

# Anti-concentration and honest, adaptive confidence bands

---

Victor Chernozhukov  
Denis Chetverikov  
Kengo Kato

The Institute for Fiscal Studies  
Department of Economics, UCL

**cemmap** working paper CWP43/16

# ANTI-CONCENTRATION AND HONEST, ADAPTIVE CONFIDENCE BANDS\*

BY VICTOR CHERNOZHUKOV<sup>†</sup>, DENIS CHETVERIKOV<sup>‡</sup> AND KENGO  
KATO<sup>§</sup>

*MIT<sup>†</sup>, UCLA<sup>‡</sup>, and University of Tokyo<sup>§</sup>*

Modern construction of uniform confidence bands for nonparametric densities (and other functions) often relies on the classical Smirnov-Bickel-Rosenblatt (SBR) condition; see, for example, Giné and Nickl (2010). This condition requires the existence of a limit distribution of an extreme value type for the supremum of a studentized empirical process (equivalently, for the supremum of a Gaussian process with the same covariance function as that of the studentized empirical process). The principal contribution of this paper is to remove the need for this classical condition. We show that a considerably weaker sufficient condition is derived from an anti-concentration property of the supremum of the approximating Gaussian process, and we derive an inequality leading to such a property for separable Gaussian processes. We refer to the new condition as a generalized SBR condition. Our new result shows that the supremum does not concentrate too fast around any value.

We then apply this result to derive a Gaussian multiplier bootstrap procedure for constructing honest confidence bands for nonparametric density estimators (this result can be applied in other nonparametric problems as well). An essential advantage of our approach is that it applies generically even in those cases where the limit distribution of the supremum of the studentized empirical process does not exist (or is unknown). This is of particular importance in problems where resolution levels or other tuning parameters have been chosen in a data-driven fashion, which is needed for adaptive constructions of the confidence bands. Finally, of independent interest is our introduction of a new, practical version of Lepski's method, which computes the optimal, non-conservative resolution levels via a Gaussian multiplier bootstrap method.

**1. Introduction.** Let  $X_1, \dots, X_n$  be i.i.d. random vectors with common unknown density  $f$  on  $\mathbb{R}^d$ . We are interested in constructing confidence

---

\*First version: October 2011. Revised April 12, 2014. V. Chernozhukov and D. Chetverikov are supported by a National Science Foundation grant. K. Kato is supported by the Grant-in-Aid for Young Scientists (B) (25780152), the Japan Society for the Promotion of Science.

*Keywords and phrases:* anti-concentration of separable Gaussian processes, honest confidence bands, Lepski's method, multiplier method, non-Donsker empirical processes

bands for  $f$  on a subset  $\mathcal{X} \subset \mathbb{R}^d$  that are *honest* to a given class  $\mathcal{F}$  of densities on  $\mathbb{R}^d$ . Typically,  $\mathcal{X}$  is a compact set on which  $f$  is bounded away from zero, and  $\mathcal{F}$  is a class of smooth densities such as a subset of a Hölder ball. A confidence band  $\mathcal{C}_n = \mathcal{C}_n(X_1, \dots, X_n)$  is a family of random intervals

$$\mathcal{C}_n := \{\mathcal{C}_n(x) = [c_L(x), c_U(x)] : x \in \mathcal{X}\}$$

that contains the graph of  $f$  on  $\mathcal{X}$  with a guaranteed probability. Following [31], a band  $\mathcal{C}_n$  is said to be *asymptotically honest with level  $\alpha \in (0, 1)$  for the class  $\mathcal{F}$*  if

$$\liminf_{n \rightarrow \infty} \inf_{f \in \mathcal{F}} \mathbb{P}_f(f(x) \in \mathcal{C}_n(x), \forall x \in \mathcal{X}) \geq 1 - \alpha.$$

Also, we say that a band  $\mathcal{C}_n$  is *asymptotically honest at a polynomial rate with level  $\alpha \in (0, 1)$  for the class  $\mathcal{F}$*  if

$$(1) \quad \inf_{f \in \mathcal{F}} \mathbb{P}_f(f(x) \in \mathcal{C}_n(x), \forall x \in \mathcal{X}) \geq 1 - \alpha - Cn^{-c}$$

for some constants  $c, C > 0$ .

Let  $\hat{f}_n(\cdot, l)$  be a generic estimator of  $f$  with a smoothing parameter  $l$ , say bandwidth or resolution level, where  $l$  is chosen from a candidate set  $\mathcal{L}_n$ ; see [26, 42, 44] for a textbook level introduction to the theory of density estimation. Let  $\hat{l}_n = \hat{l}_n(X_1, \dots, X_n)$  be a possibly data-dependent choice of  $l$  in  $\mathcal{L}_n$ . Denote by  $\sigma_{n,f}(x, l)$  the standard deviation of  $\sqrt{n}\hat{f}_n(x, l)$ , i.e.,  $\sigma_{n,f}(x, l) := (n \text{Var}_f(\hat{f}_n(x, l)))^{1/2}$ . Then we consider a confidence band of the form

$$(2) \quad \mathcal{C}_n(x) = \left[ \hat{f}_n(x, \hat{l}_n) - \frac{c(\alpha)\sigma_{n,f}(x, \hat{l}_n)}{\sqrt{n}}, \hat{f}_n(x, \hat{l}_n) + \frac{c(\alpha)\sigma_{n,f}(x, \hat{l}_n)}{\sqrt{n}} \right],$$

where  $c(\alpha)$  is a (possibly data-dependent) critical value determined to make the confidence band to have level  $\alpha$ . Generally,  $\sigma_{n,f}(x, l)$  is unknown and has to be replaced by an estimator.

A crucial point in construction of confidence bands is the computation of the critical value  $c(\alpha)$ . Assuming that  $\sigma_{n,f}(x, l)$  is positive on  $\mathcal{X} \times \mathcal{L}_n$ , define the stochastic process

$$(3) \quad Z_{n,f}(v) := Z_{n,f}(x, l) := \frac{\sqrt{n}(\hat{f}_n(x, l) - \mathbb{E}_f[\hat{f}_n(x, l)])}{\sigma_{n,f}(x, l)},$$

where  $v = (x, l) \in \mathcal{X} \times \mathcal{L}_n =: \mathcal{V}_n$ . We refer to  $Z_{n,f}$  as a “studentized process”. If, for the sake of simplicity, the bias  $|f(x) - \mathbb{E}_f[\hat{f}_n(x, l)]_{l=\hat{l}_n}|$  is sufficiently

small compared to  $\sigma_{n,f}(x, \hat{l}_n)$ , then

$$\begin{aligned} \mathbb{P}_f(f(x) \in \mathcal{C}_n(x), \forall x \in \mathcal{X}) &\approx \mathbb{P}_f\left(\sup_{x \in \mathcal{X}} |Z_{n,f}(x, \hat{l}_n)| \leq c(\alpha)\right) \\ &\geq \mathbb{P}_f\left(\sup_{v \in \mathcal{V}_n} |Z_{n,f}(v)| \leq c(\alpha)\right), \end{aligned}$$

so that the band (2) will be of level  $\alpha \in (0, 1)$  by taking

$$(4) \quad c(\alpha) = (1 - \alpha)\text{-quantile of } \|Z_{n,f}\|_{\mathcal{V}_n} := \sup_{v \in \mathcal{V}_n} |Z_{n,f}(v)|.$$

The critical value  $c(\alpha)$ , however, is infeasible since the finite sample distribution of the process  $Z_{n,f}$  is unknown. Instead, we estimate the  $(1 - \alpha)$ -quantile of  $\|Z_{n,f}\|_{\mathcal{V}_n}$ .

Suppose that one can find an appropriate centered Gaussian process  $G_{n,f}$  indexed by  $\mathcal{V}_n$  with known or estimable covariance structure such that  $\|Z_{n,f}\|_{\mathcal{V}_n}$  is close to  $\|G_{n,f}\|_{\mathcal{V}_n}$ . Then we may approximate the  $(1 - \alpha)$ -quantile of  $\|Z_{n,f}\|_{\mathcal{V}_n}$  by

$$c_{n,f}(\alpha) := (1 - \alpha)\text{-quantile of } \|G_{n,f}\|_{\mathcal{V}_n}.$$

Typically, one computes or approximates  $c_{n,f}(\alpha)$  by one of the following two methods.

1. Analytical method: derive analytically an approximated value of  $c_{n,f}(\alpha)$ , by using an explicit limit distribution or large deviation inequalities.
2. Simulation method: simulate the Gaussian process  $G_{n,f}$  to compute  $c_{n,f}(\alpha)$  numerically, by using, for example, a multiplier method.

The main purpose of this paper is to introduce a general approach to establishing the validity of the so-constructed confidence band. Importantly, our analysis does not rely on the existence of an explicit (continuous) limit distribution of any kind, which is a major difference from the previous literature. For the density estimation problem, if  $\mathcal{L}_n$  is a singleton, i.e., the smoothing parameter is chosen deterministically, the existence of such a continuous limit distribution, which is typically a Gumbel distribution, has been established for convolution kernel density estimators and *some* wavelet projection kernel density estimators [see 40, 1, 18, 20, 4, 5, 17]. We refer to the existence of the limit distribution as the Smirnov-Bickel-Rosenblatt (SBR) condition. However, the SBR condition has not been obtained for other density estimators such as non-wavelet projection kernel estimators based, for example, on Legendre polynomials or Fourier series. In addition,

to guarantee the existence of a continuous limit distribution often requires more stringent regularity conditions than a Gaussian approximation itself. More importantly, if  $\mathcal{L}_n$  is not a singleton, which is typically the case when  $\hat{l}_n$  is data-dependent, and so the randomness of  $\hat{l}_n$  has to be taken into account, it is often hard to determine an exact limit behavior of  $\|G_{n,f}\|_{\mathcal{V}_n}$ .

We thus take a different route and significantly generalize the SBR condition. Our key ingredient is the *anti-concentration* property of suprema of Gaussian processes that shows that suprema of Gaussian processes do not concentrate too fast. To some extent, this is a reverse of numerous concentration inequalities for Gaussian processes. In studying the effect of approximation and estimation errors on the coverage probability, it is required to know how the random variable  $\|G_{n,f}\|_{\mathcal{V}_n} := \sup_{v \in \mathcal{V}_n} |G_{n,f}(v)|$  concentrates or “anti-concentrates” around, say, its  $(1 - \alpha)$ -quantile. It is not difficult to see that  $\|G_{n,f}\|_{\mathcal{V}_n}$  itself has a continuous distribution, so that *with keeping  $n$  fixed*, the probability that  $\|G_{n,f}\|_{\mathcal{V}_n}$  falls into the interval with center  $c_{n,f}(\alpha)$  and radius  $\epsilon$  goes to 0 as  $\epsilon \rightarrow 0$ . However, what we need to know is the behavior of those probabilities when  $\epsilon$  depends on  $n$  and  $\epsilon = \epsilon_n \rightarrow 0$ . In other words, bounding explicitly “anti-concentration” probabilities for suprema of Gaussian processes is desirable. We will first establish bounds on the Lévy concentration function (see Definition 2.1) for suprema of Gaussian processes and then use these bounds to quantify the effect of approximation and estimation errors on the finite sample coverage probability. We say that a *generalized SBR* condition or simply an anti-concentration condition holds if  $\|G_{n,f}\|_{\mathcal{V}_n}$  concentrates sufficiently slowly, so that this effect is sufficiently small to yield asymptotically honest confidence bands.

As a substantive application of our results, we consider the problem of constructing honest *adaptive* confidence bands based on either convolution or wavelet projection kernel density estimators in Hölder classes  $\mathcal{F} \subset \cup_{t \in [\underline{t}, \bar{t}]} \Sigma(t, L)$  for some  $0 < \underline{t} < \bar{t} < \infty$  where  $\Sigma(t, L)$  is the Hölder ball of densities with radius  $L$  and smoothness level  $t$ . Following [6], we say the confidence band  $\mathcal{C}_n$  is *adaptive* if for every  $t, \epsilon > 0$  there exists  $C > 0$  such that for all  $n \geq 1$ ,

$$\sup_{f \in \mathcal{F} \cap \Sigma(t, L)} \mathbf{P}_f \left( \sup_{x \in \mathcal{X}} \lambda(\mathcal{C}_n(x)) > C r_n(t) \right) \leq \epsilon,$$

where  $\lambda$  denotes the Lebesgue measure on  $\mathbb{R}$  and  $r_n(t) := (\log n/n)^{t/(2t+d)}$ , the minimax optimal rate of convergence for estimating a density  $f$  in the function class  $\Sigma(t, L)$  in the sup-metric  $d_\infty(\hat{f}, f) = \sup_{x \in \mathcal{X}} |\hat{f}(x) - f(x)|$ . We use Lepski’s method [30, 2] to find an adaptive value of the smoothing parameter. Here our contribution is to introduce a *Gaussian multiplier*

*bootstrap implementation* of Lepski’s method. This is a practical proposal since previous implementations relied on conservative (one-sided) maximal inequalities and are not necessarily recommended for practice; see, for example, [19] for a discussion.

We should also emphasize that our techniques can also be used for constructing honest and/or adaptive confidence bands in many other nonparametric problems, but in this paper we focus on the density problem for the sake of clarity. Our techniques (anti-concentration of separable Gaussian processes (Theorem 2.1), and coupling inequalities (Theorems A.1 and A.2)) are of particular importance in non-Donsker settings since they allow to prove validity of the Gaussian multiplier bootstrap for approximating distributions of suprema of sequences of empirical processes of VC type function classes where the metric entropy of the process may increase with  $n$ . Thus, these techniques may be important in many nonparametric problems. For example, applications of our anti-concentration bounds can be found in [10] and [11], which consider the problems of nonparametric inference on a minimum of a function and nonparametric testing of qualitative hypotheses about functions, respectively.

1.1. *Related references.* Confidence bands in nonparametric estimation have been extensively studied in the literature. A classical approach, which goes back to [40] and [1], is to use explicit limit distributions of normalized suprema of studentized processes. A “Smirnov-Bickel-Rosenblatt type limit theorem” combines Gaussian approximation techniques and extreme value theory for Gaussian processes. It was argued that the convergence to normal extremes is rather slow despite that the Gaussian approximation is relatively fast [24]. To improve the finite sample coverage, bootstrap is often used in construction of confidence bands [see 12, 3]. However, to establish the validity of bootstrap confidence bands, researchers relied on the existence of continuous limit distributions of normalized suprema of original studentized processes. In the deconvolution density estimation problem, [32] considered confidence bands without using Gaussian approximation. In the current density estimation problem, their idea reads as bounding the deviation probability of  $\|\hat{f}_n - \mathbb{E}[\hat{f}_n(\cdot)]\|_\infty$  by using Talagrand’s [41] inequality and replacing the expected supremum by the Rademacher average. Such a construction is indeed general and applicable to many other problems, but is likely to be more conservative than our construction.

1.2. *Organization of the paper.* In the next section, we give a new anti-concentration inequality for suprema of Gaussian processes. Section 3 contains a theory of generic confidence band construction under high level con-

ditions. These conditions are easily satisfied both for convolution and projection kernel techniques under mild primitive assumptions, which are also presented in Section 3. Section 4 is devoted to constructing honest adaptive confidence bands in Hölder classes. Finally, most proofs are contained in the Appendix, and some proofs and discussions are put into the Supplemental Material [9].

1.3. *Notation.* In what follows, constants  $c, C, c_1, C_1, c_2, C_2, \dots$  are understood to be positive and independent of  $n$ . The values of  $c$  and  $C$  may change at each appearance but constants  $c_1, C_1, c_2, C_2, \dots$  are fixed. Throughout the paper,  $\mathbb{E}_n[\cdot]$  denotes the average over index  $1 \leq i \leq n$ , i.e., it simply abbreviates the notation  $n^{-1} \sum_{i=1}^n [\cdot]$ . For example,  $\mathbb{E}_n[g(X_i)] = n^{-1} \sum_{i=1}^n g(X_i)$ . For a set  $T$ , denote by  $\ell^\infty(T)$  the set of all bounded functions, that is, all functions  $z : T \rightarrow \mathbb{R}$  such that

$$\|z\|_T := \sup_{t \in T} |z(t)| < \infty.$$

Moreover, for a generic function  $g$ , we also use the notation  $\|g\|_\infty := \sup_x |g(x)|$  where the supremum is taken over the domain of  $g$ . For two random variables  $\xi$  and  $\eta$ , we write  $\xi \stackrel{d}{=} \eta$  if they share the same distribution. The standard Euclidean norm is denoted by  $|\cdot|$ .

**2. Anti-concentration of suprema of Gaussian processes.** The main purpose of this section is to derive an upper bound on the *Lévy concentration function* for suprema of separable Gaussian processes, where the terminology is adapted from [39]. Let  $(\Omega, \mathcal{A}, \mathbb{P})$  be the underlying (complete) probability space.

DEFINITION 2.1 (Lévy concentration function). Let  $Y = (Y_t)_{t \in T}$  be a separable stochastic process indexed by a semimetric space  $T$ . For all  $x \in \mathbb{R}$  and  $\epsilon \geq 0$ , let

$$(5) \quad p_{x,\epsilon}(Y) := \mathbb{P} \left( \left| \sup_{t \in T} Y_t - x \right| \leq \epsilon \right).$$

Then the *Lévy concentration function* of  $\sup_{t \in T} Y_t$  is defined for all  $\epsilon \geq 0$  as

$$(6) \quad p_\epsilon(Y) := \sup_{x \in \mathbb{R}} p_{x,\epsilon}(Y).$$

Likewise, define  $p_{x,\epsilon}(|Y|)$  by (5) with  $\sup_{t \in T} Y_t$  replaced by  $\sup_{t \in T} |Y_t|$  and define  $p_\epsilon(|Y|)$  by (6) with  $p_{x,\epsilon}(Y)$  replaced by  $p_{x,\epsilon}(|Y|)$ .

Let  $X = (X_t)_{t \in T}$  be a separable Gaussian process indexed by a semimetric space  $T$  such that  $\mathbb{E}[X_t] = 0$  and  $\mathbb{E}[X_t^2] = 1$  for all  $t \in T$ . Assume that  $\sup_{t \in T} X_t < \infty$  a.s. Our aim here is to obtain a qualitative bound on the concentration function  $p_\epsilon(X)$ . In a trivial example where  $T$  is a singleton, i.e.,  $X$  is a real standard normal random variable, it is immediate to see that  $p_\epsilon(X) \asymp \epsilon$  as  $\epsilon \rightarrow 0$ . A non-trivial case is that when  $T$  is not a singleton and both  $T$  and  $X$  are indexed by  $n = 1, 2, \dots$ , i.e.,  $T = T_n$  and  $X = X^n = (X_{n,t})_{t \in T_n}$ , and the complexity of the set  $\{X_{n,t} : t \in T_n\}$  (in  $L^2(\Omega, \mathcal{A}, \mathbb{P})$ ) is increasing in  $n$ . In such a case, it is typically not known whether  $\sup_{t \in T_n} X_{n,t}$  has a limiting distribution as  $n \rightarrow \infty$  and therefore it is not trivial at all whether, for any sequence  $\epsilon_n \rightarrow 0$ ,  $p_{\epsilon_n}(X^n) \rightarrow 0$  as  $n \rightarrow \infty$ .

The following is the first main result of this paper.

**THEOREM 2.1 (Anti-concentration for suprema of separable Gaussian processes).** *Let  $X = (X_t)_{t \in T}$  be a separable Gaussian process indexed by a semimetric space  $T$  such that  $\mathbb{E}[X_t] = 0$  and  $\mathbb{E}[X_t^2] = 1$  for all  $t \in T$ . Assume that  $\sup_{t \in T} X_t < \infty$  a.s. Then  $a(X) := \mathbb{E}[\sup_{t \in T} X_t] \in [0, \infty)$  and*

$$(7) \quad p_\epsilon(X) \leq 4\epsilon(a(X) + 1),$$

for all  $\epsilon \geq 0$ .

The similar conclusion holds for the concentration function of  $\sup_{t \in T} |X_t|$ .

**COROLLARY 2.1.** *Let  $X = (X_t)_{t \in T}$  be a separable Gaussian process indexed by a semimetric space  $T$  such that  $\mathbb{E}[X_t] = 0$  and  $\mathbb{E}[X_t^2] = 1$  for all  $t \in T$ . Assume that  $\sup_{t \in T} X_t < \infty$  a.s. Then  $a(|X|) := \mathbb{E}[\sup_{t \in T} |X_t|] \in [\sqrt{2/\pi}, \infty)$  and*

$$(8) \quad p_\epsilon(|X|) \leq 4\epsilon(a(|X|) + 1),$$

for all  $\epsilon \geq 0$ .

We refer to (7) and (8) as anti-concentration inequalities because they show that suprema of separable Gaussian processes can not concentrate too fast. The proof of Theorem 2.1 and Corollary 2.1 follows by extending the results in [8] where we derived anti-concentration inequalities for maxima of Gaussian random vectors. See the Appendix for a detailed exposition.



**3. Generic construction of honest confidence bands.** We go back to the analysis of confidence bands. Recall that we consider the following setting. We observe i.i.d. random vectors  $X_1, \dots, X_n$  with common unknown density  $f \in \mathcal{F}$  on  $\mathbb{R}^d$ , where  $\mathcal{F}$  is a nonempty subset of densities on  $\mathbb{R}^d$ . We denote by  $P_f$  the probability distribution corresponding to the density  $f$ . We first state the result on the construction of honest confidence bands under certain high level conditions and then show that these conditions hold for most commonly used kernel density estimators.

3.1. *Main Result.* Let  $\mathcal{X} \subset \mathbb{R}^d$  be a set of interest. Let  $\hat{f}_n(\cdot, l)$  be a generic estimator of  $f$  with a smoothing parameter  $l \in \mathcal{L}_n$  where  $\mathcal{L}_n$  is the candidate set. Denote by  $\sigma_{n,f}(x, l)$  the standard deviation of  $\sqrt{n}\hat{f}_n(x, l)$ . We assume that  $\sigma_{n,f}(x, l)$  is positive on  $\mathcal{V}_n := \mathcal{X} \times \mathcal{L}_n$  for all  $f \in \mathcal{F}$ . Define the studentized process  $Z_{n,f} = \{Z_{n,f}(v) : v = (x, l) \in \mathcal{V}_n\}$  by (3). Let

$$W_{n,f} := \|Z_{n,f}\|_{\mathcal{V}_n}$$

denote the supremum of the studentized process. We assume that  $W_{n,f}$  is a well-defined random variable. Let  $c_1, C_1$  be some positive constants. We will assume the following high level conditions.

CONDITION H1 (Gaussian approximation). *For every  $f \in \mathcal{F}$ , there exists (on a possibly enriched probability space) a sequence of random variables  $W_{n,f}^0$  such that (i)  $W_{n,f}^0 \stackrel{d}{=} \|G_{n,f}\|_{\mathcal{V}_n}$  where  $G_{n,f} = \{G_{n,f}(v) : v \in \mathcal{V}_n\}$  is a tight Gaussian random element in  $\ell^\infty(\mathcal{V}_n)$  with  $E[G_{n,f}(v)] = 0$ ,  $E[G_{n,f}(v)^2] = 1$  for all  $v \in \mathcal{V}_n$ , and  $E[\|G_{n,f}\|_{\mathcal{V}_n}] \leq C_1\sqrt{\log n}$ ; and moreover (ii)*

$$(9) \quad \sup_{f \in \mathcal{F}} P_f(|W_{n,f} - W_{n,f}^0| > \epsilon_{1n}) \leq \delta_{1n},$$

where  $\epsilon_{1n}$  and  $\delta_{1n}$  are some sequences of positive numbers bounded from above by  $C_1 n^{-c_1}$ .

Analysis of uniform confidence bands often relies on the classical Smirnov-Bickel-Rosenblatt (SBR) condition that states that for some sequences  $A_n$  and  $B_n$ ,

$$(10) \quad A_n(\|G_{n,f}\|_{\mathcal{V}_n} - B_n) \stackrel{d}{\rightarrow} Z, \text{ as } n \rightarrow \infty,$$

where  $Z$  is a Gumbel random variable; see, for example, [20]. Here both  $A_n$  and  $B_n$  are typically of order  $\sqrt{\log n}$ . However, this condition is often difficult to verify. Therefore, we propose to use a weaker condition (recall the definition of the Lévy concentration function given in Definition 2.1):

CONDITION H2 (Anti-concentration or Generalized SBR condition). *For any sequence  $\epsilon_n$  of positive numbers, we have*

$$(a) \sup_{f \in \mathcal{F}} p_{\epsilon_n}(|G_{n,f}|) \rightarrow 0 \text{ if } \epsilon_n \sqrt{\log n} \rightarrow 0; \text{ or}$$

$$(b) \sup_{f \in \mathcal{F}} p_{\epsilon_n}(|G_{n,f}|) \leq C_1 \epsilon_n \sqrt{\log n}.$$

Note that Condition H2-(a) follows trivially from H2-(b). In turn, under H1, Condition H2-(b) is a simple consequence of Corollary 2.1. Condition H2-(a) (along with Conditions H1 and H3-H6 below) is sufficient to show that the confidence bands are asymptotically honest but we will use Condition H2-(b) to show that the confidence bands are asymptotically honest at a polynomial rate. We refer to H2 as a generalized SBR condition because H2-(a) holds if (10) holds with  $A_n$  of order  $\sqrt{\log n}$ . An advantage of Condition H2 in comparison with the classical condition (10) is that H2 follows easily from Corollary 2.1.

Let  $\alpha \in (0, 1)$  be a fixed constant (confidence level). Recall that  $c_{n,f}(\alpha)$  is the  $(1 - \alpha)$ -quantile of the random variable  $\|G_{n,f}\|_{\mathcal{V}_n}$ . If  $G_{n,f}$  is pivotal, i.e., independent of  $f$ ,  $c_{n,f}(\alpha) = c_n(\alpha)$  can be directly computed, at least numerically. Otherwise, we have to approximate or estimate  $c_{n,f}(\alpha)$ . Let  $\hat{c}_n(\alpha)$  be an estimator or approximated value of  $c_{n,f}(\alpha)$ , where we assume that  $\hat{c}_n(\alpha)$  is nonnegative (which is reasonable since  $c_{n,f}(\alpha)$  is nonnegative). The following is concerned with a generic regularity condition on the accuracy of the estimator  $\hat{c}_n(\alpha)$ .

CONDITION H3 (Estimation error of  $\hat{c}_n(\alpha)$ ). *For some sequences  $\tau_n$ ,  $\epsilon_{2n}$ , and  $\delta_{2n}$  of positive numbers bounded from above by  $C_1 n^{-c_1}$ , we have*

$$(a) \sup_{f \in \mathcal{F}} \mathbb{P}_f(\hat{c}_n(\alpha) < c_{n,f}(\alpha + \tau_n) - \epsilon_{2n}) \leq \delta_{2n}; \text{ and}$$

$$(b) \sup_{f \in \mathcal{F}} \mathbb{P}_f(\hat{c}_n(\alpha) > c_{n,f}(\alpha - \tau_n) + \epsilon_{2n}) \leq \delta_{2n}.$$

In the next subsection, we shall verify this condition for the estimator  $\hat{c}_n(\alpha)$  based upon the Gaussian multiplier bootstrap method. Importantly, in this condition, we introduce the sequence  $\tau_n$  and compare  $\hat{c}_n(\alpha)$  with  $c_{n,f}(\alpha + \tau_n)$  and  $c_{n,f}(\alpha - \tau_n)$  instead of directly comparing it with  $c_{n,f}(\alpha)$ , which considerably simplifies verification of this condition. With  $\tau_n = 0$  for all  $n$ , we would need to have an upper bound on  $c_{n,f}(\alpha) - c_{n,f}(\alpha + \tau_n)$  and  $c_{n,f}(\alpha - \tau_n) - c_{n,f}(\alpha)$ , which might be difficult to obtain in general.

The discussion in the introduction presumes that  $\sigma_{n,f}(x, l)$  were known, but of course it has to be replaced by a suitable estimator in practice. Let

$\hat{\sigma}_n(x, l)$  be a generic estimator of  $\sigma_{n,f}(x, l)$ . Without loss of generality, we may assume that  $\hat{\sigma}_n(x, l)$  is nonnegative. Condition H4 below states a high-level assumption on the estimation error of  $\hat{\sigma}_n(x, l)$ . Verifying Condition H4 is rather standard for specific examples.

CONDITION H4 (Estimation error of  $\hat{\sigma}_n(\cdot)$ ). *For some sequences  $\epsilon_{3n}$  and  $\delta_{3n}$  of positive numbers bounded from above by  $C_1 n^{-c_1}$ ,*

$$\sup_{f \in \mathcal{F}} \mathbb{P}_f \left( \sup_{v \in \mathcal{V}_n} \left| \frac{\hat{\sigma}_n(v)}{\sigma_{n,f}(v)} - 1 \right| > \epsilon_{3n} \right) \leq \delta_{3n}.$$

We now consider strategies to deal with the bias term. We consider two possibilities. The first possibility is to control the bias explicitly, so that the confidence band contains the bias controlling term. This construction is inspired by [4]. The advantage of this construction is that it yields the confidence band the length of which shrinks at the minimax optimal rate with no additional inflating terms; see Theorem 4.1 below. The disadvantage, however, is that this construction yields a conservative confidence band in terms of coverage probability. We consider this strategy in Conditions H5 and H6 and Theorem 3.1. The other possibility is to undersmooth, so that the bias is asymptotically negligible, and hence the resulting confidence band contains no bias controlling terms. This is an often used strategy; see, for example, [20]. The advantage of this construction is that it sometimes yields an exact (non-conservative) confidence band, so that the confidence band covers the true function with probability  $1 - \alpha$  asymptotically exactly; see Corollary 3.1 below. The disadvantages, however, are that this method yields the confidence band that shrinks at the rate slightly slower than the minimax optimal rate, and that is centered around a non-optimal estimator. We consider the possibility of undersmoothing in Corollary 3.1 below. Note that Conditions H5 and H6 below are not assumed in Corollary 3.1.

We now consider the first possibility, that is we assume that the smoothing parameter  $\hat{l}_n := \hat{l}_n(X_1, \dots, X_n)$ , which is allowed to depend on the data, is chosen so that the bias can be controlled sufficiently well. Specifically, for all  $l \in \mathcal{L}_n$ , define

$$\Delta_{n,f}(l) := \sup_{x \in \mathcal{X}} \frac{\sqrt{n} |f(x) - \mathbb{E}_f[\hat{f}_n(x, l)]|}{\sigma_n(x, l)}.$$

We assume that there exists a sequence of random variables  $c'_n$ , which are known or can be calculated via simulations, that control  $\Delta_{n,f}(\hat{l}_n)$ . In particular, the theory in the next subsection assumes that  $c'_n$  is chosen as a multiple of the estimated high quantile of the supremum of certain Gaussian process.

CONDITION H5 (Bound on  $\Delta_{n,f}(\hat{l}_n)$ ). *For some sequence  $\delta_{4n}$  of positive numbers bounded from above by  $C_1 n^{-c_1}$ ,*

$$\sup_{f \in \mathcal{F}} \mathbb{P}_f \left( \Delta_{n,f}(\hat{l}_n) > c'_n \right) \leq \delta_{4n}.$$

In turn, we assume that  $c'_n$  can be controlled by  $u_n \sqrt{\log n}$  where  $u_n$  is a sequence of nonnegative positive numbers. Typically,  $u_n$  is either a bounded or slowly growing sequence; see, for example, our construction under primitive conditions in the next section.

CONDITION H6 (Bound on  $c'_n$ ). *For some sequences  $\delta_{5n}$  and  $u_n$  of positive numbers where  $\delta_{5n}$  is bounded from above by  $C_1 n^{-c_1}$ ,*

$$\sup_{f \in \mathcal{F}} \mathbb{P}_f \left( c'_n > u_n \sqrt{\log n} \right) \leq \delta_{5n}.$$

When  $\mathcal{L}_n$  is a singleton, conditions like H5 and H6 have to be assumed. When  $\mathcal{L}_n$  contains more than one element, that is we seek for an adaptive procedure, verification of Conditions H5 and H6 is non-trivial. In Section 4, we provide an example of such analysis.

We consider the confidence band  $\mathcal{C}_n = \{\mathcal{C}_n(x) : x \in \mathcal{X}\}$  defined by

$$(11) \quad \mathcal{C}_n(x) := \left[ \hat{f}_n(x, \hat{l}_n) - s_n(x, \hat{l}_n), \hat{f}_n(x, \hat{l}_n) + s_n(x, \hat{l}_n) \right],$$

where

$$(12) \quad s_n(x, \hat{l}_n) := (\hat{c}_n(\alpha) + c'_n) \hat{\sigma}_n(x, \hat{l}_n) / \sqrt{n}.$$

Define

$$\begin{aligned} \bar{\epsilon}_{n,f} &:= \epsilon_{1n} + \epsilon_{2n} + \epsilon_{3n}(c_{n,f}(\alpha) + u_n \sqrt{\log n}), \\ \delta_n &:= \delta_{1n} + \delta_{2n} + \delta_{3n} + \delta_{4n} + \delta_{5n}. \end{aligned}$$

We are now in position to state the main result of this section. Recall the definition of Lévy concentration function (Definition 2.1).

**THEOREM 3.1 (Honest generic confidence bands).** *Suppose that Conditions H1 and H3-H6 are satisfied. Then*

$$(13) \quad \inf_{f \in \mathcal{F}} \mathbb{P}_f(f \in \mathcal{C}_n) \geq (1 - \alpha) - \delta_n - \tau_n - p_{\bar{\epsilon}_{n,f}}(|G_{n,f}|).$$

If, in addition, Condition H2-(a) is satisfied and  $\epsilon_{3n}u_n\sqrt{\log n} \leq C_1n^{-c_1}$ , then

$$(14) \quad \liminf_{n \rightarrow \infty} \inf_{f \in \mathcal{F}} \mathbb{P}_f(f \in \mathcal{C}_n) \geq 1 - \alpha,$$

and if, in addition, Condition H2-(b) is satisfied, then

$$(15) \quad \inf_{f \in \mathcal{F}} \mathbb{P}_f(f \in \mathcal{C}_n) \geq 1 - \alpha - Cn^{-c},$$

where  $c$  and  $C$  are constants depending only on  $\alpha$ ,  $c_1$  and  $C_1$ .

COMMENT 3.1 (Honest confidence bands). Theorem 3.1 shows that the confidence band defined in (11) and (12) is asymptotically honest with level  $\alpha$  for the class  $\mathcal{F}$ . Moreover, under condition H2-(b), the coverage probability can be smaller than  $1 - \alpha$  only by a polynomially small term  $Cn^{-c}$  uniformly over the class  $\mathcal{F}$ . That is, in this case the confidence band is *asymptotically honest at a polynomial rate* as defined in (1). ■

COMMENT 3.2 (Advantages of Theorem 3.1). An advantage of Theorem 3.1 is that it does not require the classical SBR condition that is often difficult to obtain. Instead, it only requires a weaker generalized SBR condition (H2), which allows us to control the effect of estimation and approximation errors on the coverage probabilities. In the next subsection, we will show that as long as the bias  $\Delta_{n,f}(\hat{l}_n)$  can be controlled, our theorem applies when  $\hat{f}_n(\cdot)$  is defined using either convolution or projection kernels under mild conditions, and, as far as projection kernels are concerned, it covers estimators based on compactly supported wavelets, Battle-Lemarié wavelets of *any* order as well as other non-wavelet projection kernels such as those based on Legendre polynomials and Fourier series. When  $\mathcal{L}_n$  is a singleton, the SBR condition for compactly supported wavelets was obtained in [5] under certain assumptions that can be verified *numerically* for any given wavelet, for Battle-Lemarié wavelets of degree up-to 4 in [20], and for Battle-Lemarié wavelets of degree higher than 4 in [17]. To the best of our knowledge, the SBR condition for non-wavelet projection kernel functions (such as those based on Legendre polynomials and Fourier series) has not been obtained in the literature. In addition, and perhaps most importantly, there are no results in the literature on the SBR condition when  $\mathcal{L}_n$  is not a singleton. Finally, the SBR condition, being based on extreme value theory, yields only a logarithmic (in  $n$ ) rate of approximation of coverage probability; that is, this approach is *asymptotically honest at a logarithmic rate*.

In contrast, our approach can lead to confidence bands that are asymptotically honest at a polynomial rate; see (15). Note also that one can obtain confidence bands that would be asymptotically honest at a polynomial rate with level  $\alpha$  by considering confidence bands that are asymptotically honest with level  $\alpha' < \alpha$  but such confidence bands would in general be wider than those provided by our approach. ■

COMMENT 3.3. [On dependence of constants  $c, C$  on  $\alpha$  in (15)] We note that (15) is a *non-asymptotic* bound. In addition, it immediately follows from the proof of Theorem 3.1 that the constants  $c$  and  $C$  in (15) can be chosen to be independent of  $\alpha$  (thus, they depend only on  $c_1$  and  $C_1$ ) as long as

$$(16) \quad |\log \alpha| \leq C_1 \log n.$$

Therefore, (15) can be applied with  $\alpha = \alpha_n$  depending on  $n$  as long as (16) holds (and Condition H3 is satisfied for the given sequence  $\alpha = \alpha_n$ ). ■

COMMENT 3.4 (On the condition  $\epsilon_{3n} u_n \sqrt{\log n} \leq C_1 n^{-c_1}$ ). The second part of Theorem 3.1 requires the condition that  $\epsilon_{3n} u_n \sqrt{\log n} \leq C_1 n^{-c_1}$ . This is a very mild assumption. Indeed, under Condition H4,  $\epsilon_{3n} \leq C_1 n^{-c_1}$ , so that the assumption that  $\epsilon_{3n} u_n \sqrt{\log n} \leq C_1 n^{-c_1}$  is met (with possibly different constants  $c_1$  and  $C_1$ ) as long as  $u_n$  is bounded from above by a slowly growing sequence, for example,  $u_n \leq C_1 \log n$ , which is typically the case; see, for example, our construction in Section 4. ■

The confidence band defined in (11) and (12) is constructed so that the bias  $\Delta_{n,f}(\hat{l}_n)$  is controlled explicitly via the random variable  $c'_n$ . Alternatively, one can choose to undersmooth so that the bias is negligible asymptotically. To cover this possibility, we note that it follows from the proof of Theorem 3.1 that if  $u_n \log n \rightarrow 0$  or  $u_n \log n \leq C_1 n^{-c_1}$ , then conclusions (14) or (15) of Theorem 3.1 continue to hold, respectively, with  $s_n(x, \hat{l}_n)$  in (12) replaced by  $\hat{c}_n(\alpha) \hat{\sigma}_n(x, \hat{l}_n) / \sqrt{n}$ . Thus, obtaining the asymptotically honest at a polynomial rate confidence band requires polynomial undersmoothing ( $u_n \log n \leq C_1 n^{-c_1}$ ) but, on the other hand, logarithmic undersmoothing ( $u_n \log n \rightarrow 0$ ) suffices if polynomial rate is not required. Moreover, if  $\mathcal{L}_n$  is a singleton, it is possible to show that the confidence band is asymptotically exact, with a polynomial convergence rate (21) under the condition  $u_n \log n \leq C_1 n^{-c_1}$ . We collect these observations into the following corollary, the detailed proof of which can be found in the Supplemental Material.

**COROLLARY 3.1 (Honest generic confidence bands with under-smoothing).** Consider the confidence band  $\tilde{\mathcal{C}}_n = \{\tilde{\mathcal{C}}_n(x) : x \in \mathcal{X}\}$  defined by

$$\tilde{\mathcal{C}}_n(x) := \left[ \hat{f}_n(x, \hat{l}_n) - \tilde{s}_n(x, \hat{l}_n), \hat{f}_n(x, \hat{l}_n) + \tilde{s}_n(x, \hat{l}_n) \right],$$

where

$$\tilde{s}_n(x, \hat{l}_n) := \hat{c}_n(\alpha) \hat{\sigma}_n(x, \hat{l}_n) / \sqrt{n}.$$

Suppose that Conditions H1, H3, and H4 are satisfied. In addition, assume that for some sequences  $\delta_{6n}$  and  $u_n$  of positive numbers,

$$(17) \quad \sup_{f \in \mathcal{F}} \mathbb{P}_f \left( \Delta_{n,f}(\hat{l}_n) > u_n \sqrt{\log n} \right) \leq \delta_{6n},$$

where  $\delta_{6n}$  is bounded from above by  $C_1 n^{-c_1}$ . If Condition H2-(a) holds and  $u_n \log n \rightarrow 0$ , then

$$(18) \quad \liminf_{n \rightarrow \infty} \inf_{f \in \mathcal{F}} \mathbb{P}_f(f \in \tilde{\mathcal{C}}_n) \geq 1 - \alpha.$$

If Condition H2-(b) holds and  $u_n \log n \leq C_1 n^{-c_1}$ , then

$$(19) \quad \inf_{f \in \mathcal{F}} \mathbb{P}_f(f \in \tilde{\mathcal{C}}_n) \geq 1 - \alpha - C n^{-c}.$$

Moreover, assume in addition that  $\mathcal{L}_n$  is a singleton. If Condition H2-(a) holds and  $u_n \log n \rightarrow 0$ , then

$$(20) \quad \lim_{n \rightarrow \infty} \sup_{f \in \mathcal{F}} \left| \mathbb{P}_f(f \in \tilde{\mathcal{C}}_n) - (1 - \alpha) \right| = 0.$$

If Condition H2-(b) and  $u_n \log n \leq C_1 n^{-c_1}$ , then

$$(21) \quad \sup_{f \in \mathcal{F}} \left| \mathbb{P}_f(f \in \tilde{\mathcal{C}}_n) - (1 - \alpha) \right| \leq C n^{-c}.$$

Here  $c$  and  $C$  are constants depending only on  $\alpha$ ,  $c_1$  and  $C_1$ .

**COMMENT 3.5 (Other methods for controlling bias term).** In practice, there can be other methods for controlling the bias term. For example, an alternative approach is to estimate the bias function in a pointwise manner and construct bias corrected confidence bands; see, for example, [45] in the nonparametric regression case. A yet alternative approach to controlling the bias based upon bootstrap in construction of confidence bands is proposed and studied by the recent paper of [25]. ■

COMMENT 3.6 (On dependence of constants  $c, C$  on  $\alpha$  in (19) and (21)). Similar to Comment 3.3, we note that (19) and (21) are *non-asymptotic* bounds, and it immediately follows from the proof of Corollary 3.1 that these bounds apply with  $\alpha = \alpha_n$  depending on  $n$  and constants  $c$  and  $C$  depending only on  $c_1$  and  $C_1$  as long as  $|\log \alpha| \leq C_1 \log n$  (in case of (19)) and  $|\log(\alpha - \tau_n)| \leq C_1 \log n$  (in case of (21)). ■

3.2. *Verifying Conditions H1-H4 for confidence bands constructed using common density estimators via Gaussian multiplier bootstrap.* We now argue that when  $\hat{c}_n(\alpha)$  is constructed via Gaussian multiplier bootstrap, Conditions H1-H4 hold for common density estimators – specifically, both for convolution and for projection kernel density estimators under mild assumptions on the kernel function.

Let  $\{K_l\}_{l \in \mathcal{L}_n}$  be a family of kernel functions where  $K_l : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$  and  $l$  is a smoothing parameter. We consider kernel density estimators of the form

$$(22) \quad \hat{f}_n(x, l) := \mathbb{E}_n[K_l(X_i, x)] = \frac{1}{n} \sum_{i=1}^n K_l(X_i, x),$$

where  $x \in \mathcal{X}$  and  $l \in \mathcal{L}_n$ . The variance of  $\sqrt{n}\hat{f}_n(x, l)$  is given by

$$\sigma_{n,f}^2(x, l) := \mathbb{E}_f[K_l(X_1, x)^2] - (\mathbb{E}_f[K_l(X_1, x)])^2.$$

We estimate  $\sigma_{n,f}^2(x, l)$  by

$$(23) \quad \hat{\sigma}_n^2(x, l) := \frac{1}{n} \sum_{i=1}^n K_l(X_i, x)^2 - \hat{f}_n(x, l)^2.$$

This is a sample analogue estimator.

**Examples.** Our general theory covers a wide class of kernel functions, such as convolution, wavelet projection, and non-wavelet projection kernels.

(i) *Convolution kernel.* Consider a function  $K : \mathbb{R} \rightarrow \mathbb{R}$ . Let  $\mathcal{L}_n \subset (0, \infty)$ . Then for  $x = (x_1, \dots, x_d)' \in \mathbb{R}^d$ ,  $y = (y_1, \dots, y_d)' \in \mathbb{R}^d$ , and  $l \in \mathcal{L}_n$ , the convolution kernel function is defined by

$$(24) \quad K_l(y, x) := 2^{ld} \prod_{1 \leq m \leq d} K\left(2^l(y_m - x_m)\right).$$

Here  $2^{-l}$  is the bandwidth parameter.



(ii) *Wavelet projection kernel.* Consider a father wavelet  $\phi$ , i.e., a function  $\phi$  such that (a)  $\{\phi(\cdot - k) : k \in \mathbb{Z}\}$  is an orthonormal system in  $L_2(\mathbb{R})$ , (b) the spaces  $V_j = \{\sum_k c_k \phi(2^j x - k) : \sum_k c_k^2 < \infty\}$ ,  $j = 0, 1, 2, \dots$ , are nested in the sense that  $V_j \subset V_{j'}$  whenever  $j \leq j'$ , and (c)  $\cup_{j \geq 0} V_j$  is dense in  $L_2(\mathbb{R})$ . Let  $\mathcal{L}_n \subset \mathbb{N}$ . Then for  $x = (x_1, \dots, x_d)' \in \mathbb{R}^d$ ,  $y = (y_1, \dots, y_d)' \in \mathbb{R}^d$ , and  $l \in \mathcal{L}_n$ , the wavelet projection kernel function is defined by

$$(25) \quad K_l(y, x) := 2^{ld} \sum_{k_1, \dots, k_d \in \mathbb{Z}} \prod_{1 \leq m \leq d} \phi(2^l y_m - k_m) \prod_{1 \leq m \leq d} \phi(2^l x_m - k_m).$$

Here  $l$  is the resolution level. We refer to [13] and [26] as basic references on wavelet theory.

(iii) *Non-wavelet projection kernel.* Let  $\{\varphi_j : j = 1, \dots, \infty\}$  be an orthonormal basis of  $L_2(\mathcal{X})$ , the space of square integrable (with respect to Lebesgue measure) functions on  $\mathcal{X}$ . Let  $\mathcal{L}_n \subset (0, \infty)$ . Then for  $x = (x_1, \dots, x_d)' \in \mathbb{R}^d$ ,  $y = (y_1, \dots, y_d)' \in \mathbb{R}^d$ , and  $l \in \mathcal{L}_n$ , the non-wavelet projection kernel function is defined by

$$(26) \quad K_l(y, x) := \sum_{j=1}^{\lfloor 2^{ld} \rfloor} \varphi_j(y) \varphi_j(x),$$

where  $\lfloor a \rfloor$  is the largest integer that is smaller than or equal to  $a$ . Here  $\lfloor 2^{ld} \rfloor$  is the number of series (basis) terms used in the estimation. When  $d = 1$  and  $\mathcal{X} = [-1, 1]$ , examples of orthonormal bases are Fourier basis

$$(27) \quad \{1, \cos(\pi x), \cos(2\pi x), \dots\}$$

and Legendre polynomial basis

$$(28) \quad \{1, (3/2)^{1/2} x, (5/8)^{1/2} (3x^2 - 1), \dots\}.$$

When  $d > 1$  and  $\mathcal{X} = [-1, 1]^d$ , one can take tensor products of bases for  $d = 1$ . ■

We assume that the critical value  $\hat{c}_n(\alpha)$  is obtained via the multiplier bootstrap method:

**ALGORITHM 1 (Gaussian multiplier bootstrap).** *Let  $\xi_1, \dots, \xi_n$  be independent  $N(0, 1)$  random variables that are independent of the data  $X_1^n := \{X_1, \dots, X_n\}$ . Let  $\xi_1^n := \{\xi_1, \dots, \xi_n\}$ . For all  $x \in \mathcal{X}$  and  $l \in \mathcal{L}_n$ , define a Gaussian multiplier process:*

$$(29) \quad \hat{\mathbb{G}}_n(x, l) := \hat{\mathbb{G}}_n(X_1^n, \xi_1^n)(x, l) := \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i \frac{K_l(X_i, x) - \hat{f}_n(x, l)}{\hat{\sigma}_n(x, l)}.$$

Then the estimated critical value  $\hat{c}_n(\alpha)$  is defined as

$$\hat{c}_n(\alpha) = \text{conditional } (1 - \alpha)\text{-quantile of } \|\hat{\mathbb{G}}_n\|_{\mathcal{V}_n} \text{ given } X_1^n.$$

Gaussian multiplier bootstrap is a special case of a more general exchangeable bootstrap; see, for example, [37]. We refer the reader to [22] for the first systematic use of the Gaussian multipliers and to [29] and [23] for conditional multiplier central limit theorems in the Donsker setting.

Let

$$\mathcal{K}_{n,f} := \left\{ \frac{K_l(\cdot, x)}{\sigma_{n,f}(x, l)} : (x, l) \in \mathcal{X} \times \mathcal{L}_n \right\}$$

denote the class of studentized kernel functions and define

$$\sigma_n = \sup_{f \in \mathcal{F}} \sup_{g \in \mathcal{K}_{n,f}} (\mathbb{E}_f[g(X_1)^2])^{1/2}.$$

Note that  $\sigma_n \geq 1$ .

For a given class  $\mathcal{G}$  of measurable functions on a probability space  $(S, \mathcal{S}, Q)$  and  $\epsilon > 0$ , the  $\epsilon$ -covering number of  $\mathcal{G}$  with respect to the  $L_2(Q)$ -semimetric is denoted by  $N(\mathcal{G}, L_2(Q), \epsilon)$  (see Chapter 2 of [43] on details of covering numbers). We will use the following definition of VC type classes:

**DEFINITION 3.1 (VC type class).** Let  $\mathcal{G}$  be a class of measurable functions on a measurable space  $(S, \mathcal{S})$ , and let  $b > 0$ ,  $a \geq e$ , and  $v \geq 1$  be some constants. Then the class  $\mathcal{G}$  is called VC( $b, a, v$ ) type class if it is uniformly bounded in absolute value by  $b$  (i.e.,  $\sup_{g \in \mathcal{G}} \|g\|_\infty \leq b$ ) and the covering numbers of  $\mathcal{G}$  satisfy

$$\sup_Q N(\mathcal{G}, L_2(Q), b\tau) \leq (a/\tau)^v, \quad 0 < \tau < 1,$$

where the supremum is taken over all finitely discrete probability measures  $Q$  on  $(S, \mathcal{S})$ .

Then we will assume the following condition.

**CONDITION VC.** *There exist sequences  $b_n > 0$ ,  $a_n \geq e$ , and  $v_n \geq 1$  such that for every  $f \in \mathcal{F}$ , the class  $\mathcal{K}_{n,f}$  is VC( $b_n, a_n, v_n$ ) type and pointwise measurable.*

We refer to Chapter 2.3 of [43] for the definition of pointwise measurable classes of functions. We note that Condition VC is a mild assumption, which

we verify for common constructions in Appendix F (as a part of proving results for the next section; see Comment 3.5 below); see also Appendix I.

For some sufficiently large absolute constant  $A$ , take

$$K_n := Av_n (\log n \vee \log(a_n b_n / \sigma_n)).$$

We will assume without loss of generality that  $K_n \geq 1$  for all  $n$ . The following theorem verifies Conditions H1-H4 with so defined  $\hat{\sigma}_n^2(x, l)$  and  $\hat{c}_n(\alpha)$  under Condition VC, using the critical values constructed via Algorithm 1.

**THEOREM 3.2 (Conditions H1-H4 Hold for Our Construction).** *Suppose that Condition VC is satisfied and there exist constants  $c_2, C_2 > 0$  such that  $b_n^2 \sigma_n^4 K_n^4 / n \leq C_2 n^{-c_2}$ . Then Conditions H1-H4, including both H2-(a) and H2-(b), hold with some constants  $c_1, C_1 > 0$  that depend only on  $c_2, C_2$ .*

**COMMENT 3.7 (Convolution and wavelet projection kernels).** The assumption of Theorem 3.2 holds for convolution and wavelet projection kernels under mild conditions on the resolution level  $l$ . It follows from Lemma F.2 in Appendix F that, under mild regularity conditions, for convolution and wavelet projection kernel functions,  $\sigma_n \leq C$  and Condition VC holds with  $b_n \leq C 2^{l_{\max, n} d / 2}$ ,  $a_n \leq C$ , and  $v_n \leq C$  for some  $C > 0$  where  $l_{\max, n} = \sup\{\mathcal{L}_n\}$ . Hence, for these kernel functions, the assumption that  $b_n^2 \sigma_n^4 K_n^4 / n \leq C_2 n^{-c_2}$  reduces to

$$2^{l_{\max, n} d} (\log^4 n) / n \leq C_2 n^{-c_2}$$

(with possibly different constants  $c_2, C_2$ ), which is a mild requirement on the bandwidth value or resolution level. This is a very mild assumption on the possible resolution levels. Similar comments apply to non-wavelet projection kernels with Fourier and Legendre polynomial bases. See Appendix I in the Supplemental Material. ■

**COMMENT 3.8 (On Condition H3).** We note that under conditions of Theorem 3.2, Condition H3 remains true with the same constants  $c_1$  and  $C_1$  even if  $\alpha = \alpha_n$  depends on  $n$  (if we define  $c_{n, f}(\beta) = 0$  for  $\beta \geq 1$  and  $c_{n, f}(\beta) = \infty$  for  $\beta \leq 0$ ). To see this, note that according to Theorem 3.2, constants  $c_1$  and  $C_1$  in Condition H3 depend only on constants  $c_2$  and  $C_2$ , and do not depend on  $\alpha$ . ■

#### 4. Honest and Adaptive Confidence Bands in Hölder Classes.

In this section, we study the problem of constructing honest adaptive confidence bands in Hölder smoothness classes. Recall that for  $t, L > 0$ , the Hölder ball of densities with radius  $L$  and smoothness level  $t$  is defined by

$$\Sigma(t, L) := \left\{ f : \mathbb{R}^d \rightarrow \mathbb{R} : f \text{ is a } \lfloor t \rfloor\text{-times continuously differentiable density,} \right. \\ \left. \|D^\alpha f\|_\infty \leq L, \forall |\alpha| \leq \lfloor t \rfloor, \sup_{x \neq y} \frac{|D^\alpha f(x) - D^\alpha f(y)|}{|x - y|^{t - \lfloor t \rfloor}} \leq L, \forall |\alpha| = \lfloor t \rfloor \right\},$$

where  $\lfloor t \rfloor$  denotes the largest integer smaller than  $t$ , and for a multi-index  $\alpha = (\alpha_1, \dots, \alpha_d)$  with  $|\alpha| = \alpha_1 + \dots + \alpha_d$ ,  $D^\alpha f(x) := \partial^{|\alpha|} f(x) / \partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}$  [see, for example 42]. We assume that for some  $0 < \underline{t} \leq \bar{t} < \infty$  and  $L \geq 1$ ,

$$(30) \quad \mathcal{F} \subset \cup_{t \in [\underline{t}, \bar{t}]} \Sigma(t, L),$$

and consider the confidence band  $\mathcal{C}_n = \{\mathcal{C}_n(x) : x \in \mathcal{X}\}$  of the form (11) and (12), where  $\mathcal{X}$  is a (suitable) compact set in  $\mathbb{R}^d$ .

We begin by stating our assumptions. First, we restrict attention to kernel density estimators  $\hat{f}_n$  based on either convolution or wavelet projection kernel functions. Let  $r$  be an integer such that  $r \geq 2$  and  $r > \bar{t}$ .

**CONDITION L1 (Density estimator).** *The density estimator  $\hat{f}_n$  is either a convolution or wavelet projection kernel density estimator defined in (22), (24), and (25). For convolution kernels, the function  $K : \mathbb{R} \rightarrow \mathbb{R}$  has compact support and is of bounded variation, and moreover is such that  $\int K(s) ds = 1$  and  $\int s^j K(s) dx = 0$  for  $j = 1, \dots, r - 1$ . For wavelet projection kernels, the function  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  is either a compactly supported father wavelet of regularity  $r - 1$  (that is,  $\phi$  is  $(r - 1)$ -times continuously differentiable), or a Battle-Lemarié wavelet of regularity  $r - 1$ .*

The assumptions stated in Condition L1 are commonly used in the literature. See [16] for a more general class of convolution kernel functions that would suffice for our results. Details on compactly supported and Battle-Lemarié wavelets can be found in Chapters 6 and 5.4 of [13], respectively.

It is known that if the function class  $\mathcal{F}$  is sufficiently large (for example, if  $\mathcal{F} = \Sigma(t, L) \cup \Sigma(t', L)$  for  $t' > t$ ), the construction of honest adaptive confidence bands is not possible; see [33]. Therefore, following [20], we will restrict the function class  $\mathcal{F} \subset \cup_{t \in [\underline{t}, \bar{t}]} \Sigma(t, L)$  in a suitable way, as follows:

**CONDITION L2 (Bias bounds).** *There exist constants  $l_0, c_3, C_3 > 0$  such that for every  $f \in \mathcal{F} \subset \cup_{t \in [\underline{t}, \bar{t}]} \Sigma(t, L)$ , there exists  $t \in [\underline{t}, \bar{t}]$  with*

$$(31) \quad c_3 2^{-lt} \leq \sup_{x \in \mathcal{X}} |E_f[\hat{f}_n(x, l)] - f(x)| \leq C_3 2^{-lt},$$

for all  $l \geq l_0$ .

This condition is inspired by the path-breaking work of [20] (see also [36]). It can be interpreted as the requirement that the functions  $f$  in the class  $\mathcal{F}$  are “self-similar” in the sense that their regularity remains the same at large and small scales; see also [4]. To put it differently, “self-similarity” could be understood as the requirement that the bias of the kernel approximation to  $f$  with bandwidth  $2^{-l}$  remains approximately proportional to  $(2^{-l})^t$  – i.e. not much smaller or not much bigger – for all small values of the bandwidth  $2^{-l}$ .

It is useful to note that the upper bound in (31) holds for all  $f \in \Sigma(t, L)$  (for sufficiently large  $C_3$ ) under Condition L1; see, for example, Theorem 9.3 in [26]. In addition, [20] showed that under Condition L1, the restriction due to the lower bound in (31) is weak in the sense that the set of elements of  $\Sigma(t, L)$  for which the lower bound in (31) does not hold is “topologically small”. Moreover, they showed that the minimax optimal rate of convergence in the sup-norm over  $\Sigma(t, L)$  coincide with that over the set of elements of  $\Sigma(t, L)$  for which Condition L2 holds. We refer to [20] for a detailed and deep discussion of these conditions and results.

We also note that, depending on the problem, construction of honest adaptive confidence bands is often possible under somewhat weaker conditions than that in L2. For example, if we are interested in the function class  $\Sigma(t, L) \cup \Sigma(t', L)$  for some  $t' > t$ , [27] showed that it is necessary and sufficient to exclude functions  $\Sigma(t, L) \setminus \Sigma(t, L, \rho_n)$  where  $\Sigma(t, L, \rho_n) = \{f \in \Sigma(t, L) : \inf_{g \in \Sigma(t', L)} \|g - f\|_\infty \geq \rho_n\}$  and where  $\rho_n > 0$  is allowed to converge to zero as  $n$  increases but sufficiently slowly. If we are interested in the function class  $\cup_{t \in [\underline{t}, \bar{t}]} \Sigma(t, L)$ , [4] showed that (essentially) necessary and sufficient condition can be written in the form of the bound from below on the rate with which wavelet coefficients of the density  $f$  are allowed to decrease. Here we prefer to work with Condition L2 directly because it is directly related to the properties of the estimator  $\hat{f}_n$  and does not require any further specifications of the function class  $\mathcal{F}$ .

In order to introduce the next condition, we need to observe that under Condition L2, for every  $f \in \mathcal{F}$ , there exists a *unique*  $t \in [\underline{t}, \bar{t}]$  satisfying (31); indeed, if  $t_1 < t_2$ , then for any  $c, C > 0$ , there exists  $\bar{l}$  such that  $C2^{-lt_2} < c2^{-lt_1}$  for all  $l \geq \bar{l}$ , so that for each  $f \in \mathcal{F}$  condition (31) can hold for all  $l \geq l_0$  for at most one value of  $t$ . This defines the map

$$(32) \quad t : \mathcal{F} \rightarrow [\underline{t}, \bar{t}], \quad f \mapsto t(f).$$

The next condition states our assumptions on the candidate set  $\mathcal{L}_n$  of the values of the smoothing parameter:

CONDITION L3 (Candidate set). *There exist constants  $c_4, C_4 > 0$  such that for every  $f \in \mathcal{F}$ , there exists  $l \in \mathcal{L}_n$  with*

$$(33) \quad \left( \frac{c_4 \log n}{n} \right)^{1/(2t(f)+d)} \leq 2^{-l} \leq \left( \frac{C_4 \log n}{n} \right)^{1/(2t(f)+d)},$$

for the map  $t : f \mapsto t(f)$  defined in (32). In addition, the candidate set is  $\mathcal{L}_n = [l_{\min, n}, l_{\max, n}] \cap \mathbb{N}$ .

This condition thus ensures via (33) that the candidate set  $\mathcal{L}_n$  contains an appropriate value of the smoothing parameter that leads to the optimal rate of convergence for every density  $f \in \mathcal{F}$ .

Finally, we will make the following mild condition:

CONDITION L4 (Density bounds). *There exist constants  $\delta, \underline{f}, \bar{f} > 0$  such that for all  $f \in \mathcal{F}$ ,*

$$(34) \quad f(x) \geq \underline{f} \text{ for all } x \in \mathcal{X}^\delta \text{ and } f(x) \leq \bar{f} \text{ for all } x \in \mathbb{R}^d,$$

where  $\mathcal{X}^\delta$  is the  $\delta$ -enlargement of  $\mathcal{X}$ , i.e.,  $\mathcal{X}^\delta = \{x \in \mathbb{R}^d : \inf_{y \in \mathcal{X}} |x - y| \leq \delta\}$ .

We now discuss how we choose various parameters in the confidence band  $\mathcal{C}_n$ . In the previous section, we have shown how to obtain honest confidence bands as long as we can control the bias  $\Delta_{n, f}(\hat{l}_n)$  appropriately. So to construct honest adaptive confidence bands, we seek a method to choose the smoothing parameter  $\hat{l}_n \in \mathcal{L}_n$  so that the bias  $\Delta_{n, f}(\hat{l}_n)$  can be controlled and, at the same time, the confidence band  $\mathcal{C}_n$  is adaptive.

Let  $\bar{\mathcal{V}}_n := \{(x, l, l') : x \in \mathcal{X}, l, l' \in \mathcal{L}_n, l < l'\}$ , and for  $(x, l, l') \in \bar{\mathcal{V}}_n$ , denote

$$\tilde{\sigma}_n(x, l, l') := \left( \frac{1}{n} \sum_{i=1}^n (K_l(X_i, x) - K_{l'}(X_i, x))^2 - (\hat{f}_n(x, l) - \hat{f}_n(x, l'))^2 \right)^{1/2}.$$

Also, for some small  $c_\sigma > 0$ , let

$$\hat{\sigma}_n(x, l, l') := (c_\sigma \tilde{\sigma}_n(x, l')) \vee \tilde{\sigma}_n(x, l, l')$$

denote the truncated version of  $\tilde{\sigma}_n(x, l, l')$ . In practice, we suggest setting  $c_\sigma = 0.5(1 - 2^{-d/2})$  (the constant  $c_\sigma$  is chosen so that with probability approaching one,  $\hat{\sigma}_n(x, l, l') = \tilde{\sigma}_n(x, l, l')$  for all  $(x, l, l') \in \bar{\mathcal{V}}_n$  for convolution kernel estimators, and for all  $(x, l, l') \in \bar{\mathcal{V}}_n$  with  $l \leq l' - s$  for wavelet projection kernel estimators where  $s$  is some constant; see Lemmas F.2 and F.4 in the Supplemental Material).

There exist several techniques in the literature to construct  $\hat{l}_n$  so that  $\Delta_{n,f}(\hat{l}_n)$  can be controlled and the confidence band  $\mathcal{C}_n$  is adaptive; see, for example, [35] for a thorough introduction. One of the most important such techniques is the Lepski method (see [30] for a detailed explanation of the method). In this paper, we introduce a new implementation of the Lepski method, which we refer to as a multiplier bootstrap implementation of the Lepski method.

**ALGORITHM 2 (Multiplier bootstrap implementation of the Lepski method).** *Let  $\gamma_n$  be a sequence of positive numbers converging to zero. Let  $\xi_1, \dots, \xi_n$  be independent  $N(0, 1)$  random variables that are independent of the data  $X_1^n := \{X_1, \dots, X_n\}$ . Let  $\xi_1^n := (\xi_1, \dots, \xi_n)$ . For all  $(x, l, l') \in \bar{\mathcal{V}}_n$ , define a Gaussian multiplier process:*

$$\begin{aligned} \tilde{\mathbb{G}}_n(x, l, l') &:= \tilde{\mathbb{G}}_n(X_1^n, \xi_1^n)(x, l, l') \\ &:= \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i \frac{(K_l(X_i, x) - K_{l'}(X_i, x)) - (\hat{f}_n(x, l) - \hat{f}_n(x, l'))}{\hat{\sigma}_n(x, l, l')}. \end{aligned}$$

Also, define

$$\tilde{c}_n(\gamma_n) = \text{conditional } (1 - \gamma_n)\text{-quantile of } \|\tilde{\mathbb{G}}_n\|_{\bar{\mathcal{V}}_n} \text{ given } X_1^n.$$

Moreover, for all  $l \in \mathcal{L}_n$ , let

$$\mathcal{L}_{n,l} := \{l' \in \mathcal{L}_n : l' > l\}.$$

Finally, for some constant  $q > 1$ , which is independent of  $n$ , define a Lepski-type estimator

$$(35) \quad \hat{l}_n := \inf \left\{ l \in \mathcal{L}_n : \sup_{l' \in \mathcal{L}_{n,l}} \sup_{x \in \mathcal{X}} \frac{\sqrt{n} |\hat{f}_n(x, l) - \hat{f}_n(x, l')|}{\hat{\sigma}_n(x, l, l')} \leq q \tilde{c}_n(\gamma_n) \right\}.$$

**COMMENT 4.1 (On our implementation of Lepski's method).** We refer to (35) as a (Gaussian) multiplier bootstrap implementation of the Lepski method because  $\tilde{c}_n(\gamma_n)$  is obtained as the conditional  $(1 - \gamma_n)$ -quantile of  $\|\tilde{\mathbb{G}}\|_{\bar{\mathcal{V}}_n}$  given  $X_1^n$ . Previous literature on the Lepski method used Talagrand's inequality combined with some bounds on expected suprema of certain empirical processes (obtained via symmetrization and entropy methods) to choose the threshold level for the estimator (the right hand side of the inequality in (35)); see [19] and [21]. Because of the one-sided nature of the aforementioned inequalities, however, it was argued that the resulting

threshold turned out to be too high leading to limited applicability of the estimator in small and moderate samples. In contrast, an advantage of our construction is that we use  $q\tilde{c}_n(\gamma_n)$  as a threshold level, which is essentially the minimal possible value of the threshold that suffices for good properties of the estimator. ■

Once we have  $\hat{l}_n$ , to define the confidence band  $\mathcal{C}_n$ , we need to specify  $\hat{\sigma}_n(x, l)$ ,  $\hat{c}_n(\alpha)$ , and  $c'_n$ . We assume that  $\hat{\sigma}_n(x, l)$  is obtained via (23) and  $\hat{c}_n(\alpha)$  via Algorithm 1. To specify  $c'_n$ , let  $u'_n$  be a sequence of positive numbers such that  $u'_n$  is sufficiently large for large  $n$ . Specifically, for large  $n$ ,  $u'_n$  is assumed to be larger than some constant  $C(\mathcal{F})$  depending only on the function class  $\mathcal{F}$ . Set

$$c'_n := u'_n \tilde{c}_n(\gamma_n).$$

COMMENT 4.2 (On the choice of  $\gamma_n$ ,  $q$  and  $u'_n$ ). As follows from Lemmas F.7 and F.8, the parameter  $\gamma_n$  appearing in (35) determines the probability that the estimator  $\hat{l}_n$  fails to select an appropriate value of the smoothing parameter. Thus, in practice  $\gamma_n$  should be chosen small relative to the nominal coverage level  $\alpha$ . Also, for fixed  $n$  and  $\gamma_n$ , the choice of the parameters  $q$  and  $u'_n$  depends on the trade-off between the error in the coverage probability and length of the confidence bands: smaller values of  $q$  yield higher values of  $\hat{l}_n$  leading to undersmoothing and good control of the coverage probability; larger values of  $q$  yield lower values of  $\hat{l}_n$  leading to oversmoothing and narrow confidence bands; similarly, larger values of  $u'_n$  yield wider confidence bands but better control of the coverage probability. Finding the optimal value of  $q$  is a difficult theoretical problem and is beyond the scope of the paper. Also, in principle, it is possible to trace out the value  $C(\mathcal{F})$  from the proof of the theorem below and set  $u'_n = C(\mathcal{F})$ . However, since the function class  $\mathcal{F}$  is typically unknown in practice,  $u'_n$  can be set as a slowly growing sequence of positive numbers. In our small-scale simulation study presented in Section J of the Supplemental Material, we find that the values  $q = 1.1$  and  $u'_n = 0.5$  strike a good balance between coverage probability control and the length of the confidence bands in one-dimensional examples. We should note however that the empirical researchers should always test out different values of  $q$  and  $u'_n$  in Monte Carlo examples that mimic the data at hand. ■

The following theorem shows that the confidence band  $\mathcal{C}_n$  defined in this way is honest and adaptive for  $\mathcal{F}$ :

**THEOREM 4.1 (Honest and Adaptive Confidence Bands via Our Method).** *Suppose that Conditions L1-L4 are satisfied. In addition, sup-*



pose that there exist constants  $c_5, C_5 > 0$  such that (i)  $2^{l_{\max, n} d} (\log^4 n) / n \leq C_5 n^{-c_5}$ , (ii)  $l_{\min, n} \geq c_5 \log n$ , (iii)  $\gamma_n \leq C_5 n^{-c_5}$ , (iv)  $|\log \gamma_n| \leq C_5 \log n$ , (v)  $u'_n \geq C(\mathcal{F})$ , and (vi)  $u'_n \leq C_5 \log n$ . Then Conditions H1-H6 in Section 3 and (15) in Theorem 3.1 hold and

$$(36) \quad \sup_{f \in \mathcal{F}} \mathbb{P}_f \left( \sup_{x \in \mathcal{X}} \lambda(\mathcal{C}_n(x)) > C(1 + u'_n) r_n(t(f)) \right) \leq C n^{-c},$$

where  $\lambda(\cdot)$  denotes the Lebesgue measure on  $\mathbb{R}$  and  $r_n(t) := (\log n / n)^{t/(2t+d)}$ . Here the constants  $c, C > 0$  depend only on  $c_5, C_5$ , the constants that appear in Conditions L1-L4,  $c_\sigma, \alpha$ , and the function  $K$  (when convolution kernels are used) or the father wavelet  $\phi$  (when wavelet projection kernels are used). Moreover,

$$(37) \quad \sup_{f \in \mathcal{F} \cap \Sigma(t, L)} \mathbb{P}_f \left( \sup_{x \in \mathcal{X}} \lambda(\mathcal{C}_n(x)) > C(1 + u'_n) r_n(t) \right) \leq C n^{-c},$$

with the same constants  $c, C$  as those in (36).

COMMENT 4.3 (Honest and adaptive confidence bands). Equation (15) implies that the confidence band  $\mathcal{C}_n$  constructed above is asymptotically honest at a polynomial rate for the class  $\mathcal{F}$ . In addition, recall that  $r_n(t)$  is the minimax optimal rate of convergence in the sup-metric for the class  $\mathcal{F} \cap \Sigma(t, L)$ ; see [20]. Therefore, (37) implies that the confidence band  $\mathcal{C}_n$  is adaptive whenever  $u'_n$  is bounded or almost adaptive if  $u'_n$  is slowly growing; see discussion in front of Theorem 4.1 on selecting  $u'_n$ . ■

COMMENT 4.4 (On inflating terms). When  $u'_n$  is bounded, the rate of convergence of the length of the confidence band to zero  $(1 + u'_n) r_n(t)$  coincides with the minimax optimal rate of estimation of over  $\Sigma(t, L)$  with *no additional inflating terms*. This shows an advantage of the method of constructing confidence bands based on the explicit control of the bias term in comparison with the method based on undersmoothing where inflating terms seem to be necessary. This type of construction is inspired by the interesting ideas in [4]. ■

COMMENT 4.5 (Extensions). Finally, we note that the proof of (15) and (36) in Theorem 4.1 did not use (30) directly. The proof only relies on Conditions L1-L4 whereas (30) served to motivate these conditions. Therefore, results (15) and (36) of Theorem 4.1 apply more generally as long as Conditions L1-L4 hold, not just for Hölder smoothness classes. ■

APPENDIX A: COUPLING INEQUALITIES FOR SUPREMA OF  
EMPIRICAL AND RELATED PROCESSES

The purpose of this section is to provide two coupling inequalities based on Slepian-Stein methods that are useful for the analysis of uniform confidence bands. The first inequality is concerned with suprema of empirical processes and is proven in Corollary 2.2 in [7]. The second inequality is new, is concerned with suprema of Gaussian multiplier processes, and will be obtained from a Gaussian comparison theorem derived in [8].

Let  $X_1, \dots, X_n$  be i.i.d. random variables taking values in a measurable space  $(S, \mathcal{S})$ . Let  $\mathcal{G}$  be a pointwise-measurable  $\text{VC}(b, a, v)$  type function class for some  $b > 0$ ,  $a \geq e$ , and  $v \geq 1$  (the definition of VC type classes is given in Section 3). Let  $\sigma^2 > 0$  be any constant such that  $\sup_{g \in \mathcal{G}} \mathbb{E}[g(X_1)^2] \leq \sigma^2 \leq b^2$ . Define the empirical process

$$\mathbb{G}_n(g) := \frac{1}{\sqrt{n}} \sum_{i=1}^n (g(X_i) - \mathbb{E}[g(X_1)]), \quad g \in \mathcal{G},$$

and let

$$W_n := \|\mathbb{G}_n\|_{\mathcal{G}} := \sup_{g \in \mathcal{G}} |\mathbb{G}_n(g)|$$

denote the supremum of the empirical process. Note that  $W_n$  is a well-defined random variable since  $\mathcal{G}$  is assumed to be pointwise-measurable. Let  $B = \{B(g) : g \in \mathcal{G}\}$  be a tight Gaussian random element in  $\ell^\infty(\mathcal{F})$  with mean zero and covariance function

$$\mathbb{E}[B(g_1)B(g_2)] = \mathbb{E}[g_1(X_1)g_2(X_1)] - \mathbb{E}[g_1(X_1)]\mathbb{E}[g_2(X_1)],$$

for all  $g_1, g_2 \in \mathcal{G}$ . It is well known that such a process exists under the VC type assumption [see 43, p.100-101]. Finally, for some sufficiently large absolute constant  $A$ , let

$$K_n := Av(\log n \vee \log(ab/\sigma)).$$

In particular, we will assume that  $K_n \geq 1$ . The following theorem shows that  $W_n$  can be well approximated by the supremum of the corresponding Gaussian process  $B$  under mild conditions on  $b$ ,  $\sigma$ , and  $K_n$ . The proof of this theorem can be found in Corollary 2.2 in [7]

**THEOREM A.1** (Slepian-Stein type coupling for suprema of empirical processes). *Consider the setting specified above. Then for every  $\gamma \in (0, 1)$  one*

can construct on an enriched probability space a random variable  $W^0$  such that (i)  $W^0 \stackrel{d}{=} \|B\|_{\mathcal{G}}$  and (ii)

$$\begin{aligned} \mathbb{P} \left( |W_n - W^0| > \frac{bK_n}{(\gamma n)^{1/2}} + \frac{(b\sigma)^{1/2} K_n^{3/4}}{\gamma^{1/2} n^{1/4}} + \frac{b^{1/3} \sigma^{2/3} K_n^{2/3}}{\gamma^{1/3} n^{1/6}} \right) \\ \leq A' \left( \gamma + \frac{\log n}{n} \right), \end{aligned}$$

where  $A'$  is an absolute constant.

COMMENT A.1 (Comparison with the Hungarian couplings). The main advantage of the coupling provided in this theorem in comparison with, say, Hungarian coupling [28], which can be used to derive a similar result, is that our coupling does not depend on total variation norm of functions  $g \in \mathcal{G}$  leading to sharper inequalities than those obtained via Hungarian coupling when the function class  $\mathcal{G}$  consists, for example, of Fourier series or Legendre polynomials; see [7]. In addition, our coupling does not impose any side restrictions. In particular, it does not require bounded support of  $X$  and allows for point masses on the support. In addition, if the density of  $X$  exists, our coupling does not assume that this density is bounded away from zero on the support. See, for example, [38] for the construction of the Hungarian coupling and the use of aforementioned conditions. ■

Let  $\xi_1, \dots, \xi_n$  be independent  $N(0, 1)$  random variables independent of  $X_1^n := \{X_1, \dots, X_n\}$ , and let  $\xi_1^n := \{\xi_1, \dots, \xi_n\}$ . We assume that random variables  $X_1, \dots, X_n, \xi_1, \dots, \xi_n$  are defined as coordinate projections from the product probability space. Define the Gaussian multiplier process

$$\tilde{\mathbb{G}}_n(g) := \tilde{\mathbb{G}}_n(X_1^n, \xi_1^n)(g) := \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i(g(X_i) - \mathbb{E}_n[g(X_i)]), \quad g \in \mathcal{G},$$

and for  $x_1^n \in S^n$ , let  $\tilde{W}_n(x_1^n) := \|\tilde{\mathbb{G}}_n(x_1^n, \xi_1^n)\|_{\mathcal{G}}$  denote the supremum of this process calculated for fixed  $X_1^n = x_1^n$ . Note that  $\tilde{W}_n(x_1^n)$  is a well-defined random variable. In addition, let

$$\psi_n := \sqrt{\frac{\sigma^2 K_n}{n}} + \left( \frac{b^2 \sigma^2 K_n^3}{n} \right)^{1/4} \quad \text{and} \quad \gamma_n(\delta) := \frac{1}{\delta} \left( \frac{b^2 \sigma^2 K_n^3}{n} \right)^{1/4} + \frac{1}{n}.$$

The following theorem shows that  $\tilde{W}_n(X_1^n)$  can be well approximated with high probability by the supremum of the Gaussian process  $B$  under mild conditions on  $b$ ,  $\sigma$ , and  $K_n$ . The proof of this theorem can be found in the Supplemental Material.

THEOREM A.2 (Slepian-Stein type coupling for suprema of conditional multiplier processes). *Consider the setting specified above. Suppose that  $b^2 K_n \leq n\sigma^2$ . Then for every  $\delta > 0$ , there exists a set  $S_{n,0} \in \mathcal{S}^n$  such that  $\mathbb{P}(X_1^n \in S_{n,0}) \geq 1 - 3/n$  and for every  $x_1^n \in S_{n,0}$  one can construct on an enriched probability space a random variable  $W^0$  such that (i)  $W^0 \stackrel{d}{=} \|B\|_{\mathcal{G}}$  and (ii)*

$$\mathbb{P}(|\tilde{W}_n(x_1^n) - W^0| > (\psi_n + \delta)) \leq A'' \gamma_n(\delta),$$

where  $A''$  is an absolute constant.

COMMENT A.2 (On the use of Slepian-Stein couplings). Theorems A.1 and A.2 combined with anti-concentration inequalities (Theorem 2.1 and Corollary 2.1) can be used to prove validity of Gaussian multiplier bootstrap for approximating distributions of suprema of empirical processes of VC type function classes *without weak convergence arguments*. This allows us to cover cases where complexity of the function class  $\mathcal{G}$  is increasing with  $n$ , which is typically the case in nonparametric problems in general and in confidence band construction in particular. Moreover, approximation error can be shown to be polynomially (in  $n$ ) small under mild conditions. ■

## APPENDIX B: SOME TECHNICAL TOOLS

THEOREM B.1. *Let  $\xi_1, \dots, \xi_n$  be i.i.d. random variables taking values in a measurable space  $(S, \mathcal{S})$ . Suppose that  $\mathcal{G}$  is a nonempty, pointwise measurable class of functions on  $S$  uniformly bounded by a constant  $b$  such that there exist constants  $a \geq e$  and  $v > 1$  with  $\sup_Q N(\mathcal{G}, L_2(Q), b\epsilon) \leq (a/\epsilon)^v$  for all  $0 < \epsilon \leq 1$ . Let  $\sigma^2$  be a constant such that  $\sup_{g \in \mathcal{G}} \text{Var}(g) \leq \sigma^2 \leq b^2$ . If  $b^2 v \log(ab/\sigma) \leq n\sigma^2$ , then for all  $t \leq n\sigma^2/b^2$ ,*

$$\mathbb{P} \left[ \sup_{g \in \mathcal{G}} \left| \sum_{i=1}^n \{g(\xi_i) - \mathbb{E}[g(\xi_1)]\} \right| > A \sqrt{n\sigma^2 \left\{ t \vee \left( v \log \frac{ab}{\sigma} \right) \right\}} \right] \leq e^{-t},$$

where  $A > 0$  is an absolute constant.

PROOF. This version of Talagrand's inequality follows from Theorem 3 in [34] combined with a bound on expected values of suprema of empirical processes derived in [15]. See also [41] for the original version of Talagrand's inequality. ■

Proofs of the following two lemmas can be found in the Supplemental Material.

LEMMA B.1. *Let  $Y := \{Y(t) : t \in T\}$  be a separable, centered Gaussian process such that  $\mathbb{E}[Y(t)^2] \leq 1$  for all  $t \in T$ . Let  $c(\alpha)$  denote the  $(1 - \alpha)$ -quantile of  $\|Y\|_T$ . Assume that  $\mathbb{E}[\|Y\|_T] < \infty$ . Then  $c(\alpha) \leq \mathbb{E}[\|Y\|_T] + \sqrt{2|\log \alpha|}$  and  $c(\alpha) \leq M(\|Y\|_T) + \sqrt{2|\log \alpha|}$  for all  $\alpha \in (0, 1)$  where  $M(\|Y\|_T)$  is the median of  $\|Y\|_T$ .*

LEMMA B.2. *Let  $\mathcal{G}_1$  and  $\mathcal{G}_2$  be  $VC(b_1, a_1, v_1)$  and  $VC(b_2, a_2, v_2)$  type classes, respectively, on a measurable space  $(S, \mathcal{S})$ . Let  $a = (a_1^{v_1} a_2^{v_2})^{1/(v_1+v_2)}$ . Then (i)  $\mathcal{G}_1 \cdot \mathcal{G}_2 = \{g_1 \cdot g_2 : g_1 \in \mathcal{G}_1, g_2 \in \mathcal{G}_2\}$  is  $VC(b_1 b_2, 2a, v_1 + v_2)$  type class, (ii)  $\mathcal{G}_1 - \mathcal{G}_2 = \{g_1 - g_2 : g_1 \in \mathcal{G}_1, g_2 \in \mathcal{G}_2\}$  is  $VC(b_1 + b_2, a, v_1 + v_2)$  type class, and (iii)  $\mathcal{G}_1^2 = \{g_1^2 : g_1 \in \mathcal{G}_1\}$  is  $VC(b_1^2, 2a_1, v_1)$  type class.*

### APPENDIX C: PROOFS FOR SECTION 2

PROOF OF THEOREM 2.1. The fact that  $a(X) < \infty$  follows from Landau-Shepp-Fernique theorem (see, for example, Lemma 2.2.5 in [14]). In addition, since  $\sup_{t \in T} X_t \geq X_{t_0}$  for any fixed  $t_0 \in T$ ,  $a(X) \geq \mathbb{E}[X_{t_0}] = 0$ . We now prove (7).

Since the Gaussian process  $X = (X_t)_{t \in T}$  is separable, there exists a sequence of finite subsets  $T_n \subset T$  such that  $Z_n := \max_{t \in T_n} X_t \rightarrow \sup_{t \in T} X_t =: Z$  a.s. as  $n \rightarrow \infty$ . Fix any  $x \in \mathbb{R}$ . Since  $|Z_n - x| \rightarrow |Z - x|$  a.s. and a.s. convergence implies weak convergence, there exists an at most countable subset  $\mathcal{N}_x$  of  $\mathbb{R}$  such that for all  $\epsilon \in \mathbb{R} \setminus \mathcal{N}_x$ ,

$$\lim_{n \rightarrow \infty} \mathbb{P}(|Z_n - x| \leq \epsilon) = \mathbb{P}(|Z - x| \leq \epsilon).$$

But by Theorem 3 in [8],

$$\mathbb{P}(|Z_n - x| \leq \epsilon) \leq 4\epsilon(\mathbb{E}[\max_{t \in T_n} X_t] + 1) \leq 4\epsilon(a(X) + 1),$$

for all  $\epsilon \geq 0$ . Therefore,

$$(38) \quad \mathbb{P}(|Z - x| \leq \epsilon) \leq 4\epsilon(a(X) + 1),$$

for all  $\epsilon \in \mathbb{R} \setminus \mathcal{N}_x$ . By right continuity of  $\mathbb{P}(|Z - x| \leq \cdot)$ , it follows that (38) holds for all  $\epsilon \geq 0$ . Since  $x \in \mathbb{R}$  is arbitrary, we obtain (7).  $\blacksquare$

PROOF OF COROLLARY 2.1. Let  $\tilde{T}$  be a set such that there exists a bijective mapping from  $\tilde{T}$  to  $T$ . Denote this mapping  $h : \tilde{T} \rightarrow T$ . Define the process  $\tilde{X} = (\tilde{X}_t)_{t \in T \cup \tilde{T}}$  by letting  $\tilde{X}_t = X_t$  for  $t \in T$  and  $\tilde{X}_t = -X_{h(t)}$  for  $t \in \tilde{T}$ . Then  $p_\epsilon(|X|) = p_\epsilon(\tilde{X})$  and  $a(\tilde{X}) = a(|X|)$ . Hence, by Theorem 2.1,

$$p_\epsilon(|X|) = p_\epsilon(\tilde{X}) \leq 4\epsilon(a(\tilde{X}) + 1) = 4\epsilon(a(|X|) + 1).$$

The asserted claim follows.  $\blacksquare$

## APPENDIX D: PROOFS FOR SECTION 3

PROOF OF THEOREM 3.1. Pick any  $f \in \mathcal{F}$ . By the triangle inequality, we have for any  $x \in \mathcal{X}$ ,

$$\frac{\sqrt{n}|\hat{f}_n(x, \hat{l}_n) - f(x)|}{\hat{\sigma}_n(x, \hat{l}_n)} \leq \left( |Z_{n,f}(x, \hat{l}_n)| + \Delta_{n,f}(\hat{l}_n) \right) \frac{\sigma_{n,f}(x, \hat{l}_n)}{\hat{\sigma}_n(x, \hat{l}_n)},$$

by which we have

$$\begin{aligned} & \mathbb{P}_f(f(x) \in \mathcal{C}_n(x), \forall x \in \mathcal{X}) \\ & \geq \mathbb{P}_f(|Z_{n,f}(x, \hat{l}_n)| + \Delta_{n,f}(\hat{l}_n) \leq (\hat{c}_n(\alpha) + c'_n)\hat{\sigma}_n(x, \hat{l}_n)/\sigma_{n,f}(x, \hat{l}_n), \forall x \in \mathcal{X}) \\ & \geq \mathbb{P}_f(\sup_{x \in \mathcal{X}} |Z_{n,f}(x, \hat{l}_n)| + \Delta_{n,f}(\hat{l}_n) \leq (\hat{c}_n(\alpha) + c'_n)(1 - \epsilon_{3n}) - \delta_{3n}) \\ & \geq \mathbb{P}_f(\sup_{x \in \mathcal{X}} |Z_{n,f}(x, \hat{l}_n)| \leq \hat{c}_n(\alpha)(1 - \epsilon_{3n}) - c'_n\epsilon_{3n}) - \delta_{3n} - \delta_{4n} \\ & \geq \mathbb{P}_f(\|Z_{n,f}\|_{\mathcal{V}_n} \leq \hat{c}_n(\alpha)(1 - \epsilon_{3n}) - c'_n\epsilon_{3n}) - \delta_{3n} - \delta_{4n} \\ & \geq \mathbb{P}_f(\|Z_{n,f}\|_{\mathcal{V}_n} \leq \hat{c}_n(\alpha)(1 - \epsilon_{3n}) - u_n\epsilon_{3n}\sqrt{\log n}) - \delta_{3n} - \delta_{4n} - \delta_{5n}, \end{aligned}$$

where the third line follows from Condition H4, the fourth line from Condition H5, the fifth line from the inequality  $\sup_{x \in \mathcal{X}} |Z_{n,f}(x, \hat{l}_n)| \leq \|Z_{n,f}\|_{\mathcal{V}_n}$ , and the sixth line from Condition H6. Further, the probability in the last line above equals (recall that  $W_{n,f} = \|Z_{n,f}\|_{\mathcal{V}_n}$ )

$$\begin{aligned} & \mathbb{P}_f(W_{n,f} \leq \hat{c}_n(\alpha)(1 - \epsilon_{3n}) - u_n\epsilon_{3n}\sqrt{\log n}) \\ (39) \quad & \geq \mathbb{P}_f(W_{n,f} \leq c_{n,f}(\alpha + \tau_n)(1 - \epsilon_{3n}) - \epsilon_{2n} - u_n\epsilon_{3n}\sqrt{\log n}) - \delta_{2n}, \end{aligned}$$

where (39) follows from Condition H3. Now, the probability in (39) is bounded from below by Condition H1 by

$$\begin{aligned} & \mathbb{P}_f(W_{n,f}^0 \leq c_{n,f}(\alpha + \tau_n)(1 - \epsilon_{3n}) - \epsilon_{1n} - \epsilon_{2n} - u_n\epsilon_{3n}\sqrt{\log n}) - \delta_{1n} \\ (40) \quad & \geq \mathbb{P}_f(W_{n,f}^0 \leq c_{n,f}(\alpha + \tau_n)) - p_{\bar{\epsilon}_n}(|G_{n,f}|) - \delta_{1n} \end{aligned}$$

$$(41) \quad \geq 1 - \alpha - \tau_n - p_{\bar{\epsilon}_n}(|G_{n,f}|) - \delta_{1n},$$

where (40) follows from the definition of the Lévy concentration function  $p_{\bar{\epsilon}_n}(|G_{n,f}|)$  given that  $\bar{\epsilon}_n = \epsilon_{1n} + \epsilon_{2n} + \epsilon_{3n}(c_{n,f}(\alpha) + u_n\sqrt{\log n})$  and (41) follows since  $c_{n,f}(\cdot)$  is the quantile function of  $W_{n,f}^0$ . Combining these inequalities leads to (13).

To prove (14) and (15), note that  $\delta_n \leq Cn^{-c}$  and  $\tau_n \leq Cn^{-c}$  by Conditions H1 and H3-H6. Further, it follows from Lemma B.1 that  $c_{n,f}(\alpha) \leq \mathbb{E}[|G_{n,f}|_{\mathcal{V}_n}] + (2|\log \alpha|)^{1/2} \leq C\sqrt{\log n}$ , and so  $\epsilon_{3n}u_n\sqrt{\log n} \leq C_1n^{-c_1}$  implies that  $\bar{\epsilon}_{n,f} \leq Cn^{-c}$ . Therefore, (14) and (15) follow from (13) and Condition H2.  $\blacksquare$

PROOF OF COROLLARY 3.1. The proof is similar to that of Theorem 3.1. The details are provided in the Supplemental Material.  $\blacksquare$

PROOF OF THEOREM 3.2. In this proof,  $c, C > 0$  are constants that depend only on  $c_2, C_2$  but their values can change at each appearance.

Fix any  $f \in \mathcal{F}$ . Let  $G_{n,f} = \{G_{n,f}(v) : v \in \mathcal{V}_n\}$  be a tight Gaussian random element in  $\ell^\infty(\mathcal{V}_n)$  with mean zero and the same covariance function as that of  $Z_{n,f}$ . Since  $b_n^2 \sigma_n^4 K_n^4 / n \leq C_2 n^{-c_2}$ , it follows from Theorem A.1 that we can construct a random variable  $W_{n,f}^0$  such that  $W_{n,f}^0 \stackrel{d}{=} \|G_{n,f}\|_{\mathcal{V}_n}$  and (9) holds with some  $\epsilon_{1n}$  and  $\delta_{1n}$  bounded from above by  $Cn^{-c}$ . In addition, inequality  $\mathbb{E}[\|G_{n,f}\|_{\mathcal{V}_n}] \leq C\sqrt{\log n}$  follows from Corollary 2.2.8 in [43]. Condition H1 follows. Given Condition H1, Condition H2-(b) follows from Corollary 2.1, and Condition H2-(a) follows from H2-(b).

Consider Condition H4. There exists  $n_0$  such that  $C_2 n_0^{-c_2} \leq 1$ . It suffices to verify the condition only for  $n \geq n_0$ . Note that

$$(42) \quad \left| \frac{\hat{\sigma}_n(x, l)}{\sigma_{n,f}(x, l)} - 1 \right| \leq \left| \frac{\hat{\sigma}_n^2(x, l)}{\sigma_{n,f}^2(x, l)} - 1 \right|.$$

Define  $\mathcal{K}_{n,f}^2 := \{g^2 : g \in \mathcal{K}_{n,f}\}$ . Given the definition of  $\hat{\sigma}_n(x, l)$ , the right hand side of (42) is bounded by

$$(43) \quad \sup_{g \in \mathcal{K}_{n,f}^2} |\mathbb{E}_n[g(X_i)] - \mathbb{E}[g(X_1)]| + \sup_{g \in \mathcal{K}_{n,f}} |\mathbb{E}_n[g(X_i)]^2 - \mathbb{E}[g(X_1)]^2|.$$

It follows from Lemma B.2 that  $\mathcal{K}_{n,f}^2$  is VC( $b_n^2, 2a_n, v_n$ ) type class. Moreover, for all  $g \in \mathcal{K}_{n,f}^2$ ,

$$\mathbb{E}[g(X_i)^2] \leq b_n^2 \mathbb{E}[g(X_i)] \leq b_n^2 \sigma_n^2.$$

Therefore, Talagrand's inequality (Theorem B.1) with  $t = \log n$ , which can be applied because  $b_n^2 K_n / (n\sigma_n^2) \leq b_n^2 \sigma_n^4 K_n^4 / n \leq C_2 n^{-c_2} \leq 1$  and  $b_n^2 \log n / (n\sigma_n^2) \leq b_n^2 K_n / (n\sigma_n^2) \leq 1$  (recall that  $\sigma_n \geq 1$  and  $K_n \geq 1$ ), gives

$$(44) \quad \mathbb{P} \left( \sup_{g \in \mathcal{K}_{n,f}^2} |\mathbb{E}_n[g(X_i)] - \mathbb{E}[g(X_1)]| > \frac{1}{2} \sqrt{\frac{b_n^2 \sigma_n^2 K_n}{n}} \right) \leq \frac{1}{n}.$$

In addition,

$$\sup_{g \in \mathcal{K}_{n,f}} |\mathbb{E}_n[g(X_i)]^2 - \mathbb{E}[g(X_1)]^2| \leq 2b_n \sup_{g \in \mathcal{K}_{n,f}} |\mathbb{E}_n[g(X_i)] - \mathbb{E}[g(X_1)]|,$$

so that another application of Talagrand's inequality yields

$$(45) \quad \mathbb{P} \left( \sup_{g \in \mathcal{K}_{n,f}} |\mathbb{E}_n[g(X_i)]^2 - \mathbb{E}[g(X_1)]^2| > \frac{1}{2} \sqrt{\frac{b_n^2 \sigma_n^2 K_n}{n}} \right) \leq \frac{1}{n}.$$

Given that  $b_n^2 \sigma_n^2 K_n / n \leq b_n^2 \sigma_n^4 K_n^4 / n \leq C_2 n^{-c_2}$ , combining (42)-(45) gives Condition H4 with  $\epsilon_{3n} := (b_n^2 \sigma_n^2 K_n / n)^{1/2}$  and  $\delta_{3n} := 2/n$ .

Finally, we verify Condition H3. There exists  $n_1$  such that  $\epsilon_{3n_1} \leq 1/2$ . It suffices to verify the condition only for  $n \geq n_1$ , so that  $\epsilon_{3n} \leq 1/2$ . Define

$$\tilde{\mathbb{G}}_n(x, l) = \tilde{\mathbb{G}}_n(X_1^n, \xi_1^n)(x, l) := \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i \frac{K_l(X_i, x) - \hat{f}_n(x, l)}{\sigma_n(x, l)},$$

and

$$\Delta \mathbb{G}_n(x, l) = \hat{\mathbb{G}}_n(x, l) - \tilde{\mathbb{G}}_n(x, l).$$

In addition, define

$$\begin{aligned} \hat{W}_n(x_1^n) &:= \sup_{(x,l) \in \mathcal{X} \times \mathcal{L}_n} \hat{\mathbb{G}}_n(x_1^n, \xi_1^n)(x, l), \\ \tilde{W}_n(x_1^n) &:= \sup_{(x,l) \in \mathcal{X} \times \mathcal{L}_n} \tilde{\mathbb{G}}_n(x_1^n, \xi_1^n)(x, l). \end{aligned}$$

Consider the set  $S_{n,1}$  of values  $X_1^n$  such that  $|\hat{\sigma}_n(x, l) / \sigma_{n,f}(x, l) - 1| \leq \epsilon_{3n}$  for all  $(x, l) \in \mathcal{X} \times \mathcal{L}_n$  whenever  $X_1^n \in S_{n,1}$ . The previous calculations show that  $\mathbb{P}_f(X_1^n \in S_{n,1}) \geq 1 - \delta_{3n} = 1 - 2/n$ . Pick and fix any  $x_1^n \in S_{n,1}$ . Then

$$\Delta \mathbb{G}_n(x_1^n, \xi_1^n)(x, l) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i \frac{K_l(x_i, x) - \hat{f}_n(x, l)}{\sigma_n(x, l)} \left( \frac{\sigma_n(x, l)}{\hat{\sigma}_n(x, l)} - 1 \right)$$

is a Gaussian process with mean zero and

$$\text{Var}(\Delta \mathbb{G}_n(x_1^n, \xi_1^n)(x, l)) = \frac{\hat{\sigma}_n^2(x, l)}{\sigma_n^2(x, l)} \left( \frac{\sigma_n(x, l)}{\hat{\sigma}_n(x, l)} - 1 \right)^2 \leq \epsilon_{3n}^2.$$

Further, the function class

$$\tilde{\mathcal{K}}_{n,f} := \left\{ \frac{K_l(\cdot, x)}{\sigma_n(x, l)} \left( \frac{\sigma_n(x, l)}{\hat{\sigma}_n(x, l)} - 1 \right) : (x, l) \in \mathcal{X} \times \mathcal{L}_n \right\}$$

is contained in the function class

$$\left\{ \frac{a K_l(\cdot, x)}{\sigma_n(x, l)} : (x, l, a) \in \mathcal{X} \times \mathcal{L}_n \times [-1, 1] \right\},$$



and hence is  $\text{VC}(b_n, 4a_n, 1 + v_n)$  type class by Lemma B.2. In addition,

$$\begin{aligned} & \mathbb{E} \left[ \left( \Delta \mathbb{G}_n(x_1^n, \xi_1^n)(x', l') - \Delta \mathbb{G}_n(x_1^n, \xi_1^n)(x'', l'') \right)^2 \right] \\ & \leq \mathbb{E}_n \left[ \left( \frac{K_l(x_i, x')}{\sigma_n(x', l')} \left( \frac{\sigma_n(x', l')}{\hat{\sigma}_n(x', l')} - 1 \right) - \frac{K_l(x_i, x'')}{\sigma_n(x'', l'')} \left( \frac{\sigma_n(x'', l'')}{\hat{\sigma}_n(x'', l'')} - 1 \right) \right)^2 \right], \end{aligned}$$

for all  $x', x'' \in \mathcal{X}$  and  $l', l'' \in \mathcal{L}_n$ , so that covering numbers for the index set  $\mathcal{X} \times \mathcal{L}_n$  with respect to the intrinsic (standard deviation) semimetric induced from the Gaussian process  $\Delta \mathbb{G}_n(x_1^n, \xi_1^n)$  are bounded by uniform covering numbers for the function class  $\tilde{\mathcal{K}}_{n,f}$ . Therefore, an application of Corollary 2.2.8 in [43] gives

$$\mathbb{E} \left[ \sup_{(x,l) \in \mathcal{X} \times \mathcal{L}_n} |\Delta \mathbb{G}_n(x_1^n, \xi_1^n)(x, l)| \right] \leq C \epsilon_{3n} \sqrt{(1 + v_n) \log \left( \frac{4a_n b_n}{\epsilon_{3n}} \right)} \leq C n^{-c},$$

Here the second inequality follows from the definition of  $\epsilon_{3n}$  above and the following inequalities

$$\begin{aligned} \sqrt{(1 + v_n) \log \left( \frac{4a_n b_n}{\epsilon_{3n}} \right)} & \leq \sqrt{(1 + v_n) \left( \log \left( \frac{4a_n b_n}{\sigma_n} \right) + \log \left( \frac{\sigma_n}{\epsilon_{3n}} \right) \right)} \\ & \leq C \sqrt{K_n} \left( 1 + \sqrt{\log \left( \frac{\sigma_n}{\epsilon_{3n}} \right)} \right) \\ & \leq C \sqrt{K_n} \left( 1 + \sqrt{\log \left( \frac{n}{b_n^2 K_n} \right)} \right) \\ & \leq C \sqrt{K_n} \left( 1 + \sqrt{\log n} \right) \leq C K_n \end{aligned}$$

where in the last line we used  $b_n \geq \sigma_n \geq 1$ , and  $K_n \geq v_n \log n \geq \log n$ . Combining this bound with the Borell-Sudakov-Tsirel'son inequality, and using the inequality

$$|\hat{W}_n(x_1^n) - \tilde{W}_n(x_1^n)| \leq \sup_{(x,l) \in \mathcal{X} \times \mathcal{L}_n} |\Delta \mathbb{G}_n(x_1^n, \xi_1^n)(x, l)|,$$

we see that there exists  $\lambda_{1n} \leq C n^{-c}$  such that

$$(46) \quad \mathbb{P}(|\hat{W}_n(x_1^n) - \tilde{W}_n(x_1^n)| \geq \lambda_{1n}) \leq C n^{-c},$$

whenever  $x_1^n \in S_{n,1}$ . Further, since  $b_n^2 \sigma_n^4 K_n^4 / n \leq C_2 n^{-c_2}$  and  $b_n \geq \sigma_n \geq 1$ , Theorem A.2 shows that there exist  $\lambda_{2n} \leq C n^{-c}$  and a measurable set  $S_{n,2}$

of values  $X_1^n$  such that  $P_f(X_1^n \in S_{n,2}) \geq 1 - 3/n$  and for every  $x_1^n \in S_{n,2}$  one can construct a random variable  $W^0$  such that  $W^0 \stackrel{d}{=} \|G_{n,f}\|_{\mathcal{V}_n}$  and

$$(47) \quad P(|\tilde{W}_n(x_1^n) - W^0| \geq \lambda_{2n}) \leq Cn^{-c}.$$

Here  $W^0$  may depend on  $x_1^n$  but  $c, C$  can be chosen in such a way that they depend only on  $c_2, C_2$  (as noted in the beginning).

Pick and fix any  $x_1^n \in S_{n,0} := S_{n,1} \cap S_{n,2}$  and construct a suitable  $W^0 \stackrel{d}{=} \|G_{n,f}\|_{\mathcal{V}_n}$  for which (47) holds. Then by (46), we have

$$(48) \quad P(|\hat{W}_n(x_1^n) - W^0| \geq \lambda_n) \leq Cn^{-c},$$

where  $\lambda_n := \lambda_{1n} + \lambda_{2n}$ . Denote by  $\hat{c}_n(\alpha, x_1^n)$  the  $(1 - \alpha)$ -quantile of  $\hat{W}_n(x_1^n)$ . Then we have

$$\begin{aligned} P(\|G_{n,f}\|_{\mathcal{V}_n} \leq \hat{c}_n(\alpha, x_1^n) + \lambda_n) &= P(W^0 \leq \hat{c}_n(\alpha, x_1^n) + \lambda_n) \\ &\geq P(\hat{W}_n(x_1^n) \leq \hat{c}_n(\alpha, x_1^n)) - Cn^{-c} \\ &\geq 1 - \alpha - Cn^{-c}, \end{aligned}$$

by which we have  $\hat{c}_n(\alpha, x_1^n) \geq c_{n,f}(\alpha + Cn^{-c}) - \lambda_n$ . Since  $x_1^n \in S_{n,0}$  is arbitrary and  $\hat{c}_n(\alpha) = \hat{c}_n(\alpha, X_1^n)$ , we see that whenever  $X_1^n \in S_{n,0}$ ,  $\hat{c}_n(\alpha) \geq c_{n,f}(\alpha + Cn^{-c}) - \lambda_n$ . Part (a) of Condition H3 follows from the fact that  $P_f(X_1^n \in S_{n,0}) \geq 1 - 5/n$  and  $\lambda_n \leq Cn^{-c}$ . Part (b) follows similarly. ■

#### ACKNOWLEDGMENTS

The authors would like to thank Emre Barut, Enno Mammen, and Richard Nickl for very helpful discussions. We also thank the editors, and anonymous referees for their very helpful reviews that helped improve the paper.

#### SUPPLEMENTARY MATERIAL

##### Supplement to “Anti-concentration and honest, adaptive confidence bands”

( ). This supplemental file contains additional proofs omitted in the main text, some results regarding non-wavelet projection kernel estimators, and a small-scale simulation study.

#### REFERENCES

- [1] Bickel, P. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.* **26** 1826-1856.
- [2] Birge, L. (2001). An alternative point of view of Lepski’s method. *Lecture Notes-IMS Monograph Series* **36** 113-133.

- [3] Bissantz, N., Dümbgen, L., Holzmann, H. and Munk, A. (2007). Non-parametric confidence bands in deconvolution density estimation. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **69** 483-506.
- [4] Bull, A.D. (2012). Honest adaptive confidence bands and self-similar functions. *Electron. J. Stat.* **6** 1490-1516.
- [5] Bull, A.D. (2013). A Smirnov-Bickel-Rosenblatt theorem for compactly-supported wavelets. *Constr. Approx.* **37** 295-309
- [6] Cai, T. and Low, M. (2004). An adaptation theory for nonparametric confidence intervals. *Ann. Statist.* **32** 1805-1840.
- [7] Chernozhukov, V. Chetverikov, D. and Kato, K. (2012). Gaussian approximation of suprema of empirical processes. Preprint. arXiv:1212.6885v2.
- [8] Chernozhukov, V. Chetverikov, D. and Kato, K. (2012). Comparison and anti-concentration bounds for maxima of Gaussian random vectors. Preprint. arXiv:1301.4807v3. To appear in *Probab. Theory Related Fields*.
- [9] Chernozhukov, V. Chetverikov, D. and Kato, K. (2014). Supplement to “Anti-concentration and honest, adaptive confidence bands”.
- [10] Chernozhukov, V., Lee, S., and Rosen, A. (2013). Intersection bounds: estimation and inference. *Econometrica* **81** 667-737. arXiv:0907.3503v4.
- [11] Chetverikov, D. (2012). Testing regression monotonicity in econometric models. arXiv:1212.6757.
- [12] Claeskens, G. and van Keilegom, I. (2003). Bootstrap confidence bands for regression curves and their derivatives. *Ann. Statist.* **31** 1852-1884.
- [13] Daubechies, I. (1992). *Ten Lectures on Wavelets*. CBMS-NSF Regional Conference Series in Applied Mathematics.
- [14] Dudley, R.M. (1999). *Uniform Central Limit Theorems*. Cambridge University Press.
- [15] Giné, E. and Guillou, A. (2001). A law of the iterated logarithm for kernel density estimators in the presence of censoring. *Ann. Inst. Henri Poincaré Probab. Stat.* **37** 503-522.
- [16] Giné, E. and Guillou, A. (2002). Rates of strong uniform consistency for multivariate kernel density estimators. *Ann. Inst. Henri Poincaré Probab. Stat.* **38** 907-921.
- [17] Giné, E., Gunturk, C.S. and Madych W.R. (2011). On the periodized square of  $L^2$  cardinal splines. *Exp. Math.* **20** 177-188.
- [18] Giné, E., Koltchinskii, V. and Sakhnenko, L. (2004). Kernel density estimators: convergence in distribution for weighted sup-norms. *Probab. Theory Related Fields* **130** 167-198.
- [19] Giné, E. and Nickl, R. (2010). An exponential inequality for the distribution function of the kernel density estimator, with applications to adaptive estimation. *Probab. Theory Related Fields* **143** 569-596.
- [20] Giné, E. and Nickl, R. (2010). Confidence bands in density estimation. *Ann. Statist.* **38** 1122-1170.
- [21] Giné, E. and Nickl, R. (2010). Adaptive estimation of a distribution function and its density in sup-norm loss by wavelet and spline projections. *Bernoulli* **16** 1137-1163.
- [22] Giné, E. and Zinn, J. (1984). Some limit theorems for empirical processes. Special Invited Paper in *Ann. Probab.* **12** 929-989.
- [23] Giné, E. and Zinn, J. (1990). Bootstrapping general empirical processes *Ann. Probab.* **18** 851-869.
- [24] Hall, P. (1991). On convergence rates of suprema. *Probab. Theory Related Fields* **89** 447-455.
- [25] Hall, P., and Horowitz, J. (2013). A simple bootstrap method for constructing non-parametric confidence bands for functions. *Ann. Statist.* **41** 1892-1921.

- [26] Härdle, W., Kerkycharian, G., Picard, D., and Tsybakov, A. (1998). *Wavelets, Approximation, and Statistical Applications*. Springer-Verlag.
- [27] Hoffman, M. and Nickl, R. (2011). On adaptive inference and confidence bands. *Ann. Statist.* **39** 2383-2409.
- [28] Komlós, J. Major, J. Tusánady, G. (1975). An approximation of partial sums of independent rv's, and the sample df. I. *Z. Wahrsch. Verw. Gebiete* **32** 111-131.
- [29] Ledoux, M. and Talagrand, M. (1988). Un critère sur les petites boules dans le théorème limite central. *Probab. Theory Related Fields* **77** 29-47.
- [30] Lepski, O. (1991). Asymptotically minimax adaptive estimation I: Upper bounds. Optimally adaptive estimates. *Theory Probab. Appl.* **36** 682-697.
- [31] Li, K.-C. (1989). Honest confidence regions for nonparametric regression. *Ann. Statist.* **17** 1001-1008.
- [32] Lounici, K. and Nickl, R. (2011). Global uniform risk bounds for wavelet deconvolution estimators. *Ann. Statist.* **39** 201-231.
- [33] Low, M. (1997). On nonparametric confidence intervals. *Ann. Statist.* **25** 2547-2554.
- [34] Massart, P. (2000). About the constants in Talagrand's concentration inequalities for empirical processes. *Ann. Probab.* **28** 863-884.
- [35] Massart, P. (2007). *Concentration Inequalities and Model Selection*. Springer-Verlag.
- [36] Picard, D. and Tribouley, K. (2000). Adaptive confidence interval for pointwise curve estimation. *Ann. Statist.* **28** 298-335.
- [37] Praestgaard, J. and Wellner, J. (1993). Exchangeably weighted bootstraps of the general empirical process. *Ann. Probab.* **21** 2053-2086.
- [38] Rio, E. (1994). Local invariance principles and their application to density estimation. *Probab. Theory Related Fields* **98** 21-45.
- [39] Rudelson, M. and Vershynin, R. (2009). Smallest singular value of a random rectangular matrix. *Comm. Pure Appl. Math.* **62** 1707-1739.
- [40] Smirnov, N.V. (1950). On the construction of confidence regions for the density of distributions of random variables (in Russian). *Doklady Akad. Nauk SSSR* **74** 189-191.
- [41] Talagrand, M. (1996). New concentration inequalities in product spaces. *Invent. Math.* **126** 503-563.
- [42] Tsybakov, A. (2009). *Introduction to Nonparametric Estimation*. Springer Series in Statistics.
- [43] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak convergence and empirical processes: With applications to statistics*. Springer-Verlag.
- [44] Wasserman, L. (2006). *All of Nonparametric Statistics*. Springer Texts in Statistics.
- [45] Xia, Y. (1998). Bias-corrected confidence bands in nonparametric regression. *J. Roy. Stat. Soc. Ser. B Stat. Methodol.* **60** 797-811.

DEPARTMENT OF ECONOMICS AND  
OPERATIONS RESEARCH CENTER, MIT  
50 MEMORIAL DRIVE  
CAMBRIDGE, MA 02142, USA.  
E-MAIL: vchern@mit.edu

DEPARTMENT OF ECONOMICS, UCLA  
BUNCHE HALL, 8283  
315 PORTOLA PLAZA  
LOS ANGELES, CA 90095, USA.  
E-MAIL: chetverikov@econ.ucla.edu

GRADUATE SCHOOL OF ECONOMICS  
UNIVERSITY OF TOKYO  
7-3-1 HONGO, BUNKYO-KU  
TOKYO 113-0033, JAPAN.  
E-MAIL: kkato@e.u-tokyo.ac.jp

**SUPPLEMENT TO “ANTI-CONCENTRATION AND  
HONEST, ADAPTIVE CONFIDENCE BANDS”\***

BY VICTOR CHERNOZHUKOV<sup>†</sup>, DENIS CHETVERIKOV<sup>‡</sup> AND KENGO  
KATO<sup>§</sup>

*MIT<sup>†</sup>, UCLA<sup>‡</sup>, and University of Tokyo<sup>§</sup>*

This supplemental material contains additional proofs omitted in  
the main text, some results regarding non-wavelet projection kernel  
estimators, and a small-scale simulation study.

APPENDIX E: PROOF OF COROLLARY 3.1

Pick any  $f \in \mathcal{F}$ . Then

$$\begin{aligned}
 & \mathbb{P}_f(f(x) \in \tilde{\mathcal{C}}_n(x), \forall x \in \mathcal{X}) \\
 (49) \quad & \geq \mathbb{P}_f(|Z_{n,f}(x, \hat{l}_n)| + \Delta_{n,f}(\hat{l}_n) \leq \hat{c}_n(\alpha) \hat{\sigma}_n(x, \hat{l}_n) / \sigma_{n,f}(x, \hat{l}_n), \forall x \in \mathcal{X}) \\
 (50) \quad & \geq \mathbb{P}_f(\sup_{x \in \mathcal{X}} |Z_{n,f}(x, \hat{l}_n)| + \Delta_{n,f}(\hat{l}_n) \leq \hat{c}_n(\alpha)(1 - \epsilon_{3n}) - \delta_{3n}) \\
 (51) \quad & \geq \mathbb{P}_f(\sup_{x \in \mathcal{X}} |Z_{n,f}(x, \hat{l}_n)| \leq \hat{c}_n(\alpha)(1 - \epsilon_{3n}) - u_n \sqrt{\log n}) - \delta_{3n} - \delta_{6n} \\
 (52) \quad & \geq \mathbb{P}_f(\|Z_{n,f}\|_{\mathcal{V}_n} \leq \hat{c}_n(\alpha)(1 - \epsilon_{3n}) - u_n \sqrt{\log n}) - \delta_{3n} - \delta_{6n},
 \end{aligned}$$

where (49) follows by the triangle inequality, (50) by Condition H4, (51) by assumption (17), and (52) by the inequality  $\sup_{x \in \mathcal{X}} |Z_{n,f}(x, \hat{l}_n)| \leq \|Z_{n,f}\|_{\mathcal{V}_n}$ . Further, recalling that  $W_{n,f} = \|Z_{n,f}\|_{\mathcal{V}_n}$  and writing

$$\tilde{\epsilon}_n := \epsilon_{1n} + \epsilon_{2n} + \epsilon_{3n} c_{n,f}(\alpha) + u_n \sqrt{\log n},$$

the probability in (52) equals

$$\begin{aligned}
 & \mathbb{P}_f(W_{n,f} \leq \hat{c}_n(\alpha)(1 - \epsilon_{3n}) - u_n \sqrt{\log n}) \\
 (53) \quad & \geq \mathbb{P}_f(W_{n,f} \leq c_{n,f}(\alpha + \tau_n)(1 - \epsilon_{3n}) - \epsilon_{2n} - u_n \sqrt{\log n}) - \delta_{2n} \\
 (54) \quad & \geq \mathbb{P}_f(W_{n,f}^0 \leq c_{n,f}(\alpha + \tau_n) - \tilde{\epsilon}_n) - \delta_{1n} - \delta_{2n} \\
 (55) \quad & \geq 1 - \alpha - \tau_n - p_{\tilde{\epsilon}_n}(|G_{n,f}|) - \delta_{1n} - \delta_{2n},
 \end{aligned}$$

---

\*First version: October 2011. Revised April 12, 2014. V. Chernozhukov and D. Chetverikov are supported by a National Science Foundation grant. K. Kato is supported by the Grant-in-Aid for Young Scientists (B) (25780152), the Japan Society for the Promotion of Science.

where (53) follows by Condition H3, (54) by Condition H1, and (55) by the definition of the Lévy concentration function  $p_{\tilde{c}_n}(|G_{n,f}|)$ . As in the proof of Theorem 3.1,  $c_{n,f}(\alpha) \leq \mathbb{E}[\|G_{n,f}\|_{\mathcal{V}_n}] + (2|\log \alpha|)^{1/2} \leq C\sqrt{\log n}$ . Therefore,  $\tilde{c}_n = o(1/\sqrt{\log n})$  if  $u_n \log n \rightarrow 0$ , and  $\tilde{c}_n \leq Cn^{-c}$  if  $u_n \log n \leq C_1 n^{-c_1}$ . Hence (18) and (19) follow from  $\tau_n + \delta_{1n} + \delta_{2n} + \delta_{3n} + \delta_{6n} \leq Cn^{-c}$ .

To prove (20) and (21), note that when  $\mathcal{L}_n$  is a singleton, inequality  $\sup_{x \in \mathcal{X}} |Z_{n,f}(x, \hat{l}_n)| \leq \|Z_{n,f}\|_{\mathcal{V}_n}$  becomes equality. Therefore, an argument similar to that used above yields in addition to (18) and (19),

$$\begin{aligned} \mathbb{P}_f(f(x) \in \tilde{\mathcal{C}}_n(x), \forall x \in \mathcal{X}) &\leq 1 - \alpha + \tau_n + p_{\tilde{c}_n}(|G_{n,f}|) \\ &\quad + \delta_{1n} + \delta_{2n} + \delta_{3n} + \delta_{6n}, \end{aligned}$$

where  $\tilde{c}'_n := \epsilon_{1n} + \epsilon_{2n} + \epsilon_{3n}c_{n,f}(\alpha - \tau_n) + u_n\sqrt{\log n}$ . Hence, (20) and (21) follow since  $c_{n,f}(\alpha - \tau_n) \leq \mathbb{E}[\|G_{n,f}\|_{\mathcal{V}_n}] + (2|\log(\alpha - \tau_n)|)^{1/2} \leq C\sqrt{\log n}$ . This completes the proof of the corollary.  $\blacksquare$

#### APPENDIX F: PROOFS FOR SECTION 4

In this section, we prove Theorem 4.1. Here constants  $c, C > 0$  depend only on the constants appearing in the statement of Theorem 4.1 but their values may change at each appearance. Also, in the proof we will assume that  $l_{\max,n} - l_{\min,n} \rightarrow \infty$ . Otherwise, the results of the theorem follow from Lemma F.2 below and Theorems 3.2 and 3.1.

For  $(x, l, l') \in \bar{\mathcal{V}}_n$ , denote

$$(56) \quad \sigma_{n,f}(x, l, l') := (c_\sigma \sigma_n(x, l')) \vee \left( n \text{Var}_f(\hat{f}_n(x, l) - \hat{f}_n(x, l')) \right)^{1/2}.$$

The proof is long, so we start with several preliminary lemmas.

LEMMA F.1. *Let  $\phi$  be a father wavelet satisfying the latter part of Condition L1. Then there exist constants  $c_\phi, C_\phi > 0$  depending only on  $\phi$  such that for all  $x = (x_1, \dots, x_d)' \in \mathbb{R}^d$ ,*

$$(57) \quad c_\phi \leq \sum_{k_1, \dots, k_d \in \mathbb{Z}} \prod_{1 \leq m \leq d} \phi(x_m - k_m)^2 \leq C_\phi.$$

PROOF OF LEMMA F.1. Since

$$\sum_{k_1, \dots, k_d \in \mathbb{Z}} \prod_{1 \leq m \leq d} \phi(x_m - k_m)^2 = \prod_{1 \leq m \leq d} \sum_{k_m \in \mathbb{Z}} \phi(x_m - k_m)^2,$$

it suffices to consider the case  $d = 1$ . Then (57) becomes

$$c_\phi \leq \sum_{k \in \mathbb{Z}} \phi(x - k)^2 \leq C_\phi.$$

Since the function  $\sum_{k \in \mathbb{Z}} \phi(\cdot - k)^2$  has period 1, it suffices to consider  $x \in [0, 1]$ . To prove the lower bound, note that it follows from Lemma 8.6 in [7] that the series  $\sum_{k \in \mathbb{Z}} \phi(\cdot - k)^2$  converges uniformly, and hence  $\sum_{k \in \mathbb{Z}} \phi(\cdot - k)^2$  is continuous (the functions  $\sum_{k: |k| \leq m} \phi(\cdot - k)^2$  are continuous for all  $m$  by our assumptions on the regularity of the farther wavelet). Further, it follows from Corollary 8.1 in [7] that  $\sum_{k \in \mathbb{Z}} \phi(\cdot - k)$  is identically equal to some non-zero constant, so that  $\sum_{k \in \mathbb{Z}} \phi(x - k)^2 > 0$  for all  $x \in [0, 1]$ . Since the minimum of a continuous function on a compact set is achieved, the lower bound follows. The upper bound follows by noting that both for compactly supported wavelets and for Battle-Lemarié wavelets,  $|\phi(x)| \leq C \exp(-c|x|)$  for all  $x \in \mathbb{R}$ . The asserted claim follows.  $\blacksquare$

LEMMA F.2. *Under the assumptions of Theorem 4.1, there exists  $n_0$  such that for all  $n \geq n_0$ , Condition VC holds with  $b_n \leq C2^{l_{\max, n} d/2}$ ,  $a_n \leq C$ ,  $v_n \leq C$ . In addition,  $\sigma_n \leq C$  and for all  $f \in \mathcal{F}$ ,  $l \in \mathcal{L}_n$ , and  $n \geq n_0$ , we have*

$$(58) \quad \underline{\sigma}2^{ld/2} \leq \inf_{x \in \mathcal{X}} \sigma_{n,f}(x, l) \leq \sup_{x \in \mathcal{X}} \sigma_{n,f}(x, l) \leq \bar{\sigma}2^{ld/2}$$

and for all  $f \in \mathcal{F}$ ,  $l, l' \in \mathcal{L}_n$  such that  $l < l'$ , and  $n \geq n_0$ , we have

$$(59) \quad c_\sigma \underline{\sigma}2^{l'd/2} \leq \inf_{x \in \mathcal{X}} \sigma_{n,f}(x, l, l') \leq \sup_{x \in \mathcal{X}} \sigma_{n,f}(x, l, l') \leq \bar{\sigma}2^{l'd/2}.$$

Moreover, there exists  $s > 0$  such that for all  $x \in \mathcal{X}$ ,  $f \in \mathcal{F}$ ,  $l, l' \in \mathcal{L}_n$  such that  $l \leq l' - s$ , and  $n \geq n_0$ , we have

$$(60) \quad \sigma_{n,f}(x, l, l') > c_\sigma \sigma_n(x, l').$$

Here  $n_0$ ,  $s$ ,  $\underline{\sigma}$ , and  $\bar{\sigma}$  depend only on the constants appearing in the statement of Theorem 4.1.

PROOF OF LEMMA F.2. For simplicity of notation, we assume that  $d = 1$  in this proof. The case  $d > 1$  follows similarly. Note that we have

$$(61) \quad 2^{-l} |K_l(y, x)| \leq C \exp(-2^l c|y - x|),$$

for all  $l \in \mathcal{L}_n$  and  $x, y \in \mathcal{X}$ . Indeed, for convolution kernels and compactly supported wavelets, this follows from compactness of the support of  $K(\cdot)$  and  $\phi(\cdot)$ , respectively, and for Battle-Lemarié wavelets, this follows from Lemma 8.6 in [7].

Further, note that under our conditions,

$$\mathbb{E}_f[K_l(X_1, x)] \rightarrow f(x)$$

as  $n \rightarrow \infty$  uniformly over all  $f \in \mathcal{F}$ ,  $l \in \mathcal{L}_n$ , and  $x \in \mathcal{X}$  since  $l_{\min, n} \rightarrow \infty$ . In addition, it follows from (61) that under our conditions

$$\begin{aligned} & 2^{-l} \left| \int K_l(y, x)^2 (f(y) - f(x)) dy \right| \\ & \leq 2^l C \int \exp(-2^l c |y - x|) |f(y) - f(x)| dy \\ & \leq C \int \exp(-c |z - \tilde{x}|) |f(z 2^{-l}) - f(\tilde{x} 2^{-l})| dz \rightarrow 0 \end{aligned}$$

uniformly over all  $f \in \mathcal{F}$ ,  $l \in \mathcal{L}_n$ , and  $x \in \mathcal{X}$  where we denoted  $\tilde{x} = x 2^l$ . Therefore,

$$2^{-l} \mathbb{E}_f [K_l(X_1, x)^2] \rightarrow f(x) \int K(y)^2 dy$$

as  $n \rightarrow \infty$  uniformly over all  $f \in \mathcal{F}$ ,  $l \in \mathcal{L}_n$ , and  $x \in \mathcal{X}$  for convolution kernels. Moreover, for wavelet projection kernels,

$$\begin{aligned} 2^{-l} \mathbb{E}_f [K_l(X_1, x)^2] &= 2^l \int \left( \sum_{k \in \mathbb{Z}} \phi(2^l y - k) \phi(2^l x - k) \right)^2 f(y) dy \\ &= 2^l \int \left( \sum_{k \in \mathbb{Z}} \phi(2^l y - k) \phi(2^l x - k) \right)^2 f(x) dy + o(1) \\ &= f(x) \sum_{k \in \mathbb{Z}} \phi(2^l x - k)^2 + o(1) \end{aligned}$$

as  $n \rightarrow \infty$  uniformly over all  $f \in \mathcal{F}$ ,  $l \in \mathcal{L}_n$ , and  $x \in \mathcal{X}$  where the second line follows from the argument above and the third line from orthonormality in  $L^2(\mathbb{R})$  of the system  $\{\phi(\cdot - k), k \in \mathbb{Z}\}$ . Since  $\underline{f} \leq f(x) \leq \bar{f}$  for all  $x \in \mathcal{X}$  and (57) holds under our conditions, (58) follows from noting that  $\sigma_{n, f}^2 = \mathbb{E}_f [K_l(X_1, x)^2] - \mathbb{E}_f [K_l(X_1, x)]^2$ . Further, the lower bound in (59) follows from (56) and (58), and the upper bound in (59) as well as (60) follow from (58) by noting that

$$(\sigma_{n, f}(x, l') - \sigma_{n, f}(x, l))^2 \leq \sigma_{n, f}(x, l, l')^2 \leq (\sigma_{n, f}(x, l') + \sigma_{n, f}(x, l))^2.$$

Inequalities (58) also imply that  $\sigma_n \leq C$ .

Further, when  $d = 1$ , the function class  $\bar{\mathcal{K}}^d := \{2^{-ld} K_l(\cdot, x) : l \in \mathbb{N}, x \in \mathbb{R}^d\}$  is VC( $b, a, v$ ) type for some  $b, a$ , and  $v$  independent of  $n$  by discussion on p. 911 in [3] (for convolution kernels), Lemma 2 in [4] (for compactly supported wavelets), and Lemma 2 in [5] (for Battle-Lemarié wavelets). When  $d > 1$ ,  $\bar{\mathcal{K}}^d$  is VC( $b, a, v$ ) type class (with possibly different  $b, a$ , and  $v$ ) by Lemma B.2. Now, another application of Lemma B.2 shows that  $\mathcal{K}_{n, f}$  is VC( $b_n, a_n, v_n$ ) type class where  $b_n \leq C 2^{l_{\max, n} d/2}$ ,  $a_n \leq C$ , and  $v_n \leq C$ . This completes the proof of the lemma.  $\blacksquare$



LEMMA F.3. *Under the assumptions of Theorem 4.1, Conditions H1-H4 are satisfied.*

PROOF OF LEMMA F.3. The claim follows from combining Lemma F.2 and Theorem 3.2. ■

LEMMA F.4. *Under the assumptions of Theorem 4.1, there exist sequences  $\zeta_{1n}$  and  $\zeta_{2n}$  of positive numbers bounded from above by  $Cn^{-c}$  such that*

$$\sup_{f \in \mathcal{F}} \mathbb{P}_f \left( \sup_{(x,l,l') \in \bar{\mathcal{V}}_n} \left| \frac{\hat{\sigma}_n(x,l,l')}{\sigma_{n,f}(x,l,l')} - 1 \right| > \zeta_{1n} \right) \leq \zeta_{2n}.$$

PROOF OF LEMMA F.4. The proof follows from arguments similar to those used to verify Condition H4 in the proof of Theorem 3.2. ■

LEMMA F.5. *Under the assumptions of Theorem 4.1, there exist  $c > 0$  and  $C > 0$  such that for all  $f \in \mathcal{F}$ ,*

$$\lambda_{0n} := 1 - \mathbb{P}_f(c\sqrt{\log n} \leq \tilde{c}_n(\gamma_n) \leq C\sqrt{\log n}) \leq Cn^{-c}.$$

PROOF OF LEMMA F.5. Since we assume that  $l_{\max,n} - l_{\min,n} \rightarrow \infty$ , for sufficiently large  $n$ , there will be  $l, l' \in \mathcal{L}_n$  such that  $l \leq l' - s$ . For such  $n$ ,  $l$ , and  $l'$ ,  $\hat{\sigma}_n(x,l,l') > c_\sigma \hat{\sigma}_n(x,l')$  with probability at least  $1 - Cn^{-c}$  by Lemmas F.4 and F.2. On the other hand, on the event  $\hat{\sigma}_n(x,l,l') > c_\sigma \hat{\sigma}_n(x,l')$ , by construction,  $\tilde{c}_n(\gamma_n)$  is not smaller than  $(1 - \gamma_n)$ -quantile of  $N(0, 1)$  distribution. Since  $\gamma_n \leq Cn^{-c}$ , it follows that

$$\mathbb{P}_f(\tilde{c}_n(\gamma_n) < c\sqrt{\log n}) \leq Cn^{-c}.$$

Now let us verify the inequality

$$(62) \quad \mathbb{P}_f(\tilde{c}_n(\gamma_n) > C\sqrt{\log n}) \leq Cn^{-c}.$$

Recall that  $\bar{\mathcal{V}}_n = \{(x, l, l') : x \in \mathcal{X}, l, l' \in \mathcal{L}_n, l < l'\}$ . Applying Lemma B.1 conditional on the data given that  $X_1^n = x_1^n$  yields

$$\tilde{c}_n(\gamma_n) \leq \mathbb{E} \left[ \|\tilde{\mathbb{G}}_n(x_1^n, \xi_1^n)\|_{\bar{\mathcal{V}}_n} \right] + \sqrt{2|\log \gamma_n|}.$$

Also, for  $(x, l, l') \in \bar{\mathcal{V}}_n$ , with probability at least  $1 - Cn^{-c}$ ,

$$\begin{aligned} & |\tilde{\mathbb{G}}_n(x_1^n, \xi_1^n)(x, l, l')| \\ & \leq \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i \frac{K_l(X_i, x) - \hat{f}_n(x, l)}{\hat{\sigma}_n(x, l, l')} \right| + \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i \frac{K_{l'}(X_i, x) - \hat{f}_n(x, l')}{\hat{\sigma}_n(x, l, l')} \right| \\ & \leq \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i \frac{K_l(X_i, x) - \hat{f}_n(x, l)}{c_\sigma \hat{\sigma}_n(x, l')} \right| + \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i \frac{K_{l'}(X_i, x) - \hat{f}_n(x, l')}{c_\sigma \hat{\sigma}_n(x, l')} \right|, \\ & \leq C \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i \frac{K_l(X_i, x) - \hat{f}_n(x, l)}{c_\sigma \hat{\sigma}_n(x, l)} \right| + \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i \frac{K_{l'}(X_i, x) - \hat{f}_n(x, l')}{c_\sigma \hat{\sigma}_n(x, l')} \right|, \end{aligned}$$

so that

$$\mathbb{E} \left[ \|\tilde{\mathbb{G}}_n(x_1^n, \xi_1^n)\|_{\bar{\mathcal{V}}_n} \right] \leq C \mathbb{E} \left[ \|\hat{\mathbb{G}}_n(x_1^n, \xi_1^n)\|_{\mathcal{V}_n} \right]$$

on the set  $S_{n,1}$  of values  $x_1^n$  such that  $\mathbb{P}_f(X_1^n \notin S_{n,1}) \leq Cn^{-c}$ . It follows from Dudley's inequality that under Condition VC, which is verified in Lemma F.2, that

$$\mathbb{E} \left[ \|\hat{\mathbb{G}}_n(x_1^n, \xi_1^n)\|_{\mathcal{V}_n} \right] \leq C \sqrt{\log n}$$

on the set  $S_{n,2}$  of values  $x_1^n$  such that  $\mathbb{P}_f(X_1^n \notin S_{n,2}) \leq Cn^{-c}$  (see the proof of Theorem 3.2 in the main text for a similar derivation). Combining presented inequalities gives (62) and completes the proof of the lemma. ■

LEMMA F.6. *Under the assumptions of Theorem 4.1, there exist  $c > 0$  and  $C > 0$  such that*

$$\sup_{f \in \mathcal{F}} \mathbb{P}_f \left( \sup_{(x, l, l') \in \bar{\mathcal{V}}_n(s)} \frac{\sqrt{n} |B_n(x, l, l')|}{\hat{\sigma}_n(x, l, l')} \geq \tilde{c}_n(\gamma_n) \right) \leq \gamma_n + Cn^{-c}$$

where

$$B_n(x, l, l') := \left( \hat{f}_n(x, l) - \mathbb{E}_f[\hat{f}_n(x, l)] \right) - \left( \hat{f}_n(x, l') - \mathbb{E}_f[\hat{f}_n(x, l')] \right)$$

and

$$\bar{\mathcal{V}}_n(s) := \{(x, l, l') : x \in \mathcal{X}, l, l' \in \mathcal{L}_n, l \leq l' - s\}$$

where  $s$  appears in Lemma F.2.

PROOF OF LEMMA F.6. Note that  $\tilde{c}_n(\gamma_n)$  is not smaller than conditional  $(1 - \gamma_n)$ -quantile of  $\|\tilde{\mathbb{G}}_n\|_{\bar{\mathcal{V}}_n(s)}$  given  $X_1^n$ . Therefore, the proof follows from arguments similar to those used in the proof of Theorem 3.1 since for  $(x, l, l') \in \bar{\mathcal{V}}_n(s)$ ,  $\sigma_{n,f}(x, l, l') = (n \text{Var}_f(\hat{f}_n(x, l) - \hat{f}_n(x, l')))^{1/2}$  (truncation is not binding). ■

PROOF OF THEOREM 4.1. First, we note that for any  $t' > t(f)$ ,  $f \notin \Sigma(t', L)$ ; otherwise, we would have that  $\sup_{x \in \mathcal{X}} |\mathbb{E}_f[\hat{f}_n(x, l)] - f(x)| \leq C2^{-lt'}$  contradicting the lower bound in (31). Therefore, (36) implies (37), and so it suffices to verify Conditions H1-H6 and to prove (15) and (36).

By Lemma F.3, we can and will assume that Conditions H1-H4 hold. In addition, Condition H6 with some  $u_n$  satisfying  $cu'_n \leq u_n \leq Cu'_n$  follows from Lemma F.5.

We now show that under our assumptions, Conditions H5 is also satisfied. Let  $\delta_{4n} = 1$  for  $n \leq \max(n_0, n_1)$ , where  $n_0$  appears in Lemma F.2, and  $n_1$  is chosen so that for all  $n > n_1$  and  $l \in \mathcal{L}_n$ ,  $l \geq l_0$  for  $l_0$  appearing in Condition L2, so that Condition H5 holds for these  $n$ 's with  $C_1$  sufficiently large and  $c_1$  sufficiently small. Therefore, it suffices to consider  $n > \max(n_0, n_1)$ .

Let  $t := t(f)$  and let  $M_1$ , and  $M_2$  be such that  $M_1(c_32^{st}/C_3 - 1) \geq 2(q+1)$ , and  $M_2\bar{\sigma}/\underline{\sigma} \leq (q-1)/4$  where  $s$ ,  $\underline{\sigma}$ , and  $\bar{\sigma}$  are introduced in Lemma F.2, and  $q$  appears in (35) (by increasing  $s$  if necessary, without loss of generality, we assume that  $c_32^{st} > C_3$ ). For  $M > 0$ , define

$$(63) \quad l^*(M) := \inf \left\{ l \in \mathcal{L}_n : C_32^{-lt}\sqrt{n} \leq M\tilde{c}_n(\gamma_n)\bar{\sigma}2^{ld/2} \right\},$$

and let  $l_1^* := l^*(M_1)$  and  $l_2^* := l^*(M_2)$ . We will invoke the following lemmas.

LEMMA F.7. *There exist  $c, C > 0$  such that for all  $f \in \mathcal{F}$ ,*

$$\lambda_{1n} := \mathbb{P}_f(\hat{l}_n < l_1^* - s) \leq Cn^{-c}.$$

PROOF OF LEMMA F.7. Define  $\mathcal{L}_n^1 := \{l \in \mathcal{L}_n : l < l_1^* - s\}$ . If  $\mathcal{L}_n^1$  is empty, we are done. So, assume that  $\mathcal{L}_n^1$  is not empty. Let  $l' := l_1^* - 1$ . Note that  $l' \in \mathcal{L}_n$  by Condition L3. Then for any  $l \in \mathcal{L}_n^1$ , by Condition L2 and the triangle inequality,

$$\begin{aligned} c_32^{-lt} &\leq \sup_{x \in \mathcal{X}} |\mathbb{E}_f[\hat{f}_n(x, l)] - f(x)| \\ &\leq \sup_{x \in \mathcal{X}} |B_n(x, l, l')| + \sup_{x \in \mathcal{X}} |\hat{f}_n(x, l) - \hat{f}_n(x, l')| + C_32^{-l't}, \end{aligned}$$

so that

$$\sup_{x \in \mathcal{X}} |\hat{f}_n(x, l) - \hat{f}_n(x, l')| \geq c_32^{-lt} - C_32^{-l't} - \sup_{x \in \mathcal{X}} |B_n(x, l, l')|.$$

Further, for  $l \in \mathcal{L}_n^1$ , it follows from (63) that

$$\begin{aligned} c_32^{-lt} - C_32^{-l't} &\geq C_32^{-l't}(c_32^{st}/C_3 - 1) \\ &\geq n^{-1/2}M_1\tilde{c}_n(\gamma_n)\bar{\sigma}2^{ld/2}(c_32^{st}/C_3 - 1) \\ &\geq 2n^{-1/2}(q+1)\tilde{c}_n(\gamma_n)\bar{\sigma}2^{ld}. \end{aligned}$$

Therefore,

$$\begin{aligned}
\mathbb{P}_f(\hat{l}_n \in \mathcal{L}_n^1) &\leq \mathbb{P}_f\left(\inf_{l \in \mathcal{L}_n^1} \sup_{x \in \mathcal{X}} \frac{\sqrt{n}|\hat{f}_n(x, l) - \hat{f}_n(x, l')|}{\hat{\sigma}_n(x, l, l')} \leq q\tilde{c}_n(\gamma_n)\right) \\
&\leq \mathbb{P}_f\left(\inf_{l \in \mathcal{L}_n^1} \sup_{x \in \mathcal{X}} \frac{\sqrt{n}|\hat{f}_n(x, l) - \hat{f}_n(x, l')|}{2\bar{\sigma}2^{l'd/2}} \leq q\tilde{c}_n(\gamma_n)\right) + Cn^{-c} \\
&\leq \mathbb{P}_f\left(\sup_{l \in \mathcal{L}_n^1} \sup_{x \in \mathcal{X}} \frac{\sqrt{n}|B_n(x, l, l')|}{2\bar{\sigma}2^{l'd/2}} \geq \tilde{c}_n(\gamma_n)\right) + Cn^{-c} \\
&\leq \mathbb{P}_f\left(\sup_{(x, l, l') \in \mathcal{V}_n(s)} \frac{\sqrt{n}|B_n(x, l, l')|}{\hat{\sigma}_n(x, l, l')} \geq \tilde{c}_n(\gamma_n)\right) + Cn^{-c} \leq Cn^{-c}
\end{aligned}$$

where the second line follows from Lemmas F.2 and F.4, the third line from inequalities presented above, the and fourth line from Lemmas F.2, F.4, and F.6. This completes the proof of the lemma.  $\blacksquare$

LEMMA F.8. *There exist  $c, C > 0$  such that for all  $f \in \mathcal{F}$ ,*

$$\lambda_{2n} := \mathbb{P}_f(\hat{l}_n \geq l_2^* + s) \leq Cn^{-c}.$$

PROOF OF LEMMA F.8. Define  $\mathcal{L}_n^2 := \{l \in \mathcal{L}_n : l \geq l_2^* + s\}$ . If  $\mathcal{L}_n^2$  is empty, we are done. So assume that  $\mathcal{L}_n^2$  is not empty. Then for any  $l \in \mathcal{L}_n^2$ , by Condition L2 and the triangle inequality,

$$|\hat{f}_n(x, l) - \hat{f}_n(x, l_2^*)| \leq \sup_{x \in \mathcal{X}} |B_n(x, l_2^*, l)| + C_3 2^{-lt} + C_3 2^{-l_2^* t}.$$

Further, for  $l \in \mathcal{L}_n^2$ , it follows from (63) by construction of  $M_2$  that

$$C_3 2^{-lt} \sqrt{n} \leq M_2 \tilde{c}_n(\gamma_n) \bar{\sigma} 2^{ld/2} \leq 4^{-1}(q-1) \tilde{c}_n(\gamma_n) \underline{\sigma} 2^{ld/2},$$

and similarly,

$$C_3 2^{-l_2^* t} \sqrt{n} \leq M_2 \tilde{c}_n(\gamma_n) \bar{\sigma} 2^{l_2^* d/2} \leq 4^{-1}(q-1) \tilde{c}_n(\gamma_n) \underline{\sigma} 2^{ld/2}$$

since  $l > l_2^*$ . Therefore,

$$\begin{aligned}
\mathbb{P}_f(\hat{l}_n \in \mathcal{L}_n^2) &\leq \mathbb{P}_f\left(\sup_{l \in \mathcal{L}_n^2} \sup_{x \in \mathcal{X}} \frac{\sqrt{n}|\hat{f}_n(x, l) - \hat{f}_n(x, l_2^*)|}{\hat{\sigma}_n(x, l_2^*, l)} > q\tilde{c}_n(\gamma_n)\right) \\
&\leq \mathbb{P}_f\left(\sup_{l \in \mathcal{L}_n^2} \sup_{x \in \mathcal{X}} \frac{\sqrt{n}|B_n(x, l_2^*, l)|}{\hat{\sigma}_n(x, l_2^*, l)} > \tilde{c}_n(\gamma_n)\right) + Cn^{-c} \\
&\leq \mathbb{P}_f\left(\sup_{(x, l, l') \in \mathcal{V}_n(s)} \frac{\sqrt{n}|B_n(x, l, l')|}{\hat{\sigma}_n(x, l, l')} > \tilde{c}_n(\gamma_n)\right) + Cn^{-c} \leq Cn^{-c}
\end{aligned}$$

where the second line follows from inequalities presented above and Lemmas F.2 and F.4 and the third line from Lemma F.6. This completes the proof of the lemma.  $\blacksquare$

Let  $l_0^* := l_0^*(f)$  be  $l \in \mathcal{L}_n$  satisfying (33), which exists by Condition L3. Now we can verify Condition H5:

LEMMA F.9. *Under our assumptions, Condition H5 is satisfied.*

PROOF OF LEMMA F.9. We claim that with probability at least  $1 - \lambda_{0n}$ ,

$$(64) \quad \sqrt{n}2^{-l_1^*(t+d/2)} \leq C\tilde{c}_n(\gamma_n).$$

Indeed, consider the event that

$$\tilde{c}_n(\gamma_n) \geq c\sqrt{\log n}.$$

By Lemma F.5, the probability of this event is at least  $1 - \lambda_{0n}$ . On this event, if  $l_1^* \geq l_0^*$ , then

$$\sqrt{n}2^{-l_1^*(t+d/2)} \leq \sqrt{n}2^{-l_0^*(t+d/2)} \leq \sqrt{C_4 \log n} \leq C\tilde{c}_n(\gamma_n),$$

and if  $l_1^* < l_0^*$ , then

$$C_3 2^{-l_1^* t} \sqrt{n} \leq M_1 \tilde{c}_n(\gamma_n) \bar{\sigma} 2^{l_1^* d/2},$$

so that (64) holds in both cases.

Hence, with probability at least  $1 - \lambda_{0n} - \lambda_{1n}$ ,

$$(65) \quad \Delta_{n,f}(\hat{l}_n) \leq \frac{\sqrt{n}C_3 2^{-\hat{l}_n t}}{\inf_{x \in \mathcal{X}} \sigma_{n,f}(x, \hat{l}_n)} \leq \sqrt{n}C_3 2^{-\hat{l}_n t} / (\underline{\sigma} 2^{\hat{l}_n d/2})$$

$$(66) \quad = C\sqrt{n}2^{-\hat{l}_n(t+d/2)} \leq C\sqrt{n}2^{-(l_1^*-s)(t+d/2)} \leq C\tilde{c}_n(\gamma_n),$$

where (65) follows from Condition L2 and Lemma F.2, and (66) from Lemma F.7, and (64). Since  $\lambda_{0n} + \lambda_{1n} \leq Cn^{-c}$ , Condition H5 follows because  $c'_n = u'_n \tilde{c}_n(\gamma_n)$  and  $u'_n$  is sufficiently large ( $u'_n \geq C(\mathcal{F})$ ).  $\blacksquare$

Finally, to prove the theorem, we will use the following lemma:

LEMMA F.10. *There exist  $c, C > 0$  such that*

$$\lambda_{3n} := \sup_{f \in \mathcal{F}} \mathbb{P}_f \left( \sup_{x \in \mathcal{X}} \hat{\sigma}_n(x, \hat{l}_n)^2 > C \left( \frac{\log n}{n} \right)^{-d/(2t+d)} \right) \leq Cn^{-c}.$$

PROOF OF LEMMA F.10. We claim that with probability at least  $1 - \lambda_{0n}$ ,

$$(67) \quad 2^{-(l_2^* - 1)(t + d/2)} \sqrt{n} \geq c\sqrt{\log n}.$$

Indeed, consider the event that

$$\tilde{c}_n(\gamma_n) \geq c\sqrt{\log n}.$$

By Lemma F.5, the probability of this event is at least  $1 - \lambda_{0n}$ . On this event, if  $l_2^* - 1 \leq l_0^*$ , then

$$\sqrt{n}2^{-(l_2^* - s)(t + d/2)} \geq \sqrt{n}2^{-l_0^*(t + d/2)} \geq \sqrt{c_4 \log n},$$

and if  $l_2^* - 1 > l_0^*$ , then  $l' = l_2^* - 1$  satisfies

$$\sqrt{n}2^{-l'(t + d/2)} \geq c\tilde{c}_n(\gamma_n) \geq c\sqrt{\log n}.$$

Therefore, (67) holds in both cases on this event.

Hence, by Lemma F.8, with probability at least  $1 - \lambda_{0n} - \lambda_{2n}$ ,

$$2^{-\hat{l}_n(t + d/2)} \sqrt{n} \geq c\sqrt{\log n}.$$

Conclude that with probability at least  $1 - \delta_{3n} - \lambda_{0n} - \lambda_{2n}$ ,

$$\sup_{x \in \mathcal{X}} \hat{\sigma}_n(x, \hat{l}_n)^2 \leq (1 + \epsilon_{3n})^2 \sup_{x \in \mathcal{X}} \sigma_{n,f}(x, \hat{l}_n)^2 \leq C2^{\hat{l}_n d} \leq C \left( \frac{\log n}{n} \right)^{-d/(2t+d)}.$$

Since  $\delta_{3n} + \lambda_{0n} + \lambda_{2n} \leq Cn^{-c}$ , the asserted claim follows.  $\blacksquare$

We now finish the proof of the theorem. We have by now verified Conditions H1-H6. Since Conditions H1-H6 hold with Condition H6 being satisfied with  $u_n \leq Cu'_n \leq C \log n$ , Theorem 3.1 applies, so that (15) from the main text holds. Further, by construction,

$$\sup_{x \in \mathcal{X}} \lambda(\mathcal{C}_n(x)) = 2(\hat{c}_n(\alpha) + c'_n) \sup_{x \in \mathcal{X}} \hat{\sigma}_n(x, \hat{l}_n) / \sqrt{n}.$$

Therefore, combining Conditions H3 and H6 and Lemma F.10, we have

$$(68) \quad \sup_{f \in \mathcal{F}} \mathbb{P}_f \left( \sup_{x \in \mathcal{X}} \lambda(\mathcal{C}_n(x)) > C\bar{c}_n \frac{r_n(t(f))}{\sqrt{\log n}} \right) \leq \delta_{2n} + \delta_{5n} + \lambda_{3n},$$

where

$$\bar{c}_n := c_{n,f}(\alpha - \tau_n) + \epsilon_{2n} + u'_n \sqrt{\log n}.$$

Since  $\tau_n$  and  $\epsilon_{2n}$  are both bounded by  $C_1 n^{-c_1}$  (Condition H3), there exists  $n_2$  such that  $\tau_n \leq \alpha/2$  and  $\epsilon_{2n} \leq 1$  for  $n \geq n_2$ . For  $n < n_2$ , (36) from the main text holds by choosing sufficiently large  $C$ . Consider  $n \geq n_2$ . Then

$$c_{n,f}(\alpha - \tau_n) + \epsilon_{2n} \leq c_{n,f}(\alpha/2) + 1$$

By Lemma B.1,  $c_{n,f}(\alpha/2) \leq \mathbb{E}[\|G_{n,f}\|_{\mathcal{V}_n}] + \sqrt{2|\log(\alpha/2)|}$ . By Condition H1,  $\mathbb{E}[\|G_{n,f}\|_{\mathcal{V}_n}]$  is bounded from below, and so  $c_{n,f}(\alpha/2) + 1 \leq CE[\|G_{n,f}\|_{\mathcal{V}_n}]$ . Further,  $\mathbb{E}[\|G_{n,f}\|_{\mathcal{V}_n}] \leq C_1\sqrt{\log n}$  (Condition H1) gives

$$\bar{c}_n \leq C(1 + u'_n)\sqrt{\log n}.$$

Substituting this expression into (68) yields (36) from the main text. This completes the proof of the theorem.  $\blacksquare$

## APPENDIX G: PROOFS FOR APPENDIX A

The proof of Theorem A.2 uses the following technical results.

**THEOREM G.1.** *Let  $X$  and  $Y$  be Gaussian random vectors in  $\mathbb{R}^p$  with mean zero and covariance matrices  $\Sigma^X$  and  $\Sigma^Y$ , respectively. Then for every  $g \in C^2(\mathbb{R})$ ,*

$$\left| \mathbb{E} \left[ g \left( \max_{1 \leq j \leq p} X_j \right) \right] - \mathbb{E} \left[ g \left( \max_{1 \leq j \leq p} Y_j \right) \right] \right| \leq \|g''\|_\infty \Delta/2 + 2\|g'\|_\infty \sqrt{2\Delta \log p},$$

where  $\Delta = \max_{1 \leq j, k \leq p} |\Sigma_{jk}^X - \Sigma_{jk}^Y|$ .

**PROOF.** See Theorem 1 in [2].  $\blacksquare$

**THEOREM G.2.** *Let  $\mu$  and  $\nu$  be Borel probability measures on  $\mathbb{R}$ . Let  $\varepsilon > 0$  and  $\delta > 0$ . Suppose that  $\mu(A) \leq \nu(A^\delta) + \varepsilon$  for every Borel subset  $A$  of  $\mathbb{R}$ . Let  $V$  be a random variable with distribution  $\mu$ . Then there is a random variable  $W$  with distribution  $\nu$  such that  $\mathbb{P}(|V - W| > \delta) \leq \varepsilon$ .*

**PROOF.** See Lemma 4.1 in [1].  $\blacksquare$

**THEOREM G.3.** *Let  $\beta > 0$  and  $\delta > 1/\beta$ . For every Borel subset  $B$  of  $\mathbb{R}$ , there is a smooth function  $g : \mathbb{R} \rightarrow \mathbb{R}$  and absolute constant  $A > 0$  such that  $\|g'\|_\infty \leq \delta^{-1}$ ,  $\|g''\|_\infty \leq A\beta\delta^{-1}$ , and for all  $t \in \mathbb{R}$*

$$(1 - \varepsilon)1_B(t) \leq g(t) \leq \varepsilon + (1 - \varepsilon)1_{B^{3\delta}}(t),$$

where  $\varepsilon = \varepsilon_{\beta, \delta}$  is given by

$$\varepsilon = \sqrt{e^{-\alpha}(1 + \alpha)} < 1, \quad \alpha = \beta^2\delta^2 - 1.$$

PROOF. See Lemma 4.2 in [1]. ■

We are now in position to prove Theorem A.2.

PROOF OF THEOREM A.2. In this proof,  $C$  is an absolute constant but its value can change at each appearance. Define  $\mathcal{G} \cdot \mathcal{G} = \{g \cdot \tilde{g} : g, \tilde{g} \in \mathcal{G}\}$  and  $(\mathcal{G} - \mathcal{G})^2 = \{(g - \tilde{g})^2 : g, \tilde{g} \in \mathcal{G}\}$ . Lemma B.2 implies that  $\mathcal{G} \cdot \mathcal{G}$  is VC( $b^2, 2a, 2v$ ) type and  $(\mathcal{G} - \mathcal{G})^2$  is VC( $4b^2, 2a, 4v$ ) type function classes. In addition,  $\mathbb{E}[g^2] \leq b^2\sigma^2$  for all  $g \in \mathcal{G} \cdot \mathcal{G}$  and  $\mathbb{E}[g^2] \leq 16b^2\sigma^2$  for all  $g \in (\mathcal{G} - \mathcal{G})^2$ . Together with the assumed condition  $b^2K_n \leq n\sigma^2$ , this justifies an application of Talagrand's inequality (Theorem B.1) with  $t = \log n$ , which gives

$$(69) \quad \mathbb{P} \left( \sup_{g \in \mathcal{G}} |\mathbb{E}_n[g(X_i)] - \mathbb{E}[g(X_1)]| \leq \sqrt{\frac{\sigma^2 K_n}{n}} \right) \geq 1 - \frac{1}{n},$$

$$(70) \quad \mathbb{P} \left( \sup_{g \in \mathcal{G} \cdot \mathcal{G}} |\mathbb{E}_n[g(X_i)] - \mathbb{E}[g(X_1)]| \leq \sqrt{\frac{b^2 \sigma^2 K_n}{n}} \right) \geq 1 - \frac{1}{n},$$

$$(71) \quad \mathbb{P} \left( \sup_{g \in (\mathcal{G} - \mathcal{G})^2} |\mathbb{E}_n[g(X_i)] - \mathbb{E}[g(X_1)]| \leq \sqrt{\frac{b^2 \sigma^2 K_n}{n}} \right) \geq 1 - \frac{1}{n}.$$

Let  $S_{n,0} \in \mathcal{S}^n$  be the intersection of events in (69)-(71). Then  $\mathbb{P}(X_1^n \in S_{n,0}) \geq 1 - 3/n$ .

Fix any  $x_1^n \in S_{n,0}$ . Let  $\tau = \sigma/(bn^{1/2})$ , and let  $\{g_1, \dots, g_N\} \subset \mathcal{G}$  be a subset of elements of  $\mathcal{G}$  such that for any  $g \in \mathcal{G}$  there exists  $j = j(g) \in \{1, \dots, N\}$  such that  $\mathbb{E}[(g(X_1) - g_j(X_1))^2] \leq b^2\tau^2$ . We can and will assume that  $N \leq (a/\tau)^v$ . Define

$$W(x_1^n)(\tau) := \max_{1 \leq j \leq N} |\tilde{\mathbb{G}}_n(x_1^n, \xi_1^n)(g_j)|,$$

$$W^0(\tau) := \max_{1 \leq j \leq N} |B(g_j)|.$$

In addition, define  $\tilde{W}^0 := \|B\|_{\mathcal{G}}$  and

$$\mathcal{G}(\tau) := \{g - \tilde{g} : g, \tilde{g} \in \mathcal{G}, \mathbb{E}[(g(X_1) - \tilde{g}(X_1))^2] \leq b^2\tau^2\}.$$

Clearly, we have  $|\tilde{W}_n(x_1^n) - W(x_1^n)(\tau)| \leq \|\tilde{\mathbb{G}}_n(x_1^n, \xi_1^n)\|_{\mathcal{G}(\tau)}$  and  $|\tilde{W}^0 - W^0(\tau)| \leq \|B\|_{\mathcal{G}(\tau)}$ . The rest of the proof consists of 3 steps. Steps 1 and 2 provide bounds on  $\|\tilde{\mathbb{G}}_n(x_1^n, \xi_1^n)\|_{\mathcal{G}(\tau)}$  and  $\|B\|_{\mathcal{G}(\tau)}$ , respectively. Step 3 gives a coupling inequality and finishes the proof using a method for comparing  $W(x_1^n)(\tau)$  and  $W^0(\tau)$ .



*Step 1 (Bound on  $\|\tilde{\mathbb{G}}_n(x_1^n, \xi_1^n)\|_{\mathcal{G}(\tau)}$ ).* Here we show that with probability at least  $1 - 2/n$ ,

$$\|\tilde{\mathbb{G}}_n(x_1^n, \xi_1^n)\|_{\mathcal{G}(\tau)} \leq L \left\{ \sqrt{\frac{\sigma^2 K_n}{n}} + \left( \frac{b^2 \sigma^2 K_n^3}{n} \right)^{1/4} \right\} = L\psi_n$$

for some absolute constant  $L$ .

Note that

$$\sup_{g \in \mathcal{G}(\tau)} |\mathbb{E}_n[g(x_i)^2] - (\mathbb{E}_n[g(x_i)])^2| \leq \sup_{g \in \mathcal{G}(\tau)} \mathbb{E}_n[g(x_i)^2] =: D(\tau).$$

Then  $D(\tau) \leq p_1 + p_2 \leq \sigma^2/n + \sqrt{b^2 \sigma^2 K_n/n}$  where

$$\begin{aligned} p_1 &:= \sup_{g \in \mathcal{G}(\tau)} \mathbb{E}[g(X_1)^2] \leq b^2 \tau^2 = \sigma^2/n, \\ p_2 &:= \sup_{g \in \mathcal{G}(\tau)} |\mathbb{E}_n[g(x_i)^2] - \mathbb{E}[g(X_1)^2]| \leq \sup_{g \in (\mathcal{G}-\mathcal{G})^2} |\mathbb{E}_n[g(x_i)] - \mathbb{E}[g(X_1)]| \\ &\leq \sqrt{b^2 \sigma^2 K_n/n}. \end{aligned}$$

By the Borell-Sudakov-Tsirel'son inequality (see Theorem A.2.1 in [10]), with probability at least  $1 - 2/n$ ,

$$\|\tilde{\mathbb{G}}_n(x_1^n, \xi_1^n)\|_{\mathcal{G}(\tau)} \leq \mathbb{E} \left[ \|\tilde{\mathbb{G}}_n(x_1^n, \xi_1^n)\|_{\mathcal{G}(\tau)} \right] + \sqrt{2D(\tau) \log n}.$$

Further,  $\mathbb{E}[\|\tilde{\mathbb{G}}_n(x_1^n, \xi_1^n)\|_{\mathcal{G}(\tau)}] \leq C(r_1 + r_2)$  where

$$\begin{aligned} r_1 &:= \mathbb{E} \left[ \sup_{g \in \mathcal{G}(\tau)} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i g(x_i) \right| \right], \\ r_2 &:= \sup_{g \in \mathcal{G}(\tau)} |\mathbb{E}_n[g(x_i)]|. \end{aligned}$$

To bound  $r_1$ , let  $\varphi = \sigma/(bn^{1/2}) + (\sigma^2 K_n/(b^2 n))^{1/4}$ . Note that  $\sqrt{D(\tau)}/b \leq \varphi$  and  $\varphi \leq 1 + (K_n/n)^{1/4} \leq 2 < a$ . Hence, by Corollary 2.2.8 in [10],

$$\begin{aligned} r_1 &\leq Cb \int_0^\varphi \sqrt{\sup_Q \log N(\mathcal{G}, L_2(Q), b\varepsilon)} d\varepsilon \leq Cb\varphi \sqrt{v \log(a/\varphi)} \\ &\leq \sqrt{K_n} \left( \frac{\sigma}{\sqrt{n}} + \left( \frac{b^2 \sigma^2 K_n}{n} \right)^{1/4} \right). \end{aligned}$$

To bound  $r_2$ , we have

$$\begin{aligned} r_2 &\leq 2 \sup_{g \in \mathcal{G}} |\mathbb{E}_n[g(x_i)] - \mathbb{E}[g(X_1)]| + \sup_{g \in \mathcal{G}(\tau)} \mathbb{E}[|g(X_1)|] \\ &\leq 2\sqrt{\sigma^2 K_n/n} + b\tau \leq 3\sqrt{\sigma^2 K_n/n}. \end{aligned}$$

Combining these inequalities gives the claim of step 1.

*Step 2 (Bound on  $\|B\|_{\mathcal{G}(\tau)}$ ).* We show that with probability at least  $1 - 2/n$ ,

$$\|B\|_{\mathcal{G}(\tau)} \leq \sqrt{\frac{\sigma^2 K_n}{n}} \leq \psi_n.$$

By the Borell-Sudakov-Tsirel'son inequality, with probability at least  $1 - 2/n$ ,

$$\|B\|_{\mathcal{G}(\tau)} \leq \mathbb{E}[\|B\|_{\mathcal{G}(\tau)}] + b\tau \sqrt{2 \log n}.$$

By Corollary 2.2.8 in [10],

$$\mathbb{E}[\|B\|_{\mathcal{G}(\tau)}] \leq Cb \int_0^\tau \sqrt{\sup_Q \log N(\mathcal{G}, L_2(Q), b\varepsilon) d\varepsilon} \leq Cb\tau \sqrt{v \log(a/\tau)}.$$

Substituting  $\tau = \sigma/(bn^{1/2})$  into these inequalities gives the claim of step 2.

*Step 3 (Coupling Inequality).* This is the main step of the proof. Let  $\delta > 0$  and  $\beta = 2\sqrt{\log n}/\delta$ . Then

$$\varepsilon := \sqrt{e^{1-\beta^2\delta^2} \beta^2 \delta^2} \leq C/n.$$

Take any Borel subset  $B$  of  $\mathbb{R}$  and apply Theorem G.3 to define a function  $f$  corresponding to the set  $B^{L\psi_n}$ ,  $L\psi_n$ -enlargement of the set  $B$ , with chosen  $\beta$  and  $\delta$ . We have for all  $t \in \mathbb{R}$ ,

$$(1 - \varepsilon)1_{B^{L\psi_n}}(t) \leq f(t) \leq \varepsilon + (1 - \varepsilon)1_{B^{L\psi_n+3\delta}}(t).$$

Further, since  $x_1^n \in S_{n,0}$ ,

$$\Delta := \sup_{g_1, g_2 \in \mathcal{G}} |\Delta_{g_1, g_2}| \leq C \sqrt{\frac{b^2 \sigma^2 K_n}{n}},$$

where

$$\begin{aligned} \Delta_{g_1, g_2} &:= (\mathbb{E}_n[g_1(x_i)g_2(x_i)] - \mathbb{E}_n[g_1(x_i)]\mathbb{E}_n[g_2(x_i)]) \\ &\quad - (\mathbb{E}[g_1(X_1)g_2(X_1)] - \mathbb{E}[g_1(X_1)]\mathbb{E}[g_2(X_1)]). \end{aligned}$$

So, applying Theorem G.1 to  $W(x_1^n)(\tau)$  and  $W^0(\tau)$  with chosen  $f$  gives

$$|\mathbb{E}[f(W(x_1^n)(\tau))] - \mathbb{E}[f(W^0(\tau))]| \leq \frac{C}{\delta^2} \sqrt{\frac{b^2 \sigma^2 K_n \log n}{n}} + \frac{C}{\delta} \left( \frac{b^2 \sigma^2 K_n^3}{n} \right)^{1/4}.$$

We will assume that  $b^2 \sigma^2 K_n^3 / (n \delta^4) \leq 1$  (otherwise, the bound claimed in the statement of the theorem is trivial). Then

$$|\mathbb{E}[f(W(x_1^n)(\tau))] - \mathbb{E}[f(W^0(\tau))]| \leq \frac{C}{\delta} \left( \frac{b^2 \sigma^2 K_n^3}{n} \right)^{1/4} \leq C \gamma_n(\delta).$$

Therefore,

$$\begin{aligned} \mathbb{E}[1_B(\tilde{W}_n(x_1^n))] &\leq \mathbb{E}[1_{B^{L\psi_n}}(W(x_1^n)(\tau))] + C/n \\ &\leq \mathbb{E}[f(W(x_1^n)(\tau))]/(1 - \varepsilon) + C/n \\ &\leq \mathbb{E}[f(W^0(\tau))]/(1 - \varepsilon) + C \gamma_n(\delta) \\ &\leq \mathbb{E}[1_{B^{L\psi_n+3\delta}}(W^0(\tau))] + C \gamma_n(\delta) \\ &\leq \mathbb{E}[1_{B^{(L+1)\psi_n+3\delta}}(\tilde{W}^0)] + C \gamma_n(\delta), \end{aligned}$$

where  $C$  is varying from line to line. The claim of the theorem follows by applying Theorem G.2.  $\blacksquare$

## APPENDIX H: PROOFS OF LEMMAS B.1 AND B.2

**PROOF OF LEMMA B.1.** Pick any  $\alpha \in (0, 1)$ . Since  $\mathbb{E}[Y(t)^2] \leq 1$  for all  $t \in T$ , the Borel-Sudakov-Tsirel'son inequality (see Theorem A.2.1 in [10]) gives for all  $r > 0$ ,

$$\mathbb{P}(\|Y\|_T \geq \mathbb{E}[\|Y\|_T] + r) \leq e^{-r^2/2}.$$

Setting  $r = \sqrt{2|\log \alpha|}$  gives

$$\mathbb{P}\left(\|Y\|_T \geq \mathbb{E}[\|Y\|_T] + \sqrt{2|\log \alpha|}\right) \leq \alpha.$$

This implies that  $c(\alpha) \leq \mathbb{E}[\|Y\|_T] + \sqrt{2|\log \alpha|}$ . The result with  $M(\|Y\|_T)$  follows similarly because the Borel-Sudakov-Tsirel'son inequality also applies with  $M(\|Y\|_T)$  replacing  $\mathbb{E}[\|Y\|_T]$ .  $\blacksquare$

**PROOF OF LEMMA B.2.** Consider part (i). Clearly, for any  $g \in \mathcal{G}_1 \cdot \mathcal{G}_2$ ,  $\|g\|_S \leq b_1 b_2$ . Further, for any finitely discrete probability measure  $Q$  on

$(S, \mathcal{S})$  and  $j = 1, 2$ , let  $g_{j1}, \dots, g_{jN_j}$  be a set of functions from the class  $\mathcal{G}_j$  such that for any  $g_j \in \mathcal{G}_j$ , there is some  $k(g_j)$  such that

$$\mathbb{E}_Q[(g_j - g_{jk(g_j)})^2]^{1/2} \leq b_j \tau / 2.$$

By assumption, we can and will assume that  $N_j \leq (2a_j/\tau_j)^{v_j}$ . Then the set  $\{g_{1k}g_{2l} : k = 1, \dots, N_1; l = 1, \dots, N_2\}$  contains

$$N_1 N_2 \leq \left( \frac{2(a_1^{v_1} a_2^{v_2})^{1/(v_1+v_2)}}{\tau} \right)^{v_1+v_2}$$

elements. At the same time,

$$\begin{aligned} \mathbb{E}_Q[(g_1 g_2 - g_{1k(g_1)} g_{2k(g_2)})^2]^{1/2} &\leq \mathbb{E}_Q[g_1^2 (g_2 - g_{2k(g_2)})^2]^{1/2} \\ &\quad + \mathbb{E}_Q[(g_1 - g_{1k(g_1)})^2 g_{2k(g_2)}^2]^{1/2} \\ &\leq b_1 b_2 \tau / 2 + b_1 b_2 \tau / 2 = b_1 b_2 \tau. \end{aligned}$$

The claim of part (i) follows. Parts (ii) and (iii) follow similarly.  $\blacksquare$

## APPENDIX I: ON USE OF NON-WAVELET PROJECTION KERNELS

In Section 3, we provided weak conditions for the construction of honest confidence bands in density estimation. In particular, we demonstrated that, as long as the bias can be controlled, our confidence bands are honest if assumptions of Theorem 3.2 hold. In this section, we verify these assumptions for Fourier and Legendre polynomial projection kernels. We show that these conditions hold under weak conditions on the number of series terms for Fourier projection kernel and under somewhat stronger conditions on the number of series terms for Legendre polynomial projection kernel.

**Fourier projection kernel.** Here we show that assumptions of Theorem 3.2 hold for Fourier projection kernel under weak conditions on the number of series terms. Assume that  $d = 1$ ,  $\mathcal{X} = [-1, 1]$ , and  $K_l(\cdot, \cdot)$  is the projection kernel function based on the Fourier basis as defined in (26) and (27). Assume in addition that the density  $f$  is supported on  $\mathcal{X}$ , and is bounded from below and from above on  $\mathcal{X}$  uniformly over  $\mathcal{F}$ . Finally, assume that  $\sup_{x \in \mathcal{X}} |\mathbb{E}_f[K_l(X_1, x)] - f(x)| \leq C$  uniformly over  $f \in \mathcal{F}$  and  $l \in \mathcal{L}_n$  for some  $C > 0$ . The last assumption holds if  $\sup_{x \in \mathcal{X}} |\mathbb{E}_f[K_l(X_1, x)] - f(x)| \rightarrow 0$  as  $l \rightarrow \infty$  uniformly over  $f \in \mathcal{F}$  so that Fourier projection kernel estimator is asymptotically unbiased, which is necessary for consistency of the estimator. Then we obtain the following bounds.

First, since  $\varphi_1(x) = 1$ ,  $\varphi_{j+1}(x) = \cos(\pi j x)$ ,  $j = 1, 2, \dots$ , we have

$$(72) \quad cm \leq \sum_{j=1}^m \varphi_j(x)^2 \leq Cm,$$

uniformly over all  $x \in [-1, 1]$  and  $m \geq 1$  for some  $c, C > 0$ . The upper bound in (72) is trivial because  $|\cos(\pi j x)| \leq 1$ . To prove the lower bound, we have

$$(73) \quad \begin{aligned} \sum_{j=1}^m \cos^2(\pi j x) &= \frac{1}{2} \sum_{j=1}^m (1 + \cos(2\pi j x)) = \frac{m}{2} - \frac{1}{4} + \frac{1}{2} \left( \frac{1}{2} + \sum_{j=1}^m \cos(2\pi j x) \right) \\ &= \frac{m}{2} - \frac{1}{4} + \frac{\sin((2m+1)\pi x)}{4 \sin(\pi x)}, \end{aligned}$$

and the last term in (73) is bounded from below by some absolute constant yielding the lower bound in (72). Therefore,  $|K_l(y, x)| \leq C2^l$ ,  $|\mathbb{E}_f[K_l(X_1, x)]| \leq C$  and

$$c2^l \leq c \sum_{j=1}^{\lfloor 2^l \rfloor} \varphi_j(x)^2 \leq \mathbb{E}_f[K_l(X_1, x)^2] \leq C \sum_{j=1}^{\lfloor 2^l \rfloor} \varphi_j(x)^2 \leq C2^l,$$

uniformly over  $f \in \mathcal{F}$ ,  $l \in \mathcal{L}_n$ , and  $x \in \mathcal{X}$ . This implies that  $c2^l \leq \sigma_{n,f}(x, l)^2 \leq C2^l$  uniformly over  $f \in \mathcal{F}$ ,  $l \in \mathcal{L}_n$ , and  $x \in \mathcal{X}$  and so  $\sigma_n \leq C$ . Further, uniformly over  $x_1, x_2, y \in \mathcal{X}$ ,

$$\begin{aligned} &\left| \sum_{j=1}^{\lfloor 2^l \rfloor} \varphi_j(y) \varphi_j(x_1) - \sum_{j=1}^{\lfloor 2^l \rfloor} \varphi_j(y) \varphi_j(x_2) \right| \leq \left| \sum_{j=1}^{\lfloor 2^l \rfloor} \varphi_j(y) (\varphi_j(x_1) - \varphi_j(x_2)) \right| \\ &\leq \left( \sum_{j=1}^{\lfloor 2^l \rfloor} \varphi_j(y)^2 \sum_{j=1}^{\lfloor 2^l \rfloor} (\varphi_j(x_1) - \varphi_j(x_2))^2 \right)^{1/2} \leq C2^{l/2} |x_1 - x_2| \left( \sum_{j=1}^{\lfloor 2^l \rfloor} j^2 \right)^{1/2} \\ &= C2^{2l} |x_1 - x_2|. \end{aligned}$$

Therefore, it follows from Example 19.7 in [9] that the function class  $\{K_l(\cdot, x) : x \in \mathcal{X}\}$  is VC( $b_l, a_l, v_l$ ) type class with  $b_l \leq C2^l$ ,  $a_l \leq C2^l$ , and  $v_l \leq C$ , and so Lemma B.2 implies that  $\mathcal{K}_{n,f}$  is VC( $b_n, a_n, v_n$ ) type class with  $b_n \leq C2^{l_{\max, n}/2}$ ,  $a_n \leq C2^{2l_{\max, n}}$ , and  $v_n \leq C$  where  $l_{\max, n} = \sup\{\mathcal{L}_n\}$ . Hence, the assumption that  $b_n^2 \sigma_n^4 K_n^4 / n \leq C_2 n^{-c_2}$  becomes  $2^{l_{\max, n}} (\log^4 n) / n \leq C_2 n^{-c_2}$  (with possibly different  $c_2, C_2$ ) as long as  $l_{\max, n} \leq C \log n$  for some  $C > 0$ .

When  $d > 1$  and  $\mathcal{X} = [-1, 1]^d$ , a similar argument shows that the assumption that  $b_n^2 \sigma_n^4 K_n^4/n \leq C_2 n^{-c_2}$  becomes  $2^{l_{\max, n} d} (\log^4 n)/n \leq C_2 n^{-c_2}$  (with possibly different  $c_2, C_2$ ) as long as  $l_{\max, n} \leq C \log n$  for some  $C > 0$ .

**Legendre polynomial projection kernel.** Here we provide primitive conditions that suffice for assumptions of Theorem 3.2 in the case of Legendre polynomial projection kernel. Assume that  $d = 1$ ,  $\mathcal{X} = [-1, 1]$ , and  $K_l(\cdot, \cdot)$  is the projection kernel function based on the Legendre polynomial basis as defined in (26) and (28). Assume in addition that the density  $f$  is supported on  $\mathcal{X}$  and is bounded from above on  $\mathcal{X}$  uniformly over  $\mathcal{F}$ . Further, assume that  $\sup_{x \in \mathcal{X}} |\mathbb{E}_f[K_l(X_1, x)] - f(x)| \leq C$  uniformly over  $f \in \mathcal{F}$  and  $l \in \mathcal{L}_n$  for some  $C > 0$ . See discussion of this assumption for the case of Fourier projection kernel above.

Note that when  $\varphi_1(\cdot), \varphi_2(\cdot), \dots$  are Legendre polynomials, it is known that  $\sum_{j=1}^K \varphi_j(x)^2 \leq CK^2$  for some  $C > 0$ ; see, for example, [8]. Therefore, under our assumptions,  $|K_l(y, x)| \leq C2^{2l}$ ,  $|\mathbb{E}_f[K_l(X_1, x)]| \leq C$ , and  $\mathbb{E}_f[K_l(X_1, x)^2] \leq C2^{2l}$  uniformly over  $f \in \mathcal{F}$ ,  $l \in \mathcal{L}_n$ , and  $x \in \mathcal{X}$ . Assume also that  $\mathbb{E}_f[K_l(X_1, x)^2] \geq c2^{2l}$  uniformly over  $f \in \mathcal{F}$ ,  $l \in \mathcal{L}_n$ , and  $x \in \mathcal{X}$ . Given the upper bound on  $\mathbb{E}_f[K_l(X_1, x)^2]$  above, the last assumption can be interpreted as that the variance of the kernel estimator is of the same order for all  $x \in \mathcal{X}$ . These bounds imply that  $c2^{2l} \leq \sigma_{n, f}(x, l)^2 \leq C2^{2l}$  uniformly over  $f \in \mathcal{F}$ ,  $l \in \mathcal{L}_n$ , and  $x \in \mathcal{X}$  and so  $\sigma_n \leq C$ . Further, the same argument as that applied in the case of Fourier series shows that  $\{K_l(\cdot, x) : x \in \mathcal{X}\}$  is VC( $b_l, a_l, v_l$ ) type class with  $b_l \leq C2^{2l}$ ,  $a_l \leq C2^{cl}$ , and  $v_l \leq C$ , and so Lemma B.2 implies that  $\mathcal{K}_{n, f}$  is VC( $b_n, a_n, v_n$ ) type class with  $b_n \leq C2^{l_{\max, n}}$ ,  $a_n \leq C2^{cl_{\max, n}}$ , and  $v_n \leq C$  where  $l_{\max, n} = \sup\{\mathcal{L}_n\}$ . Hence, the assumption that  $b_n^2 \sigma_n^4 K_n^4/n \leq C_2 n^{-c_2}$  becomes  $2^{2l_{\max, n}} (\log^4 n)/n \leq C_2 n^{-c_2}$  (with possibly different  $c_2$  and  $C_2$ ) as long as  $l_{\max, n} \leq C \log n$  for some  $C > 0$ . When  $d > 1$  and  $\mathcal{X} = [-1, 1]^d$ , a similar argument shows that the assumption that  $b_n^2 \sigma_n^4 K_n^4/n \leq C_2 n^{-c_2}$  becomes  $2^{2l_{\max, n} d} (\log^4 n)/n \leq C_2 n^{-c_2}$  (with possibly different  $c_2$  and  $C_2$ ) as long as  $l_{\max, n} \leq C \log n$  for some  $C > 0$ .

## APPENDIX J: MONTE CARLO SIMULATION STUDY

In this section, we provide results of a small scale simulation study. The main purpose of this simulation study is to calibrate parameters for our procedure of constructing honest and adaptive confidence bands considered in Section 4. We consider two design settings. In the first design, random variables  $X_1, \dots, X_n$  are drawn from the distribution of  $1.5N(0, 1)$ , that is, the density  $f$  has the form  $f(x) = (1/1.5)\phi(x/1.5)$  for all  $x \in \mathbb{R}$  where  $\phi(x)$  is the pdf of standard normal distribution. In the second design, random

variables  $X_1, \dots, X_n$  are drawn from the mixture of normal distributions; specifically, the pdf of  $X_i$ 's is given by  $f(x) = 0.5\phi(x-1.5) + 0.5\phi(x+1.5)$  for all  $x \in \mathbb{R}$ . The same distributions are considered, for example, in [6]. In all simulations, we set  $n = 1000$  and we consider uniform confidence bands on the interval  $(-3, +3)$ . Also, we use the kernel function  $K(x) = 0.75(1 - x^2)$  for  $x \in (-1, +1)$  and  $K(x) = 0$  otherwise, and we consider the confidence bands with level  $\alpha = 0.10$ . For our implementation of the Lepski estimator, we set  $\gamma_n = 0.02$  and, depending on the experiment, we set  $q$  in the range between 1.1 and 1.5. We choose the set of smoothing parameters  $\mathcal{L}$  for the Lepski estimator so that the set of bandwidth values  $\{2^{-l}, l \in \mathcal{L}\}$  is equal to  $\{0.5, 0.7, 0.8, 0.9, 1.0, 1.2, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0\}$ . We check that the Lepski estimator is not on the boundary of this set in most cases. Both  $\hat{c}_n(\alpha)$  and  $\tilde{c}_n(\gamma_n)$  are based on the Gaussian Multiplier bootstrap procedure described in Algorithms 1 and 2 with the number of bootstrap samples  $B = 500$ . Finally, depending on the experiment, we set  $u'_n$  in the range between 0.5 and 1.0. For each experiment, the results are based on 500 simulations.

The results of the simulation study are presented in Tables 1 and 2. Table 1 contains information on coverage probability, and Table 2 gives the average value over 500 simulations of the average length of the confidence bands over the interval  $(-3, +3)$ . From these tables, we conclude that the parameter values  $q = 1.1$  and  $u'_n = 0.5$  strike a good balance between control of the coverage probability and the average length of the confidence bands. Also, it seems that setting  $q = 1.1$  and  $u = 1.0$  should be a rather safe choice from the point of view of coverage probability control in many settings. However, as a cautionary remark, we note that empirical researchers should always test out different values of  $q$  and  $u'_n$  in Monte Carlo examples that mimic the data at hand.

Finally, to get an idea how our confidence bands with the proposed values of the parameters ( $q = 1.1$  and  $u'_n = 0.5$ ) look in practice, we draw confidence bands with this choice of values of the parameters for 4 simulated data sets as well as true density functions in Figure 1. Two of the data sets are based on the normal density and the other two are based on the mixture of normal densities. In all data sets, we have  $n = 1000$ . Both for the case of the normal density and for the case of the mixture of normal densities, these data sets are the first two used in the simulations behind Tables 1 and 2. Overall, we conclude that the confidence bands follow the shape of the true density functions.

TABLE 1  
*Results of Monte Carlo experiments for coverage probability.*

Density	$u'_n$	$q$				
		1.1	1.2	1.3	1.4	1.5
normal	0.5	0.89	0.84	0.88	0.83	0.78
	0.6	0.97	0.94	0.92	0.88	0.79
	0.7	0.98	0.96	0.97	0.93	0.81
	0.8	0.99	1.00	0.98	0.94	0.85
	0.9	1.00	1.00	0.98	0.94	0.84
	1.0	1.00	1.00	0.99	0.95	0.82
mixture	0.5	0.93	0.83	0.68	0.51	0.38
	0.6	0.94	0.87	0.70	0.57	0.52
	0.7	0.95	0.92	0.84	0.79	0.64
	0.8	0.96	0.94	0.89	0.86	0.81
	0.9	0.98	0.97	0.94	0.91	0.89
	1.0	0.99	0.99	0.98	0.95	0.94

TABLE 2  
*Results of Monte Carlo experiments for average length of confidence bands.*

Density	$u'_n$	$q$				
		1.1	1.2	1.3	1.4	1.5
normal	0.5	0.064	0.060	0.059	0.056	0.055
	0.6	0.069	0.065	0.063	0.061	0.059
	0.7	0.074	0.070	0.068	0.066	0.063
	0.8	0.079	0.075	0.072	0.070	0.068
	0.9	0.084	0.080	0.076	0.075	0.073
	1.0	0.089	0.084	0.082	0.079	0.077
mixture	0.5	0.076	0.072	0.069	0.066	0.064
	0.6	0.082	0.078	0.074	0.071	0.068
	0.7	0.088	0.084	0.080	0.077	0.074
	0.8	0.093	0.089	0.085	0.082	0.078
	0.9	0.100	0.095	0.090	0.087	0.083
	1.0	0.105	0.100	0.096	0.092	0.089



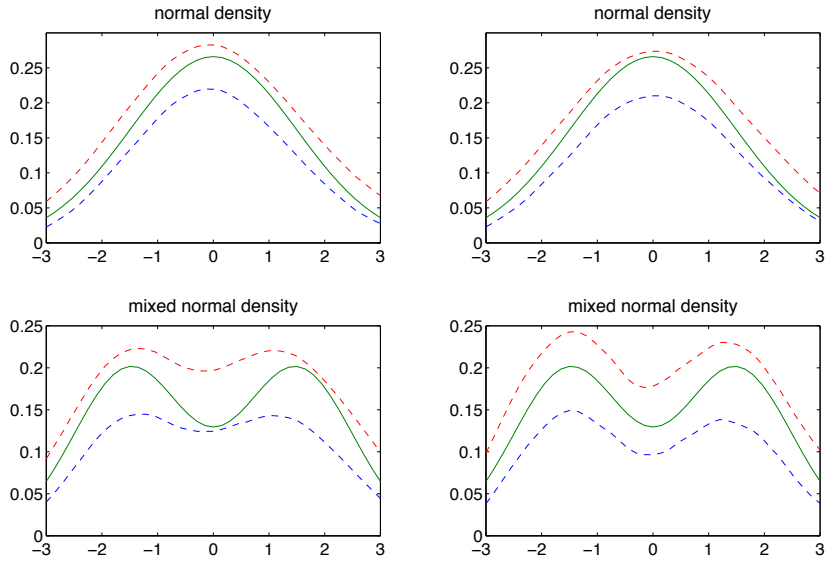


FIG 1. Confidence bands with the proposed values of the parameters in 4 simulated data sets. In two data sets, denoted “normal density”,  $X_i$ 's have the density  $f(x) = (1/1.5)\phi(x/1.5)$ , and in the other two, denoted “mixed normal density”,  $X_i$ 's have the density  $f(x) = 0.5\phi(x-1.5) + 0.5\phi(x+1.5)$ . Nominal coverage probability is 0.9 ( $\alpha = 0.1$ ). Continuous line is the true density function; dashed lines are confidence bands.

## REFERENCES

- [1] Chernozhukov, V. Chetverikov, D. and Kato, K. (2012). Gaussian approximation of suprema of empirical processes. Preprint. arXiv:1212.6885v2.
- [2] Chernozhukov, V. Chetverikov, D. and Kato, K. (2012). Comparison and anti-concentration bounds for maxima of Gaussian random vectors. Preprint. arXiv:1301.4807v3. To appear in *Probab. Theory Related Fields*.
- [3] Giné, E. and Guillou, A. (2002). Rates of strong uniform consistency for multivariate kernel density estimators. *Ann. Inst. Henri Poincaré Probab. Stat.* **38** 907-921.
- [4] Giné, E. and Nickl, R. (2009). Uniform limit theorems for wavelet density estimators. *Ann. Probab.* **37** 1605-1646.
- [5] Giné, E. and Nickl, R. (2010). Adaptive estimation of a distribution function and its density in sup-norm loss by wavelet and spline projections. *Bernoulli* **16** 1137-1163.
- [6] Hall, P., Sheather, S., Jones, M., and Marron, J. (1991). On optimal data-based bandwidth selection in kernel density estimation. *Biometrika* **78** 263-269.
- [7] Härdle, W., Kerkycharian, G., Picard, D., and Tsybakov, A. (1998). *Wavelets, Approximation, and Statistical Applications*. Springer-Verlag.
- [8] Newey, W. (1997). Convergence rates and asymptotic normality for series estimators. *J. Econometrics* **79** 147-168.
- [9] van der Vaart, A.W. (1998). *Asymptotic statistics*. Cambridge Series in Statistical and Probabilistic Mathematics.
- [10] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak convergence and empirical processes: With applications to statistics*. Springer-Verlag.

DEPARTMENT OF ECONOMICS AND  
OPERATIONS RESEARCH CENTER, MIT  
50 MEMORIAL DRIVE  
CAMBRIDGE, MA 02142, USA.  
E-MAIL: vchern@mit.edu

DEPARTMENT OF ECONOMICS, UCLA  
BUNCHE HALL, 8283  
315 PORTOLA PLAZA  
LOS ANGELES, CA 90095, USA.  
E-MAIL: chetverikov@econ.ucla.edu

GRADUATE SCHOOL OF ECONOMICS  
UNIVERSITY OF TOKYO  
7-3-1 HONGO, BUNKYO-KU  
TOKYO 113-0033, JAPAN.  
E-MAIL: kkato@e.u-tokyo.ac.jp