cemmap
centre for microdata methods and practice

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 CWP69/13

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL
An ESRC Research Centre

Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE **Tel:** 44 (0)20 7291 4800 **Email:** info@cemmap.ac.uk **www.cemmap.ac.uk**

ANTI-CONCENTRATION AND HONEST, ADAPTIVE CONFIDENCE BANDS

VICTOR CHERNOZHUKOV, DENIS CHETVERIKOV, AND KENGO KATO

Abstract. Modern construction of uniform confidence bands for nonparametric densities (and other functions) often relies on the 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. Furthermore, our approach is asymptotically honest at a polynomial rate – namely, the error in coverage level converges to zero at a fast, polynomial speed (with respect to the sample size). In sharp contrast, the approach based on extreme value theory is asymptotically honest only at a logarithmic rate – the error converges to zero at a slow, logarithmic speed. 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.

Date: November 27, 2013. First version: October 2011.

Key words and phrases. Anti-concentration of separable Gaussian processes, honest confidence bands, Lepski's method, multiplier method.

We are grateful to Emre Barut, Enno Mammen, and Richard Nickl for useful discussions. 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.

1. Introduction

Let $X_1,\dots,X_n$ be i.i.d. random variables with common unknown density f on $\mathbb{R}^d$. We are interested in constructing confidence 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 [27], a band $\mathcal{C}_n$ is said to be *asymptotically honest with level* $\alpha\in(0,1)$ *for the class* $\mathcal{F}$ if

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

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 [23, 39, 42] 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., $\sigma_{n,f}(x,l):=\sqrt{n\operatorname{Var}_f(\hat{f}_n(x,l))}$. Then we consider a confidence band of the form

$$\mathcal{C}_n(x)=\left[\hat{f}_n(x,\hat{l}_n)-\frac{c(\alpha)\sigma_{n,f}(x,\hat{l}_n)}{\sqrt{n}},\hat{f}_n(x,\hat{l}_n)+\frac{c(\alpha)\sigma_{n,f}(x,\hat{l}_n)}{\sqrt{n}}\right], \quad (1.1)$$

where $c(\alpha)$ is a (possibly data-dependent) critical value determined to make the confidence band to have level α. 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

$$Z_{n,f}(v):=Z_{n,f}(x,l):=\frac{\sqrt{n}(\hat{f}_n(x,l)-\mathrm{E}_f[\hat{f}_n(x,l)])}{\sigma_{n,f}(x,l)}, \quad (1.2)$$

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 $|f(x)-\mathrm{E}_f[\hat{f}_n(x,l)]_{l=\hat{l}_n}|$ is sufficiently small compared to $\sigma_{n,f}(x,\hat{l}_n)$, then

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

so that the band (1.1) will be of level $\alpha \in (0,1)$ by taking

$$c(\alpha) = (1-\alpha)\text{-quantile of } \|Z_{n,f}\|_{\mathcal{V}_n} := \sup_{v \in \mathcal{V}_n} |Z_{n,f}(v)|. \tag{1.3}$$

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. If, for some normalizing constants A_n and B_n, $A_n(\|G_{n,f}\|_{\mathcal{V}_n} - B_n)$ has a continuous limit distribution, the validity of the confidence band would follow via the continuity of the limit distribution. 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 37, 1, 16, 19, 4, 5, 15]. 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 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 ϵ goes to 0 as $\epsilon \to 0$. However, what we need to know is the behavior of those probabilities when ϵ is n-dependent and $\epsilon = \epsilon_n \to 0$. 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 radius L and smoothness level t. Following [6], we say the confidence band $\mathcal{C}_n$ is *adaptive* if for every $t, \varepsilon > 0$ there exists $C > 0$ such that for all $n \geq 1$,

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

where λ 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 [26, 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, [18] 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 result on the anti-concentration of separable Gaussian processes is also of independent interest in many other problems. For example, applications of our anti-concentration bounds can be found in [9] and [10], 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 [37] 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 considerably slow despite that the Gaussian approximation is relatively fast [21]. To improve the finite sample coverage, bootstrap is often used in construction of confidence bands [see 11, 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, [28] considered confidence bands without using Gaussian approximation. In the current density estimation problem, their idea reads as bounding the deviation probability of $\|\hat{f}_n - \mathrm{E}[\hat{f}_n(\cdot)]\|_\infty$ by using Talagrand's [38] 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 conditions. 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.

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 $z : T \to \mathbb{R}$ 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 ξ and η, 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 [35]. Let $(\Omega, \mathcal{A}, \mathrm{P})$ 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

$$p_{x,\epsilon}(Y) := \mathrm{P}\left(\left| \sup_{t\in T} Y_t - x \right| \leq \epsilon \right). \tag{2.1}$$

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

$$p_\epsilon(Y) := \sup_{x\in\mathbb{R}} p_{x,\epsilon}(Y). \tag{2.2}$$

Likewise, define $p_{x,\epsilon}(|Y|)$ by (2.1) with $\sup_{t\in T} Y_t$ replaced by $\sup_{t\in T}|Y_t|$ and define $p_\epsilon(|Y|)$ by (2.2) 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 $\mathrm{E}[X_t] = 0$ and $\mathrm{E}[X_t^2] = 1$ 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 $\epsilon \to 0$. A non-trivial case is that when T is not a singleton and both T and X are indexed by $n = 1, 2, \ldots$, 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 $L^2(\Omega, \mathcal{A}, \mathrm{P})$) 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 $n \to \infty$ and therefore it is not trivial at all whether, for any sequence $\epsilon_n \to 0$, $p_{\epsilon_n}(X^n) \to 0$ as $n \to \infty$, which is in fact generally not true as Example 1 in [8] shows. 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 $\mathrm{E}[X_t] = 0$ and $\mathrm{E}[X_t^2] = 1$ for all $t \in T$. Assume that $\sup_{t\in T} X_t < \infty$ a.s. Then $a(X) := \mathrm{E}[\sup_{t\in T} X_t] \in [0, \infty)$ and*

$$p_\epsilon(X) \leq 4\epsilon\,(a(X) + 1)\,, \tag{2.3}$$

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 $\mathrm{E}[X_t] = 0$ and $\mathrm{E}[X_t^2] = 1$ for all $t \in T$. Assume that $\sup_{t\in T} X_t < \infty$ a.s. Then $a(|X|) := \mathrm{E}[\sup_{t\in T}|X_t|] \in [\sqrt{2/\pi}, \infty)$ and*

$$p_\epsilon(|X|) \leq 4\epsilon a(|X| + 1), \tag{2.4}$$

for all $\epsilon \geq 0$.

We refer to (2.3) and (2.4) 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 variables $X_1, \dots, X_n$ 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 band 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 (1.2). 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 make 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 $W^0_{n,f}$ such that (i) $W^0_{n,f} \stackrel{d}{=} \|G_{n,f}\|_{\mathcal{V}_n}$ 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 $\mathrm{E}[G_{n,f}(v)] = 0, \mathrm{E}[G_{n,f}(v)^2] = 1$ for all $v \in \mathcal{V}_n$, and $\mathrm{E}[\|G_{n,f}\|_{\mathcal{V}_n}] \le C_1\sqrt{\log n}$; and moreover (ii)*

$$\sup_{f \in \mathcal{F}} \mathrm{P}_f(|W_{n,f} - W^0_{n,f}| > \epsilon_{1n}) \le \delta_{1n}, \tag{3.1}$$

where ϵ_{1n} and δ_{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,

$$A_n(\|G_{n,f}\|_{\mathcal{V}_n} - B_n) \stackrel{d}{\to} Z, \text{ as } n \to \infty, \tag{3.2}$$

where Z is a Gumbel random variable; see, for example, [19]. 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 H 2 (Anti-concentration or Generalized SBR condition)**.** *For any sequence ϵ_n of positive numbers, we have*

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

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

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 prove polynomial (in n) rate of approximation. We refer to H2 as a generalized SBR condition because H2-(a) holds if (3.2) holds with A_n of order $\sqrt{\log n}$. An advantage of Condition H2 in comparison with the classical condition (3.2) 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 H 3 (Estimation error of $\hat{c}_n(\alpha)$)**.** *For some sequences τ_n, ϵ_{2n}, and δ_{2n} of positive numbers bounded from above by $C_1 n^{-c_1}$, we have*

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

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

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 τ_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 ϵ_{3n} and δ_{3n} of positive numbers bounded from above by $C_1 n^{-c_1}$,*

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

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, [19]. 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 $\hat{l}_n := \hat{l}_n(X_1, \ldots, X_n)$, 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

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

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 $\|G_{n,f}\|_{\mathcal{V}_n}$.

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

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

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 δ_{5n} and u_n of positive numbers where δ_{5n} is bounded from above by $C_1 n^{-c_1}$,*

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

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

$$\mathcal{C}_n(x) := \left[\hat{f}_n(x,\hat{l}_n) - s_n(x,\hat{l}_n), \hat{f}_n(x,\hat{l}_n) + s_n(x,\hat{l}_n)\right], \tag{3.3}$$

where

$$s_n(x,\hat{l}_n) := (\hat{c}_n(\alpha) + c_n')\hat{\sigma}_n(x,\hat{l}_n)/\sqrt{n}. \tag{3.4}$$

Define

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

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*

$$\inf_{f\in\mathcal{F}} \mathrm{P}_f(f \in \mathcal{C}_n) \geq (1-\alpha) - \delta_n - \tau_n - p_{\bar{\epsilon}_{n,f}}(|G_{n,f}|). \tag{3.5}$$

If, in addition, Condition H2-(a) is satisfied and $\epsilon_{3n}u_n\sqrt{\log n} \leq C_1 n^{-c_1}$, then

$$\liminf_{n\to\infty} \inf_{f\in\mathcal{F}} \mathrm{P}_f(f \in \mathcal{C}_n) \geq 1-\alpha, \tag{3.6}$$

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

$$\inf_{f\in\mathcal{F}} \mathrm{P}_f(f \in \mathcal{C}_n) \geq 1-\alpha - Cn^{-c}, \tag{3.7}$$

where c and C are constants depending only on c_1 and C_1.

Comment 3.1 (Honest confidence bands)**.** Theorem 3.1 shows that the confidence band defined in (3.3) and (3.4) is asymptotically honest with level α for the class $\mathcal{F}$. Moreover, under condition H2-(b), since the constants c and C in the statement (3.7) depend only on c_1, C_1, 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.* ∎

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 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 wavelet 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 [19], and for Battle-Lemarié wavelets of degree higher than 4 in [15]. 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 (3.7). ■

Comment 3.3 (On the condition $\epsilon_{3n} u_n \sqrt{\log n} \le C_1 n^{-c_1}$)**.** The second part of Theorem 3.1 requires the condition that $\epsilon_{3n} u_n \sqrt{\log n} \le C_1 n^{-c_1}$. This is a very mild assumption. Indeed, under Condition H4, $\epsilon_{3n} \le C_1 n^{-c_1}$, so that the assumption that $\epsilon_{3n} u_n \sqrt{\log n} \le C_1 n^{-c_1}$ is met as long as u_n is bounded from above by a slowly growing sequence, for example, $u_n \le C_1 \log n$, which is typically the case; see, for example, our construction in Section 4. ■

The confidence band defined in (3.3) and (3.4) 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 $u_n \log n \le C_1 n^{-c_1}$, then conclusions (3.6) and (3.7) of Theorem 3.1 continue to hold with $s_n(x, \hat{l}_n)$ in (3.4) replaced by $\hat{c}_n(\alpha)\hat{\sigma}_n(x, \hat{l}_n)/\sqrt{n}$. Moreover, if $\mathcal{L}_n$ is a singleton, it is possible to show that the confidence band is asymptotically exact. 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 undersmoothing**)**.** *Consider the confidence band* $\widetilde{\mathcal{C}}_n = \{\widetilde{\mathcal{C}}_n(x) : x \in \mathcal{X}\}$ *defined by*

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

where

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

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

$$\sup_{f\in\mathcal{F}} \mathrm{P}_f\left(\Delta_{n,f}(\hat{l}_n) > u_n\sqrt{\log n}\right) \leq \delta_{6n}, \tag{3.8}$$

where δ_{6n} is bounded from above by $C_1 n^{-c_1}$ and u_n satisfies $u_n \log n \leq C_1 n^{-c_1}$. Then under Condition H2-(a),

$$\liminf_{n\to\infty} \inf_{f\in\mathcal{F}} \mathrm{P}_f(f\in\widetilde{\mathcal{C}}_n) \geq 1-\alpha, \tag{3.9}$$

and under Condition H2-(b),

$$\inf_{f\in\mathcal{F}} \mathrm{P}_f(f\in\widetilde{\mathcal{C}}_n) \geq 1-\alpha - Cn^{-c}. \tag{3.10}$$

Moreover, if $\mathcal{L}_n$ is a singleton, then under condition H2-(a),

$$\lim_{n\to\infty}\sup_{f\in\mathcal{F}} \left| \mathrm{P}_f(f\in\widetilde{\mathcal{C}}_n) - (1-\alpha)\right| \to 0, \tag{3.11}$$

and under Condition H2-(b),

$$\sup_{f\in\mathcal{F}} \left| \mathrm{P}_f(f\in\widetilde{\mathcal{C}}_n) - (1-\alpha)\right| \leq Cn^{-c}. \tag{3.12}$$

Here c and C are constants depending only on c_1 and C_1.

Comment 3.4 (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, [43] 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 [22]. ■

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 the convolution and 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 $K_l : \mathbb{R}^d\times\mathbb{R}^d \to \mathbb{R}$ and l is a smoothing parameter. We consider kernel density estimators of the form

$$\hat{f}_n(x,l) := \mathbb{E}_n[K_l(X_i,x)] = \frac{1}{n}\sum_{i=1}^n K_l(X_i,x), \tag{3.13}$$

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

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

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

$$\hat{\sigma}^2_n(x,l) := \frac{1}{n}\sum_{i=1}^n K_l(X_i,x)^2 - \hat{f}_n(x,l)^2. \tag{3.14}$$

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 $K : \mathbb{R} \to \mathbb{R}$. 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

$$K_l(y,x) := 2^{ld} \prod_{1\le m\le d} K\left(2^l(y_m - x_m)\right). \tag{3.15}$$

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

(ii) *Wavelet projection kernel.* Consider a father wavelet ϕ, i.e., a function ϕ 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 $j \le j'$, and (c) $\cup_{j\ge 0} V_j$ 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

$$K_l(y,x) := 2^{ld} \sum_{k_1,\dots,k_d\in\mathbb{Z}} \prod_{1\le m\le d} \phi(2^l y_m - k_m) \prod_{1\le m\le d} \phi(2^l x_m - k_m). \tag{3.16}$$

Here l is the resolution level. We refer to [12] and [23] 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

$$K_l(y,x) := \sum_{j=1}^{[2^{ld}]} \varphi_j(y)\varphi_j(x), \tag{3.17}$$

where $[a]$ is the largest integer that is smaller than or equal to a. Here $[2^{ld}]$ 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

$$\{1, \cos(\pi x), \cos(2\pi x), \dots\} \tag{3.18}$$

and Legendre polynomial basis

$$\{1, (3/2)^{1/2}x, (5/8)^{1/2}(3x^2 - 1), \dots\}. \tag{3.19}$$

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 $X_1^n := \{X_1, \dots, X_n\}$. 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:*

$$\hat{\mathbb{G}}_n(x,l) := \hat{\mathbb{G}}_n(X_1^n, \xi_1^n)(x,l) := \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i \frac{K_l(X_i, x) - \hat{f}_n(x,l)}{\hat{\sigma}_n(x,l)}. \qquad (3.20)$$

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

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

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

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

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 ϵ-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 [41] 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

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

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

Then we will make 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.*[1]

Here we note this is a mild assumption, which we verify for common constructions in Appendix 4 (as a part of proving results for the next section; see Comment 3.5 below); see also Appendix J.

[1] We refer to Chapter 2.3 of [41] for the definition of pointwise measurable classes of functions.

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 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 $b_n^2\sigma_n^4K_n^4/n \leq C_2 n^{-c_2}$. Then Conditions H1-H4 hold with some constants $c_1, C_1 > 0$ that depend only on c_2, C_2.*

Comment 3.5 (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 E.2 in Appendix E that, under mild regularity conditions, for convolution and wavelet projection kernel functions, $\sigma_n \leq C$ and Condition VC holds with $b_n \leq C2^{l_{\max,n}d/2}$, $a_n \leq C$, 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 $b_n^2\sigma_n^4K_n^4/n \leq C_2n^{-c_2}$ 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) as long as $l_{\max,n} \leq C\log n$ for some $C > 0$. 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 J in the Supplemental Material. ■

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 radius L and smoothness level t is defined by

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

where $\lfloor t \rfloor$ denotes the largest integer smaller than t, and for a multi-index $\alpha = (\alpha_1, \dots, \alpha_d)$ with $|\alpha| = \alpha_1 + \cdots + \alpha_d$, $D^\alpha f(x) := \partial^{|\alpha|} f(x)/\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}$ [see, for example 39]. We assume that for some $0 < \underline{t} \leq \bar{t} < \infty$ and $L \geq 1$,

$$\mathcal{F} \subset \cup_{t \in [\underline{t}, \bar{t}]} \Sigma(t, L), \tag{4.1}$$

and consider the confidence band $\mathcal{C}_n = \{\mathcal{C}_n(x) : x \in \mathcal{X}\}$ of the form (3.3) and (3.4), where $\mathcal{X}$ is a (suitable) compact set in $\mathbb{R}^d$.

We begin with 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 (3.13), (3.15), and (3.16). For convolution kernels, the function $K : \mathbb{R} \to \mathbb{R}$ 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 $\phi : \mathbb{R} \to \mathbb{R}$ is either a compactly supported father wavelet of regularity $r-1$ (that is, ϕ 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 [14] 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 [12], 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 [29]. Therefore, following [19], 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*

$$c_3 2^{-lt} \leq \sup_{x \in \mathcal{X}} |\mathrm{E}_f[\hat{f}_n(x, l)] - f(x)| \leq C_3 2^{-lt}, \tag{4.2}$$

for all $l \geq l_0$.

This condition is inspired by the path-breaking work of [19] (see also [33]). 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 (4.2) holds for all $f \in \Sigma(t, L)$ (for sufficiently large C_3) under Condition L1; see, for example, Theorem 9.3 in [23]. In addition, [19] showed that under Condition L1, the restriction due to the lower bound in (4.2) is weak in the sense that the set of elements of $\Sigma(t, L)$ for which the lower bound in (4.2) 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 [19] 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 $t' > t$, [24] showed that it is necessary and sufficient to exclude functions $\Sigma(t, L) \backslash \Sigma(t, L, \rho_n)$ 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 the *unique* $t \in [\underline{t}, \bar{t}]$ satisfying (4.2); 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 (4.2) can hold for all $l \geq l_0$ for at most one value of t. This defines the map

$$t : \mathcal{F} \to [\underline{t}, \bar{t}],\ f \mapsto t(f). \tag{4.3}$$

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*

$$\left(\frac{c_4 \log n}{n}\right)^{1/(2t(f)+d)} \leq 2^{-l} \leq \left(\frac{C_4 \log n}{n}\right)^{1/(2t(f)+d)}, \tag{4.4}$$

for the map $t : f \mapsto t(f)$ defined in (4.3). In addition, the candidate set is either $\mathcal{L}_n = [l_{\min,n}, l_{\max,n}]$ (for convolution kernels) or $\mathcal{L}_n = [l_{\min,n}, l_{\max,n}] \cap \mathbb{N}$ (for convolution or wavelet projection kernels).

This condition thus ensures via (4.4) 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 $f \in \mathcal{F}$,*

$$f(x) \geq \underline{f} \text{ for all } x \in \mathcal{X}^\delta \text{ and } f(x) \leq \bar{f} \text{ for all } x \in \mathbb{R}^d, \tag{4.5}$$

where $\mathcal{X}^\delta$ is the δ-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. There exist several techniques in the literature to achieve these goals; see, for example, [31] for a thorough introduction. One of the most important such techniques is the Lepski method (see [26] 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 γ_n be a sequence of positive numbers converging to zero. Let $c_{n,f}(\gamma_n)$ be the $(1-\gamma_n)$-quantile of the random variable $\|G_{n,f}\|_{\mathcal{V}_n}$ appearing in Condition H1, and let $\hat{c}_n(\gamma_n)$ be an estimator of $c_{n,f}(\gamma_n)$ defined by Algorithm 1 with α replaced by γ_n, that is $\hat{c}_n(\gamma_n)$ is the conditional $(1-\gamma_n)$-quantile of the supremum of the Gaussian multiplier process $\|\hat{\mathbb{G}}_n\|_{\mathcal{V}_n}$ defined in (3.20) given the data X_1^n. For all $l \in \mathcal{L}_n$, let*

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

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

$$\hat{l}_n := \inf \left\{ l \in \mathcal{L}_n : \sup_{l' \in \mathcal{L}_{n,l}} \sup_{x \in \mathcal{X}} \frac{\sqrt{n}|\hat{f}_n(x,l) - \hat{f}_n(x,l')|}{\hat{\sigma}_n(x,l) + \hat{\sigma}_n(x,l')} \leq q\hat{c}_n(\gamma_n) \right\}. \quad (4.6)$$

Comment 4.1 (On our implementation of Lepski's method)**.** We refer to (4.6) as a (Gaussian) multiplier bootstrap implementation of the Lepski method because $\hat{c}_n(\gamma_n)$ is obtained as the conditional $(1-\gamma_n)$-quantile of $\|\hat{\mathbb{G}}\|_{\mathcal{V}_n}$ given X_1^n. Previous literature on the Lepski method used Talagrand's inequality combined with some bounds on expectations of suprema of certain empirical processes (obtained via entropy methods and Rademacher averages) to choose the threshold level for the estimator (the right hand side of the inequality in (4.6)); see [18] and [20]. 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 $q\hat{c}_n(\gamma_n)$ 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 (3.14) 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 $\mathcal{F}$. In principle, the value $C(\mathcal{F})$ can be traced out from the proof of the theorem below. 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. The problem of selecting u'_n in practice is like that of selecting the level of undersmoothing, and solving this problem is beyond the scope of this paper. Set

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

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, suppose 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 (3.7) in Theorem 3.1 hold and*

$$\sup_{f \in \mathcal{F}} \mathrm{P}_f \left(\sup_{x \in \mathcal{X}} \lambda(\mathcal{C}_n(x)) > C(1 + u_n') r_n(t(f)) \right) \leq C n^{-c}, \tag{4.7}$$

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, and on the function* K *(when convolution kernels are used) or the father wavelet* ϕ *(when wavelet projection kernels are used). Moreover,*

$$\sup_{f \in \mathcal{F} \cap \Sigma(t,L)} \mathrm{P}_f \left(\sup_{x \in \mathcal{X}} \lambda(\mathcal{C}_n(x)) > C(1 + u_n') r_n(t) \right) \leq C n^{-c}, \tag{4.8}$$

with the same constants c, C *as those in (4.7).*

Comment 4.2 (Honest and adaptive confidence bands)**.** Equation (3.7) 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 [19]. Therefore, (4.8) 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 4.3 (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 4.4 (Extensions)**.** Finally, we note that the proof of (3.7) and (4.7) in Theorem 4.1 did not use (4.1) directly. The proof only relies on Conditions L1-L4 whereas (4.1) served to motivate these conditions. Therefore, results (3.7) and (4.7) 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 new 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 a direct corollary of Theorem 2.1 in [7]. The second inequality 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 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 $\sup_{g \in \mathcal{G}} \mathrm{E}[g(X_1)^2] \leq \sigma^2 \leq b^2$. Define the empirical process

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

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

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

for all $g_1, g_2 \in \mathcal{G}$. It is well known that such a process exists under the VC type assumption [see 41, 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, σ, and K_n.

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)*

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

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 [25], 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, [34] 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

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

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

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

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, σ, and K_n.

Theorem A.2 (Slepian-Stein type coupling for suprema of 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* $\mathrm{P}(X_1^n \in S_{n,0}) \geq 1 - 3/n$ *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)*

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

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 σ^2 be a constant such that $\sup_{g \in \mathcal{G}} \mathrm{Var}(g) \leq \sigma^2 \leq b^2$. If $b^2 v \log(ab/\sigma) \leq n\sigma^2$, then for all $t \leq n\sigma^2/b^2$,*

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

where $A > 0$ is an absolute constant.

Proof. This version of Talagrand's inequality follows from Theorem 3 in [30] combined with a bound on expected values of suprema of empirical processes derived in [14]. ■

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 $\mathrm{E}[Y(t)^2] = 1$ for all $t \in T$. Let $c(\alpha)$ denote the $(1-\alpha)$-quantile of $\|Y\|_T$. Assume that $\mathrm{E}[\|Y\|_T] < \infty$. Then $c(\alpha) \leq \mathrm{E}[\|Y\|_T] + \sqrt{2|\log \alpha|}$ 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 $(S, \mathcal{S})$. Then with $a = (a_1^{v_1} a_2^{v_2})^{1/(v_1+v_2)}$, (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 [13]). Since $\sup_{t \in T} X_t \geq X_{t_0}$ for any fixed $t_0 \in T$, $a(X) \geq \mathrm{E}[X_{t_0}] = 0$. We now prove (2.3).

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 $Z_n := \max_{t \in T_n} X_t \to \sup_{t \in T} X_t =: Z$ a.s. as $n \to \infty$. Fix any $x \in \mathbb{R}$. Since $|Z_n - x| \to |Z - x|$ 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 $\epsilon \in \mathbb{R} \backslash \mathcal{N}_x$,

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

But by Theorem 3 in [8],

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

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

$$\mathrm{P}(|Z - x| \leq \epsilon) \leq 4\epsilon(a(X) + 1), \tag{C.1}$$

for all $\epsilon \in \mathbb{R} \backslash \mathcal{N}_x$. By right continuity of $\mathrm{P}(|Z - x| \leq \cdot)$, it follows that (C.1) holds for all $\epsilon \geq 0$. Since $x \in \mathbb{R}$ is arbitrary, we obtain (2.3). ∎

Proof of Corollary 2.1. The proof is analogous to that of Theorem 2.1 and therefore is omitted. ∎

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

$$\begin{aligned}
&\mathrm{P}_f(f(x) \in \mathcal{C}_n(x), \forall x \in \mathcal{X}) \\
&\geq \mathrm{P}_f(|Z_{n,f}(x, \hat{l}_n)| + \Delta_{n,f}(\hat{l}_n) \leq (\hat{c}_n(\alpha) + c_n')\hat{\sigma}_n(x, \hat{l}_n)/\sigma_{n,f}(x, \hat{l}_n), \forall x \in \mathcal{X}) \\
&\geq \mathrm{P}_f(\sup_{x \in \mathcal{X}} |Z_{n,f}(x, \hat{l}_n)| + \Delta_{n,f}(\hat{l}_n) \leq (\hat{c}_n(\alpha) + c_n')(1 - \epsilon_{3n})) - \delta_{3n} &\text{(D.1)} \\
&\geq \mathrm{P}_f(\sup_{x \in \mathcal{X}} |Z_{n,f}(x, \hat{l}_n)| \leq \hat{c}_n(\alpha)(1 - \epsilon_{3n}) - c_n'\epsilon_{3n}) - \delta_{3n} - \delta_{4n} &\text{(D.2)} \\
&\geq \mathrm{P}_f(\|Z_{n,f}\|_{\mathcal{V}_n} \leq \hat{c}_n(\alpha)(1 - \epsilon_{3n}) - c_n'\epsilon_{3n}) - \delta_{3n} - \delta_{4n} &\text{(D.3)} \\
&\geq \mathrm{P}_f(\|Z_{n,f}\|_{\mathcal{V}_n} \leq \hat{c}_n(\alpha)(1 - \epsilon_{3n}) - u_n\epsilon_{3n}\sqrt{\log n}) - \delta_{3n} - \delta_{4n} - \delta_{5n}, &\text{(D.4)}
\end{aligned}$$

where (D.1) follows from Condition H4, (D.2) from Condition H5, (D.3) from the inequality $\sup_{x \in \mathcal{X}} |Z_{n,f}(x, \hat{l}_n)| \leq \|Z_{n,f}\|_{\mathcal{V}_n}$, and (D.4) from Condition H6. Further, the probability in (D.4) equals (recall that $W_{n,f} = \|Z_{n,f}\|_{\mathcal{V}_n}$)

$$\begin{aligned}
&\mathrm{P}_f(W_{n,f} \leq \hat{c}_n(\alpha)(1 - \epsilon_{3n}) - u_n\epsilon_{3n}\sqrt{\log n}) \\
&\quad \geq \mathrm{P}_f(W_{n,f} \leq c_{n,f}(\alpha + \tau_n)(1 - \epsilon_{3n}) - \epsilon_{2n} - u_n\epsilon_{3n}\sqrt{\log n}) - \delta_{2n}, &\text{(D.5)}
\end{aligned}$$

where (D.5) follows from Condition H3. Now, the probability in (D.5) is bounded from below by Condition H1 by

$$\begin{aligned}
&\mathrm{P}_f(W_{n,f}^0 \leq c_{n,f}(\alpha + \tau_n)(1 - \epsilon_{3n}) - \epsilon_{1n} - \epsilon_{2n} - u_n\epsilon_{3n}\sqrt{\log n}) - \delta_{1n} \\
&\geq \mathrm{P}_f(W_{n,f}^0 \leq c_{n,f}(\alpha + \tau_n)) - p_{\bar{\epsilon}_n}(|G_{n,f}|) - \delta_{1n} &\text{(D.6)} \\
&\geq 1 - \alpha - \tau_n - p_{\bar{\epsilon}_n}(|G_{n,f}|) - \delta_{1n}, &\text{(D.7)}
\end{aligned}$$

where (D.6) 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 (D.7) follows since $c_{n,f}(\cdot)$ is the quantile function of $W_{n,f}^0$. Combining these inequalities leads to (3.5).

To prove (3.6) and (3.7), note that $\delta_n \leq Cn^{-c}$ and $\tau_n \leq Cn^{-c}$ by Conditions H1 and H3-H6. Further, by Markov's inequality, $c_{n,f}(\alpha) \leq$

$\mathrm{E}[\|G_{n,f}\|_{\mathcal{V}_n}]/\alpha \leq C\sqrt{\log n}$, so that $\bar{\epsilon}_{n,f} \leq Cn^{-c}$ because $\epsilon_{3n}u_n\sqrt{\log n} \leq C_1 n^{-c_1}$. Therefore, (3.6) and (3.7) follow from (3.5) and Condition H2. ■

Proof of Corollary 3.1. The proof is similar to that of Theorem 3.1. The details are provided in the Supplemental Material. ■

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 (3.1) holds with some ϵ_{1n} and δ_{1n} bounded from above by Cn^{-c}. In addition, inequality $\mathrm{E}[\|G_{n,f}\|_{\mathcal{V}_n}] \leq C\sqrt{\log n}$ follows from Corollary 2.2.8 in [41]. 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

$$\left|\frac{\hat{\sigma}_n(x,l)}{\sigma_{n,f}(x,l)} - 1\right| \leq \left|\frac{\hat{\sigma}_n^2(x,l)}{\sigma_{n,f}^2(x,l)} - 1\right|. \tag{D.8}$$

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 (D.8) is bounded by

$$\sup_{g \in \mathcal{K}_{n,f}^2} |\mathbb{E}_n[g(X_i)] - \mathrm{E}[g(X_1)]| + \sup_{g \in \mathcal{K}_{n,f}} \left|\mathbb{E}_n[g(X_i)]^2 - \mathrm{E}[g(X_1)]^2\right|. \tag{D.9}$$

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

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

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

$$\mathrm{P}\left(\sup_{g \in \mathcal{K}_{n,f}^2} |\mathbb{E}_n[g(X_i)] - \mathrm{E}[g(X_1)]| > \frac{1}{2}\sqrt{\frac{b_n^2\sigma_n^2 K_n}{n}}\right) \leq \frac{1}{n}. \tag{D.10}$$

In addition,

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

so that another application of Talagrand's inequality yields

$$\mathrm{P}\left(\sup_{g \in \mathcal{K}_{n,f}} \left|\mathbb{E}_n[g(X_i)]^2 - \mathrm{E}[g(X_1)]^2\right| > \frac{1}{2}\sqrt{\frac{b_n^2\sigma_n^2 K_n}{n}}\right) \leq \frac{1}{n}. \tag{D.11}$$

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 (D.8)-(D.11) 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

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

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 $\mathrm{P}_f(X_1^n \in S_{n,1}) \geq 1 - \delta_{3n} = 1 - 2/n$. 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

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

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

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

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

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

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 [41] gives

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

where the last inequality follows from the definition of ϵ_{3n} above. 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 $\lambda_{1n} \leq Cn^{-c}$ such that

$$\mathrm{P}(|\hat{W}_n(x_1^n) - \tilde{W}_n(x_1^n)| \geq \lambda_{1n}) \leq Cn^{-c}, \tag{D.12}$$

whenever $x_1^n \in S_{n,1}$. Further, since $b_n^2\sigma_n^4K_n^4/n \leq C_2n^{-c_2}$, Theorem A.2 shows that there exist $\lambda_{2n} \leq Cn^{-c}$ and a measurable set $S_{n,2}$ of values X_1^n such that $\mathrm{P}_f(X_1^n \in S_{n,2}) \geq 1-3/n$ 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

$$\mathrm{P}(|\tilde{W}_n(x_1^n) - W^0| \geq \lambda_{2n}) \leq Cn^{-c}. \tag{D.13}$$

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 (D.13) holds. Then by (D.12), we have

$$\mathrm{P}(|\hat{W}_n(x_1^n) - W^0| \geq \lambda_n) \leq Cn^{-c}, \tag{D.14}$$

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

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

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 $\mathrm{P}_f(X_1^n \in S_{n,0}) \geq 1-5/n$ and $\lambda_n \leq Cn^{-c}$. Part (b) follows similarly. ■

Appendix E. 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 and in Conditions L1-L4 but their values may change at each appearance. When $\mathcal{L}_n = [l_{\min,n}, l_{\max,n}]$, let $s \in (0,1]$ be some number. When $\mathcal{L}_n = [l_{\min,n}, l_{\max,n}] \cap \mathbb{N}$, let $s = 1$. The role of s in the proof is to make sure that for any $l \in \mathcal{L}_n$, either there is some $l' \in \mathcal{L}_n$ such that $l' \in [l-s, l)$ or there is no $l' \in \mathcal{L}_n$ such that $l' < l$. The proof is long, so we start with several preliminary lemmas.

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

$$\sum_{k_1,\dots,k_d \in \mathbb{Z}} \prod_{1 \le m \le d} \phi(x_m - k_m)^2 \ge c_\phi. \tag{E.1}$$

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

$$\underline{\sigma} 2^{ld/2} \le \inf_{x \in \mathcal{X}} \sigma_{n,f}(x,l) \le \sup_{x \in \mathcal{X}} \sigma_{n,f}(x,l) \le \bar{\sigma} 2^{ld/2}. \tag{E.2}$$

Here n_0, $\underline{\sigma}$, and $\bar{\sigma}$ depend only on the constants appearing in the statement of Theorem 4.1 and in Conditions L1-L4.

Proofs of Lemmas E.1 and E.2 can be found in the Supplemental Material.

Lemma E.3. *Under the assumptions of Theorem 4.1, Conditions H1-H4 are satisfied. Moreover, Condition H3 holds uniformly over all $\alpha \in (0,1)$.*

Proof of Lemma E.3. The result follows from combining Lemma E.2 and Theorem 3.2. ■

Lemma E.4. *Under the assumptions of Theorem 4.1, there exist $c > 0$ and $C > 0$ such that*

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

Proof of Lemma E.4. Lemma E.3 implies that the result of Theorem 3.2 applies under our assumptions. Therefore, we can and will assume that Conditions H1-H4 hold with Condition H3 being satisfied uniformly over all $\alpha \in (0,1)$. Inequality

$$\mathrm{P}_f(\hat{c}_n(\gamma_n) < c\sqrt{\log n}) \le Cn^{-c}$$

follows from Condition H3 and the fact that $c_{n,f}(\gamma_n + \tau_n)$ is bounded from below by $(1-\gamma_n-\tau_n)$ quantile of $N(0,1)$ distribution where $\gamma_n + \tau_n \le Cn^{-c}$.

Now let us verify inequality

$$\mathrm{P}_f(\hat{c}_n(\gamma_n) > C\sqrt{\log n}) \le Cn^{-c}. \tag{E.3}$$

Let $M(x_1^n)$ denote the median of $\hat{W}_n(x_1^n) = \|\hat{\mathbb{G}}_n(x_1^n, \xi_1^n)\|_{\mathcal{V}_n}$. Applying Lemma B.1 conditional on the data gives

$$\hat{c}_n(\gamma_n) \le M(X_1^n) + \sqrt{2|\log \gamma_n|}. \tag{E.4}$$

Further, in the proof of Theorem 3.2, it was shown that there exists a measurable set $S_{n,0}$ of values of X_1^n such that $\mathrm{P}_f(X_1^n \notin S_{n,0}) \le Cn^{-c}$ and for each $x_1^n \in S_{n,0}$ one can construct a random variable W^0 such that

$\mathrm{P}(|\hat{W}_n(x_1^n) - W^0| \geq \zeta_{1n}) \leq \zeta_{2n}$ for some ζ_{1n} and ζ_{2n} both bounded by Cn^{-c} and $W^0 \stackrel{d}{=} \|G_{n,f}\|_{\mathcal{V}_n}$; see (D.14). Therefore,

$$\mathrm{P}_f(M(X_1^n) > c_{n,f}(1/2 - \zeta_{2n}) + \zeta_{1n}) \leq Cn^{-c}. \tag{E.5}$$

Since $\mathrm{E}[\|G_{n,f}\|_{\mathcal{V}_n}] \leq C_1\sqrt{\log n}$ (assumed in Condition H1), Markov's inequality implies that $c_{n,f}(1/2 - \zeta_{2n}) \leq C\sqrt{\log n}$. Combining this inequality with (E.4) and (E.5) and using $|\log \gamma_n| \leq C_5 \log n$ give (E.3). This completes the proof of the lemma. ∎

Proof of Theorem 4.1. First, we note that for any $t' > t(f)$, $f \notin \Sigma(t', L)$; otherwise, we would have that $\sup_{x\in\mathcal{X}} |\mathrm{E}_f[\hat{f}_n(x,l)] - f(x)| \leq C2^{-lt'}$ contradicting the lower bound in (4.2). Therefore, (4.7) implies (4.8), and so it suffices to verify Conditions H1-H6 and to prove (3.7) and (4.7).

By Lemma E.3, we can and will assume that Conditions H1-H4 hold with Condition H3 being satisfied uniformly over all $\alpha \in (0,1)$. In addition, Condition H6 with some u_n satisfying $cu_n' \leq u_n \leq Cu_n'$ follows from Lemma E.4.

We now show that under our assumptions, Conditions H5 is also satisfied. By Condition H4, ϵ_{3n} is bounded by $C_1 n^{-c_1}$. So, there exists n_1 such that $\epsilon_{3n} \leq 1/2$ for all $n \geq n_1$. Let $\delta_{4n} = 1$ for $n < \max(n_0, n_1)$, where n_0 is chosen so that for all $n \geq n_0$ 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 $n \geq \max(n_0, n_1)$.

Let $t := t(f)$. Let $m > s$ be such that $c_3 2^{(m-s)\underline{t}} > C_3$, $M_1 > 0$ be such that $M_1(1 - C_3 2^{-(m-s)\underline{t}}/c_3) > 2(q+1)C_3/c_3$, and $M_2 > 0$ be such that $4M_2 < (q-1)\underline{\sigma}/\bar{\sigma}$ where s is introduced in the beginning of Appendix E, q in (4.6), and $\underline{\sigma}$ and $\bar{\sigma}$ in Lemma E.2. For $M > 0$, define

$$l^\star(M) := \inf\left\{ l \in \mathcal{L}_n : C_3 2^{-lt}\sqrt{n} \leq M\hat{c}_n(\gamma_n) \sup_{x\in\mathcal{X}} \hat{\sigma}_n(x,l) \right\},$$

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

Lemma E.5. *There exist $c, C > 0$ such that*

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

Proof of Lemma E.5. Define $\mathcal{L}_n^1 := \{l \in \mathcal{L}_n : l < l_1^\star - m\}$. If there is no $l' \in \mathcal{L}_n$ such that $l' < l_1^\star$, we are done. Otherwise, since $l_1^\star \in \mathcal{L}_n$ by $\mathcal{L}_n$ being closed, there exists some $l' \in \mathcal{L}_n$ such that $l' \in [l_1^\star - s, l_1^\star)$ (Condition L3). Fix this l'. Then

$$\mathrm{P}_f(\hat{l}_n < l_1^\star - m) \leq \mathrm{P}_f\left(\inf_{l\in\mathcal{L}_n^1} \sup_{x\in\mathcal{X}} \frac{\sqrt{n}|\hat{f}_n(x,l) - \hat{f}_n(x,l')|}{\hat{\sigma}_n(x,l) + \hat{\sigma}_n(x,l')} \leq q\hat{c}_n(\gamma_n) \right) \tag{E.6}$$

By Condition L2 and the triangle inequality,

$$\begin{aligned}
c_3 2^{-lt} &\le \sup_{x\in\mathcal{X}} |\mathrm{E}_f[\hat{f}_n(x,l)] - f(x)| \\
&\le \sup_{x\in\mathcal{X}} |\mathrm{E}_f[\hat{f}_n(x,l)] - \hat{f}_n(x,l)| + \sup_{x\in\mathcal{X}} |\hat{f}_n(x,l) - \hat{f}_n(x,l')| \\
&\quad + \sup_{x\in\mathcal{X}} |\hat{f}_n(x,l') - \mathrm{E}_f[\hat{f}_n(x,l')]| + \sup_{x\in\mathcal{X}} |\mathrm{E}_f[\hat{f}_n(x,l')] - f(x)|.
\end{aligned}$$

Hence applying Condition L2 one more time and letting

$$B_n(l) := \sup_{x\in\mathcal{X}} |\hat{f}_n(x,l) - \mathrm{E}_f[\hat{f}_n(x,l)]|,$$

we have

$$\sup_{x\in\mathcal{X}} |\hat{f}_n(x,l) - \hat{f}_n(x,l')| \ge c_3 2^{-lt} - C_3 2^{-l't} - B_n(l) - B_n(l').$$

Further, for $l \in \mathcal{L}_n^1$, by the definition of $l_1^\star$, construction of M_1, and since $t \ge \underline{t}$,

$$\begin{aligned}
\frac{c_3 2^{-lt} - C_3 2^{-l't}}{2} &\ge \frac{c_3}{2^{lt+1}} \left(1 - C_3 2^{-(m-s)\underline{t}}/c_3\right) \\
&\ge \frac{c_3}{2C_3\sqrt{n}} M_1 \hat{c}_n(\gamma_n) \sup_{x\in\mathcal{X}} \hat{\sigma}_n(x,l) \left(1 - C_3 2^{-(m-s)\underline{t}}/c_3\right) \\
&\ge (q+1)\hat{c}_n(\gamma_n) \sup_{x\in\mathcal{X}} \hat{\sigma}_n(x,l)/\sqrt{n},
\end{aligned}$$

and

$$\begin{aligned}
\frac{c_3 2^{-lt} - C_3 2^{-l't}}{2} &\ge \frac{C_3}{2^{l't+1}} \left(c_3 2^{(m-s)\underline{t}}/C_3 - 1\right) \\
&\ge \frac{1}{2\sqrt{n}} M_1 \hat{c}_n(\gamma_n) \sup_{x\in\mathcal{X}} \hat{\sigma}_n(x,l') \left(c_3 2^{(m-s)\underline{t}}/C_3 - 1\right) \\
&\ge (q+1)\hat{c}_n(\gamma_n) \sup_{x\in\mathcal{X}} \hat{\sigma}_n(x,l')/\sqrt{n}.
\end{aligned}$$

Combining these inequalities yields for $l \in \mathcal{L}_n^1$,

$$\begin{aligned}
&\sup_{x\in\mathcal{X}} \frac{\sqrt{n}|\hat{f}_n(x,l) - \hat{f}_n(x,l')|}{\hat{\sigma}_n(x,l) + \hat{\sigma}_n(x,l')} \ge \frac{\sup_{x\in\mathcal{X}} \sqrt{n}|\hat{f}_n(x,l) - \hat{f}_n(x,l')|}{\sup_{x\in\mathcal{X}} \hat{\sigma}_n(x,l) + \sup_{x\in\mathcal{X}} \hat{\sigma}_n(x,l')} \\
&\ge (q+1)\hat{c}_n(\gamma_n) - \frac{B_n(l) + B_n(l')}{\sup_{x\in\mathcal{X}} \hat{\sigma}_n(x,l) + \sup_{x\in\mathcal{X}} \hat{\sigma}_n(x,l')}.
\end{aligned}$$

Therefore, (E.6) gives

$$\begin{aligned}
\mathrm{P}_f(\hat{l}_n < l_1^\star - m) &\le \mathrm{P}_f\left(\sup_{v\in\mathcal{V}_n} \frac{\sqrt{n}|\hat{f}_n(v) - \mathrm{E}[\hat{f}_n(v)]|}{\hat{\sigma}_n(v)} \ge \hat{c}_n(\gamma_n)\right) \\
&\le \gamma_n + Cn^{-c} \le Cn^{-c},
\end{aligned}$$

where the inequalities in the second line follow from an argument similar to that used in the proof of Theorem 3.1. This gives the asserted claim. ∎

Lemma E.6. *There exist $c, C > 0$ such that*

$$\lambda_{2n} := \mathrm{P}_f(\hat{l}_n > l_2^\star) \le Cn^{-c}.$$

Proof of Lemma E.6. Define $\mathcal{L}_n^2 := \{l \in \mathcal{L}_n : l > l_2^\star\}$. Consider the event $\mathcal{A}_n$ that $\sup_{v \in \mathcal{V}_n} |\hat{\sigma}_n(v)/\sigma_n(v) - 1| \le \epsilon_{3n}$. By Condition H4, the probability of this event is at least $1 - \delta_{3n}$. On the event $\mathcal{A}_n$, for all $l \in \mathcal{L}_n$,

$$\sup_{x \in \mathcal{X}} \hat{\sigma}_n(x, l) \le (1 + \epsilon_{3n}) \sup_{x \in \mathcal{X}} \sigma_{n,f}(x, l) \le (1 + \epsilon_{3n}) \bar{\sigma} 2^{ld/2},$$

by Lemma E.2. Therefore, on the event $\mathcal{A}_n$, for all $l \in \mathcal{L}_n^2$,

$$C_3 2^{-lt} \sqrt{n} \le M_2 \hat{c}_n(\gamma_n)(1 + \epsilon_{3n}) \bar{\sigma} 2^{ld/2}. \tag{E.7}$$

Indeed, if (E.7) does not hold for $l = l_2^\star$, then $\mathcal{L}_n^2$ is empty by the definition of $l_2^\star$; otherwise, (E.7) holds for all $l > l_2^\star$, and, in particular, for all $l \in \mathcal{L}_n^2$. Hence, on the event $\mathcal{A}_n$, for all $l \in \mathcal{L}_n^2$,

$$\begin{aligned} C_3 2^{-lt} \sqrt{n} &\le M_2 \hat{c}_n(\gamma_n)(1 + \epsilon_{3n}) \bar{\sigma} 2^{ld/2} \\ &\le M_2 \hat{c}_n(\gamma_n)(1 + \epsilon_{3n})(\bar{\sigma}/\underline{\sigma}) \inf_{x \in \mathcal{X}} \sigma_{n,f}(x, l) \quad \text{(E.8)} \\ &\le M_2 \hat{c}_n(\gamma_n)(1 + 2\epsilon_{3n})^2 (\bar{\sigma}/\underline{\sigma}) \inf_{x \in \mathcal{X}} \hat{\sigma}_n(x, l) \quad \text{(E.9)} \\ &\le (q - 1)\hat{c}_n(\gamma_n) \inf_{x \in \mathcal{X}} \hat{\sigma}_n(x, l), \quad \text{(E.10)} \end{aligned}$$

where (E.8) follows from Lemma E.2, (E.9) from the definition of the event $\mathcal{A}_n$ and $\epsilon_{3n} \le 1/2$, so that $1/(1+2\epsilon_{3n}) \le 1 - \epsilon_{3n}$, and (E.10) from the choice of M_2 and $(1 + 2\epsilon_{3n})^2 \le 4$, which holds for all $n \ge n_1$. Hence

$$\hat{\Delta}_{n,f}(l) := \sup_{x \in \mathcal{X}} \frac{\sqrt{n}|f(x) - \mathrm{E}_f[\hat{f}_n(x, l)]|}{\hat{\sigma}_n(x, l)}$$

satisfies

$$\mathrm{P}_f\left(\sup_{l \in \mathcal{L}_n^2} \hat{\Delta}_{n,f}(l) > (q - 1)\hat{c}_n(\gamma_n)\right) \le \delta_{3n} \tag{E.11}$$

by Condition L2. Further, by the definition of $\hat{l}_n$ and the triangle inequality,

$$\begin{aligned} \mathrm{P}_f(\hat{l}_n > l_2^\star) &\le \mathrm{P}_f\left(\sup_{l, l' \in \mathcal{L}_n^2} \sup_{x \in \mathcal{X}} \frac{\sqrt{n}|\hat{f}_n(x, l) - \hat{f}_n(x, l')|}{\hat{\sigma}_n(x, l) + \hat{\sigma}_n(x, l')} > q\hat{c}_n(\gamma_n)\right) \\ &\le \mathrm{P}_f\left(\sup_{l \in \mathcal{L}_n^2} \sup_{x \in \mathcal{X}} \frac{\sqrt{n}|\hat{f}_n(x, l) - f(x)|}{\hat{\sigma}_n(x, l)} > q\hat{c}_n(\gamma_n)\right) \end{aligned}$$

Using the definition of $\hat{\Delta}_{n,f}(l)$ and applying the triangle inequality once again, the probability in the last expression can be further bounded from above by

$$
\begin{aligned}
&\mathrm{P}_f\left(\sup_{l\in\mathcal{L}_n^2}\sup_{x\in\mathcal{X}}\frac{\sqrt{n}|\hat{f}_n(x,l)-\mathrm{E}_f[\hat{f}_n(x,l)]|}{\hat{\sigma}_n(x,l)}+\hat{\Delta}_{n,f}(l)>q\hat{c}_n(\gamma_n)\right)\\
&\le \mathrm{P}_f\left(\sup_{v\in\mathcal{V}_n}\frac{\sqrt{n}|\hat{f}_n(v)-\mathrm{E}_f[\hat{f}_n(v)]|}{\hat{\sigma}_n(v)}>\hat{c}_n(\gamma_n)\right)+\delta_{3n} \qquad \text{(E.12)}\\
&\le \gamma_n+\delta_{3n}+Cn^{-c}\le Cn^{-c}. \qquad \text{(E.13)}
\end{aligned}
$$

where (E.12) follows from (E.11), and (E.13) from an argument similar to that used in the proof of Theorem 3.1 and because γ_n and δ_{3n} are both bounded by Cn^{-c}. This gives the asserted claim. ∎

Let $l_0^\star := l_0^\star(f)$ be $l \in \mathcal{L}_n$ satisfying (4.4), which exists by Condition L3. Now we can verify Condition H5:

Lemma E.7. *Under our assumptions, Condition H5 is satisfied.*

Proof of Lemma E.7. We claim that with probability at least $1-\delta_{3n}-\lambda_{0n}$,

$$\sqrt{n}2^{-l_1^\star(t+d/2)}\le C\hat{c}_n(\gamma_n). \qquad \text{(E.14)}$$

Indeed, consider the event that for all $l\in\mathcal{L}_n$,

$$\hat{c}_n(\gamma_n)\ge c\sqrt{\log n}\ \text{and}\ \sup_{x\in\mathcal{X}}\hat{\sigma}_n(x,l)\le(1+\epsilon_{3n})\bar{\sigma}2^{ld/2}.$$

By Lemmas E.2 and E.4 and Condition H4, the probability of this event is greater than or equal to $1-\delta_{3n}-\lambda_{0n}$. On this event, if $l_1^\star\ge l_0^\star$, then

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

and if $l_1^\star<l_0^\star$, then the set

$$\{l\in\mathcal{L}_n : C_3 2^{-lt}\sqrt{n}\le M_1\hat{c}_n(\gamma_n)\sup_{x\in\mathcal{X}}\hat{\sigma}_n(x,l)\}$$

is nonempty, so that $l_1^\star$ belongs to the set

$$\{l\in\mathcal{L}_n : C_3 2^{-lt}\sqrt{n}\le M_1\hat{c}_n(\gamma_n)(1+\epsilon_{3n})\bar{\sigma}2^{ld/2}\},$$

and so (E.14) holds in both cases on this event because $1+\epsilon_{3n}\le 3/2$.

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

$$
\begin{aligned}
\Delta_{n,f}(\hat{l}_n)&\le\frac{\sqrt{n}C_3 2^{-\hat{l}_n t}}{\inf_{x\in\mathcal{X}}\sigma_{n,f}(x,\hat{l}_n)}\le\sqrt{n}C_3 2^{-\hat{l}_n t}/(\underline{\sigma}2^{\hat{l}_n d/2}) \qquad \text{(E.15)}\\
&=C\sqrt{n}2^{-\hat{l}_n(t+d/2)}\le C\sqrt{n}2^{-(l_1^\star-m)(t+d/2)}\le C\hat{c}_n(\gamma_n), \qquad \text{(E.16)}
\end{aligned}
$$

where (E.15) follows from Condition L2 and Lemma E.2, and (E.16) from Lemma E.5, and (E.14). Since $\delta_{3n}+\lambda_{0n}+\lambda_{1n}\le Cn^{-c}$, Condition H5 follows because $c_n'=u_n'\hat{c}_n(\gamma_n)$ and u_n' is sufficiently large ($u_n'\ge C(\mathcal{F})$). ∎

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

Lemma E.8. *There exist $c, C > 0$ such that*

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

Proof of Lemma E.8. We claim that with probability at least $1 - \delta_{3n} - \lambda_{0n}$,

$$2^{-(l_2^\star - s)(t+d/2)} \sqrt{n} \geq c\sqrt{\log n}. \tag{E.17}$$

Indeed, consider the event that for all $l \in \mathcal{L}_n$,

$$\hat{c}_n(\gamma_n) \leq C\sqrt{\log n} \text{ and } \sup_{x \in \mathcal{X}} \hat{\sigma}_n(x, l) \geq (1 - \epsilon_{3n}) \underline{\sigma} 2^{ld/2}.$$

By Lemmas E.2 and E.4 and Condition H4, the probability of this event is greater than or equal to $1 - \delta_{3n} - \lambda_{0n}$. On this event, if $l_2^\star - s \leq l_0^\star$, then

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

and if $l_2^\star - s > l_0^\star$, then there is an element $l' \in \mathcal{L}_n$ such that $l' \in [l_2^\star - s, l_2^\star)$, and

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

since l' does not belong to the set

$$\{l \in \mathcal{L}_n : C_3 2^{-lt} \sqrt{n} \leq M_2 \hat{c}_n(\gamma_n) \sup_{x \in \mathcal{X}} \hat{\sigma}_n(x, l)\},$$

and so also does not belong to the set

$$\{l \in \mathcal{L}_n : C_3 2^{-lt} \sqrt{n} \leq M_2 \hat{c}_n(\gamma_n)(1 - \epsilon_{3n}) \underline{\sigma} 2^{ld/2})\}.$$

Therefore, (E.17) holds in both cases on this event.

Hence, by Lemma E.6, with probability at least $1 - \delta_{3n} - \lambda_{0n} - \lambda_{2n}$,

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

Conclude that with the same probability

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

Since $\delta_{3n} + \lambda_{0n} + \lambda_{2n} \leq C n^{-c}$, the result follows. ∎

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 $u_n \leq C u_n' \leq C \log n$, Theorem 3.1 applies, so that (3.7) 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 E.8, we have

$$\sup_{f \in \mathcal{F}} \mathrm{P}_f \left(\sup_{x \in \mathcal{X}} \lambda(\mathcal{C}_n(x)) > C\bar{c}_n \frac{r_n(t(f))}{\sqrt{\log n}} \right) \leq \delta_{2n} + \delta_{5n} + \lambda_{3n}, \tag{E.18}$$

where

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

Since τ_n and ϵ_{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 $n < n_2$, (4.7) 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, $c_{n,f}(\alpha/2) \leq \mathrm{E}[\|G_{n,f}\|_{\mathcal{V}_n}] + \sqrt{2|\log(\alpha/2)|}$. By Condition H1, $\mathrm{E}[\|G_{n,f}\|_{\mathcal{V}_n}]$ is bounded from below, and so $c_{n,f}(\alpha/2) + 1 \leq C\mathrm{E}[\|G_{n,f}\|_{\mathcal{V}_n}]$. Further, $\mathrm{E}[\|G_{n,f}\|_{\mathcal{V}_n}] \leq C_1\sqrt{\log n}$ (Condition H1) gives

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

Substituting this expression into (E.18) yields (4.7). This completes the proof of the theorem. ■

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). Nonparametric 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.

[9] Chernozhukov, V., Lee, S., and Rosen, A. (2013). Intersection bounds: estimation and inference. *Econometrica* **81** 667-737. arXiv:0907.3503v4.

[10] Chetverikov, D. (2012). Testing regression monotonicity in econometric models. arXiv:1212.6757.

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

[12] Daubechies, I. (1992). *Ten Lectures on Wavelets.* CBMS-NSF Regional Conference Series in Applied Mathematics.

[13] Dudley, R.M. (1999). *Uniform Central Limit Theorems.* Cambridge University Press.

[14] 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.

[15] Giné, E., Gunturk, C.S. and Madych W.R. (2011). On the periodized square of L^2 cardinal splines. *Exp. Math.* **20** 177-188.

[16] Giné, E., Koltchinskii, V. and Sakhanenko, L. (2004). Kernel density estimators: convergence in distribution for weighted sup-norms. *Probab. Theory Related Fields* **130** 167-198.

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

[18] 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.

[19] Giné, E. and Nickl, R. (2010). Confidence bands in density estimation. *Ann. Statist.* **38** 1122-1170.

[20] 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.

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

[22] Hall, P., and Horowitz, J. (2013). A simple bootstrap method for constructing nonparametric confidence bands for functions. *Ann. Statist.* **41** 1892-1921.

[23] Härdle, W., Kerkyacharian, G., Picard, D., and Tsybakov, A. (1998). *Wavelets, Approximation, and Statistical Applications.* Springer-Verlag.

[24] Hoffman, M. and Nickl, R. (2011). On adaptive inference and confidence bands. *Ann. Statist.* **39** 2383-2409.

[25] Komlós, J. Major, J. Tusanády, G. (1975). An approximation of partial sums of independent rv's, and the sample df. I. *Z. Wahrsch. Verw. Gebiete* **32** 111-131.

[26] Lepski, O. (1991). Asymptotically minimax adaptive estimation I: Upper bounds. Optimally adaptive estimates. *Theory Probab. Appl.* **36** 682-697.

[27] Li, K.-C. (1989). Honest confidence regions for nonparametric regression. *Ann. Statist.* **17** 1001-1008.

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

[29] Low, M. (1997). On nonparametric confidence intervals. *Ann. Statist.* **25** 2547-2554.

[30] Massart, P. (2000). About the constants in Talagrand's concentration inequalities for empirical processes. *Ann. Probab.* **28** 863-884.

[31] Massart, P. (2007). *Concentration Inequalities and Model Selection.* Springer-Verlag.

[32] Newey, W. (1997). Convergence rates and asymptotic normality for series estimators. *J. Econometrics* **79** 147-168.

[33] Picard, D. and Tribouley, K. (2000). Adaptive confidence interval for pointwise curve estimation. *Ann. Statist.* **28** 298-335.

[34] Rio, E. (1994). Local invariance principles and their application to density estimation. *Probab. Theory Related Fields* **98** 21-45.

[35] Rudelson, M. and Vershynin, R. (2009). Smallest singular value of a random rectangular matrix. *Comm. Pure Appl. Math.* **62** 1707-1739.

[36] Silverman, B.W. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *Ann. Statist.* **6** 177-184.

[37] 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.

[38] Talagrand, M. (1996). New concentration inequalities in product spaces. *Invent. Math.* **126** 503-563.

[39] Tsybakov, A. (2009). *Introduction to Nonparametric Estimation.* Springer Series in Statistics.

[40] van der Vaart, A.W. (1998). *Asymptotic statistics.* Cambridge Series in Statistical and Probabilistic Mathematics.

[41] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak convergence and empirical processes: With applications to statistics.* Springer-Verlag.

[42] Wasserman, L. (2006). *All of Nonparametric Statistics.* Springer Texts in Statistics.

[43] Xia, Y. (1998). Bias-corrected confidence bands in nonparametric regression. *J. Roy. Stat. Soc. Ser. B Stat. Methodol.* **60** 797-811.

Supplemental Material

Deferred Proofs and Discussions

V. Chernozhukov, D. Chetverikov, and K. Kato

Appendix F. Proof of Corollary 3.1

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

$$
\begin{aligned}
&\mathrm{P}_f(f(x) \in \widetilde{\mathcal{C}}_n(x), \forall x \in \mathcal{X}) \\
&\geq \mathrm{P}_f(|Z_{n,f}(x,\hat{l}_n)| + \Delta_{n,f}(\hat{l}_n) \leq \hat{c}_n(\alpha)\hat{\sigma}_n(x,\hat{l}_n)/\sigma_{n,f}(x,\hat{l}_n), \forall x \in \mathcal{X}) \quad \text{(F.1)} \\
&\geq \mathrm{P}_f(\sup_{x\in\mathcal{X}} |Z_{n,f}(x,\hat{l}_n)| + \Delta_{n,f}(\hat{l}_n) \leq \hat{c}_n(\alpha)(1-\epsilon_{3n})) - \delta_{3n} \quad \text{(F.2)} \\
&\geq \mathrm{P}_f(\sup_{x\in\mathcal{X}} |Z_{n,f}(x,\hat{l}_n)| \leq \hat{c}_n(\alpha)(1-\epsilon_{3n}) - u_n\sqrt{\log n}) - \delta_{3n} - \delta_{6n} \quad \text{(F.3)} \\
&\geq \mathrm{P}_f(\|Z_{n,f}\|_{\mathcal{V}_n} \leq \hat{c}_n(\alpha)(1-\epsilon_{3n}) - u_n\sqrt{\log n}) - \delta_{3n} - \delta_{6n}, \quad \text{(F.4)}
\end{aligned}
$$

where (F.1) follows by the triangle inequality, (F.2) by Condition H4, (F.3) by assumption (3.8), and (F.4) 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

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

the probability in (F.4) equals

$$
\begin{aligned}
&\mathrm{P}_f(W_{n,f} \leq \hat{c}_n(\alpha)(1-\epsilon_{3n}) - u_n\sqrt{\log n}) \\
&\geq \mathrm{P}_f(W_{n,f} \leq c_{n,f}(\alpha+\tau_n)(1-\epsilon_{3n}) - \epsilon_{2n} - u_n\sqrt{\log n}) - \delta_{2n} \quad \text{(F.5)} \\
&\geq \mathrm{P}_f(W^0_{n,f} \leq c_{n,f}(\alpha+\tau_n) - \widetilde{\epsilon}_n) - \delta_{1n} - \delta_{2n} \quad \text{(F.6)} \\
&\geq 1 - \alpha - \tau_n - p_{\widetilde{\epsilon}_n}(|G_{n,f}|) - \delta_{1n} - \delta_{2n}, \quad \text{(F.7)}
\end{aligned}
$$

where (F.5) follows by Condition H3, (F.6) by Condition H1, and (F.7) by the definition of the Lévy concentration function $p_{\widetilde{\epsilon}_n}(|G_{n,f}|)$. As in the proof of Theorem 3.1, $c_{n,f}(\alpha) \leq C\sqrt{\log n}$, so that $\widetilde{\epsilon}_n \leq Cn^{-c}$. Hence (3.9) and (3.10) follow from $\tau_n + \delta_{1n} + \delta_{2n} + \delta_{3n} + \delta_{6n} \leq Cn^{-c}$.

To prove (3.11) and (3.12), 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 (3.9) and (3.10),

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

where $\widetilde{\epsilon}'_n := \epsilon_{1n} + \epsilon_{2n} + \epsilon_{3n}c_{n,f}(\alpha-\tau_n) + u_n\sqrt{\log n}$. Hence, (3.11) and (3.12) follow. This completes the proof of the corollary. ■

Appendix G. Proofs of Lemmas B.1 and B.2

Proof of Lemma B.1. Pick any $\alpha \in (0,1)$. Since $\mathrm{E}[Y(t)^2] = 1$ for all $t \in T$, the Borel-Sudakov-Tsirel'son inequality (see Theorem A.2.1 in [41]) gives for all $r > 0$,

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

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

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

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

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

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

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,

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

The claim of part (i) follows. Parts (ii) and (iii) follow similarly. ■

Appendix H. Proofs for Appendix A

Proof of Theorem A.1. In this proof, C is an absolute constant but its value can change at each appearance. The proof consists of applying Theorem 2.1 in [7]. Standard calculations show that for any $\varepsilon \in (0,1)$,

$$J(\varepsilon) := \int_0^\varepsilon \sup_Q \sqrt{1 + \log N(\mathcal{G}, L_2(Q), b\tau)}\, d\tau \leq C\varepsilon \sqrt{\log (a/\varepsilon)^v}.$$

Further, for some sufficiently large C, let $\kappa_n := C(b\sigma^2 + b^3 K_n/n)^{1/3}$. Lemma 2.2 in [7] implies

$$\mathrm{E}\left[\sup_{g\in\mathcal{G}} \mathbb{E}_n[|g(X_i)|^3]\right] \le \sup_{g\in\mathcal{G}} \mathrm{E}\left[|g(X_1)|^3\right] + Cn^{-1/2}b^{3/2}\left(b^{3/2}J(\delta_3^{3/2}) + \frac{b^{3/2}J^2(\delta_3^{3/2})}{\sqrt{n}\delta_3^3}\right),$$

for any $\delta_3 \ge \sup_{g\in\mathcal{G}} \mathrm{E}[|g(X_1)|^3]^{1/3}/b$. Setting $\delta_3 = b^{1/3}\sigma^{2/3}/b = (\sigma/b)^{2/3}$ gives

$$\begin{aligned}\mathrm{E}\left[\sup_{g\in\mathcal{G}} \mathbb{E}_n[|g(X_i)|^3]\right] &\le b\sigma^2 + Cn^{-1/2}b^{3/2}\left(b^{3/2-1}\sigma K_n^{1/2} + b^{3/2}K_n n^{-1/2}\right)\\ &\le C(b\sigma^2 + b^3K_n/n) \le \kappa_n^3.\end{aligned}$$

Let $\varepsilon_n = \sigma/(bn^{1/2})$. Then $H_n(\varepsilon_n) := \log(\sup_Q N(\mathcal{G}, L_2(Q), b\varepsilon_n) \vee n) \le K_n$ and $J(\varepsilon_n) \le C\sigma K_n^{1/2}/(bn^{1/2})$. Note that since C in the definition of κ_n is sufficiently large, $b/\kappa_n < \gamma^{-1/3}n^{1/3}H_n(\varepsilon)^{-1/3}$. Therefore, Theorem 2.1 combined with Lemma 2.2 in [7]) shows for any $\gamma \in (0,1)$, $q \ge 4$, and $\delta_4 \ge \sup_{g\in\mathcal{G}} \mathrm{E}[g(X_1)^4]^{1/4}/b$, one can construct a random variable W^0 such that $W^0 \stackrel{d}{=} \|B\|_{\mathcal{G}}$ and

$$\mathrm{P}\left(|W_n - W^0| > C_q\Delta_n(\varepsilon_n,\gamma)\right) \le \gamma + C(\log n)/n, \tag{H.1}$$

where C_q is an absolute constant that depends only on q, and

$$\begin{aligned}\Delta_n(\varepsilon_n,\gamma) &:= \phi_n(\varepsilon_n) + \gamma^{-1/q}\varepsilon_n b + \gamma^{-1/q}bn^{-1/2} + \gamma^{-2/q}bn^{-1/2}\\ &\quad + \gamma^{-1/2}\mathcal{E}_n^{1/2}H_n^{1/2}(\varepsilon_n)n^{-1/4} + \gamma^{-1/3}\kappa_n H_n^{2/3}(\varepsilon_n)n^{-1/6},\\ \phi_n(\varepsilon_n) &\le C\left(bJ(\varepsilon_n) + \varepsilon_n^{-2}bJ^2(\varepsilon_n)n^{-1/2}\right),\\ \mathcal{E}_n &\le C\left(b^2J(\delta_4^2) + \delta_4^{-4}b^2J^2(\delta_4^2)n^{-1/2}\right).\end{aligned}$$

Using the bound derived above, we have

$$\phi_n(\varepsilon_n) \le C(\sigma K_n^{1/2}n^{-1/2} + bK_nn^{-1/2}) \le CbK_nn^{-1/2},$$

and setting $\delta_4 = (b^2\sigma^2)^{1/4}/b = (\sigma/b)^{1/2}$,

$$\mathcal{E}_n \le C(b\sigma K_n^{1/2} + b^2K_nn^{-1/2}).$$

Setting $q = 4$ and using $K_n \ge 1$ and $\gamma < 1$ gives

$$\begin{aligned}&\gamma^{-1/q}\varepsilon_n b + \gamma^{-1/q}bn^{-1/2} + \gamma^{-2/q}bn^{-1/2} \le CbK_n/(\gamma n)^{1/2},\\ &\gamma^{-1/2}\mathcal{E}_n^{1/2}H_n^{1/2}(\varepsilon_n)n^{-1/4} \le C(\gamma^{-1/2}(b\sigma)^{1/2}K_n^{3/4}n^{-1/4} + bK_n/(\gamma n)^{1/2}),\\ &\gamma^{-1/3}\kappa_nH_n^{2/3}(\varepsilon_n)n^{-1/6} \le C(\gamma^{-1/3}bK_nn^{-1/2} + \gamma^{-1/3}b^{1/3}\sigma^{2/3}K_n^{2/3}n^{-1/6}).\end{aligned}$$

Substituting these bounds into (H.1) and using the definition of $\Delta_n(\varepsilon_n,\gamma_n)$, we obtain the asserted claim. ∎

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

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

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

where $\Delta = \max_{1\le j,k\le p} |\Sigma^X_{jk} - \Sigma^Y_{jk}|$.

Proof. See Theorem 1 in [8]. ∎

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

Proof. See Lemma 4.1 in [7]. ∎

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

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

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

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

Proof. See Lemma 4.2 in [7]. ∎

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 $\mathrm{VC}(b^2, 2a, 2v)$ type and $(\mathcal{G}-\mathcal{G})^2$ is $\mathrm{VC}(4b^2, 2a, 4v)$ type function classes. In addition, $\mathrm{E}[g^2] \le b^2\sigma^2$ for all $g \in \mathcal{G}\cdot\mathcal{G}$ and $\mathrm{E}[g^2] \le 16b^2\sigma^2$ for all $g \in (\mathcal{G}-\mathcal{G})^2$. Together with the assumed condition $b^2 K_n \le n\sigma^2$, this justifies an application of Talagrand's inequality (Theorem B.1) with $t = \log n$, which gives

$$\mathrm{P}\left(\sup_{g\in\mathcal{G}} |\mathbb{E}_n[g(X_i)] - \mathrm{E}[g(X_1)]| \le \sqrt{\frac{\sigma^2 K_n}{n}} \right) \ge 1 - \frac{1}{n}, \tag{H.2}$$

$$\mathrm{P}\left(\sup_{g\in\mathcal{G}\cdot\mathcal{G}} |\mathbb{E}_n[g(X_i)] - \mathrm{E}[g(X_1)]| \le \sqrt{\frac{b^2\sigma^2 K_n}{n}} \right) \ge 1 - \frac{1}{n}, \tag{H.3}$$

$$\mathrm{P}\left(\sup_{g\in(\mathcal{G}-\mathcal{G})^2} |\mathbb{E}_n[g(X_i)] - \mathrm{E}[g(X_1)]| \le \sqrt{\frac{b^2\sigma^2 K_n}{n}} \right) \ge 1 - \frac{1}{n}. \tag{H.4}$$

Let $S_{n,0} \in \mathcal{S}^n$ be the intersection of events in (H.2)-(H.4). Then $\mathrm{P}(X_1^n \in S_{n,0}) \ge 1 - 3/n$.

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 $\mathrm{E}[(g(X_1) - g_j(X_1))^2] \leq b^2\tau^2$. 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

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

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

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

By the Borell-Sudakov-Tsirel'son inequality (see Theorem A.2.1 in [41]), with probability at least $1 - 2/n$,

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

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

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

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 [41],

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

To bound r_2, we have

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

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 [41],

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

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 H.3 to define a function f corresponding to the set $B^{L\psi_n}$, $L\psi_n$-enlargement of the set B, with chosen β and δ. 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,

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

where

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

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

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

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

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

Therefore,

$$\begin{aligned}
\mathrm{E}[1_B(\tilde{W}_n(x_1^n))] &\leq \mathrm{E}[1_{B^{L\psi_n}}(W(x_1^n)(\tau))] + 2/n \\
&\leq \mathrm{E}[f(W(x_1^n)(\tau))]/(1-\varepsilon) + 2/n \\
&\leq \mathrm{E}[f(W^0(\tau))]/(1-\varepsilon) + C\gamma_n(\delta) \\
&\leq \mathrm{E}[1_{B^{L\psi_n+3\delta}}(W^0(\tau))] + C\gamma_n(\delta) \\
&\leq \mathrm{E}[1_{B^{(L+1)\psi_n+3\delta}}(\tilde{W}^0)] + C\gamma_n(\delta),
\end{aligned}$$

where C is varying from line to line. The claim of the theorem follows by applying Theorem H.2. ■

Appendix I. Proof of Lemmas E.1 and E.2

Proof of Lemma E.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 (E.1) becomes

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

Since the function $\sum_{k\in\mathbb{Z}}\phi(\cdot-k)^2$ has period 1, it suffices to consider $x\in[0,1]$. It follows from Lemma 8.6 in [23] 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 [23] 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 asserted claim follows. ■

Proof of Lemma E.2. Fix $f\in\mathcal{F}$, $l\in\mathcal{L}_n$, and $x\in\mathcal{X}$. Note that we have

$$2^{-ld}|K_l(y,x)| \leq C\exp(-2^l c|y-x|), \tag{I.1}$$

for all $y\in\mathcal{X}$. 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 [23]. Therefore, $|K_l(y,x)| \leq 2^{ld}C$ for all $y \in \mathcal{X}$, $|\mathrm{E}_f[K_l(X_1,x)]| \leq C$, and $\mathrm{E}_f[K_l(X_1,x)^2] \leq 2^{ld}C$. Further,

$$\mathrm{E}_f[K_l(X_1,x)^2] = \int_{\mathbb{R}^d} K_l(y,x)^2 f(y)dy \geq \int_{y:|y-x|\leq\delta} K_l(y,x)^2 f(y)dy$$

$$\geq \underline{f}\int_{y:|y-x|\leq\delta} K_l(y,x)^2 dy \geq \underline{f}\left(\int_{\mathbb{R}^d} K_l(y,x)^2 dy - \int_{y:|y-x|>\delta} K_l(y,x)^2 dy\right).$$

By (I.1),

$$\int_{y:|y-x|>\delta} K_l(y,x)^2 dy \leq 2^{2ld}C \int_{y:|y-x|>\delta} \exp(-2^{l+1}c|y-x|)dy$$
$$= 2^{ld}C \int_{y:|y|>2^l\delta} \exp(-2c|y|)dy \leq C.$$

In addition, for convolution kernels,

$$\int_{\mathbb{R}^d} K_l(y,x)^2 dy = 2^{ld}\int_{\mathbb{R}^d} K(y)^2 dy \geq 2^{ld}c,$$

and for wavelet projection kernels, for $x = (x_1,\dots,x_m)'$,

$$\int_{\mathbb{R}^d} K_l(y,x)^2 dy = 2^{ld}\sum_{k_1,\dots,k_m\in\mathbb{Z}}\prod_{1\leq m\leq d}\phi(2^l x_m - k_m)^2 \geq 2^{ld}c,$$

where the equality follows from orthonormality in $L^2(\mathbb{R})$ of the system $\{\phi(\cdot - k), k\in\mathbb{Z}\}$ and the inequality follows from Lemma E.1. Therefore, for sufficiently large n, so that $l \geq l_{\min,n} \geq c_5 \log n$ is sufficiently large, we have $\mathrm{E}_f[K_l(X_1,x)^2] \geq 2^{ld}c$. Hence, for sufficiently large n, (E.2) holds for all $f \in \mathcal{F}$ and $l \in \mathcal{L}_n$ and $\sigma_n \leq C$. 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 [14] (for convolution kernels), Lemma 2 in [17] (for compactly supported wavelets), and Lemma 2 in [20] (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 $b_n \leq C2^{l_{\max,n}d/2}$, $a_n \leq C$, and $v_n \leq C$. This completes the proof of the lemma. ∎

Appendix J. 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 (3.17) and (3.18). 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 $\sup_{x\in\mathcal{X}} |\mathrm{E}_f[K_l(X_1, x)] - f(x)| \leq C$ uniformly over $f \in \mathcal{F}$ and $l \in \mathcal{L}_n$ for some $C > 0$. The last assumption holds if $\sup_{x\in\mathcal{X}} |\mathrm{E}_f[K_l(X_1, x)] - f(x)| \to 0$ as $l \to \infty$ 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

$$cm \leq \sum_{j=1}^{M} \varphi_j(x)^2 \leq Cm, \tag{J.1}$$

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

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

and the last term in (J.2) is bounded from below by some absolute constant yielding the lower bound in (J.1). Therefore, $|K_l(y, x)| \leq C2^l$, $|\mathrm{E}_f[K_l(X_1, x)]| \leq C$ and

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

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 $x_1, x_2, y \in \mathcal{X}$,

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

Therefore, it follows from Example 19.7 in [40] 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 $\mathrm{VC}(b_n, a_n, v_n)$ 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 (3.17) and (3.19). 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 $\sup_{x\in\mathcal{X}} |\mathrm{E}_f[K_l(X_1,x)] - f(x)| \leq C$ 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, [32]. Therefore, under our assumptions, $|K_l(y,x)| \leq C2^{2l}$, $|\mathrm{E}_f[K_l(X_1,x)]| \leq C$, and $\mathrm{E}_f[K_l(X_1,x)^2] \leq C2^{2l}$ uniformly over $f \in \mathcal{F}$, $l \in \mathcal{L}_n$, and $x \in \mathcal{X}$. Assume also that $\mathrm{E}_f[K_l(X_1,x)^2] \geq c2^{2l}$ uniformly over $f \in \mathcal{F}$, $l \in \mathcal{L}_n$, and $x \in \mathcal{X}$. Given the upper bound on $\mathrm{E}_f[K_l(X_1,x)^2]$ 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 $\mathrm{VC}(b_l, a_l, v_l)$ 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 $\mathrm{VC}(b_n, a_n, v_n)$ 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 $C > 0$.

(V. Chernozhukov) Department of Economics and Operations Research Center, MIT, 50 Memorial Drive, Cambridge, MA 02142, USA.
E-mail address: vchern@mit.edu

(D. Chetverikov) Department of Economics, UCLA, Bunche Hall 8283, 315 Portola Plaza, Los Angeles, CA 90095, USA.
E-mail address: chetverikov@econ.ucla.edu

(K. Kato) Graduate School of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan.
E-mail address: kkato@e.u-tokyo.ac.jp