles in deconvolution problems, *Annals of Statistics*, 36, 2110-2134.

[18] Hall, P. and A. Meister (2007) A ridge-parameter approach to deconvolution, *Annals of Statistics*, 35, 1535-1558.

[19] Kato, K. and Y. Sasaki (2016) Uniform confidence bands in deconvolution with unknown error distribution, Working paper.

[20] Komlós, J., Major, P. and G. Tusnády (1975) An approximation of partial sums of independent RV'-s and the sample DF. I, *Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete*, 32, 111-131.

[21] Lepski, O. V. (1990) A problem of adaptive estimation in Gaussian white noise, *Teor. Veroyatnost. i Primenen.*, 35, 459-470.

[22] Levy, H. (2016) *Stochastic Dominance*, Springer.

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

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

[25] Neumann, M.H (1997) On the effect of estimating the error density in nonparametric deconvolution, *Journal of Nonparametric Statistics*, 7, 307-330.

[26] OECD (2008) *Growing unequal? Income distribution and poverty in OECD countries*, Paris: OECD.

[27] Schennach, S. (2013) Convolution without independence, Cemmap Working Paper CWP46/13.

[28] Söhl, J. and M. Trabs (2012) A uniform central limit theorem and efficiency for deconvolution estimators, *Electronic Journal of Statistics*, 6, 2486-2518.

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

[30] van Es, A. J. and H.-W. Uh (2005) Asymptotic normality for kernel type deconvolution estimators, *Scandinavian Journal of Statistics*, 32, 467-483.

## Appendix A. Proofs of theorems

**Notation:** Hereafter, let $P^{\#}$ and $E^{\#}$ be the conditional probability and expectation under the bootstrap distribution given $\{X_i\}_{i=1}^n$, respectively. Also, denote

$$\begin{aligned}
\bar{\mathbb{L}}(u) &= \frac{1}{\pi}\int_0^1 \frac{\sin(\omega u)}{\omega}\frac{K^{\mathrm{ft}}(\omega)}{f_\epsilon^{\mathrm{ft}}(\omega/h)}\mathbb{I}\{|\omega|\geq h\omega_0\}d\omega,\\
\mathcal{G}_n(t) &= r(h)\int \bar{\mathbb{L}}\left(\frac{t-a}{h}\right) f_X(a)^{1/2}dW(a),\\
p_\varepsilon(\mathcal{G}_n) &= \sup_x P\left\{\left|\sup_{t\in\mathcal{T}}\mathcal{G}_n(t)-x\right|\leq\varepsilon\right\},
\end{aligned}$$

where $W$ is a (two-sided) Wiener process on $\mathbb{R}$, $f_X$ is the pdf of $X$, and

$$r(h) = \begin{cases} h^{\beta-\frac{1}{2}} & \text{under Assumption OS} \\ 1/\varsigma(h) & \text{under Assumption SS} \end{cases},$$

with $\varsigma(h)$ defined in eq. (8) of the paper. Note that analogous to $\bar{\mathbb{K}}$ (defined in Assumptions OS (ii) and SS (ii)), $\bar{\mathbb{L}}$ is considered as a trimmed version of $\mathbb{L}$. Due to the trimming, properties of the Fourier transform guarantee $\bar{\mathbb{L}} \in L_2(\mathbb{R})$ for each $h$ under the assumption $f_\epsilon^{\mathrm{ft}} \neq 0$, and this guarantees existence of the stochastic integral in the definition of $\mathcal{G}_n$.

Also, for any $a \in (0,1)$, let $c_a$ denote the constant such that $\sqrt{n}h^{\beta-\frac{1}{2}}c_a$ is the $(1-a)$-th quantile of $\sup_{t\in\mathcal{T}}|\mathcal{G}_n(t)|$.

A.1. **Proof of Theorem 1.** We only prove the statement under Assumption OS (i.e., the ordinary smooth case). The statement under Assumption SS is shown by a similar argument using Lemmas 4-6.

First, we prove

$$P\left\{\sqrt{n}h^{\beta-\frac{1}{2}}\hat{c}_\alpha > \sqrt{n}h^{\beta-\frac{1}{2}}c_{\alpha+\delta_{1n}}-\epsilon_{1n}\right\}\geq 1-\delta_{2n}, \tag{18}$$

for some $\epsilon_{1n},\delta_{1n},\delta_{2n} = O(n^{-c})$ with $c>0$. Lemma 2 implies that with probability greater than $1-\delta_{2n}$,

$$\begin{aligned}
1-\alpha &= P^{\#}\left\{\sqrt{n}h^{\beta-\frac{1}{2}}\sup_{t\in\mathcal{T}}\left|\hat{F}^{\#}_{X^*}(t)-\hat{F}_{X^*}(t)\right|\leq\sqrt{n}h^{\beta-\frac{1}{2}}\hat{c}_\alpha\right\}\\
&\leq P^{\#}\left\{\sup_{t\in\mathcal{T}}\left|\tilde{\mathcal{G}}_n(t)\right|\leq\sqrt{n}h^{\beta-\frac{1}{2}}\hat{c}_\alpha+\epsilon_{1n}\right\}+\delta_{1n},
\end{aligned}$$

for some $\epsilon_{1n},\delta_{1n},\delta_{2n} = O(n^{-c})$ with $c>0$, where $\tilde{\mathcal{G}}_n$ has the same distribution as $\mathcal{G}_n$ under $P^{\#}$. Since $\sqrt{n}h^{\beta-\frac{1}{2}}c_a$ is also the $(1-a)$-th quantile of $\sup_{t\in\mathcal{T}}\left|\tilde{\mathcal{G}}_n(t)\right|$ under $P^{\#}$, the above inequality implies

$$P^{\#}\left\{\sup_{t\in\mathcal{T}}\left|\tilde{\mathcal{G}}_n(t)\right|\leq\sqrt{n}h^{\beta-\frac{1}{2}}c_{\alpha+\delta_{1n}}\right\}\leq P^{\#}\left\{\sup_{t\in\mathcal{T}}\left|\tilde{\mathcal{G}}_n(t)\right|\leq\sqrt{n}h^{\beta-\frac{1}{2}}\hat{c}_\alpha+\epsilon_{1n}\right\},$$

with probability greater than $1-\delta_{2n}$. Thus, we obtain (18).

The main result is thus obtained from the following sequence of inequalities

$$
\begin{aligned}
P\{T_n \le \hat{c}_\alpha\} &\ge P\left\{\sup_{t\in\mathcal{T}} |\mathcal{G}_n(t)| \le \sqrt{n}h^{\beta-\frac{1}{2}}\hat{c}_\alpha - \epsilon_n\right\} - \delta_n \\
&\ge P\left\{\sup_{t\in\mathcal{T}} |\mathcal{G}_n(t)| \le \sqrt{n}h^{\beta-\frac{1}{2}}c_{\alpha+\delta_{1n}} - \epsilon_{1n} - \epsilon_n\right\} - \delta_n - \delta_{2n} \\
&\ge P\left\{\sup_{t\in\mathcal{T}} |\mathcal{G}_n(t)| \le \sqrt{n}h^{\beta-\frac{1}{2}}c_{\alpha+\delta_{1n}}\right\} - 2p_{\bar{\epsilon}_n}(\mathcal{G}_n) - \delta_n - \delta_{2n} \\
&= 1-\alpha-\delta_{1n} - 2p_{\bar{\epsilon}_n}(\mathcal{G}_n) - \delta_n - \delta_{2n} \\
&\ge 1-\alpha-\delta_{1n} - M\bar{\epsilon}_n\sqrt{\log(1/h)} - \delta_n - \delta_{2n},
\end{aligned}
$$

where the first inequality follows from Lemma 1, the second inequality follows from (18), the third inequality follows from the definitions of $\bar{\epsilon}_n = \epsilon_{1n} + \epsilon_n$ and $p_\varepsilon(\mathcal{G}_n)$, along with the fact $\mathcal{G}_n$ and $-\mathcal{G}_n$ have the same distribution (which ensures $p_\varepsilon(|\mathcal{G}_n|) \le 2p_\varepsilon(\mathcal{G}_n)$), the equality follows from the definition that $\sqrt{n}h^{\beta-\frac{1}{2}}c_{\alpha+\delta_{1n}}$ is the $(1-\alpha-\delta_{1n})$-th quantile of $\sup_{t\in\mathcal{T}}|\mathcal{G}_n(t)|$, and the last inequality follows from Lemma 3. Therefore, letting $\delta_{3n} = \delta_{1n} + M\bar{\epsilon}_n\sqrt{\log(1/h)} + \delta_n + \delta_{2n}$, we have

$$
P\{T_n \le \hat{c}_\alpha\} \ge 1-\alpha-\delta_{3n}.
$$

Since $\delta_n, \delta_{1n}, \delta_{2n}, \bar{\epsilon}_n$ are all positive sequences of order $O(n^{-a})$ with some $a > 0$ and $\sqrt{\log(1/h)}$ is a log-rate, we obtain eq. (11) in the paper.

A.2. **Proof of Theorem 3.** For simplicity, we restrict attention to the case of $N_i = 2$. For more general situations where $N_i$ is arbitrary but bounded above by $C$, the proof follows by similar arguments after accounting for the dependence structure in $\hat{f}_\epsilon^{\text{ft}}$.

We first make the following preliminary observations. Note that $\tilde{F}_{X^*}(t)$ can be alternatively written as

$$
\tilde{F}_{X^*}(t) = \frac{1}{2\pi}\int_{-1/h}^{1/h} \frac{\operatorname{Im}\{e^{\mathrm{i}\omega t}\hat{f}_X^{\text{ft}}(\omega)\}}{-\omega}\frac{K^{\text{ft}}(h\omega)}{\hat{f}_\epsilon^{\text{ft}}(\omega)}d\omega. \tag{19}
$$

where $\hat{f}_X^{\text{ft}}(\omega) = N^{-1}\sum_{i,j} e^{\mathrm{i}\omega X_{i,j}}$ denotes the empirical characteristic function. A similar expression holds for $\hat{F}_{X^*}$. Let $\xi = (f_\epsilon^{\text{ft}})^2$ and $\hat{\xi} = (\hat{f}_\epsilon^{\text{ft}})^2$. We note the following properties for $\hat{\xi}$

$$
E\left[\int_{\omega_0}^{h^{-1}} \omega^{-a}|\hat{\xi}(\omega) - \xi(\omega)|^2 d\omega\right] = \begin{cases} n^{-1}h^{-(1-a)} & \text{if } a < 1 \\ n^{-1} & \text{if } a > 1 \\ n^{-1}\log(1/h) & \text{if } a = 1 \end{cases} \tag{20}
$$

$$
\sup_{|\omega|\le h^{-1}} |\xi/\hat{\xi}| \le 1 + o_p(1). \tag{21}
$$

The results in (20) can be shown by expanding the expectations. To show (21), we use Yukich (1987, Theorem 6.3) which assures that under Assumption B (i), $\sup_{|\omega|\le h^{-1}}|\hat{\xi} - \xi| = O_p(\sqrt{\log n/n})$ for $h = O(n^{-c})$ with some $c > 0$. Combined with Assumption B (ii), this implies

$\{\min_{|\omega|\leq h^{-1}}|\hat{\xi}|\}^{-1} = O_p(h^{-2\beta})$. Thus we obtain

$$\sup_{|\omega|\leq h^{-1}}|\xi/\hat{\xi}| \leq 1 + \sup_{|\omega|\leq h^{-1}}|(\hat{\xi}-\xi)/\hat{\xi}| = 1 + O_p\left(\left(\frac{\log n}{nh^{4\beta}}\right)^{1/2}\right) = 1 + o_p(1),$$

thereby proving (21).

Pick any $\eta \in (1/2, \gamma - \beta)$. Under Assumptions C (iii) and OS (i), it can be verified that

$$\int_{-1/h}^{1/h}\left|\frac{\omega^\eta f_{X^*}^{\mathrm{ft}}(\omega)}{\xi(\omega)^{1/2}}\right|^2 d\omega = O(1). \tag{22}$$

We shall also make frequent use of the following algebraic inequality:

$$|\hat{\xi}^{1/2} - \xi^{1/2}| \leq \xi^{-1/2}|\hat{\xi} - \xi|. \tag{23}$$

We now proceed to the main part of the proof. By (19), we can expand

$$\begin{aligned}
\tilde{F}_{X^*}(t) - \hat{F}_{X^*}(t) &= \frac{1}{\pi}\int_0^{\omega_0}\frac{\mathrm{Im}\{e^{-\mathrm{i}\omega t}\hat{f}_X^{\mathrm{ft}}(\omega)\}}{-\omega}\{\hat{\xi}(\omega)^{-1/2} - \xi(\omega)^{-1/2}\}K^{\mathrm{ft}}(h\omega)d\omega \\
&\quad + \frac{1}{\pi}\int_{\omega_0}^{1/h}\frac{\mathrm{Im}\{e^{-\mathrm{i}\omega t}\hat{f}_X^{\mathrm{ft}}(\omega)\}}{-\omega}\{\hat{\xi}(\omega)^{-1/2} - \xi(\omega)^{-1/2}\}K^{\mathrm{ft}}(h\omega)d\omega \\
&= B_{1n}(t) + B_{2n}(t).
\end{aligned}$$

For the term $B_{1n}(t)$, using (23), we have

$$|B_{1n}(t)| \leq \frac{1}{\pi}\int_0^{\omega_0}\left|\frac{\mathrm{Im}\{e^{-\mathrm{i}\omega t}\hat{f}_X^{\mathrm{ft}}(\omega)\}}{-\omega}\right|\left|\frac{\xi(\omega)}{\hat{\xi}(\omega)}\right|^{1/2}\frac{|\hat{\xi}(\omega) - \xi(\omega)|}{\xi(\omega)^{3/2}}d\omega.$$

By the fact $\sup_{|\omega|\leq\omega_0}|\hat{\xi} - \xi| = O_p(n^{-1/2})$ and (21), we obtain

$$\sup_{t\in\mathcal{T}}|B_{1n}(t)| = O_p(n^{-1/2})\sup_{t\in\mathcal{T}}I(t),$$

where

$$\begin{aligned}
I(t) &= \int_0^{\omega_0}\left|\frac{\mathrm{Im}\{e^{-\mathrm{i}\omega t}\hat{f}_X^{\mathrm{ft}}(\omega)\}}{-\omega}\right|d\omega \\
&\leq \int_0^{\omega_0}\left|\frac{\sin(\omega t)}{\omega}\mathrm{Re}\{\hat{f}_X^{\mathrm{ft}}(\omega)\}\right|d\omega + \int_0^{\omega_0}\left|\frac{\cos(\omega t)}{\omega}\mathrm{Im}\{\hat{f}_X^{\mathrm{ft}}(\omega)\}\right|d\omega \\
&\leq \int_0^{\omega_0}\left|\frac{\sin(\omega t)}{\omega}\right|d\omega + \int_0^{\omega_0}\left|\frac{\mathrm{Im}\{\hat{f}_X^{\mathrm{ft}}(\omega)\}}{\omega}\right|d\omega \\
&= I_1(t) + I_2.
\end{aligned}$$

Since $\mathcal{T}$ is a compact set, it holds $\sup_{t\in\mathcal{T}} I_1(t) < \infty$. By the definition of $\hat{f}_X^{\mathrm{ft}}$, the random variable $I_2$ can be bounded as

$$I_2 \leq \frac{1}{N}\sum_{i,j}\int_0^{\omega_0}\left|\frac{\sin(\omega X_{i,j})}{\omega}\right|d\omega \equiv \frac{1}{N}\sum_{i,j}T_{i,j}.$$

Since

$$E[T_{i,j}] = E\int_0^{\omega_0|X_{i,j}|}\left|\frac{\sin(t)}{t}\right|dt \leq C_1 + E[\log|X_{i,j}|] < \infty$$

for some $C_1 > 0$, it holds $I_2 = O_p(1)$. Combining these results, we obtain $\sup_{t\in\mathcal{T}} |B_{1n}(t)| = O_p(n^{-1/2})$.

For the term $B_{2n}(t)$, we further expand

$$\begin{aligned} B_{2n}(t) &= -\frac{1}{\pi}\int_{\omega_0}^{1/h} \frac{\mathrm{Im}\{e^{-\mathrm{i}\omega t} f_X^{\mathrm{ft}}(\omega)\}}{-\omega\xi(\omega)} \{\hat{\xi}(\omega)^{1/2} - \xi(\omega)^{1/2}\} K^{\mathrm{ft}}(h\omega) \frac{\xi(\omega)^{1/2}}{\hat{\xi}(\omega)^{1/2}} d\omega \\ &\quad + \frac{1}{\pi}\int_{\omega_0}^{1/h} \frac{\mathrm{Im}\left\{e^{-\mathrm{i}\omega t}\{\hat{f}_X^{\mathrm{ft}}(\omega) - f_X^{\mathrm{ft}}(\omega)\}\right\}}{-\omega} \{\hat{\xi}(\omega)^{-1/2} - \xi(\omega)^{-1/2}\} K^{\mathrm{ft}}(h\omega) d\omega \\ &= B_{21n}(t) + B_{22n}(t). \end{aligned}$$

For the term $B_{21n}(t)$, we have

$$\begin{aligned} \sup_{t\in\mathcal{T}} |B_{21n}(t)| &\leq \frac{1}{\pi}\int_{\omega_0}^{1/h} \left|\frac{\omega^\eta f_{X^*}^{\mathrm{ft}}(\omega)}{\xi(\omega)^{1/2}}\right| \left|\frac{\hat{\xi}(\omega) - \xi(\omega)}{\omega^{1+\eta}\xi(\omega)^{1/2}}\right| \left|\frac{\xi(\omega)}{\hat{\xi}(\omega)}\right|^{1/2} d\omega \\ &\leq C_2(1 + o_p(1)) \left(\int_{\omega_0}^{1/h} \left|\frac{\omega^\eta f_{X^*}^{\mathrm{ft}}(\omega)}{\xi(\omega)^{1/2}}\right|^2 d\omega\right)^{1/2} \left(\int_{\omega_0}^{1/h} \omega^{2(\beta-\eta-1)} |\hat{\xi}(\omega) - \xi(\omega)|^2 d\omega\right)^{1/2} \\ &= O(n^{-1/2} h^{(\eta-\beta+1/2)\wedge 0}), \end{aligned}$$

for some $C_2 > 0$, where the first inequality follows from the fact $|\mathrm{Im}\{e^{-\mathrm{i}\omega t} f_X^{\mathrm{ft}}(\omega)\}| \leq |f_X^{\mathrm{ft}}(\omega)| = |f_{X^*}^{\mathrm{ft}}(\omega)|\xi(\omega)^{1/2}$ and (23), the second inequality follows from (21) and Assumption OS (i), and the equality follows from (20) and (22).

Now consider the term $B_{22n}(t)$. Applying (23) and Assumption OS (i), we can write

$$\begin{aligned} \sup_{t\in\mathcal{T}} |B_{22n}(t)| &\leq \frac{1}{\pi}\int_{\omega_0}^{1/h} |\hat{f}_X^{\mathrm{ft}}(\omega) - f_X^{\mathrm{ft}}(\omega)||\hat{\xi}(\omega) - \xi(\omega)||\xi(\omega)/\hat{\xi}(\omega)|^{1/2} \frac{1}{\omega\xi(\omega)^{3/2}} d\omega \\ &\leq \frac{1}{c^3\pi}\int_{\omega_0}^{1/h} |\hat{f}_X^{\mathrm{ft}}(\omega) - f_X^{\mathrm{ft}}(\omega)||\hat{\xi}(\omega) - \xi(\omega)||\xi(\omega)/\hat{\xi}(\omega)|^{1/2} \omega^{3\beta-1} d\omega, \end{aligned}$$

for some $c > 0$. As in (22), it can be shown after expanding the expectation that

$$E\left[\int_{\omega_0}^{1/h} \omega^{-a} |\hat{f}_X^{\mathrm{ft}}(\omega) - f_X^{\mathrm{ft}}(\omega)|^2 d\omega\right] = O((nh^{1-a})^{-1}),$$

for all $a < 1$. Thus, by (21) and (22), it follows

$$\begin{aligned} \sup_{t\in\mathcal{T}} |B_{22n}(t)| &\leq \frac{1 + o_p(1)}{\pi}\int_{\omega_0}^{1/h} \omega^{3\beta-1} |\hat{f}_X^{\mathrm{ft}}(\omega) - f_X^{\mathrm{ft}}(\omega)||\hat{\xi}(\omega) - \xi(\omega)| d\omega \\ &= \frac{1 + o_p(1)}{\pi}\left(\int_{\omega_0}^{1/h} \omega^{3\beta-1} |\hat{f}_X^{\mathrm{ft}}(\omega) - f_X^{\mathrm{ft}}(\omega)|^2 d\omega\right)^{1/2} \left(\int_{\omega_0}^{1/h} \omega^{3\beta-1} |\hat{\xi}(\omega) - \xi(\omega)|^2 d\omega\right)^{1/2} \\ &= O((nh^{3\beta})^{-1}). \end{aligned}$$

Combining these results, we obtain

$$\sqrt{n} h^{\beta-1/2} \sup_{t\in\mathcal{T}} |\tilde{F}_{X^*}(t) - \hat{F}_{X^*}(t)| = O_p\left(h^{\eta\wedge(\beta-1/2)} + \frac{1}{\sqrt{n} h^{2\beta+1/2}}\right) = o_p(1),$$

under Assumption B (ii) and the condition $\eta > 1/2$.

A.3. **Proof of Theorem 4.** Define

$$\hat{D}_n^{\#}(t) = \sqrt{n}h^{\beta-1/2}\{\hat{F}_{X^*}^{\#}(t) - \hat{F}_{X^*}(t)\}, \qquad \tilde{D}_n^{\#}(t) = \sqrt{n}h^{\beta-1/2}\{\tilde{F}_{X^*}^{\#}(t) - \tilde{F}_{X^*}(t)\}.$$

Also, let $\hat{f}_X^{\mathrm{ft}\#}(\omega) = N^{-1}\sum_{i,j} e^{\mathrm{i}\omega X_{i,j}^{\#}}$ be the bootstrap counterpart of the empirical characteristic function $\hat{f}_X^{\mathrm{ft}}(\omega) = N^{-1}\sum_{i,j} e^{\mathrm{i}\omega X_{i,j}}$.

We first show that there exist $c, C > 0$ such that

$$P^{\#}\left\{\sup_{t\in\mathcal{T}} |\tilde{D}_n^{\#}(t) - \hat{D}_n^{\#}(t)| \geq Cn^{-c}\right\} = o_p(1). \tag{24}$$

By Theorem 3, it is enough for (24) to guarantee that there exist $c, C > 0$ satisfying

$$P^{\#}\left\{\sqrt{n}h^{\beta-1/2}\sup_{t\in\mathcal{T}} |\tilde{F}_{X^*}^{\#}(t) - \hat{F}_{X^*}^{\#}(t)| \geq Cn^{-c}\right\} = o_p(1).$$

To this end, note that

$$\begin{aligned}
\tilde{F}_{X^*}^{\#}(t) - \hat{F}_{X^*}^{\#}(t) &= \frac{1}{2\pi}\int_{-1/h}^{1/h} \frac{\mathrm{Im}\left\{e^{-\mathrm{i}\omega t}\left\{\hat{f}_X^{\mathrm{ft}\#}(\omega) - \hat{f}_X^{\mathrm{ft}}(\omega)\right\}\right\}}{-\omega}\{\hat{\xi}(\omega)^{-1/2} - \xi(\omega)^{-1/2}\}K^{\mathrm{ft}}(h\omega)d\omega \\
&\quad + \frac{1}{2\pi}\int_{-1/h}^{1/h} \frac{\mathrm{Im}\{e^{-\mathrm{i}\omega t}\hat{f}_X^{\mathrm{ft}}(\omega)\}}{-\omega}\{\hat{\xi}(\omega)^{-1/2} - \xi(\omega)^{-1/2}\}K^{\mathrm{ft}}(h\omega)d\omega \\
&= C_{1n}(t) + C_{2n}(t).
\end{aligned}$$

The second term $C_{2n}(t)$ equals to $\tilde{F}_{X^*}(t) - \hat{F}_{X^*}(t)$ whose bound is given in Theorem 3. Thus, we only need to consider the first term $C_{1n}(t)$. By expanding the expectations, it can be shown

$$E^{\#}\left[\int_{\omega_0}^{1/h} \omega^{-a}|\hat{f}_X^{\mathrm{ft}\#}(\omega) - \hat{f}_X^{\mathrm{ft}}(\omega)|^2 d\omega\right] = O_p((nh^{1-a})^{-1}),$$

for all $a < 1$, and analogous arguments as in the proof of Theorem 3 yield $\sup_{t\in\mathcal{T}}|C_{1n}(t)| = O_{p^{\#}}((nh^{3\beta})^{-1})$ with probability approaching one. Therefore, by paralleling the arguments in the proof of Theorem 3, we obtain (24).

We now proceed by verifying the conditions in the proof of Theorem 1. Lemma 1 and Theorem 3 ensure existence of a sequence $\epsilon_n = O(n^{-c})$ with some $c > 0$ such that

$$P\left\{\sup_{t\in\mathcal{T}}\left|\sqrt{n}h^{\beta-1/2}\{\tilde{F}_X(t) - F_X(t)\} - \mathcal{G}_n(t)\right| > \epsilon_n\right\} = o_p(1). \tag{25}$$

Furthermore by Lemma 2, combined with (24), we have that

$$P^{\#}\left\{\sup_{t\in\mathcal{T}}\left|\sqrt{n}h^{\beta-1/2}\{\tilde{F}_{X^*}^{\#}(t) - \tilde{F}_{X^*}(t)\} - \tilde{\mathcal{G}}_n(t)\right| > \epsilon_n\right\} = o_p(1). \tag{26}$$

Therefore, by (25) and (26), the conclusion follows by paralleling the arguments in the proof of Theorem 1.

A.4. **Proof of Theorem 5.** We only prove the theorem under Assumption OS (i.e., the ordinary smooth case). The proof under Assumption SS follows by a similar argument using Lemmas 4-6.

We make the following preliminary observations. First, by the techniques employed in Lemmas 1-3, we can show[4]

$$\sup_{t\in\mathcal{H}} |\hat{f}_{X^*}(t) - f_{X^*}(t)| = O_p(n^{-c}). \tag{27}$$

Next by Dattner, Reiß and Trabs (2016, Proposition 2.1), $\left\|\hat{f}_{X^*}\right\|_1 < \infty$ and $\int_{-\infty}^{\infty} \hat{f}_{X^*}(t)dt = 1$ under Assumption C. Thus, we have $\hat{F}_{X^*}(t) = \int_{-\infty}^{t} \hat{f}_{X^*}(v)dv$ or equivalently $\hat{F}'_{X^*}(t) = \hat{f}_{X^*}(t)$. The latter ensures $\hat{F}_{X^*}$ is continuous.

We now show that[5]

$$\sup_{u\in[u_1,u_2]} |\hat{Q}(u) - Q(u)| = o_p(n^{-c_1}), \tag{28}$$

for some $c_1 > 0$. By Hall and Lahiri (2008, Theorem 3.7), $\hat{Q}(u)$ converges to $Q(u)$ for each $u \in [u_1, u_2]$. Now $Q_n(u)$ is monotone at each $n$ by construction while $Q(u)$ is continuous by Assumption Q (i). Hence we can modify the proof of the Glivenko-Cantelli theorem (see, Billingsley, 1995, p. 233), to strengthen the pointwise consistency to a uniform one, i.e.,

$$\sup_{u\in[u_1,u_2]} |\hat{Q}(u) - Q(u)| = o_p(1), \tag{29}$$

(see also, Bassett and Koenker, 1986, Theorem 3.1). As $\hat{F}_{X^*}$ is continuous, it follows that $\hat{F}_{X^*}(\hat{Q}(u)) = u$ for all $0 < u < 1$. Consequently,

$$\hat{F}_{X^*}(\hat{Q}(u)) = F_{X^*}(Q(u)) = F_{X^*}(\hat{Q}(u)) - f_{X^*}(\tilde{Q}(u))(\hat{Q}(u) - Q(u)),$$

for some $\tilde{Q}(u)$ such that $|\tilde{Q}(u) - Q(u)| \leq |\hat{Q}(u) - Q(u)|$, and we obtain

$$\sup_{u\in[u_1,u_2]} |\hat{Q}(u) - Q(u)| \leq \left(\inf_{u\in[u_1,u_2]} |f_{X^*}(\tilde{Q}(u))|\right)^{-1} \sup_{u\in[u_1,u_2]} |\hat{F}_{X^*}(\hat{Q}(u)) - F_{X^*}(\hat{Q}(u))|$$

By (29) and Assumption Q (i) ($\inf_{x\in\mathcal{H}} f_{X^*}(x) > 0$), we can verify $\inf_{u\in[u_1,u_2]} |f_{X^*}(\tilde{Q}(u))| > 0$ with probability approaching one. Furthermore, we have

$$\sup_{u\in[u_1,u_2]} |\hat{F}_{X^*}(\hat{Q}(u)) - F_{X^*}(\hat{Q}(u))| \leq n^{-\frac{1}{2}} h^{-\beta+\frac{1}{2}} \sup_{t\in\mathcal{H}} |\mathcal{G}_n(t)| + o_p(1) = O_p\left(\left(\frac{\log(1/h)}{nh^{2\beta-1}}\right)^{1/2}\right),$$

where the inequality follows from Lemma 1 after employing the fact $\{\hat{Q}(u) : u \in [u_1, u_2]\} \subset \mathcal{H}$ with probability approaching one due to Assumption Q (i) and (29). The equality follows from $E[\sup_{t\in\mathcal{H}} |\mathcal{G}_n(t)|] = O(\sqrt{\log(1/h)})$ (by the proof of Lemma 3). Combining these results, we obtain (28) under Assumptions OS (iii) and B (ii).

[4] An analogous result applies for the super smooth case by Lemmas 4-6 with the rate replaced by $O_p((\log n)^{-c})$ for some $c > 1$ under the assumption $\gamma > \lambda$ and an MSE optimal bandwidth choice.

[5] For the super smooth case, we can employ similar arguments to show that $\sup_{u\in[u_1,u_2]} |\hat{Q}(u) - Q(u)| = o_p((\log n)^{-c_1})$ for some $c_1 > 1$.

We now proceed to the main part of the proof. Noting that $\hat{Q}(u)-Q(u) = f_{X^*}(\tilde{Q}(u))^{-1}\{\hat{F}_{X^*}(\hat{Q}(u))-F_{X^*}(\hat{Q}(u))\}$, we have

$$\begin{aligned}
& P\left\{\hat{Q}(u) - \frac{\hat{c}_\alpha}{\hat{f}_{X^*}(\hat{Q}(u))} \leq Q(u) \leq \hat{Q}(u) + \frac{\hat{c}_\alpha}{\hat{f}_{X^*}(\hat{Q}(u))} \quad \text{for all } u \in [u_1, u_2]\right\} \\
= \; & P\left\{\sup_{u\in[u_1,u_2]} |\hat{f}_{X^*}(\hat{Q}(u))\{\hat{Q}(u) - Q(u)\}| \leq \hat{c}_\alpha\right\} \geq P\left\{\sup_{t\in\mathcal{H}} |\hat{F}_{X^*}(t) - F_{X^*}(t)| \leq \hat{c}_\alpha(1-\Delta_n)\right\} - o(1),
\end{aligned}$$

where $\Delta_n = \sup_{u\in[u_1,u_2]}\left|\frac{\hat{f}_{X^*}(\hat{Q}(u))-f_{X^*}(\tilde{Q}(u))}{\hat{f}_{X^*}(\hat{Q}(u))}\right|$ and the inequality follows from the fact $P\left\{\{\hat{Q}(u) : u \in [u_1, u_2]\} \subset \mathcal{H}\right\} \to 1$ by Assumption Q (i) and (28). Also note that $\Delta_n = O_p(n^{-c})$ by Assumption Q (i)-(ii), (27), and (28). We now have the following sequence of inequalities

$$\begin{aligned}
& P\left\{\sup_{t\in\mathcal{H}} |\hat{F}_{X^*}(t) - F_{X^*}(t)| \leq \hat{c}_\alpha(1-\Delta_n)\right\} \geq P\left\{\sup_{t\in\mathcal{H}} |\mathcal{G}_n(t)| \leq \sqrt{n}h^{\beta-\frac{1}{2}}\hat{c}_\alpha(1-\Delta_n) - \epsilon_n\right\} - \delta_n \\
\geq \; & P\left\{\sup_{t\in\mathcal{H}} |\mathcal{G}_n(t)| \leq \left(\sqrt{n}h^{\beta-\frac{1}{2}}c_{\alpha+\delta_{1n}} - \epsilon_{1n}\right)(1-\Delta_n) - \epsilon_n\right\} - \delta_n - \delta_{2n} \\
\geq \; & P\left\{\sup_{t\in\mathcal{H}} |\mathcal{G}_n(t)| \leq \sqrt{n}h^{\beta-\frac{1}{2}}c_{\alpha+\delta_{1n}}\right\} - 2p_{\bar{\epsilon}_n}(\mathcal{G}_n) - \delta_n - \delta_{2n} \geq 1 - \alpha - \delta_{1n} - \delta_n - \delta_{2n} - 2p_{\bar{\epsilon}_n}(\mathcal{G}_n),
\end{aligned}$$

where the first inequality follows from Lemma 1, the second inequality can be derived by Lemma 2 and a similar argument in the proof of Theorem 1, the third inequality follows from the definitions of $\bar{\epsilon}_n = \epsilon_n + \epsilon_{1n}(1-\Delta_n) + \sqrt{n}h^{\beta-\frac{1}{2}}c_{\alpha+\delta_{1n}}\Delta_n$ and the concentration function. Note that Lemma 3 implies $p_{\bar{\epsilon}_n}(\mathcal{G}_n) \leq C\bar{\epsilon}_n\sqrt{\log n}$. Recalling that $\sqrt{n}h^{\beta-\frac{1}{2}}c_{\alpha+\delta_{1n}}$ is the $(\alpha+\delta_{1n})$-th quantile of $\sup_{t\in\mathcal{H}}|\mathcal{G}_n(t)|$, by Chernozhukov, Chetverikov and Kato (2014, Lemma B1),

$$\sqrt{n}h^{\beta-\frac{1}{2}}c_{\alpha+\delta_{1n}} \leq E\left[\sup_{t\in\mathcal{H}} |\mathcal{G}_n(t)|\right] + \sqrt{2|\log(\alpha+\delta_{1n})|}.$$

Since $E[\sup_{t\in\mathcal{H}}|\mathcal{G}_n(t)|] = O(\sqrt{\log(1/h)})$, this implies $\sqrt{n}h^{\beta-\frac{1}{2}}c_{\alpha+\delta_{1n}} = O(\sqrt{\log n})$ under Assumptions OS (iii) and B (ii). By the above and the rates of $\epsilon_n, \epsilon_{1n}$, it follows $p_{\bar{\epsilon}_n}(\mathcal{G}_n) = O_p(n^{-c_2})$ for some $c_2 > 0$. Furthermore, by Lemmas 1 and 2, $\delta_n$, $\delta_{1n}$, and $\delta_{2n}$ are also $O(n^{-c_3})$ for some $c_3 > 0$. Combining these results, the conclusion follows.

A.5. **Proof of Theorem 6.** We shall assume for simplicity that $f_\epsilon = f_\delta$, and consequently that the bandwidth choices for both estimators are the same. We only prove for the case of ordinary smooth error density as the proof for super-smooth density follows by the same arguments. Assume that that the smoothness parameter in the former case is $\beta$. Let

$$\mathcal{G}^D_{n,m}(t) = h^{\beta-1/2}\left\{\int \bar{\mathbb{L}}\left(\frac{t-a}{h}\right) f_X(a)^{1/2}dW_1(a) - \sqrt{\frac{n}{m}}\int \bar{\mathbb{L}}\left(\frac{t-a}{h}\right) f_Y(a)^{1/2}dW_2(a)\right\},$$

where $W_1$ and $W_2$ are two independent (two-sided) Wiener processes on $\mathbb{R}$ (for $f_\epsilon \neq f_\delta$ or unequal bandwidths, the $\bar{\mathbb{L}}$ functions in the above integrals would also be different). Also define

$$\begin{aligned}
\Psi_{n,m}(t) &= \{\hat{F}_{X^*}(t) - F_{X^*}(t)\} - \{\hat{F}_{Y^*}(t) - F_{Y^*}(t)\}, \\
\Psi^{\#}_{n,m}(t) &= \{\hat{F}^{\#}_{X^*}(t) - \hat{F}_{X^*}(t)\} - \{\hat{F}^{\#}_{Y^*}(t) - \hat{F}_{Y^*}(t)\}.
\end{aligned}$$

A.5.1. *Proof of (i).* Since the samples $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^m$ are independent of each other, by the arguments of Lemmas (1)-(3), we can show the following: For some sequences $\epsilon_n, \delta_n = O(n^{-c})$,

$$P\left\{\sup_{t\in\mathcal{T}}\left|\sqrt{n}h^{\beta-1/2}\Psi_{n,m}(t) - \mathcal{G}_{n,m}^D(t)\right| > \epsilon_n\right\} < \delta_n. \quad (30)$$

Furthermore with probability greater than $1-\delta_{2n}$, $\delta_{2n} = O(n^{-c})$, there exist sequences $\epsilon_{1n}, \delta_{1n} = O(n^{-c})$ such that

$$P^{\#}\left\{\sup_{t\in\mathcal{T}}\left|\sqrt{n}h^{\beta-1/2}\Psi_{n,m}^{\#}(t) - \tilde{\mathcal{G}}_{n,m}^{D\#}(t)\right| > \epsilon_{1n}\right\} < \delta_{1n}, \quad (31)$$

where $\tilde{\mathcal{G}}_{n,m}^{D\#}$ is a tight Gaussian process with the same distribution as $\mathcal{G}_{n,m}^D$ under $P^{\#}$. Finally it also holds that

$$p_{\epsilon_n}(\mathcal{G}_{n,m}^D) \le M\epsilon_n\sqrt{\log(1/h)}, \quad (32)$$

for any sequence $\epsilon_n = O(n^{-c})$ and some $M < \infty$. Now

$$P\left\{D_{n,m} \le \hat{c}_\alpha^D\right\} \ge P\left\{\sup_{t\in\mathcal{T}}\Psi_{n,m}(t) - \sup_t\{F_{X^*}(t) - F_{Y^*}(t)\} \le \hat{c}_\alpha^D\right\} \ge P\left\{\sup_{t\in\mathcal{T}}\Psi_{n,m}(t) \le \hat{c}_\alpha^D\right\},$$

where the last equality follows from $\sup_t\{F_{X^*}(t) - F_{Y^*}(t)\} \le 0$ under $H_0$. Using equations (30)-(32), by paralleling the arguments in the proof of Theorem 1, we can show that

$$P\left\{\sup_{t\in\mathcal{T}}\Psi_{n,m}(t) \le \hat{c}_\alpha^D\right\} \ge 1 - \alpha - \varrho_{n,m}.$$

Hence the claim follows immediately.

A.5.2. *Proof of (ii).* It is enough to show that $\rho_{n,m}$ does not depend on $P \in \mathcal{P}_0$. To this end, it is enough to show uniform validity of equations (30)-(32). Since these equations are essentially two-sample counterparts of Lemmas (1)-(3), it suffices to check uniform validity of the latter.

Note that for Lemma (1), uniformity of the bias term follows by the argument in Hall and Lahiri (2008, Theorem 3.2) using the uniform version of the Sobolev condition (i.e. the constants $M_X$ and $M_Y$ do not depend of $(F_{X^*}, F_{Y^*})$). For the stochastic term, the constants appearing in the KMT coupling in the proof of Lemma (1) are universal, and constants and sequences in other parts do not depend on $P \in \mathcal{P}_0$. Thus, $\delta_n$ in Lemma (1) does not depend on $P \in \mathcal{P}_0$. Similarly, uniformity of Lemma (2) is also verified.

For Lemma (3), it is enough to guarantee that $\sigma_n(t)$ is bounded away from zero and above by universal constants that do not depend on $P \in \mathcal{P}_0$. This is guaranteed by the assumption that $f_X$ and $f_Y$ are bounded away from zero and above by universal constants that do not depend on $P \in \mathcal{P}_0$.

A.5.3. *Proof of (iii).* Let $c_a^D$ be a constant such that $\sqrt{n}h^{\beta-1/2}c_a^D$ is the $(1-a)$-th quantile of $\sup_{t\in\mathcal{T}}\mathcal{G}_{n,m}^D(t)$. Using equation (31) and mirroring the arguments in the proof of Theorem 1, we have that

$$P\left\{\sqrt{n}h^{\beta-1/2}\hat{c}_\alpha^D < \sqrt{n}h^{\beta-1/2}c_{\alpha-\delta_{1n}}^D + \epsilon_{1n}\right\} \ge 1 - \delta_{2n}. \quad (33)$$

Under $H_1$, there exists $t^* \in \mathcal{T}$ such that $\mu = F_{X^*}(t^*) - F_{Y^*}(t^*) > 0$. Then we obtain

$$\begin{aligned} P\{D_{n.m} > \hat{c}_\alpha^D\} &\geq P\left\{\sqrt{n}h^{\beta-1/2}D_{n,m} > \sqrt{n}h^{\beta-1/2}c^D_{\alpha-\delta_{1n}} + \epsilon_{1n}\right\} - \delta_{2n} \\ &\geq P\left\{\mathcal{G}^D_{n,m}(t^*) > \sqrt{n}h^{\beta-1/2}c^D_{\alpha-\delta_{1n}} - \sqrt{n}h^{\beta-\frac{1}{2}}\mu + \epsilon_{1n} + \epsilon_n\right\} - \delta_{2n} - \delta_n, \end{aligned}$$

for some $\epsilon_n, \delta_n = O(n^{-c'})$ with some $c' > 0$, where the first inequality follows from (33) and the second inequality follows from (30). By analogous arguments as in the proof of Theorem 5, we can show $\sqrt{n}h^{\beta-1/2}c^D_{\alpha-\delta_{1n}} = O(\sqrt{\log(1/h)})$. However under Assumption OS (iii), $\sqrt{n}h^{\beta-1/2}\log^{-1/2}(1/h)\mu \to +\infty$; hence the conclusion follows immediately.

## Appendix B. Lemmas

Hereafter we use the following notation. By the Ito isometry, the variance function of the Gaussian process $\mathcal{G}_n$ can be shown to be

$$\sigma_n(t) = hr^2(h) \int \bar{\mathbb{L}}^2(a) f_X(t - ha)da.$$

Let $\bar{\sigma}_n = \sup_t \sigma_n(t)$ and $\underline{\sigma}_n = \inf_t \sigma_n(t)$. Assumption C (i) ($\inf_{t\in\mathcal{T}} f_X(t) > c > 0$) guarantees that $\underline{\sigma}_n > 0$ for all $n \in \mathbb{N}$.

Also, define the variance sub-metric $d_n(s,t) = Var(\mathcal{G}_n(s) - \mathcal{G}_n(t))$ on $\mathcal{T}$.

### B.1. **Lemmas for Theorem 1 under Assumption OS.**

**Lemma 1.** *Under Assumptions C and OS, there exist sequences $\epsilon_n, \delta_n = O(n^{-c})$ for some $c > 0$ such that*

$$P\left\{\sup_{t\in\mathcal{T}} \left|\sqrt{n}h^{\beta-1/2}\{\hat{F}_{X^*}(t) - F_{X^*}(t)\} - \mathcal{G}_n(t)\right| > \epsilon_n\right\} < \delta_n.$$

*Proof.* By applying the argument in Hall and Lahiri (2008), the bias of the estimator $\hat{F}_{X^*}$ satisfies $\sup_{t\in\mathcal{T}} |E[\hat{F}_{X^*}(t)] - F_{X^*}(t)| = O(h^\gamma)$. Thus, Assumption OS (iii) guarantees

$$\sqrt{n}h^{\beta-1/2} \sup_{t\in\mathcal{T}} |E[\hat{F}_{X^*}(t)] - F_{X^*}(t)| = o(n^{-\xi}).$$

So, the bias term is negligible and it is enough to show that

$$P\left\{\sup_{t\in\mathcal{T}} \left|\sqrt{n}h^{\beta-1/2}\{\hat{F}_{X^*}(t) - E[\hat{F}_{X^*}(t)]\} - \mathcal{G}_n(t)\right| > \epsilon_n\right\} < \delta_n, \tag{34}$$

for some $\epsilon_n, \delta_n = O(n^{-c})$ with $c > 0$. Let $F^{EDF}_{X,n}$ be the empirical distribution function by $\{X_i\}_{i=1}^n$, $\alpha_n(x) = \sqrt{n}\{F^{EDF}_{X,n}(x) - F_X(x)\}$ be the empirical process, and

$$D_n(t) = \sqrt{n}h^{\beta-1/2}\{\hat{F}_{X^*}(t) - E[\hat{F}_{X^*}(t)]\} = h^{\beta-1/2} \int \mathbb{L}\left(\frac{t-a}{h}\right) d\alpha_n(a).$$

Then (34) is rewritten as

$$P\left\{\sup_{t\in\mathcal{T}} |D_n(t) - \mathcal{G}_n(t)| > \epsilon_n\right\} < \delta_n, \tag{35}$$

for some $\epsilon_n, \delta_n = O(n^{-c})$ with $c > 0$.

First, we approximate $D_n(t)$ by

$$D_{n,0}(t) = h^{\beta-1/2} \int \bar{\mathbb{L}}\left(\frac{t-a}{h}\right) d\alpha_n(a),$$

Note that both $D_n(t)$ and $D_{n,0}(t)$ are well defined as Lebesgue-Steltjes integrals.[6] From integration by parts,

$$\begin{aligned} D_n(t) &= h^{\beta-3/2}\int \mathbb{K}\left(\frac{t-a}{h}\right)\alpha_n(a)da \\ &\quad + h^{\beta-1/2}\lim_{a\to\infty}\left\{\mathbb{L}\left(\frac{t-a}{h}\right)\alpha_n(a)\right\} - h^{\beta-1/2}\lim_{a\to-\infty}\left\{\mathbb{L}\left(\frac{t-a}{h}\right)\alpha_n(a)\right\} \\ &= h^{\beta-3/2}\int \mathbb{K}\left(\frac{t-a}{h}\right)\alpha_n(a)da, \end{aligned} \tag{36}$$

for all $n \in \mathbb{N}$, where the second equality follows from the facts $\lim_{a\to\pm\infty}\alpha_n(a) = 0$ and $\sup_u |\mathbb{L}(u)| < \infty$ for each $h$. Since a similar expression applies for $D_{n,0}(t)$, there exists $C > 0$ such that

$$D_n(t) - D_{n,0}(t) = h^{\beta-1/2}\int \{\mathbb{K}(u) - \bar{\mathbb{K}}(u)\}\alpha_n(u-th)du \le Ch^s \sup_u |\alpha_n(u)|,$$

for all $n$ large enough and $t \in \mathcal{T}$, where the inequality follows from Assumption OS (ii). Now by the strong approximation (Komlós, Major and Tusnády, 1975), there exists a tight Brownian bridge $B(t) = W(t) - tW(1)$ and universal constants $C_1, C_2 > 0$ such that

$$P\left\{\sup_u |\alpha_n(u)| \le \sup_u |B(F_X(u))| + C_1\frac{\log n}{\sqrt{n}}\right\} \ge 1 - \frac{C_2}{n},$$

for all $n \in \mathbb{N}$. Combining theses results and using the properties of $\sup_u |B(F_X(u))|$ (in particular, $P\{\sup_u |B(F_X(u))| \ge x\} \le 2\exp(-2x^2)$ for $x > 0$), there exists $C_3 > 0$ such that

$$P\left\{\sup_{t\in\mathcal{T}} |D_n(t) - D_{n,0}(t)| > h^{s/2}\right\} \le C_3\exp(-2h^{-s}) + \frac{C_2}{n},$$

for all $n$ large enough. Note that $h^{s/2} = O(n^{-c_1})$ for some $c_1 > 0$ due to Assumption OS (iii) $(n^\nu h \to 0)$. Thus, it is enough for (35) to show that

$$P\left\{\sup_{t\in\mathcal{T}} |D_{n,0}(t) - \mathcal{G}_n(t)| > \epsilon_n\right\} < \delta_n,$$

for some $\epsilon_n, \delta_n = O(n^{-c})$ with $c > 0$.

Second, we approximate $D_{n,0}(t)$ by

$$D_{n,1}(t) = h^{\beta-1/2}\int \bar{\mathbb{L}}\left(\frac{t-a}{h}\right)dB(F_X(a)).$$

Since $\bar{\mathbb{L}} \in L_2(\mathbb{R})$, this integral exists for all $t \in \mathbb{R}$. Analogous to the integration by parts formula in (36), a similar result applies for $D_{n,1}(t)$ based on stochastic integration by parts using the facts $\lim_{u\to\pm\infty}\bar{\mathbb{L}}(u) = 0$ and $\sup_a |B(F_X(a))| < \infty$ almost surely. Thus, we have

$$\begin{aligned} D_{n,0}(t) - D_{n,1}(t) &= h^{\beta-3/2}\int \bar{\mathbb{K}}\left(\frac{t-a}{h}\right)\{\alpha_n(a) - B(F_X(a))\}da \\ &\le h^{\beta-1/2}\sup_a |\alpha_n(a) - B(F_X(a))| \int |\bar{\mathbb{K}}(u)|du, \end{aligned}$$

[6] This is verified as follows. By the definition $\mathbb{L}(u) = \int_0^u \mathbb{K}(v)dv$ and Assumption OS (ii), we have $\sup_u |\mathbb{L}(u)| < \infty$. Also, by $\bar{\mathbb{L}}(u) = \int_0^u \bar{\mathbb{K}}(v)dv$ (follows from Fubini's theorem) and Assumption OS (ii), we have $\sup_u |\bar{\mathbb{L}}(u)| < \infty$. Therefore, bounded variation of the empirical process $\alpha_n$ guarantees that both $D_n(t)$ and $D_{n,0}(t)$ are well defined.

for all $n \in \mathbb{N}$, almost surely. Now by Komlós, Major and Tusnády (1975), there exist Brownian bridge $B$ with continuous sample paths and universal constants $C_4, C_5 > 0$ such that

$$P\left\{\sup_{a\in\mathbb{R}} |\alpha_n(a) - B(F_X(a))| > C_4 \frac{\log n}{\sqrt{n}}\right\} \leq \frac{C_5}{n},$$

for all $n \in \mathbb{N}$. Combining this with Assumption OS (ii) (eq. (5) in the paper), there exist $c_2, C_6 > 0$ such that

$$P\left\{\sup_{t\in\mathcal{T}} |D_{n,0}(t) - D_{n,1}(t)| > C_6 n^{-c_2}\right\} \leq \frac{C_5}{n},$$

for all $n$ large enough. Thus, it is enough for (35) to show that

$$P\left\{\sup_{t\in\mathcal{T}} |D_{n,1}(t) - \mathcal{G}_n(t)| > \epsilon_n\right\} < \delta_n,$$

for some $\epsilon_n, \delta_n = O(n^{-c})$ with $c > 0$.

Third, we approximate $D_{n,1}(t)$ by

$$D_{n,2}(t) = h^{\beta-1/2} \int \bar{\mathbb{L}}\left(\frac{t-a}{h}\right) dW(F_X(a)).$$

By the definition $B(t) = W(t) - tW(1)$, we have

$$|D_{n,1}(t) - D_{n,2}(t)| \leq h^{\beta-1/2}|W(1)| \left|\int \bar{\mathbb{L}}\left(\frac{t-a}{h}\right) f_X(a)da\right|, \tag{37}$$

for all $n \in \mathbb{N}$. Therefore, for the rate of $\sup_{t\in\mathcal{T}} |D_{n,1}(t) - D_{n,2}(t)|$, we need to characterize the order of $I_{n1}(t) = \int \bar{\mathbb{L}}\left(\frac{t-a}{h}\right) f_X(a)da$. By the definition of $\bar{\mathbb{L}}$ and

$$\int_{-\infty}^{\infty} \sin(\omega(t-a)) f_X(a)da = \frac{1}{2\mathrm{i}}\{e^{\mathrm{i}\omega t} f_X^{\mathrm{ft}}(-\omega) - e^{-\mathrm{i}\omega t} f_X^{\mathrm{ft}}(\omega)\},$$

an application of Fubini's theorem assures

$$\begin{aligned} |I_{n1}(t)| &= \left|\frac{1}{2\mathrm{i}\pi}\int_{\omega_0}^{1/h} \{e^{\mathrm{i}\omega t} f_X^{\mathrm{ft}}(-\omega) - e^{-\mathrm{i}\omega t} f_X^{\mathrm{ft}}(\omega)\} \frac{K^{\mathrm{ft}}(h\omega)}{\omega f_\epsilon^{\mathrm{ft}}(\omega)} d\omega\right| \\ &\leq \frac{1}{\pi}\int_{\omega_0}^{1/h} \omega^{-1} d\omega = O(\log(1/h)). \end{aligned}$$

where the inequality follows from $|f_X^{\mathrm{ft}}| = |f_{X^*}^{\mathrm{ft}}||f_\epsilon^{\mathrm{ft}}| \leq |f_\epsilon^{\mathrm{ft}}|$ and $f_\epsilon^{\mathrm{ft}}(\omega) = f_\epsilon^{\mathrm{ft}}(-\omega)$. Substituting this bound for $I_{n1}(t)$ into (37), we obtain

$$P\left\{\sup_{t\in\mathcal{T}} |D_{n,1}(t) - D_{n,2}(t)| > M_n h^{\beta-1/2}\log(1/h)\right\} = O(n^{-c_3}),$$

for some $c_3 > 0$ and sequence $M_n = \log n$. By Assumption OS (i) ($\beta > 1/2$), it holds $M_n h^{\beta-1/2}\log(1/h) = O(n^{-c_4})$ for some $c_4 > 0$. Therefore, it is enough for (35) to show that

$$P\left\{\sup_{t\in\mathcal{T}} |D_{n,2}(t) - \mathcal{G}_n(t)| > \epsilon_n\right\} < \delta_n,$$

for some $\epsilon_n, \delta_n = O(n^{-c})$ with $c > 0$. But we can see that the process $D_{n,2}(t)$ has the same finite dimensional distributions as the process $\mathcal{G}_n(t)$. Therefore, this trivially holds true and the conclusion is obtained. □

**Lemma 2.** *Under Assumptions C and OS, there exist sequences* $\epsilon_{1n}, \delta_{1n}, \delta_{2n} = O(n^{-c})$ *for some* $c > 0$ *such that with probability greater than* $1 - \delta_{2n}$,

$$P^{\#}\left\{\sup_{t\in\mathcal{T}}\left|\sqrt{n}h^{\beta-1/2}\{\hat{F}^{\#}_{X^*}(t) - \hat{F}_{X^*}(t)\} - \tilde{\mathcal{G}}_n(t)\right| > \epsilon_{1n}\right\} \quad < \quad \delta_{1n},$$

*where* $\tilde{\mathcal{G}}_n$ *is a tight Gaussian process with the same distribution as* $\mathcal{G}_n$ *under* $P^{\#}$.

*Proof.* The proof is essentially a reformulation of that of Bissantz, Dümbgen, Holzmann and Munk (2007, Theorem 2.1). Let $\alpha_n^{\#}(t) = \sqrt{n}\{F^{EDF}_{X^{\#},n} - F^{EDF}_{X,n}(t)\}$ denote the bootstrap empirical process. As shown in the proof of Bissantz, Dümbgen, Holzmann and Munk (2007, eq. (21)), based on Shorack (1982), there exist a Brownian bridge $B_n^{\#}$ and universal constants $C, C_1 > 0$ such that for all $n \in \mathbb{N}$,

$$P^{\#}\left\{\sup_{t\in\mathbb{R}}|\alpha_n^{\#}(t) - B_n^{\#}(F^{EDF}_{X,n}(t))| > C\frac{\log n}{\sqrt{n}}\right\} \leq \frac{C_1}{n},$$

almost surely. Now it is known that the Brownian bridge is Hölder continuous for every exponent $b \in (0, 1/2)$ almost surely. Furthermore, by Komlós, Major and Tusnády's (1975) coupling, along with the fact $P\{\sup_t |B(F_X(t))| \geq \log n\} \leq 2\exp(-2(\log n)^2)$, there exist universal constants $C_2, C_3 > 0$ such that

$$P\left\{\sup_{t\in\mathbb{R}}|F^{EDF}_{X,n}(t) - F_X(t)| > C_2\frac{\log n}{\sqrt{n}}\right\} \leq \frac{C_3}{n},$$

for all $n \in \mathbb{N}$, which consequently implies

$$P\left\{\sup_{t\in\mathbb{R}}|B_n^{\#}(F^{EDF}_{X,n}(t)) - B_n^{\#}(F_X(t))| > C_4\frac{\log n}{n^{b/2}}\right\} \leq \frac{C_5}{n},$$

for some universal constants $C_4, C_5 > 0$. Combining these results, there exist universal constants $C_6, C_7, C_8 > 0$ such that with probability greater than $1 - C_6/n$, it holds

$$P^{\#}\left\{\sup_{t\in\mathbb{R}}|\alpha_n^{\#}(t) - B_n^{\#}(F_X(t))| > C_7\frac{\log n}{n^{b/2}}\right\} \leq \frac{C_8}{n},$$

for all $n \in \mathbb{N}$. Based on this, the conclusion follows by similar arguments to the proof of Lemma 1. □

**Lemma 3.** *Suppose that Assumptions C and OS hold true. Then for any sequence* $\epsilon_n = O(n^{-c})$ *with* $c > 0$, *there exists a constant* $M > 0$ *such that*

$$p_{\epsilon_n}(\mathcal{G}_n) \leq M\epsilon_n\sqrt{\log(1/h)},$$

*for all* $n$ *large enough.*

*Proof.* Pick any $\varepsilon > 0$. By Chernozhukov, Chetverikov and Kato (2015, Theorem 3) and separability of $\mathcal{G}_n$, there exists $C > 0$ such that

$$p_\varepsilon(\mathcal{G}_n) \leq C\varepsilon\left\{\underline{\sigma}_n^{-1}E\left[\sup_{t\in\mathcal{T}}|\mathcal{G}_n(t)|\right] + \sqrt{1 \vee \log(\underline{\sigma}_n/\varepsilon)}\right\},$$

for all $n \in \mathbb{N}$. Thus, it is enough to show that

$$E\left[\sup_{t\in\mathcal{T}}|\mathcal{G}_n(t)|\right] = O(\sqrt{\log(1/h)}).$$

Now,

$$d_n^2(s,t) = h^{2\beta}\int\left\{\bar{\mathbb{L}}\left(\frac{s}{h}-a\right)-\bar{\mathbb{L}}\left(\frac{t}{h}-a\right)\right\}^2 f_X(ha)da$$

by the Ito isometry. Note that $\bar{\mathbb{L}}$ is Lipschitz continuous because its derivative $\bar{\mathbb{K}}$ is uniformly bounded on $\mathbb{R}$ (because $h^\beta \sup_u|\bar{\mathbb{K}}(u)| \leq C$ for some $C > 0$ by Assumption OS (i)). Thus, it holds

$$d_n(s,t) \leq C_1 h^{-3/2}|s-t|, \tag{38}$$

for some $C_1 > 0$ that is independent of $s$ and $t$.

Let $D(\varepsilon, d_n)$ be the $\varepsilon$-packing number for the set $\mathcal{T}$ under the sub-metric $d_n$. By (38), it holds $D(\varepsilon, d_n) \leq 2C_1 h^{-3/2}/\varepsilon$. Pick any $\delta \in (0,1)$. By van der Vaart and Wellner (1996, Corollary 2.2.8), there exist universal constants $C_2, C_3 > 0$ such that

$$\begin{aligned}
&E\left[\sup_{d_n(s,t)\leq\delta}|\mathcal{G}_n(s)-\mathcal{G}_n(t)|\right]\\
\leq\ & C_2\int_0^\delta\sqrt{\log D(\varepsilon,d_n)}d\varepsilon \leq C_2\delta\sqrt{\log(2C_1h^{-3/2})}+C_2\int_0^\delta\sqrt{\log(1/\varepsilon)}d\varepsilon \leq C_3\sqrt{\log(1/h)},
\end{aligned}$$

for all $n \in \mathbb{N}$. Thus, there exists a collection of Gaussian random variables $\{\mathcal{G}_n(t_j)\}_{i=1}^{p_n}$ with $p_n = \left\lceil \frac{1}{h^{3/2}\delta}\right\rceil$ such that

$$E\left[\sup_{t\in\mathcal{T}}|\mathcal{G}_n(t)|\right] \leq E\left[\max_{1\leq j\leq p_n}|\mathcal{G}_n(t_j)|\right] + C_3\sqrt{\log(1/h)},$$

for all $n \in \mathbb{N}$. Now the properties of the maximum of Gaussian random variables yields

$$E\left[\max_{1\leq j\leq p_n}|\mathcal{G}_n(t_j)|\right] \leq 2\bar{\sigma}_n\sqrt{1+\log p_n}.$$

Combining these results, the conclusion follows. □

B.2. **Lemmas for Theorem 1 under Assumption SS.**

**Lemma 4.** *Under Assumptions C and SS, there exist sequences* $\epsilon_n = O(\log n)^{-(1+c)}$ *and* $\delta_n = O(n^{-c})$ *with* $c > 0$ *such that*

$$P\left\{\sup_{t\in\mathcal{T}}\left|\frac{\sqrt{n}}{\varsigma(h)}\{\hat{F}_{X^*}(t)-F_{X^*}(t)\}-\mathcal{G}_n(t)\right| > \epsilon_n\right\} < \delta_n.$$

**Lemma 5.** *Under Assumptions C and SS, there exist sequences* $\epsilon_{1n}, \delta_{1n}, \delta_{2n} = O(n^{-c})$ *with* $c > 0$ *such that with probability greater than* $1-\delta_{2n}$,

$$P^{\#}\left\{\sup_{t\in\mathcal{T}}\left|\frac{\sqrt{n}}{\varsigma(h)}\{\hat{F}_X^{\#}(t)-\hat{F}_X(t)\}-\tilde{\mathcal{G}}_n(t)\right| > \epsilon_{1n}\right\} \ < \ \delta_{1n},$$

*where* $\tilde{\mathcal{G}}_n$ *is a tight Gaussian process with the same distributions as* $\mathcal{G}_n$ *under* $P^{\#}$.

These lemmas can be shown in the same way as Lemmas 1 and 2. The log rate of $\epsilon_n$ in Lemma 4 is due to the bias term. Recall that under Assumption C (ii), the bias of the estimator $\hat{F}_{X^*}$ is given by

$$\sup_{t\in\mathcal{T}}|E[\hat{F}_{X^*}(t)]-F_{X^*}(t)|=O(h^\gamma).$$

Then due to Assumption SS (iii), it holds $\sqrt{n}h^\gamma/\varsigma(h)=C(\log n)^{-c}$ for some $c>1$.

**Lemma 6.** *Suppose that Assumptions C and SS hold true. Then for any sequence $\epsilon_n = O(\log n)^{-c}$ with $c>1$ there exists $M>0$ such that*

$$p_{\epsilon_n}(\mathcal{G}_n)\le M\epsilon_n(\log n)^{1+r},$$

*for all $n$ large enough and any $r>0$ independent of $n$.*

*Proof.* Pick any $\varepsilon>0$. By Chernozhukov, Chetverikov and Kato (2015, Theorem 3) and separability of the Gaussian process $\mathcal{G}_n$, there exists $C>0$ such that

$$p_\varepsilon(\mathcal{G}_n)\le C\varepsilon\left\{\underline{\sigma}_n^{-1}E\left[\sup_{t\in\mathcal{T}}|\mathcal{G}_n(t)|\right]+\sqrt{1\vee\log(\underline{\sigma}_n/\varepsilon)}\right\},$$

for all $n\in\mathbb{N}$. By Lemmas 7 and 8 shown below, the following hold true:

$$\text{there exist } c_1>0 \text{ such that } \underline{\sigma}_n\ge c_1h^{\lambda+\nu} \text{ for all } \nu>0 \text{ and } n \text{ large enough,} \tag{39}$$

$$\text{there exist } C_1>0 \text{ such that } \bar{\sigma}_n\le C_1 \text{ for all } n \text{ large enough.} \tag{40}$$

Observe that

$$d_n^2(s,t) = \frac{h}{\varsigma^2(h)}\int\left\{\bar{\mathbb{L}}\left(\frac{s}{h}-a\right)-\bar{\mathbb{L}}\left(\frac{t}{h}-a\right)\right\}^2 f_X(ha)da$$

by the Ito isometry. Note that $\bar{\mathbb{L}}$ is Lipschitz continuous because its derivative $\bar{\mathbb{K}}$ is uniformly bounded on $\mathbb{R}$ (because $\sqrt{h}\varsigma^{-1}(h)\sup_u|\bar{\mathbb{K}}(u)|\le C_2h^{-c_2}$ for some $C_2,c_2>0$ by Assumption SS (i)). Thus, it holds $d_n(s,t)\le C_3h^{-c_2-3/2}|s-t|$ for some $C_3>0$ that is independent of $s$ and $t$. Using (40), an analogous argument as in the proof of Lemma 3 shows that $E\left[\sup_{t\in\mathcal{T}}|\mathcal{G}_n(t)|\right]=O(\sqrt{\log(1/h)})$. Combining this with (39) and Assumption SS (iii), the conclusion follows. □

**Lemma 7.** *Under Assumptions C and SS, there exists $c>0$ such that $\underline{\sigma}_n\ge ch^{\lambda+\nu}$ for all $\nu>0$ and $n$ large enough.*

*Proof.* We only prove the case of $\lambda_0\ge 0$. The proof for the case of $\lambda_0<0$ is similar. Pick any $\varepsilon>0$. By Assumption C (i), we provide a lower bound for $\underline{\sigma}_n$ via

$$\underline{\sigma}_n=\inf_{t\in\mathcal{T}}\frac{h}{\varsigma^2(h)}\int\bar{\mathbb{L}}^2(a)f_X(t-ha)da\ge\frac{c_1h}{\varsigma^2(h)}\int_{|a|\le h^\varepsilon}\bar{\mathbb{L}}^2(a)da,$$

for some $c_1>0$. Let

$$\Phi_\epsilon(\omega)=f_\epsilon^{\text{ft}}(\omega)^{-1}\mathbb{I}\{|\omega|\ge\omega_0\}.$$

Using the fact $\sin(x)=x+R(x)$ with $|R(x)|\le c_2|x|^2$ for some $c_2>0$, it follows that for all $|a|\le h^\varepsilon$,

$$|\bar{\mathbb{L}}(a)|\ge\frac{1}{\pi}\left|a\int_0^1K^{\text{ft}}(\omega)\Phi_\epsilon\left(\frac{\omega}{h}\right)d\omega\right|-\frac{c_2}{\pi}\left|a\int_0^1|a\omega|K^{\text{ft}}(\omega)\Phi_\epsilon\left(\frac{\omega}{h}\right)d\omega\right|\ge C\{1-O(h^\varepsilon)\}|aI_n|,$$

where $I_n=\int_0^1K^{\text{ft}}(\omega)\Phi_\epsilon\left(\frac{\omega}{h}\right)d\omega$ and the last inequality follows from the fact $\sup\{|a\omega|:|a|\le h^\varepsilon,\omega\in[0,1]\}=h^\varepsilon$.

We now provide a lower bound for $I_n$. Pick any $\delta > 0$. Observe that

$$
\begin{aligned}
h^{\frac{1-\lambda}{2}}\varsigma(h)^{-\frac{1}{2}}|I_n| &= \frac{\exp(-1/\mu h^\lambda)}{h^{\lambda(s+1)+\lambda_0}}\int_{h\omega_0}^1 K^{\mathrm{ft}}(\omega)\Phi_\epsilon\left(\frac{\omega}{h}\right)d\omega \\
&\geq c_3\frac{\exp(-1/\mu h^\lambda)}{h^{\lambda(s+1)}}\int_{h\omega_0}^1 K^{\mathrm{ft}}(\omega)\omega^{-\lambda_0}\exp\left(\frac{|\omega|^\lambda}{h^\lambda\mu}\right)d\omega \\
&\geq c_3\frac{\exp(-1/\mu h^\lambda)}{h^{\lambda(s+1)}}\int_{\delta}^1 K^{\mathrm{ft}}(\omega)\exp\left(\frac{|\omega|^\lambda}{h^\lambda\mu}\right)d\omega \\
&= c_3\int_0^{(1-\delta)h^{-\lambda}}\frac{K^{\mathrm{ft}}(1-h^\lambda v)}{(h^\lambda v)^s}v^s\exp\left(\frac{|1-h^\lambda v|^\lambda-1}{h^\lambda\mu}\right)dv \\
&\to c_3 r^s\int v^s\exp(-\lambda v/\mu)dv > 0,
\end{aligned}
$$

for some $c_3 > 0$, where the first inequality follows from the fact $\Phi_\epsilon(\omega) \geq c_3|\omega|^{-\lambda_0}\exp(|\omega|^\lambda/\mu)$, the second inequality holds since all the terms inside the integral are positive and $\omega^{-\lambda_0}\mathbb{I}\{h\omega_0 \leq \omega \leq 1\} \geq 1$ for $\lambda_0 \geq 0$, the second equality follows from a change of variables, and the convergence follows from the dominated convergence theorem after noting

$$
\begin{aligned}
&\frac{K^{\mathrm{ft}}(1-h^\lambda v)}{(h^\lambda v)^s}v^s\exp\left(\frac{|1-h^\lambda v|^\lambda-1}{h^\lambda\mu}\right)\mathbb{I}\{0 \leq v \leq (1-\delta)h^{-\lambda}\} \\
\leq\ &\begin{cases} \sup_{0\leq t\leq 1}\{t^{-s}K^{\mathrm{ft}}(1-t)\}v^s\exp(-v/\mu) & \text{if } \lambda \geq 1, \\ \sup_{0\leq t\leq 1}\{t^{-s}K^{\mathrm{ft}}(1-t)\}v^s\exp(-\lambda v/\mu) & \text{if } 0 < \lambda < 1. \end{cases}
\end{aligned}
$$

Thus, it holds $h^{1/2}\varsigma(h)^{-1/2}|I_n| > c_3 h^{\lambda/2}$ for all $n$ large enough.

Combining these results, there exists $c > 0$ such that

$$
\underline{\sigma}_n \geq ch^\lambda\int_{|a|\leq h^\varepsilon}|a|^2da \geq ch^{\lambda+3\varepsilon},
$$

for all $n$ large enough, and the conclusion follows. □

**Lemma 8.** *Under Assumptions C and SS, there exists $C > 0$ such that $\bar{\sigma}_n \leq C$ for all $n$ large enough.*

*Proof.* We only prove the case of $\lambda_0 \geq 0$. The proof for the case of $\lambda_0 < 0$ is similar. Pick any $\varepsilon \in (0, 2^{-1/\lambda})$. Since $f_X$ is bounded (Assumption C (ii)), there exists $C_1, C_2 > 0$ such that

$$
\begin{aligned}
\bar{\sigma}_n &\leq C_1\frac{\exp(-2/\mu h^\lambda)}{h^{\lambda(2s+1)+2\lambda_0}}\int\bar{\mathbb{L}}^2(a)da = C_2\frac{\exp(-2/\mu h^\lambda)}{h^{\lambda(2s+1)+2\lambda_0}}\int_{h\omega_0}^1\left|\frac{K^{\mathrm{ft}}(\omega)}{\omega}\Phi_\epsilon\left(\frac{\omega}{h}\right)\right|^2d\omega \\
&\leq C_2\omega_0^{-2}\frac{\exp(-2/\mu h^\lambda)}{h^{\lambda(2s+1)+2(1+\lambda_0)}}\int_{h\omega_0}^1\left|K^{\mathrm{ft}}(\omega)\Phi_\epsilon\left(\frac{\omega}{h}\right)\right|^2d\omega \\
&\leq C_2\omega_0^{-2}\frac{\exp(-2/\mu h^\lambda)}{h^{\lambda(2s+1)+2(1+\lambda_0)}}\int_{h\omega_0}^{\varepsilon}\left|K^{\mathrm{ft}}(\omega)\left(\frac{\omega}{h}\right)^{-(1+\lambda_0)}\exp\left(\frac{|\omega|^\lambda}{h^\lambda\mu}\right)\right|^2d\omega \\
&\quad +C_2\omega_0^{-2}\frac{\exp(-2/\mu h^\lambda)}{h^{\lambda(2s+1)}}\int_{|\omega|>\varepsilon}\left|K^{\mathrm{ft}}(\omega)\omega^{-(1+\lambda_0)}\exp\left(\frac{|\omega|^\lambda}{h^\lambda\mu}\right)\right|^2d\omega \\
&= T_{1n}+T_{2n},
\end{aligned}
$$

for all $n$ large enough, where the first equality follows from Plancherel's isometry,[7] and the second inequality follows from $\Phi_\epsilon(\omega) \leq C|\omega|^{-\lambda_0} \exp(|\omega|^\lambda/\mu)$. For $T_{1n}$, Assumption SS (iii) and the restriction $\varepsilon \in (0, 2^{-1/\lambda})$ guarantee

$$\begin{aligned} T_{1n} &\leq C_3 \omega_0^{-(1+\lambda_0)} \frac{\exp(-2/\mu h^\lambda)}{h^{\lambda(2s+1)+2(1+\lambda_0)}} \int_{h\omega_0}^{\varepsilon} \left| K^{\mathrm{ft}}(\omega) \exp\left(\frac{|\omega|^\lambda}{h^\lambda \mu}\right) \right|^2 d\omega \\ &\leq C_4 \frac{\exp(-1/\mu h^\lambda)}{h^{\lambda(2s+1)+2(1+\lambda_0)}} = O(n^{-c_1}), \end{aligned}$$

for some $C_3, C_4, c_1 > 0$. For $T_{2n}$, note that

$$T_{2n} \leq C_5 \varepsilon^{-(1+\lambda_0)} \frac{\exp(-2/\mu h^\lambda)}{h^{\lambda(2s+1)}} \int_{|\omega|>\varepsilon} \left| K^{\mathrm{ft}}(\omega) \exp\left(\frac{|\omega|^\lambda}{h^\lambda \mu}\right) \right|^2 d\omega,$$

for some $C_5 > 0$. By an analogous dominated convergence argument used in the proof of Lemma 7, we can show $T_{2n}$ converges to some finite constant. Combining these results, the conclusion follows. □

[7] Note that $\bar{\mathbb{L}}$ is written as $\bar{\mathbb{L}}(u) = \frac{1}{2\pi} \int_{-1}^{1} \frac{e^{-\mathrm{i}\omega u}}{\omega} \frac{K^{\mathrm{ft}}(\omega)}{f_\epsilon^{\mathrm{ft}}(\omega/h)} \mathbb{I}\{|\omega| \geq \omega_0\} d\omega$. This integral exists due to the truncation.

## Appendix C. Assumptions and proofs for Theorem 2

In this appendix we prove Theorem 2, the asymptotic distribution of $t_n$ in eq. (12) of the paper. Basic steps of our proof follow the recipe laid down by Bissantz, Dümbgen, Holzmann and Munk (2007). Importantly, we impose tail conditions on $f_\epsilon^{\rm ft}$ of the form $f_\epsilon^{\rm ft}(\omega)|\omega|^\beta \to C_\epsilon$ as $|\omega| \to \infty$. Based on this, we define

$$\begin{aligned} \mathcal{K}(u) &= \frac{1}{2\pi C_\epsilon}\int_0^\infty e^{-\mathrm{i}\omega u}\omega^\beta K^{\rm ft}(\omega)d\omega + \frac{1}{2\pi C_\epsilon}\int_{-\infty}^0 e^{-\mathrm{i}\omega u}|\omega|^\beta K^{\rm ft}(\omega)d\omega, \\ \mathcal{L}(u) &= \frac{1}{2\pi C_\epsilon}\int_0^\infty \sin(\omega u)\omega^{\beta-1} K^{\rm ft}(\omega)d\omega + \frac{1}{2\pi C_\epsilon}\int_{-\infty}^0 \sin(\omega u)|\omega|^\beta \omega^{-1} K^{\rm ft}(\omega)d\omega. \end{aligned} \quad (41)$$

These are the pointwise limits of $h^\beta\mathbb{K}(u)$ and $h^\beta\mathbb{L}(u)$ as $h \to 0$ under some assumptions on $f_\epsilon^{\rm ft}$. In addition to Assumptions OS, we impose the following conditions.

**Assumption G.**

- **(i):** *$f_\epsilon^{\rm ft}(\omega)|\omega|^\beta \to C_\epsilon$ as $|\omega| \to \infty$ for some $\beta > 1/2$.*
- **(ii):** *$h^\beta \int |\mathbb{K}(u)|du < M$ for some $M > 0$ independent of $h$. $\int |u|^{3/2}\sqrt{\log(\log^+|u|)}|\mathcal{K}(u)|du < \infty$. For some $\delta > 0$, $\int |h^\beta \bar{\mathbb{K}}(u) - \mathcal{K}(u)|du = O(h^{1/2+\delta})$.*
- **(iii):** *$\lim_{u\to\pm\infty} |\mathcal{L}(u)\sqrt{|u|\log(\log^+|u|)}| = 0$. For some $\delta_1 \in (0,1)$, $\int |\mathcal{L}(u)|^{2-\delta_1}du < \infty$. For some $\delta > 0$, $\sup_u |h^\beta\bar{\mathbb{L}}(u/h) - \mathcal{L}(u/h)| = O(h^{1/2+\delta})$.*
- **(iv):** *$f_X$ and its derivative $f_X'$ are bounded and continuous on $\mathbb{R}$ such that $\lim_{x\to\pm\infty} |xf_X(x)\log(\log^+|x|)| = 0$. Also, $\sup_x |f_X'(x)f_X(x)^{-1/2}\sqrt{|x|\log(\log^+|x|)}| < \infty$. Furthermore it holds $\int |f_X'(x)f_X(x)^{-1/2}\sqrt{|x|\log(\log^+|x|)}|dx < \infty$.*

These conditions are generalizations and simplifications of the ones in Bissantz, Dümbgen, Holzmann and Munk (2007). Assumption G (i) is stronger than the usual assumption $f_\epsilon^{\rm ft}(\omega)|\omega|^\beta < C_\epsilon$ as $|\omega| \to \infty$ but is required for explicit derivation of the limiting distribution.

Assumption G (ii) contains conditions for the deconvolution kernel $\mathbb{K}$. The first condition ensures that $\mathbb{K}$ is $L_1$-integrable. A sufficient condition for this is that $1/f_\epsilon^{\rm ft}(\omega)$ is a polynomial function in $\omega$. Indeed in this case it can be shown from the properties of the Fourier transform that $|\mathbb{K}(u)| \sim |u|^{-q}$ as $|u| \to \infty$ under some conditions on $f_\epsilon^{\rm ft}$. For instance, the choice $r > 2$ for $K$ assures $|\mathbb{K}(u)| \sim |u|^{-2}$ under the assumption

$$\int \left| \left\{ \frac{K^{\rm ft}(\omega)}{f_\epsilon^{\rm ft}(\omega/h)} \right\}'' \right| d\omega = O(h^{-\beta}).$$

A similar condition is given in, for example, Bissantz, Dümbgen, Holzmann and Munk (2007, eq. (13)). $\mathcal{K}$ in (41) is the limit of $\bar{\mathbb{K}}$ as $h \to \infty$ obtained by Assumption G (i). Recall that by Assumption OS (ii), $h^{\beta-\frac{1}{2}}\int |\mathbb{K}(u) - \bar{\mathbb{K}}(u)|du = O(h^s)$. Additionally, it can be shown from the previous assumptions and properties of the Fourier transform of $\omega^\beta K^{\rm ft}(\omega)$ that $\int |h^\beta\mathbb{K}(u) - \mathcal{K}(u)|du < \infty$. To obtain the rate $h^{1/2+\delta}$ for the latter, we need some additional conditions on the decay of $f_\epsilon^{\rm ft}$. Denote $R(\omega) = f_\epsilon^{\rm ft}(\omega)\omega^\beta - C_\epsilon$. Then a sufficient condition for the third condition in Assumption G (ii) is that $R(\omega) \sim \omega^{-1/2-\delta}$ as $|\omega| \to \infty$.

Assumption G (iii) contains conditions on the integrated kernel function $\mathbb{L}$. On the first two conditions in Assumption G (iii), we can in fact show the stronger statement that for all the commonly used kernel functions, $\mathcal{L}(u) \sim |u|^{-\beta\wedge 1}$ as $u \to \pm\infty$. Regarding the third condition in Assumption G (iii), note that we can expand

$$h^{\beta}\bar{\mathbb{L}}\left(\frac{u}{h}\right) - L\left(\frac{u}{h}\right) = \frac{h^{\beta}}{\pi C_{\epsilon}} \int_{\omega_1}^{1/h} \frac{\sin(\omega u)}{\omega} K^{\mathrm{ft}}(h\omega)\frac{R(\omega)}{\psi^{\mathrm{ft}}(\omega)} d\omega - \frac{h^{\beta}}{\pi C_{\epsilon}} \int_{0}^{\omega_1} \sin(\omega u)\omega^{\beta-1} K^{\mathrm{ft}}(h\omega) d\omega.$$

Standard arguments show that this is of the order $h^{1/2+\delta}$ under the assumption $R(\omega) \sim \omega^{-1/2-\delta}$ as $|\omega| \to \infty$.

Assumption G (iv) provides conditions on the decay rates of the pdf $f_X$ and its derivative $f'_X$. Similar assumptions are adopted in the literature (e.g., Bickel and Rosenblatt, 1973).

Based on these conditions, we obtain Theorem 2 with

$$B = \int \mathcal{L}(a)^2 da, \qquad b_n = (-2\log h)^{1/2} + (-2\log h)^{-1/2} \log\left(\frac{\int \{\mathcal{L}'(a)\}^2 da}{4\pi B}\right), \tag{42}$$

Furthermore, if we consider the simple hypothesis

$$H_0 : F_{X^*}(t) = F_0(t) \quad \text{for } t \in \mathcal{T},$$

for some $F_0$, a test statistic for $H_0$ is $t_n^0 = \sup_{t\in\mathcal{T}} |f_X(t)^{-1/2}\{\hat{F}_{X^*}(t) - F_0(t)\}|$. Consider the sequence of local alternatives

$$H_{1n} : F_{X^*}(t) = F_0(t) + \gamma_n \eta(t) \quad \text{for } t \in \mathcal{T},$$

where $\eta(t)$ is a continuous function and $\gamma_n = \sqrt{n}h^{\beta-1/2}(2\log(1/h))^{1/2}$. By an analogous argument, we can obtain

$$P\left\{(-2\log h)^{1/2}(B^{-1/2}t_n^0 - b_n) \leq c\right\} \to \exp(-s(\eta)\exp(-c)),$$

for all $c \in \mathbb{R}$, where $s(\eta) = \int_0^1 \exp((Bf_{X^*}(a))^{-1/2}\eta(a)) + \exp(-(Bf_{X^*}(a))^{-1/2}\eta(a))da$.

C.1. **Proof of Theorem 2.** We show that

$$\sup_{t\in\mathcal{T}} |f_X(t)^{-1/2}\{\hat{F}_{X^*}(t) - F_{X^*}(t)\} - \mathcal{Y}_n(t)| = o_p((-\log(h))^{-1/2}), \tag{43}$$

where $\mathcal{Y}_n = h^{-1/2}\int \mathcal{L}\left(\frac{t-a}{h}\right) dW(a)$ is a Gaussian process. Once we obtain (43), the conclusion follows by applying the arguments of Bickel and Rosenblatt (1973, Theorem A1). The rate $o_p((-\log(h))^{-1/2})$ is required because later we scale by $(-\log(h))^{1/2}$ to obtain the limiting distribution as in Bickel and Rosenblatt (1973).

First, as in the proof of Lemma 1, the bias term in $Q_n(t)$ is negligible and we can restrict attention to the mean zero process

$$D_n(t) \;= Q_n(t) - E[Q_n(t)] = \; h^{\beta-1/2}\int \mathbb{L}\left(\frac{t-a}{h}\right) d\alpha_n(a),$$

where $\alpha_n(a) = \sqrt{n}\{F_{X,n}^{EDF}(a) - F_X(a)\}$ is the empirical process, and $F_{X,n}^{EDF}$ is the empirical distribution function by $\{X_i\}_{i=1}^n$. We approximate $D_n(t)$ by

$$D_{n,1}(t) = h^{\beta-1/2} \int \bar{\mathbb{L}}\left(\frac{t-a}{h}\right) dW(F_X(a)).$$

Indeed the arguments in the proof of Lemma 1 allow us to show

$$\sup_{t\in\mathcal{T}} |D_n(t) - D_{n,1}(t)| = O_p((nh)^{-1/2}\log n)).$$

Also, $D_{n,1}(t)$ has the same finite dimensional distribution as

$$D_{n,2}(t) = h^{\beta-1/2} \int \bar{\mathbb{L}}\left(\frac{t-a}{h}\right) f_X(a)^{1/2} dW(a).$$

Next, we approximate $D_{n,2}(t)$ by

$$D_{n,3}(t) = h^{-3/2} \int \mathcal{K}\left(\frac{t-a}{h}\right) f_X(a)^{1/2} W(a) da.$$

To this end, note that for any $h > 0$,

$$\lim_{a\to\pm\infty} \mathcal{K}\left(\frac{t-a}{h}\right) f_X(a)^{1/2} W(a) \leq \sup_u |\mathcal{K}(u)| \lim_{a\to\pm\infty} |af_X(a)\log(\log^+|a|)|^{1/2} = 0,$$

where the inequality follows from the law of the iterated logarithm for the Wiener process and the equality follows from the facts $\sup_u |\mathcal{K}(u)| = O(h^{-\beta-1})$ and Assumption G (iv). Thus, using stochastic integration by parts, we can write

$$D_{n,2}(t) = h^{\beta-1/2} \int \left\{ f_X(t-hu)^{1/2}\bar{\mathbb{K}}(u) + hf'_X(t-hu)f_X(t-hu)^{-1/2}\bar{\mathbb{L}}(u) \right\} W(t-hu)du.$$

and obtain

$$\begin{aligned} |D_{n,2}(t) - D_{n,3}(t)| &\leq h^{-1/2} \int \{h^\beta \bar{\mathbb{K}}(u) - \mathcal{K}(u)\} f_X(t-hu)^{1/2} W(t-hu)du \\ &\quad + h^{1/2} \int h^\beta \bar{\mathbb{L}}(u) f'_X(t-hu) f_X(t-hu)^{-1/2} W(t-hu)du \\ &= T_{n,4}(t) + T_{n,5}(t) \end{aligned}$$

Now by the law of the iterated logarithm and Assumption G (ii) and (iv), it follows $\sup_{t\in\mathcal{T}} |T_{n,4}(t)| = O_p(h^\delta)$. For the term $T_{n,5}(t)$,

$$\begin{aligned} |T_{n,5}(t)| &\leq h^{-1/2} \sup_u |h^\beta \bar{\mathbb{L}}(u/h) - \mathcal{L}(u/h)| \int |f'_X(t-z) f_X(t-z)^{-1/2} W(t-z)|dz \\ &\quad + h^{1/2} \left| \int \mathcal{L}(u) f'_X(t-hu) f_X(t-hu)^{-1/2} W(t-hu)du \right| \\ &= T_{n,51}(t) + T_{n,52}(t). \end{aligned}$$

Using Assumption G (iii)-(iv), an application of the law of the iterated logarithm proves $\sup_{t\in\mathcal{T}} T_{n,51}(t) = O(h^\beta)$. Next, for the term $T_{n,52}(t)$, Hölder's inequality and the law of the iterated logarithm imply

$$T_{n,52}(t) \leq h^{\delta_1/(4-2\delta_1)} \|\mathcal{L}(u)\|_{2-\delta_1} \left\| f'_X(u) f_X(u)^{-1/2} \sqrt{|u|\log(\log^+|u|)} \right\|_{2+\delta_1/(1-\delta_1)}.$$

By this expression and Assumption G (iii)-(iv), we are able to show $\sup_{t\in\mathcal{T}} |T_{n,52}(t)| = o_p((-\log(h))^{-1/2})$. Combining these results, the claim $\sup_{t\in\mathcal{T}} |D_{n,2}(t) - D_{n,3}(t)| = o_p((-\log(h))^{-1/2})$ follows.

Third, we approximate the process $f_X(t)^{-1/2}D_{n,3}(t)$ with the process

$$D_{n,4}(t) = h^{-3/2}\int \mathcal{K}\left(\frac{t-a}{h}\right) W(a)da.$$

Note that

$$f_X(t)^{-1/2}D_{n,3}(t) - D_{n,4}(t) = h^{-1/2}\int \{f_X(t)^{-1/2}f_X(t-hu)^{1/2} - 1\}\mathcal{K}(u)W(t-hu)du.$$

By the law of the iterated logarithm and Assumption G (ii) and (iv), it follows

$$\sup_{t\in\mathcal{T}} |f_X(t)^{-1/2}D_{n,3}(t) - D_{n,4}(t)| = O_p(h^{1/2}).$$

Fourth, let

$$D_{n,5}(t) = h^{-1/2}\int \mathcal{L}\left(\frac{t-a}{h}\right) dW(a).$$

By stochastic integration by parts formula and Assumption G (ii),

$$D_{n,4}(t) - D_{n,5}(t) = \left\{\lim_{a\to\infty} L\left(\frac{t-a}{h}\right) W(a)\right\} - \left\{\lim_{a\to-\infty} L\left(\frac{t-a}{h}\right) W(a)\right\} = 0,$$

for each $h$, which implies that $D_{n,4}(t) = D_{n,5}(t)$ for all $t \in \mathcal{T}$. Since $D_{n,5}(t)$ has the same finite dimensional distributions as the process $\mathcal{Y}_n$, the claim in (43) follows.

## References


[1] Bassett, G. W. and R. Koenker (1986) Strong consistency of regression quantiles and related empirical processes, *Econometric Theory*, 2, 191-201.

[2] Bickel, P. J. and M. Rosenblatt (1973) On some global measures of the deviations of density function estimates, *Annals of Statistics*, 1, 1071-1095.

[3] Billingsley, P. (1995) *Probability and Measure*, 3rd edition, Wiley-Blackwell.

[4] Bissantz, N., Dümbgen, L., Holzmann, H. and A. Munk (2007) Non-parametric confidence bands in deconvolution density estimation, *Journal of the Royal Statistical Society*, B 69, 483-506.

[5] Chernozhukov, V., Chetverikov, D. and K. Kato (2014) Anti-concentration and honest, adaptive confidence bands, *Annals of Statistics*, 42, 1787-1818.

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

[7] Dattner, I., Reiß, M. and M. Trabs (2016) Adaptive quantile estimation in deconvolution with unknown error distribution, *Bernoulli*, 22, 143-192.

[8] Hall, P. and S. N. Lahiri (2008) Estimation of distributions, moments and quantiles in deconvolution problems, *Annals of Statistics*, 36, 2110-2134.

[9] Komlós, J., Major, P. and G. Tusnády (1975) An approximation of partial sums of independent RV'-s and the sample DF. I, *Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete*, 32, 111-131.

[10] Shorack, G. R. (1982) Bootstrapping robust regression, *Communications in Statistics: Theory and Methods*, 11, 961–972.

[11] van der Vaart, A. W. and J. Wellner (1996) *Weak Convergence and Empirical Processes*, Springer.

[12] Yukich, J. E. (1987) Some limit theorems for the empirical process indexed by functions, *Probability Theory and Related Fields*, 74, 71-90.



Department of Economics, London School of Economics, Houghton Street, London, WC2A 2AE, UK.

*E-mail address*: k.adusumilli@lse.ac.uk

Department of Economics, London School of Economics, Houghton Street, London, WC2A 2AE, UK.

*E-mail address*: t.otsu@lse.ac.uk

Department of Economics, Seoul National University, 1 Gwankro Gwanakgu, Seoul, 08826, Korea.

*E-mail address*: whang@snu.ac.kr