Adaptation Bounds for Confidence Bands under Self-Similarity

Timothy B. Armstrong*
Yale University

October 23, 2018

Abstract

We derive bounds on the scope for a confidence band to adapt to the unknown regularity of a nonparametric function that is observed with noise, such as a regression function or density, under the self-similarity condition proposed by Giné and Nickl (2010). We find that adaptation can only be achieved up to a term that depends on the choice of the constant used to define self-similarity, and that this term becomes arbitrarily large for conservative choices of the self-similarity constant. We construct a confidence band that achieves this bound, up to a constant term that does not depend on the self-similarity constant. Our results suggest that care must be taken in choosing and interpreting the constant that defines self-similarity, since the dependence of adaptive confidence bands on this constant cannot be made to disappear asymptotically.

1 Introduction

Consider the problem of constructing a confidence band for a function that is observed with noise, such as a regression function or density. It will be convenient to state our results in the white noise model

$$Y(t) = \int_0^t f(s)\, ds + \sigma_n W(t), \quad \sigma_n = \sigma/\sqrt{n}$$

*email: timothy.armstrong@yale.edu.

which maps to the regression or density setting with n playing the role of sample size (Brown and Low, 1996; Nussbaum, 1996). Here $f : \mathbb{R} \to \mathbb{R}$ is an unknown function, $W(t)$ is a standard Brownian motion and $Y(t)$ is observed with σ_n treated as known. To obtain good estimates and confidence bands, one must impose some regularity on the function f. This is typically done by assuming that f is in a derivative smoothness class, such as the Hölder class $\mathcal{F}_{\text{Höl}}(\gamma, B)$, which formalizes the notion that the γth derivative is bounded by B:

$$\mathcal{F}_{\text{Höl}}(\gamma, B) = \{f\colon \text{ for all } t, t' \in \mathbb{R}, |f^{(\lfloor\gamma\rfloor)}(t) - f^{(\lfloor\gamma\rfloor)}(t')| \leq B|t - t'|^{\gamma - \lfloor\gamma\rfloor}\}$$

where $\lfloor\gamma\rfloor$ denotes the greatest integer strictly less than γ. We are interested in constructing a confidence band for f on an interval, which we take to be $[0, 1]$. A confidence band is a collection of random intervals $\mathcal{C}_n(x) = \mathcal{C}_n(x; Y)$ for $x \in [0, 1]$ that depend on the data Y observed at noise level $\sigma_n = \sigma/\sqrt{n}$. Following the standard definition, we say that $\mathcal{C}_n(\cdot)$ is a confidence band with coverage $1 - \alpha$ over the class $\mathcal{F}$ if

$$\inf_{f \in \mathcal{F}} P_f \left(\text{for all } x \in [0, 1],\ f(x) \in \mathcal{C}_n(x)\right) \geq 1 - \alpha \tag{1}$$

where P_f denotes probability when $Y(t)$ is drawn according to f. Although we focus on the interval $[0, 1]$, to avoid boundary issues, we will assume that $Y(t)$ is observed on the entire real line (our results will also hold if $Y(t)$ is observed on an open set containing $[0, 1]$).

Using knowledge of the class $\mathcal{F}_{\text{Höl}}(\gamma, B)$, one can construct estimators and confidence bands that are near-optimal in a minimax sense. In practice, however, it can be difficult to specify γ and B a priori. This has lead to the paradigm of adaptation: one seeks estimators and confidence bands that are nearly optimal for all γ and B in some range without a priori knowledge of γ or B. Such procedures are called "adaptive." Unfortunately, while it is possible to construct estimators that adapt to the unknown value of γ and B, (see Tsybakov, 1998, and references therein), it follows from Low (1997) that adaptive confidence band construction over derivative smoothness classes is impossible.

To recover the possibility of adaptive confidence band construction, Giné and Nickl (2010) propose an additional condition known as "self-similarity" (see also Picard and Tribouley, 2000), which uses a constant $\varepsilon > 0$ to rule out functions such that the level of regularity is statistically difficult to detect. Imposing these additional conditions leads to a class $\mathcal{F}_{\text{self-sim}}(\gamma, B, \varepsilon) \subsetneq \mathcal{F}_{\text{Höl}}(\gamma, B)$. Giné and Nickl (2010) derive confidence bands that are rate-adaptive to the unknown parameter γ over these smaller classes, and they show that the set $\mathcal{F}_{\text{Höl}}(\gamma, B) \setminus \cup_{\varepsilon>0} \mathcal{F}_{\text{self-sim}}(\gamma, B, \varepsilon)$ of functions ruled out by this assumption (as $\varepsilon \to 0$) is small in a certain topological sense. A subsequent literature has further examined the use of self-similarity and related assumptions in forming adaptive confidence bands (see references below).

These results provide a promising approach to constructing a confidence band such that the width reflects the unknown regularity γ of the function f. However, these confidence bands require a priori knowledge of other regularity parameters, including ε, either explicitly or through unspecified constants and sequences that must be chosen in a way that depends on ε in order to guarantee coverage for a given sample size or noise level. Furthermore, these choices have a first order asymptotic effect on the width of the confidence band, and making an asymptotically conservative choice by taking $\varepsilon = \varepsilon_n \to 0$ leads to a slightly slower rate of convergence. This has led to some concern about whether self-similarity assumptions can lead to a "practical" approach to confidence band construction (see, for example, the discussion on pp. 2388-2389 of Hoffmann and Nickl, 2011): while self-similarity removes the need to specify the order γ of the derivative, currently available methods still require specifying other regularity parameters. Can one construct a confidence band that is fully adaptive without specifying any of the regularity parameters γ, B or ε?

An implication of the results in this paper is that it is impossible to achieve such a goal. In particular, we show that a confidence band that is adaptive over classes $\mathcal{F}_{\text{self-sim}}(\gamma, B, \varepsilon)$ over a range of γ or B must necessarily pay an adaptation penalty proportional to $\varepsilon^{-1/(2\gamma+1)}$. As a consequence, adaptive confidence bands in self-similarity classes require explicit specification of the self-similarity constant ε, and taking $\varepsilon = \varepsilon_n \to 0$ requires paying a penalty in the rate. On the other hand, once ε is given, one can construct a confidence band that is "practical" in the sense that it is valid for a fixed sample size or noise level in Gaussian settings, and it does not depend on additional unspecified constants or sequences once ε is given.

To describe these results formally, let $\mathcal{I}_{n,\alpha,\mathcal{F}}$ denote the set of confidence bands that satisfy the coverage requirement (1). Subject to this coverage requirement, we compare worst-case length of $\mathcal{C}_n$ over a possibly smaller class $\mathcal{G}$. Letting $\text{length}(\mathcal{A}) = \sup \mathcal{A} - \inf \mathcal{A}$ denote the length of a set $\mathcal{A}$, let

$$R_\beta(\mathcal{C}_n; \mathcal{G}) = \sup_{f \in \mathcal{G}} q_{\beta,f} \left(\sup_{x \in [0,1]} \text{length}(\mathcal{C}_n(x)) \right)$$

where $q_{\beta,f}$ denotes the β quantile when $Y \sim f$. Following Cai and Low (2004), define

$$R^*_{n,\alpha,\beta}(\mathcal{G},\mathcal{F}) = \inf_{\mathcal{C}_n(\cdot)\in\mathcal{I}_{n,\alpha,\mathcal{F}}} R_\beta(\mathcal{C}_n;\mathcal{G})$$

to be the optimal worst-case length over $\mathcal{G}$ of a band with coverage over $\mathcal{F}$, where $\mathcal{G} \subseteq \mathcal{F}$. A minimax confidence band over the set $\mathcal{F}$ is one that achieves the bound $R^*_{n,\alpha,\beta}(\mathcal{F},\mathcal{F})$. Given a family $\mathcal{F}(j)$ of function classes indexed by a regularity parameter $j \in \mathcal{J}$, the goal of adaptive confidence band construction is to find a single confidence band $\mathcal{C}_n(\cdot)$ that is close to achieving this bound for each $\mathcal{F}(j)$, while also maintaining coverage $1-\alpha$ for each $\mathcal{F}(j)$ (so that $\mathcal{C}_n(\cdot) \in \mathcal{I}_{n,\alpha,\cup_{j\in\mathcal{J}}\mathcal{F}(j)}$). Suppose that a confidence band $\mathcal{C}_n(\cdot)$ achieves this goal up to a factor $A_n(j)$:

$$\mathcal{C}_n(\cdot) \in \mathcal{I}_{n,\alpha,\cup_{j\in\mathcal{J}}\mathcal{F}(j)} \quad \text{and} \quad R_\beta(\mathcal{C}_n;\mathcal{F}(j)) \le A_n(j) R^*_{n,\alpha,\beta}(\mathcal{F}(j),\mathcal{F}(j)) \quad \text{all} \quad j \in \mathcal{J}.$$

We will call such a band *adaptive* to j up to the *adaptation penalty* $A_n(j)$. If the adaptation penalty is bounded as a function of n, we will say that the confidence band is (rate) adaptive (this corresponds to what Cai and Low (2004) call "strongly adaptive"). Note that $R^*_{n,\alpha,\beta}(\mathcal{F}(j),\cup_{j\in\mathcal{J}}\mathcal{F}(j))/R^*_{n,\alpha,\beta}(\mathcal{F}(j),\mathcal{F}(j))$ provides a lower bound for the adaptation penalty of any confidence band $\mathcal{C}_n(\cdot)$.

For Hölder classes, $R^*_{n,\alpha,\beta}(\mathcal{F}_{\text{Höl}}(\gamma,B),\mathcal{F}_{\text{Höl}}(\gamma,B))$ decreases at the rate $(n/\log n)^{-\gamma/(2\gamma+1)}$. A confidence band that is rate adaptive to γ would achieve this rate simultaneously for all γ in some set $[\underline{\gamma},\overline{\gamma}]$ while maintaining coverage over $\cup_{\gamma\in[\underline{\gamma},\overline{\gamma}]}\mathcal{F}_{\text{Höl}}(\gamma,B)$. However, as noted above, the results of Low (1997) imply that this is impossible. Indeed, $R^*_{n,\alpha,\beta}(\mathcal{F}_{\text{Höl}}(\gamma,B),\cup_{\gamma'\in[\underline{\gamma},\overline{\gamma}]}\mathcal{F}_{\text{Höl}}(\gamma',B))$ decreases at the rate $(n/\log n)^{-\underline{\gamma}/(2\underline{\gamma}+1)}$ for each $\gamma \in [\underline{\gamma},\overline{\gamma}]$, so the adaptation penalty for Hölder classes is of order $(n/\log n)^{\gamma/(2\gamma+1)-\underline{\gamma}/(2\underline{\gamma}+1)}$, which is quite severe.

To salvage the possibility of adaptation, Giné and Nickl (2010) proposed augmenting the Hölder condition with an auxiliary condition. We will focus here on the version of this condition based on convolution kernels. Let $r = \lfloor\gamma\rfloor$, and let K be a convolution kernel with finite support, satisfying Condition 1(a) in Giné and Nickl (2010):

K is symmetric, integrable of bounded variation with finite support and $\int_{\mathbb{R}} K(u)\,du = 1$, $\int_{\mathbb{R}} u^\ell K(u)\,du = 0$ all $\ell = 1,\ldots,r$. (2)

The last condition means that K is a kernel of order (at least) r. The kernel estimate of $f(x_0)$ with bandwidth h, given by $\hat{f}(x_0,h) = \int \frac{1}{h}K((x-x_0)/h)\,dY(x_0)$, has bias $\int \frac{1}{h}K((x-x_0)/h)f(x)\,dx - f(x_0)$. An upper bound on this bias for functions in $\mathcal{F}_{\text{Höl}}(\gamma,B)$ follows from standard calculations (see Lemma 3.4):

$$\sup_{x_0\in[0,1]}\left|\int \frac{1}{h}K((x-x_0)/h)f(x)\,dx - f(x_0)\right| \le \overline{C}_{K,\gamma}Bh^{\gamma} \tag{3}$$

where $\overline{C}_{K,\gamma} = \frac{1}{(r-1)!}\int |K(u)|\int_{s=0}^1 |u|^{\gamma}s^{\gamma-r}(1-s)^{r-1}\,ds\,du$ for $r \ge 1$ and $\overline{C}_{K,\gamma} = \int |K(u)||u|^{\gamma}\,du$ for $r = 0$. Giné and Nickl (2010) impose such a bound on bias directly, along with an analogous lower bound. For K satisfying (2) with $r = \lfloor \gamma \rfloor$ and $\overline{h}, b_1, b_2 > 0$, let $\mathcal{F}_{\text{GN}}(\gamma, b_1, b_2) = \mathcal{F}_{\text{GN}}(\gamma, b_1, b_2; K, \overline{h})$ denote the set of functions f satisfying Condition 3 of Giné and Nickl (2010): for all $h \le \overline{h}$,

$$b_1h^{\gamma} \le \sup_{x_0\in[0,1]}\left|\int \frac{1}{h}K((x-x_0)/h)f(x)\,dx - f(x_0)\right| \le b_2h^{\gamma}. \tag{4}$$

Since we will also be imposing Hölder conditions, which, as noted above, satisfy the upper bound with $b_2 = \overline{C}_{K,\gamma}B$, it is natural to make the lower bound proportional to B as well, by taking $b_1 = \varepsilon B$ for some $\varepsilon > 0$. To this end, let $\mathcal{F}_{\text{self-sim}}(\gamma, B, \varepsilon) = \mathcal{F}_{\text{self-sim}}(\gamma, B, \varepsilon; K, \overline{h})$ be the set of functions in $\mathcal{F}_{\text{Höl}}(\gamma, B)$ such that the lower bound in (4) holds with $b_1 = \varepsilon B$ for all $h \le \overline{h}$. By the discussion above, this is equivalent to defining $\mathcal{F}_{\text{self-sim}}(\gamma, B, \varepsilon; K, \overline{h}) = \mathcal{F}_{\text{Höl}}(\gamma, B) \cap \mathcal{F}_{\text{GN}}(\gamma, \varepsilon B, CB; K, \overline{h})$ for any $C \ge \overline{C}_{K,\gamma}$. We will refer to ε as a "self-similarity constant," and we will call the class $\mathcal{F}_{\text{self-sim}}$ a "self-similarity class." Note that, by defining ε to be (up to a constant) the ratio of the upper and lower bounds on the bias, we are separating the role of self-similarity and the smoothness constant. In particular, the self-similarity constant is scale invariant. See Section 2.3 for alternative formulations of the notion of a "self-similarity constant."

Our main results are efficiency bounds that have implications for the adaptation penalty $A_n(\gamma, B)$ for confidence bands that adapt to the regularity parameters (γ, B) over a rich enough set $\mathcal{J}$ in the self-similarity class $\mathcal{F}_{\text{self-sim}}(\varepsilon, \gamma, B)$. In particular, our results imply the existence of a constant $C_* > 0$ such that, for large enough n, the adaptation penalty for any confidence band must satisfy the lower bound $C_*\varepsilon^{-1/(2\gamma+1)} < A_n(\gamma, B)$. Furthermore, we construct a confidence band with adaptation penalty $A_n(\gamma, B) < C^*\varepsilon^{-1/(2\gamma+1)}$, where $C^* < \infty$ (the constants C_* and C^* do not depend on ε but may depend on the set $\mathcal{J}$ over which adaptation is required). For the lower bounds, we consider separately the cases of adaptation to B with γ known (i.e. $\mathcal{J} = \gamma \times [\underline{B}, \overline{B}]$) and adaptation to γ with B known (i.e. $\mathcal{J} = [\underline{\gamma}, \overline{\gamma}] \times B$). In both cases, the lower bound gives the same $\varepsilon^{-1/(2\gamma+1)}$ term. We also consider the possibility of "adapting to the self-similarity constant" and find that that this is not possible: if we allow ε to be in some set $[\underline{\varepsilon}, \overline{\varepsilon}]$, then we obtain a lower bound proportional to $\underline{\varepsilon}^{-1/(2\gamma+1)}$.

To our knowledge, this paper is the first to derive lower bounds for adaptative confidence bands under self-similarity conditions. A related question, addressed by Hoffmann and Nickl (2011) and Bull (2012), is whether the self-similarity conditions themselves can be weakened. Regarding upper bounds, our results relate to the literature deriving confidence bands under self-similarity conditions. Giné and Nickl (2010) propose a confidence band that has coverage over $f \in \mathcal{F}_{\text{self-sim}}(\varepsilon_n, \gamma, B)$ for a range of (γ, B), where $\varepsilon_n \to 0$ with the sample size, and they show that it is adaptive up to a penalty $A_n(\gamma, B)$ where $A_n(\gamma, B) \to \infty$ slowly with the sample size n. Our lower bounds show that a penalty of this form is unavoidable if one takes $\varepsilon_n \to 0$. Bull (2012) and Chernozhukov et al. (2014) propose confidence bands have coverage over self-similarity classes with ε fixed, and they show that these confidence bands are fully rate adaptive (i.e. the adaptation penalty $A_n(\gamma, B)$ is bounded as n increases). Checking whether the adaptation penalty for these confidence bands takes the optimal form $C^*\varepsilon^{-1/(2\gamma+1)}$ for small ε appears to be difficult, and we derive upper bounds using a different confidence band (although the confidence band we propose builds on ideas in these papers; see Section 2.4).

2 Adaptation Bounds for Self-Similar Functions

This section states our main results. We first give lower bounds for adaptation, separating the role of adaptation to the constant B and the exponent γ. We then construct a confidence band that achieves these bounds, up to a constant that does not depend on the self-similarity constant ε, simultaneously for all γ and B on bounded intervals. Finally, we provide lower bounds for an alternative formulation of the problem, and a discussion of our results.

2.1 Lower Bounds

We now give bounds for adaptation over the classes $\mathcal{F}_{\text{self-sim}}(\gamma, B, \varepsilon)$. Proofs of the lower bounds in this section are given in Section 3. We first consider adaptation to the constant B.

Theorem 2.1. *Let $\gamma > 0$, $0 < \underline{B} \le B \le \overline{B}$ and let $0 < 2\alpha < \beta < 1$. Let K be a kernel that satisfies $\int K(u)|u|^\gamma \, du \neq 0$ as well as (2). There exist $\overline{h}_K$ depending only on the kernel K as* *well as $C_{K,\gamma,*} > 0$ and $\eta_{K,\gamma} > 0$ depending only on K and γ such that, for any $\varepsilon \le \varepsilon' < \eta_{K,\gamma}$,*

$$
\begin{aligned}
&R^*_{n,\alpha,\beta}\left(\mathcal{F}_{\text{self-sim}}(\gamma, B, \varepsilon'; K, \overline{h}_K), \cup_{B' \in [\underline{B}, \overline{B}]} \mathcal{F}_{\text{self-sim}}(\gamma, B', \varepsilon; K, \overline{h}_K)\right) \\
&\ge (1 + o(1)) C_{K,\gamma,*} \min\{\varepsilon^{-1} B, \overline{B}\}^{\frac{1}{2\gamma+1}} \left(\sigma_n^2 \log(1/\sigma_n)\right)^{\gamma/(2\gamma+1)}.
\end{aligned}
$$

We now consider adaptation to γ with B known. To avoid notational clutter, we normalize B to one.

Theorem 2.2. *Let $0 < \underline{\gamma} < \gamma \le \overline{\gamma} \le 1$ and let $0 < 2\alpha < \beta < 1$. Let K be a nonnegative kernel that satisfies (2). There exist $C_{K,*} > 0$, $\overline{h}_K > 0$ and $\eta_K > 0$ depending only on the kernel K such that, for all $0 < \varepsilon \le \varepsilon' \le \eta_K$,*

$$
\begin{aligned}
&R^*_{n,\alpha,\beta}\left(\mathcal{F}_{\text{self-sim}}(\gamma, 1, \varepsilon'; K, \overline{h}_K), \cup_{\gamma' \in [\underline{\gamma}, \overline{\gamma}]} \mathcal{F}_{\text{self-sim}}(\gamma', 1, \varepsilon; K, \overline{h}_K)\right) \\
&\ge (1 + o(1)) C_{K,*} \varepsilon^{\frac{-1}{2\gamma+1}} \left(\sigma_n^2 \log(1/\sigma_n)\right)^{\gamma/(2\gamma+1)}.
\end{aligned}
$$

It follows from Theorems 2.1 and 2.2 that adaptive confidence bands must pay an adaptation penalty proportional to $\varepsilon^{-1/(2\gamma+1)}$. Furthermore, these results show that one cannot "adapt to the self-similarity constant:" if we require coverage for ε-self-similarity, then the adaptation penalty is proportional to $\varepsilon^{-1/(2\gamma+1)}$, even for functions that are ε'-self-similar with $\varepsilon' > \varepsilon$.

Note that both of these results place additional on the kernel K. Theorem 2.1 imposes the condition $\int K(u)|u|^\gamma \, du \ne 0$. Theorem 2.2 requires a positive kernel, and restricts attention to $\gamma \le 1$. Such assumptions are helpful in constructing functions that satisfy the lower bound in (4), since they guarantee that functions that behave like $x \mapsto |x - x_0|^\gamma$ for some x_0 satisfy this bound. Given that the lower bound in (4) depends heavily on the kernel K (see the discussion on pp. 1134-1135 of Giné and Nickl, 2010), it appears difficult to proceed without some conditions along these lines.

2.2 Achieving the Bound

We now turn to upper bounds. Both of these bounds can be achieved simultaneously for all $\gamma \in [\underline{\gamma}, \overline{\gamma}]$ and $B \in [\underline{B}, \overline{B}]$ by a single confidence band, up to an additional term that depends only on K and the range $[\underline{\gamma}, \overline{\gamma}]$. We first state the upper bound, and then describe the confidence band that achieves it. Let $\|K\| = \sqrt{\int K(t)^2 \, dt}$ denote the L_2 norm of K.

Theorem 2.3. *Let $0 < \underline{B} < \overline{B}$ and $0 < \underline{\gamma} < \overline{\gamma}$ be given, and let K be a kernel that satisfies (4). Let $\overline{C}_K = \sup_{\gamma\in(0,\infty)} \overline{C}_{K,\gamma}$ where $\overline{C}_{K,\gamma}$ is the constant defined after (3) (note that $\overline{C}_K$ is finite). There exists a confidence band $\mathcal{C}_n(\cdot)$ and a sequence $\delta_n \to 0$ such that, with probability converging to one uniformly over $\cup_{\gamma\in[\underline{\gamma},\overline{\gamma}]} \cup_{B\in[\underline{B},\overline{B}]} \mathcal{F}_{\text{self-sim}}(\varepsilon,\gamma,B)$, $\sup_{x\in[0,1]} \text{length}(\mathcal{C}_n(x))$ is bounded by $(1+\delta_n)$ times*

$$2\left(\frac{\log n}{n}\right)^{\gamma/(2\gamma+1)} \left(\frac{2\sigma^2\|K\|^2}{2\gamma+1}\right)^{\gamma/(2\gamma+1)} (\overline{C}_K^2 B/\varepsilon)^{1/(2\gamma+1)} \left[(2\gamma)^{1/(2\gamma+1)} + (2\gamma)^{-2\gamma/(2\gamma+1)}\right]$$

and $f(x) \in \mathcal{C}_n(x)$ all $x \in [0,1]$.

We will construct a confidence band such that the conclusions of this theorem hold with $\mathcal{F}_{\text{self-sim}}(\gamma,B,\varepsilon)$ replaced by $\mathcal{F}_{\text{GN}}(\gamma,\varepsilon B,\overline{C}_K B)$. Since $\mathcal{F}_{\text{self-sim}}(\gamma,B,\varepsilon) \subseteq \mathcal{F}_{\text{GN}}(\gamma,\varepsilon B,\overline{C}_K B)$ for each γ and B, this gives a stronger result.

Let $\hat{f}(x,h) = \int \frac{1}{h}K((t-x)/h)\,dY(t)$ and let $K_h f(x) = \int \frac{1}{h}K((t-x)/h)f(t)\,dt$. Let $\Delta(h,h';f) = \sup_{x\in[0,1]} |K_h f(x) - K_{h'}f(x)|$ and $\hat{\Delta}(h,h') = \sup_{x\in[0,1]} |\hat{f}(x,h) - \hat{f}(x,h')|$. For $f \in \mathcal{F}_{\text{GN}}(\gamma,\varepsilon B,\overline{C}_K B)$, we can obtain a bound on the bias using $\Delta(h_1,h_2)$. First, note that

$$\begin{aligned}
&\overline{C}_K B[(\varepsilon/\overline{C}_K)h_1^\gamma - h_2^\gamma] \le \sup_{x\in[0,1]} |K_{h_1}f(x) - f(x)| - \sup_{x\in[0,1]} |K_{h_2}f(x) - f(x)| \le \Delta(h_1,h_2;f) \\
&\le \sup_{x\in[0,1]} |K_{h_1}f(x) - f(x)| + \sup_{x\in[0,1]} |K_{h_2}f(x) - f(x)| \le \overline{C}_K B[h_1^\gamma + h_2^\gamma]
\end{aligned} \tag{5}$$

where the second and third inequalities are applications of the triangle inequality. For $0 < \gamma_\ell < \gamma_u$, define

$$a(\varepsilon,h_1,h_2,h,\gamma_\ell,\gamma_u) = \max\left\{(\varepsilon/\overline{C}_K)\min\{(h_1/h)^{\gamma_u},(h_1/h)^{\gamma_\ell}\} - \max\{(h_2/h)^{\gamma_u},(h_2/h)^{\gamma_\ell}\}, 0\right\}.$$

If $\gamma_\ell \le \gamma \le \gamma_u$ and $a(\varepsilon,h_1,h_2,h,\gamma_\ell,\gamma_u) > 0$, then $a(\varepsilon,h_1,h_2,h,\gamma_\ell,\gamma_u) \le \frac{(\varepsilon/\overline{C}_K)h_1^\gamma - h_2^\gamma}{h^\gamma}$ so that, for any $f \in \mathcal{F}_{\text{GN}}(\gamma,\varepsilon B,\overline{C}_K B)$,

$$|K_h f(x) - f(x)| \le \overline{C}_K B h^\gamma \le \overline{C}_K B \frac{(\varepsilon/\overline{C}_K)h_1^\gamma - h_2^\gamma}{a(\varepsilon,h_1,h_2,h,\gamma_\ell,\gamma_u)} \le \frac{\Delta(h_1,h_2;f)}{a(\varepsilon,h_1,h_2,h,\gamma_\ell,\gamma_u)} \tag{6}$$

where the last inequality uses (5).

The upper bound on the bias in Equation (6) allows the construction of a confidence band for f from a confidence interval $[\hat{\gamma}_\ell,\hat{\gamma}_u]$ for γ along with confidence bands for $K_h f(x)$ and $\Delta(h_1,h_2)$. If these confidence bands are uniform over the bandwidths h, h_1 and h_2 as well, then we can minimize the width of the band over these bandwidths. Let $c(h)$ and $c(h, h')$ be critical values satisfying

$$|\hat{f}(x, h) - K_h f(x)| \leq c(h) \text{ all } x \in [0, 1],\ h \in (0, \overline{h}] \tag{7}$$

and

$$|\hat{\Delta}(h, h') - \Delta(h, h'; f)| \leq \tilde{c}(h, h') \text{ all } h, h' \in (0, \overline{h}] \tag{8}$$

with some prespecified probability for all $f \in \cup_{\gamma \in [\underline{\gamma}, \overline{\gamma}]} \cup_{B \in [\underline{B}, \overline{B}]} \mathcal{F}_{\text{GN}}(\gamma, \varepsilon B, \overline{C}_K B)$. Let $\hat{h}$, $\hat{h}_1$ and $\hat{h}_2$ be data dependent bandwidths that are contained in $(0, \overline{h}]$ with probability one. It follows from (6) that, on the event that $\gamma \in [\hat{\gamma}_\ell, \hat{\gamma}_u]$ and (7) and (8) both hold, the band

$$\hat{f}(x, \hat{h}) \pm \left[c(\hat{h}) + \frac{\hat{\Delta}(\hat{h}_1, \hat{h}_2) + \tilde{c}(\hat{h}_1, \hat{h}_2)}{a(\varepsilon, \hat{h}_1, \hat{h}_2, \hat{h}, \hat{\gamma}_\ell, \hat{\gamma}_u)} \right]$$

contains $f(x)$ for all $x \in [0, 1]$. Since the bandwidths can be data dependent, we can simply choose them to be approximate minimizers of the length of this band. Let $\hat{h}$, $\hat{h}_1$ and $\hat{h}_2$ be chosen so that

$$c(\hat{h}) + \frac{\hat{\Delta}(\hat{h}_1, \hat{h}_2) + \tilde{c}(\hat{h}_1, \hat{h}_2)}{a(\varepsilon, \hat{h}_1, \hat{h}_2, \hat{h}, \hat{\gamma}_\ell, \hat{\gamma}_u)} \leq (1 + \eta_n) \inf_{h, h_1, h_2 \in (0, \overline{h}]} \left[c(h) + \frac{\hat{\Delta}(h_1, h_2) + \tilde{c}(h_1, h_2)}{a(\varepsilon, h_1, h_2, h, \hat{\gamma}_\ell, \hat{\gamma}_u)} \right],$$

where η_n is a sequence of positive constants converging to zero, and we use the convention that $\frac{\hat{\Delta}(h_1, h_2) + \tilde{c}(h_1, h_2)}{a(\varepsilon, h_1, h_2, h, \hat{\gamma}_\ell, \hat{\gamma}_u)}$ is equal to $+\infty$ if $a(\varepsilon, h_1, h_2, h, \hat{\gamma}_\ell, \hat{\gamma}_u) = 0$, so that the infimum is only over h, h_1, h_2 such that $a(\varepsilon, h_1, h_2, h, \hat{\gamma}_\ell, \hat{\gamma}_u) > 0$. The half-length of this band is then bounded by

$$(1 + \eta_n) \inf_{h, h_1, h_2 \in (0, \overline{h}]} \left[c(h) + \frac{\overline{C}_K B(h_1^\gamma + h_2^\gamma) + 2\tilde{c}(h_1, h_2)}{a(\varepsilon, h_1, h_2, h, \hat{\gamma}_\ell, \hat{\gamma}_u)} \right] \tag{9}$$

on the event that $\gamma \in [\hat{\gamma}_\ell, \hat{\gamma}_u]$ and (7) and (8) both hold (here we use the upper bound in (5)).

It remains to choose $c(h)$, $\tilde{c}(h, h')$, $\hat{\gamma}_\ell$ and $\hat{\gamma}_u$. By Theorem 2.1 of Dumbgen and Spokoiny

(2001), there exists a random variable Z (which is finite almost surely) such that

$$\begin{aligned}&|\hat{f}(x,h) - K_h f(x)| \\ &\leq \left\{Z(\log e/(2Rh))^{-1/2}[\log\log(e^e/(2Rh))] + (2\log(1/(2Rh)))^{1/2}\right\} \Big/ \left[\sqrt{nh}/(\sigma\|K\|)\right]\end{aligned}$$

for all $x \in [0,1]$ and $h \in (0, \overline{h}]$ almost surely, where $[-R, R]$ is the support of the kernel K. Since the distribution of Z does not depend on f, we can use it to form a critical value. Let $\tilde{\alpha}_n$ be a given sequence and let q_n denote the $1 - \tilde{\alpha}_n$ quantile of Z. Let $c(h)$ be given by the right hand side of the above display with Z replaced by q_n. For the confidence band for $\Delta(h, h'; f)$, let $\tilde{c}(h, h') = c(h) + c(h')$. It then follows from simple arguments using the triangle inequality that the event (7) implies (8).

To form a confidence interval for γ, let $\tilde{h}_{1,n}$, $\tilde{h}_{2,n}$ and $\tilde{b}_n$ satisfy the conditions of Theorem A.1 in Appendix A (for example, we can take $\tilde{h}_{1,n} = 1/\log n$, $\tilde{h}_{2,n} = 1/(\log n)^2$ and $\tilde{b}_n = \log\log\log n$.). Let $\hat{\gamma} = -\hat{\Delta}(\tilde{h}_{1,n}, \tilde{h}_{2,n})/\log \tilde{h}_{1,n}^{-1}$. Let $\hat{\gamma}_\ell = \hat{\gamma} - \tilde{b}_n/\log \tilde{h}_{1,n}^{-1}$ and let $\hat{\gamma}_u = \hat{\gamma} + \tilde{b}_n/\log \tilde{h}_{1,n}^{-1}$. Then $[\hat{\gamma}_u, \hat{\gamma}_\ell]$ contains γ with probability approaching one uniformly over $f \in \cup_{B\in[\underline{B},\overline{B}],\gamma\in[\underline{\gamma},\overline{\gamma}]}\mathcal{F}_{\text{GN}}(\gamma, \varepsilon B, \overline{C}_K B)$. Furthermore, if q_n increases slowly enough, it can be shown using the upper bound (9) for the width of this confidence band that the bound in Theorem 2.3 holds for this band. We provide details of the calculations in Appendix A, as well as a proof that $[\hat{\gamma}_\ell, \hat{\gamma}_u]$ is a valid confidence band.

2.3 Alternative Definition of Self-Similarity Constant

We have defined $\mathcal{F}_{\text{self-sim}}(\gamma, B, \varepsilon; K, \overline{h}) = \mathcal{F}_{\text{Höl}}(\gamma, B) \cap \mathcal{F}_{\text{GN}}(\gamma, \varepsilon B, \overline{C}_{K,\gamma}B; K, \overline{h})$, where $\overline{C}_{K,\gamma}$ is a constant that is large enough so that the upper bound in (4) is implied by the Hölder condition. With this definition, the self-similarity constant ε gives the ratio between the upper and lower bounds in (4). The coverage condition takes the union of these classes with ε fixed, so that large values of the Hölder constant require proportionally large values of the lower bound.

Alternatively, one could fix the lower bound $b_1 = \varepsilon B$ when taking the union of these classes. This leads to the class $\overline{\mathcal{F}}_{\text{self-sim}}(\gamma, B, b_1) = \mathcal{F}_{\text{Höl}}(\gamma, B) \cap \mathcal{F}_{\text{GN}}(\gamma, b_1, \overline{C}_{K,\gamma}B; K, \overline{h}) = \mathcal{F}_{\text{self-sim}}(\gamma, B, b_1/B)$. Of course, this does not change the conclusion of Theorem 2.2 (adaptation to γ with B fixed) since the formulation of this problem remains the same. For adaptation to B, however, we obtain a different formulation, with coverage required over the class $\cup_{B\in[\underline{B},\overline{B}]}\overline{\mathcal{F}}_{\text{self-sim}}(\gamma, B, b_1) = \mathcal{F}_{\text{Höl}}(\gamma, \overline{B}) \cap \mathcal{F}_{\text{GN}}(\gamma, b_1, \overline{C}_{K,\gamma}B; K, \overline{h}) = \overline{\mathcal{F}}_{\text{self-sim}}(\gamma, \overline{B}, b_1) = \mathcal{F}_{\text{self-sim}}(\gamma, \overline{B}, b_1/\overline{B})$. As the next theorem shows, this leads to a much more negative result:

adaptation to the Hölder constant is completely impossible.

Theorem 2.4. *Under the conditions of Theorem 2.1, there exist $\overline{h}_K$ depending only on the kernel K as well as $C_{K,\gamma,*} > 0$ and $\eta_{K,\gamma} > 0$ depending only on K and γ such that, for any $b_1 < \eta_{K,\gamma} B$,*

$$\begin{aligned} &R^*_{n,\alpha,\beta}\left(\mathcal{F}_{\text{Höl}}(\gamma, B) \cap \mathcal{F}_{\text{GN}}(\gamma, b_1, \overline{C}_{K,\gamma} B; K, \overline{h}_K), \mathcal{F}_{\text{Höl}}(\gamma, \overline{B}) \cap \mathcal{F}_{\text{GN}}(\gamma, b_1, \overline{C}_{K,\gamma} \overline{B}; K, \overline{h}_K)\right) \\ &\geq (1 + o(1)) C_{K,\gamma,*} \overline{B}^{\frac{1}{2\gamma+1}} \left(\sigma_n^2 \log(1/\sigma_n)\right)^{\gamma/(2\gamma+1)}. \end{aligned}$$

2.4 Discussion

The confidence band in Section 2.2 follows Bull (2012) and Chernozhukov et al. (2014) in constructing an upper bound on bias and using this to widen the confidence interval (see also Donoho (1994) and Schennach (2015) for applications of this idea in other settings). Our approach builds on this important work. In contrast to these papers, which derive bounds on the bias of an estimator with bandwidth selected using Lepski's method, we bound the bias directly for each bandwidth and use the width of the resulting confidence band to choose the bandwidth (although the two approaches are related, since the bound on the bias ultimately comes from comparisons of estimates at different bandwidths, either explicitly in our approach, or implicitly through the use of Lepski's method to choose the bandwidth). This makes it easier to derive explicit bounds, and it may be needed to get the optimal form $C\varepsilon^{-1/(2\gamma+1)}$ of the adaptation penalty (Bull (2012) and Chernozhukov et al. (2014) show that their procedures are adaptive up to a constant, but do not derive how this constant depends on ε).

An alternative approach to ensuring coverage, used by Giné and Nickl (2010), is undersmoothing, which uses a bandwidth sequence for which variance slightly dominates bias. As noted by Bull (2012) and Chernozhukov et al. (2014), this leads to a slightly slower rate of convergence, so that the confidence band is not fully adaptive. Our lower bounds shed some light on this question: one must always pay an adaptation penalty of order $\varepsilon^{-1/(2\gamma+1)}$ when ε is fixed, which means that letting $\varepsilon = \varepsilon_n \to 0$ requires paying a penalty in the rate. In practice, however, for any given finite sample size n, one only achieves coverage over a class $\mathcal{F}_{\text{self-sim}}$ corresponding to some $\varepsilon_n > 0$; undersmoothed confidence bands choose such a sequence implicitly. To make this transparent, one can explicitly specify ε_n, and report a confidence band that is valid for the given self-similarity constant and noise level, even if the "asymptotic promise" states that $\varepsilon_n \to 0$ (while our arguments do not formally cover the case where $\varepsilon = \varepsilon_n \to 0$, it appears that they could be extended to allow $\varepsilon_n \to 0$ at a slow enough rate).

There has been some discussion in the literature of whether or how self-similarity conditions can lead to a practical approach to constructing confidence bands. If "practical" means that the confidence band should not require the user to choose any regularity constants a priori, then our results show that the answer is "no." On the other hand, if one sees the self-similarity constant as an interpretable object, then we need not be so pessimistic. Indeed, the confidence band we construct is "practical" in the sense that it has valid coverage for a given noise level without relying on conservative constants or sequences (except, perhaps, in forming an initial confidence interval for γ; this can be replaced with a priori bounds $[\underline{\gamma}, \overline{\gamma}]$, although it appears that a consistent estimate is needed to get the optimal form of the adaptation penalty).

It is helpful to contrast the role of self-similarity conditions in our setting with regularity conditions used to construct confidence intervals for the mean of a univariate random variable. To form a non-trivial confidence interval for the mean of a univariate random variable, one must place some conditions on the tails of the distribution (Bahadur and Savage, 1956). One approach is to choose some $\delta > 0$, and assume that the $2 + \delta$ moment is bounded by $1/\delta$. Subject to this coverage requirement, the optimal width of the confidence interval does not depend on δ asymptotically: adding and subtracting the $1 - \alpha/2$ quantile of a normal distribution times the sample standard deviation leads to an asymptotically valid confidence interval regardless of the particular choice of $\delta > 0$. Thus, one can state that this confidence interval is asymptotically valid and optimal under a bounded $2 + \delta$ moment, without worrying about the exact choice of δ. Our results show that this is not the case with self-similarity constants: no single confidence band is asymptotically valid and optimal under ε-self-similarity for all ε.

3 Proofs of Lower Bounds

This section proves Theorems 2.1, 2.2 and 2.4. To obtain these results, we begin with bounds on minimax testing.

3.1 Bounds Based on Minimax Testing

For sets $\mathcal{F}$ and $\mathcal{G}$, let $d_{\text{test}}(\mathcal{F},\mathcal{G})$ denote the maximum difference between minimax power and size of a test of $H_0 : \mathcal{F}$ vs $H_1 : \mathcal{G}$:

$$d_{\text{test}}(\mathcal{F},\mathcal{G}) = \sup_{\phi} \inf_{f\in\mathcal{F},\, g\in\mathcal{G}} |E_g\phi(Y) - E_f\phi(Y)|$$

where E_f denotes expectation under the function f, and the supremum is over all tests ϕ based on Y observed at noise level σ_n (i.e. all measurable functions with range $[0,1]$). The following lemma allows us to obtain bounds on $R^*_{n,\alpha,\beta}$ using bounds on d_{test}. The lemma is essentially Lemma 6.1 in Robins and van der Vaart (2006), with the conclusion of the argument stated nonasymptotically.

Lemma 3.1. *Let α, β and $\tilde{R}$ be given and let $\mathcal{G} \subseteq \mathcal{F}$. Suppose that*

$$\text{for some } f_0 \in \mathcal{G},\, d_{\text{test}}\left(\{f_0\}, \mathcal{F} \cap \{f : \sup_{x\in[0,1]} |f(x) - f_0(x)| \geq \tilde{R}\}\right) < \beta - 2\alpha.$$

Then $R^*_{n,\alpha,\beta}(\mathcal{G},\mathcal{F}) \geq R^*_{n,\alpha,\beta}(\{f_0\},\mathcal{F}) \geq \tilde{R}$.

Proof. Suppose, to get a contradiction, that $R^*_{n,\alpha,\beta}(\{f_0\},\mathcal{F}) < \tilde{R}$. Then there exists a confidence band $\mathcal{C}_n(\cdot) \in \mathcal{I}_{n,\alpha,\mathcal{F}}$ with $R = R_\beta(\mathcal{C}_n;\{f_0\}) = q_{\beta,f_0}\left(\sup_{x\in[0,1]} \text{length}(\mathcal{C}_n(x))\right) < \tilde{R}$, so that

$$P_{f_0}\left(\sup_{x\in[0,1]} \text{length}\left(\mathcal{C}_n(x)\right) > R\right) = 1 - P_{f_0}\left(\sup_{x\in[0,1]} \text{length}\left(\mathcal{C}_n(x)\right) \leq R\right) \leq 1 - \beta. \tag{10}$$

Let us abuse notation slightly and let $\mathcal{C}_n$ denote the set of functions f contained in the confidence band $\mathcal{C}_n(\cdot)$, so that $f \in \mathcal{C}_n$ iff. $f(t) \in \mathcal{C}_n(t)$ all $t \in [0,1]$. Let $\phi = 1$ if there exists a function f satisfying $f \in \mathcal{F} \cap \{f : \sup_{x\in[0,1]} |f(x) - f_0(x)| \geq \tilde{R}\}$ with $f \in \mathcal{C}_n$. It is immediate from the definition of this test and the assumption that $\mathcal{C}_n(\cdot) \in \mathcal{I}_{n,\alpha,\mathcal{F}}$ that

$$\inf_{f\in\mathcal{F}\cap\{f:\sup_{x\in[0,1]}|f(x)-f_0(x)|\geq\tilde{R}\}} E_f\phi \geq 1 - \alpha \tag{11}$$

(i.e. the test has minimax power at least $1-\alpha$ for $H_1 : \mathcal{F} \cap \{f : \sup_{x\in[0,1]} |f(x) - f_0(x)| \geq \tilde{R}\}$).

Now consider the level of the test for $H_0 : \{f_0\}$. We have

$$E_{f_0}\phi(Y) = E_{f_0}\phi(Y)I(f_0 \in \mathcal{C}_n) + E_{f_0}\phi(Y)I(f_0 \notin \mathcal{C}_n) \leq E_{f_0}\phi(Y)I(f_0 \in \mathcal{C}_n) + \alpha$$

by the converage condition. The event $\phi(Y)I(f_0 \in \mathcal{C}_n)$ implies that $\mathcal{C}_n$ contains both f_0 and a function f_1 with $f_1 \in \mathcal{F}$ and $\sup_{x\in[0,1]} |f_1(x) - f_0(x)| \geq \tilde{R}$. This, in turn, implies that $\sup_{x\in[0,1]} \text{length}(\mathcal{C}_n(x)) \geq \tilde{R} > R$ on this event so that, by (10), the probability of this event under f_0 is bounded by $1-\beta$. Thus, by the above display, $E_{f_0}\phi(Y) \leq 1-\beta+\alpha$. Combining this with (11), it follows that $\inf_{f\in\mathcal{F}\cap\{f:\sup_{x\in[0,1]}|f(x)-f_0(x)|\geq\tilde{R}\}} E_f\phi - E_{f_0}\phi \geq 1-\alpha-1+\beta-\alpha = \beta - 2\alpha$, which contradicts the assumptions of the theorem. □

We will use bounds in this testing problem where, for some interval $[a, b] \subseteq [0, 1]$, f_0 and a set of alternative functions $f_{n,1}, \ldots, f_{n,M_n}$ are constructed on $[a, b]$ so that $f_0(x) = 0$ for $x \in [a, b]$ and, for each k, $f_{n,k}$ is in the Hölder class with larger constant or smaller exponent, and $\sup_{x\in[a,b]} |f_{n,k}(x)| = c_n$, where c_n is a sequence converging to zero. This follows arguments in Lepski and Tsybakov (2000). We then extend these functions so that their behavior on another interval ensures self-similarity. For adaptation to the Hölder constant B, we can take the functions to be equal outside of the interval $[a, b]$, so that the result follows immediately.

For the first step, we use the following result, which is immediate from slight modifications of arguments in Lepski and Tsybakov (2000).

Lemma 3.2. *Let a, b, $\underline{\gamma}$, $\overline{\gamma}$, $\underline{B}$ and $\overline{B}$ be given with $a < b$, $0 < \underline{\gamma} \leq \overline{\gamma} < \infty$ and $0 < \underline{B} \leq \overline{B} < \infty$, and let κ be a function with $\kappa \in \mathcal{F}_{\text{Höl}}(1, \gamma)$ for all $\gamma \in [\underline{\gamma}, \overline{\gamma}]$, with $\kappa(0) > 0$ and with finite support. Let $\widetilde{\mathcal{F}}(\gamma, B, a, b)$ denote the class of functions in $\mathcal{F}_{\text{Höl}}(\gamma, B)$ that are equal to zero outside of $[a, b]$. Let $\eta > 0$ be given and let $c_n(\gamma, B) = (1-\eta)C(\gamma, B, \kappa)\left(\sigma_n^2 \log(1/\sigma_n)\right)^{\gamma/(2\gamma+1)}$ where $C(\gamma, B, \kappa) = \left[\frac{4}{2\gamma+1} B^{1/\gamma} \left(\int \kappa(u)^2\, du\right)^{-1}\right]^{\frac{\gamma}{2\gamma+1}} \kappa(0)$. Then*

$$\lim_{n\to\infty} \sup_{\gamma\in[\underline{\gamma},\overline{\gamma}],B\in[\underline{B},\overline{B}]} d_{\text{test}}(\{0\}, \widetilde{\mathcal{F}}(\gamma, B, a, b) \cap \{f : \sup_{x\in[a,b]} |f(x)| = c_n(\gamma, B)\}) = 0.$$

Proof. Let $[-A_\kappa, A_\kappa]$ denote a set containing the support of κ. Following p. 34 of Lepski and Tsybakov (2000), let $C = (1-\eta)C(\gamma, B, \kappa)$, and let

$$h_n = \left(\frac{(1-\eta)C(\gamma, B, \kappa)}{B\kappa(0)}\right)^{1/\gamma} \left(\sigma_n^2 \log(1/\sigma_n)\right)^{1/(2\gamma+1)},$$

$$M_n = \left\lfloor \frac{b-a}{2A_\kappa h_n} \right\rfloor - 1, \quad x_{n,k} = a + (2k-1)A_\kappa h_n, \quad k = 1, \ldots, M_n$$

$$f_{k,n}(x) = Bh_n^\gamma \kappa\left(\frac{x - x_{n,k}}{h_n}\right).$$

By construction, the support of each $f_{k,n}$ is nonoverlapping with and contained in $[a, b]$.

Also, the variance of $\int_a^b f_{k,n}(x)\,dY(x)$ is

$$B^2 h_n^{2\gamma} \int \kappa\left(\frac{x - x_{n,k}}{h_n}\right) dx = B^2 h_n^{2\gamma+1} \int \kappa(u)^2\, du =: s_n^2.$$

The problem of testing between $H_0 : f = 0$ and $H_1 : f \in \{f_{n,1}, f_{n,2}, \ldots, f_{n,M_n}\}$ is therefore equivalent to the problem of testing

$$H_0 : \mu = 0_{M_n} \text{ vs } H_1 : \mu \in \{(s_n/\sigma_n)e_1, \ldots, (s_n/\sigma_n)e_{M_n}\} \tag{12}$$

where e_k is the kth basis vector in $\mathbb{R}^{M_n}$. Since each $f_{k,n}$ is contained in the set $\widetilde{\mathcal{F}}(\gamma, B, a, b) \cap \{f : \sup_{x\in[a,b]} |f(x)| = c_n(\gamma, B)\}$, bounds in this testing problem translate to bounds on $d_{\text{test}}(\{0\}, \widetilde{\mathcal{F}}(\gamma, B, a, b) \cap \{f : \sup_{x\in[a,b]} |f(x)| = c_n(\gamma, B))$.

For n larger than a constant that depends only on $(b-a)/(2A_\kappa h_n)$, we have $M_n \geq (b-a)/(3A_\kappa h_n)$ so that

$$2\log M_n \geq 2\log h_n^{-1} - 2\log[(b-a)/(3A_\kappa)] = \left(\frac{4}{2\gamma+1} + \tilde{K}_n(\gamma, B, \kappa, a, b)\right)\log(1/\sigma_n)$$

where $\tilde{K}_n(\gamma, B, \kappa, a, b)$ is a term with $\sup_{\gamma\in[\underline{\gamma},\overline{\gamma}], B\in[\underline{B},\overline{B}]} \tilde{K}_n(\gamma, B, \kappa, a, b) \to 0$. We have

$$\frac{s_n^2}{\sigma_n^2} = \left[B^2 \int \kappa(u)^2\, du\right] h_n^{2\gamma+1}\sigma_n^{-2} = \left[B^2 \int \kappa(u)^2\, du\right]\left(\frac{(1-\eta)C(\gamma, B, \kappa)}{B\kappa(0)}\right)^{(2\gamma+1)/\gamma} \log(1/\sigma_n)$$
$$= (1-\eta)^{(2\gamma+1)/\gamma}\frac{4}{2\gamma+1}\log(1/\sigma_n).$$

Thus, for δ smaller than a constant that depends only on $\overline{\gamma}$ and $\underline{\gamma}$, we have, for n greater than some constant that depends only on $\overline{\gamma}$, $\underline{\gamma}$, $\overline{B}$, $\underline{B}$, κ, a and b, $(s_n^2/\sigma_n^2)/(2\log M_n) \leq (1-\delta)$. Once this holds, $\sup_{\gamma\in[\underline{\gamma},\overline{\gamma}], B\in[\underline{B},\overline{B}]} d_{\text{test}}(\{0\}, \widetilde{\mathcal{F}}(\gamma, B, a, b) \cap \{f : \sup_{x\in[a,b]} |f(x)| = c_n(\gamma, B))$ is bounded by the largest possible difference between level and minimax power for the testing problem (12) with s_n^2/σ_n^2 given by $2(1-\delta)\log M_n$. This converges to zero by arguments on pp. 35-36 of Lepski and Tsybakov (2000). $\square$

Note that optimizing the function κ for a given γ gives the sharp asymptotic testing constant in Lepski and Tsybakov (2000). In particular, when applying the lemma with a fixed γ, we can let $\kappa = \kappa_\gamma^*$ be the function that solves the optimal recovery problem (4) in Lepski and Tsybakov (2000). (Note that this simplifies the expression since $\int \kappa_\gamma^*(u)^2\, du = 1$ by definition.)

Lemma 3.2 gives a bound for testing $\{0\}$ (the singleton set with the zero function) vs $\widetilde{\mathcal{F}}(\gamma, B, a, b) \cap \{f : \sup_{x\in[a,b]} |f(x)| = c\}$. This is not immediately useful for our purposes, since these sets contain functions that do not satisfy the lower bound required for inclusion in $\mathcal{F}_{\mathrm{GN}}(\gamma, b_1, b_2; K, \overline{h})$ for any $\overline{h}$, γ, b_1 and b_2. Instead, we will consider testing problems in which a function that is zero on $[a, b]$ but sufficiently nonsmooth outside of $[a, b]$ is added to each of these sets. For this, the following lemma will be useful. For a function $f : \mathbb{R} \to \mathbb{R}$, let $\|f\| = \sqrt{\int f(t)^2\, dt}$ denote the L_2 norm of the function f.

Lemma 3.3. *For any functions f_0 and g_0 and sets $\mathcal{F}$ and $\mathcal{G}$,*

$$\begin{aligned}
&d_{\mathrm{test}}(\mathcal{F} + \{f_0\}, \mathcal{G} + \{g_0\}) = d_{\mathrm{test}}(\mathcal{F}, \mathcal{G} + \{g_0 - f_0\}) \\
&\leq d_{\mathrm{test}}(\mathcal{F}, \mathcal{G}) + \sup_{\alpha} \left[\Phi\left(\|f_0 - g_0\|/\sigma_n - z_{1-\alpha}\right) - \alpha\right] \leq d_{\mathrm{test}}(\mathcal{F}, \mathcal{G}) + \|f_0 - g_0\|/\sigma_n.
\end{aligned}$$

Proof. The first equality follows since f_0 can be added or subtracted from Y before performing any test, so that the supremum over tests $\phi(Y)$ is the same as the supremum over tests $\phi(Y - f_0)$. For the first inequality, note that

$$\begin{aligned}
d_{\mathrm{test}}(\mathcal{F}, \mathcal{G} + \{g_0 - f_0\}) &= \sup_{\phi} \inf_{f\in\mathcal{F},\, g\in\mathcal{G}} |E_{g+f_0-g_0}\phi(Y) - E_f\phi(Y)| \\
\leq \sup_{\phi} \inf_{f\in\mathcal{F},\, g\in\mathcal{G}} &\left[|E_{g+f_0-g_0}\phi(Y) - E_g\phi(Y)| + |E_g\phi(Y) - E_f\phi(Y)|\right].
\end{aligned}$$

For any g, the first term is bounded by $\sup_\phi |E_{g+f_0-g_0}\phi(Y) - E_g\phi(Y)|$ which, using the Neyman-Pearson lemma and some calculations (see Example 2.1 in Ingster and Suslina, 2003), can be seen to be equal to

$$\sup_{\alpha} \left[\Phi\left(\|f_0 - g_0\|/\sigma_n - z_{1-\alpha}\right) - \Phi(z_{1-\alpha})\right] \leq \|f_0 - g_0\|/\sigma_n,$$

where the inequality follows from Taylor's theorem, since the derivative of the standard normal cdf is bounded by $1/\sqrt{2\pi} \leq 1$. $\square$

3.2 Constructing Functions in Self-Similarity Classes

We now construct functions contained in the classes $\mathcal{F}_{\mathrm{GN}}(\gamma, b_1, b_2; K, \overline{h})$. Let $\tilde{g}_{0,\gamma,1}(t)$ denote the function that is zero $(-\infty, 0]$ and has $\lfloor\gamma\rfloor$th derivative equal to $t^{\gamma-\lfloor\gamma\rfloor}$ on $[0, 1]$, and $\lfloor\gamma\rfloor$th derivative equal to 1 on $[1,\infty)$:

$$\tilde{g}_{0,\gamma,1}(t)=\begin{cases}0 & t<0\\ \frac{1}{\gamma(\gamma-1)\dots(\gamma-\lfloor\gamma\rfloor+1)}t^{\gamma} & 0\le t<1\\ \frac{1}{\gamma(\gamma-1)\dots(\gamma-\lfloor\gamma\rfloor+1)}t^{\lfloor\gamma\rfloor} & t\ge 1\end{cases}$$

(we define $\gamma(\gamma-1)\dots(\gamma-\lfloor\gamma\rfloor+1)$ to be equal to 1 when $\lfloor\gamma\rfloor=0$). For $t_0\in\mathbb{R}$ and $A>0$, let $\tilde{g}_{t_0,\gamma,A}(t)=A\tilde{g}_{0,\gamma,1}(t-t_0)$. Given $\gamma,\delta,\varepsilon$ and A with $0<\delta<\gamma\le 1$ and $0<\varepsilon\le 1$, let

$$\tilde{f}_{0,\gamma,\delta,\varepsilon,1}(t)=\begin{cases}0 & t<0\\ \max\{t^{\gamma},\varepsilon t^{\gamma-\delta}\} & 0\le t<1\\ 1 & t\ge 1\end{cases}=\begin{cases}0 & t<0\\ \varepsilon t^{\gamma-\delta} & 0<t<\tilde{t}\\ t^{\gamma} & \tilde{t}\le t<1\\ 1 & t\ge 1\end{cases}$$

where $\tilde{t}=\tilde{t}(\varepsilon,\delta)=\varepsilon^{1/\delta}$ and let $\tilde{f}_{t_0,\gamma,\delta,\varepsilon,A}(t)=A\tilde{f}_{0,\gamma,\delta,\varepsilon,1}(t-t_0)$ for any $t_0\in\mathbb{R}$ and $A>0$. Note that $\tilde{g}_{t_0,\gamma,A}\in\mathcal{F}_{\text{Höl}}(\gamma,A)$ and, for $0<\gamma\le 1$, $\tilde{f}_{t_0,\gamma,\delta,\varepsilon,1}\in\mathcal{F}_{\text{Höl}}(\gamma-\delta,A)$. For $\gamma\le 1$, $\tilde{g}_{t_0,\gamma,A}(t)$ and $\tilde{f}_{t_0,\gamma,\delta,\varepsilon,A}$ are equal outside of the set $[t_0,t_0+\varepsilon^{1/\delta}]$.

We now show that adding functions in $\widetilde{\mathcal{F}}(\gamma,B,a,b)$ to these functions gives functions that are in the class $\mathcal{F}_{\text{Höl}}$ and $\mathcal{F}_{\text{GN}}$ for appropriate constants. First, as noted by Giné and Nickl (2010), the upper bound in (4) holds for an appropriate constant for any Hölder continuous function. We record this fact in the following lemma.

Lemma 3.4. *Let K be a kernel that satisfies (2) with $r=\lfloor\gamma\rfloor$. Then, for $f\in\mathcal{F}_{\text{Höl}}(\gamma,B)$, $\overline{h}>0$ and any x_0,*

$$\left|\int\frac{1}{h}K((x-x_0)/h)f(x)\,dx-f(x_0)\right|\le\overline{C}_{K,\gamma}Bh^{\gamma}$$

where $\overline{C}_{K,\gamma}=\frac{1}{(r-1)!}\int|K(u)|\int_{s=0}^{1}|u|^{\gamma}s^{\gamma-r}(1-s)^{r-1}\,ds\,du$ for $r\ge 1$ and $\overline{C}_{K,\gamma}=\int|K(u)||u|^{\gamma}\,du$ for $r=0$.

Proof. In the case where $r\ge 1$,

$$\begin{aligned}&\int\frac{1}{h}K((x-x_0)/h)f(x)\,dx-f(x_0)\\&=\frac{1}{(r-1)!}\int K(u)\int_{s=0}^{1}[f^{(r)}(x_0+suh)-f^{(r)}(x_0)](uh)^r(1-s)^{r-1}\,ds\,du\end{aligned}\tag{13}$$

where we use a Taylor expansion and the assumptions on the kernel (see Equation 1.18, p. 14 in Tsybakov, 2009). If $f \in \mathcal{F}_{\text{Höl}}(\gamma, B)$, then the absolute value of (13) is bounded from above by

$$\frac{1}{(r-1)!} \int |K(u)| \int_{s=0}^{1} B|suh|^{\gamma-r}|uh|^{r}(1-s)^{r-1}\, ds\, du = \overline{C}_{K,\gamma} B h^{\gamma}$$

as required. For $r = 0$, we have $\int \frac{1}{h} K((x-x_0)/h) f(x)\, dx - f(x_0) = \int K(u)(f(uh+x_0) - f(x_0))\, du$ which is bounded in absolute value by $Bh^{\gamma} \int |K(u)||u|^{\gamma}\, du = Bh^{\gamma}\overline{C}_{K,\gamma}$ for $f \in \mathcal{F}_{\text{Höl}}(\gamma, B)$. $\square$

The next two lemmas show that adding $\tilde{g}_{t_0,\gamma,A}$ or $\tilde{f}_{t_0,\gamma,\delta,\varepsilon,A}$ to functions in $\widetilde{\mathcal{F}}(\gamma, B, a, b)$ gives Hölder continuous functions that satisfy the self-similarity condition (4).

Lemma 3.5. *Let $0 \leq a < b < t_0 < 1$ and $\gamma > 0$. Let K be a kernel that satisfies $\int K(u)|u|^{\gamma} \neq 0$ as well as (2). There exists a $\overline{h}_{K,t_0,b}$ depending only on t_0, b and the kernel K, as well as a constant $\underline{C}_{K,\gamma}$ depending only on K and γ such that for any $A > 0$, $B > 0$ and $f \in \widetilde{\mathcal{F}}(\gamma, B, a, b)$,*

$$f + \tilde{g}_{t_0,\gamma,A} \in \mathcal{F}_{\text{Höl}}(\gamma, B+A) \cap \mathcal{F}_{\text{GN}}(\gamma, \underline{C}_{K,\gamma}A, \overline{C}_{K,\gamma}(A+B); K, \overline{h}_{K,t_0,b})$$

where $\overline{C}_{K,\gamma}$ is defined in Lemma 3.4

Proof. Note that $\tilde{g}_{t_0,\gamma,A} \in \mathcal{F}_{\text{Höl}}(\gamma, A)$ and $f \in \mathcal{F}_{\text{Höl}}(\gamma, B)$, which implies $\tilde{g}_{t_0,\gamma,A} + f \in \mathcal{F}_{\text{Höl}}(\gamma, A+B)$. This also gives the upper bound in (4) with $b_2 = \overline{C}_{\gamma,K}(A+B)$, as required. Thus, it remains to verify the lower bound in (4) with $b_1 = \underline{C}_{K,\gamma}A$ for an appropriate constant $\underline{C}_{K,\gamma} > 0$. This follows with $\underline{C}_{K,\gamma} = \int_0^{\infty} K(u)u^{\gamma}\, du/[\gamma(\gamma-1)\ldots(\gamma-\lfloor\gamma\rfloor+1)] = \int K(u)|u|^{\gamma}\, du/[2\gamma(\gamma-1)\ldots(\gamma-\lfloor\gamma\rfloor+1)]$ since, for h small enough so that the support of K is bounded by $\max\{t_0-b, 1-t_0\}/h$,

$$\int \frac{1}{h} K((x-t_0)/h)[f(x) + \tilde{g}_{t_0,\gamma,A}(x)]\, dx - [f(x) + \tilde{g}_{t_0,\gamma,A}(t_0)]$$
$$= \int \frac{1}{h} K((x-t_0)/h)\tilde{g}_{t_0,\gamma,A}(x)\, dx - \tilde{g}_{t_0,\gamma,A}(t_0) = \frac{A}{\gamma(\gamma-1)\ldots(\gamma-\lfloor\gamma\rfloor+1)} \int_0^{\infty} K(u)(uh)^{\gamma}\, du.$$

$\square$

Lemma 3.6. *Let $0 \leq a < b < t_0 < 1$. Let K be nonnegative kernel that satisfies (2). There exists a $\overline{h}_{K,t_0,b}$ depending only on t_0, b and the kernel K, as well as constants $\overline{C}_K$ and $\underline{C}_K$ depending only on K such that the following holds.*

For any $A > 0$, $B > 0$, $\varepsilon > 0$, $0 < \gamma - \delta < \gamma \leq 1$ and $f \in \widetilde{\mathcal{F}}(\gamma, B, a, b)$,

$$f + \tilde{f}_{t_0,\gamma,\delta,\varepsilon,A} \in \mathcal{F}_{\text{Höl}}(\gamma - \delta, A + B) \cap \mathcal{F}_{\text{GN}}(\gamma - \delta, \underline{C}_K A\varepsilon, \overline{C}_K(A + B); K, \overline{h}_{K,t_0,b})$$

and

$$f + \tilde{g}_{t_0,\gamma,A} \in \mathcal{F}_{\text{Höl}}(\gamma, B + A) \cap \mathcal{F}_{\text{GN}}(\gamma, \underline{C}_K A, \overline{C}_K(A + B); K, \overline{h}_{K,t_0,b}).$$

Proof. The upper bound follows from the same arguments as in the proof of Lemma 3.5, since $\tilde{f}_{t_0,\gamma,\delta,\varepsilon,A} \in \mathcal{F}_{\text{Höl}}(\gamma - \delta, A)$, and by noting that the constant $\overline{C}_{K,\gamma}$ in Lemma 3.4 is bounded uniformly over $0 \leq \gamma \leq 1$. For the lower bound, we have, for h small enough that the support of K is bounded by $\max\{t_0 - b, 1 - t_0\}/h$,

$$\begin{aligned}
&\int \frac{1}{h} K((x - t_0)/h)[f(x) + \tilde{f}_{t_0,\gamma,\delta,\varepsilon,A}(x)]\, dx - [f(t_0) + \tilde{f}_{t_0,\gamma,\delta,\varepsilon,A}(t_0)] \\
&= A \int_{x=t_0}^{\infty} \frac{1}{h} K((x - t_0)/h) \max\{(x - t_0)^\gamma, \varepsilon(x - t_0)^{\gamma-\delta}\}\, dx \geq h^{\gamma-\delta} A\varepsilon \int_0^\infty K(u)u^{\gamma-\delta}\, du,
\end{aligned}$$

where the inequality uses the fact that K is nonnegative. This is bounded from below by $A\varepsilon\underline{C}_K h^{\gamma-\delta}$ where $\underline{C}_K = \int_0^\infty K(u)u\, du$ (again using nonnegativity of K). The lower bound for $f + \tilde{g}_{t_0,\gamma,A}$ follows by noting that the constant $\underline{C}_{K,\gamma}$ in Lemma 3.5 is bounded from below uniformly over $0 < \gamma \leq 1$. $\square$

3.2.1 Testing Bounds for Self-Similar Functions

According to Lemma 3.5, we can obtain bounds for adaptation to the Hölder constant subject to coverage over self-similarity classes using the classes $\widetilde{\mathcal{F}}(\gamma, B, a, b) + \{\tilde{g}_{t_0,\gamma,A}\}$. Similarly, Lemma 3.6 allows us to obtain bounds for adaptation to the Hölder exponent using the classes $\widetilde{\mathcal{F}}(\gamma, B, a, b) + \{\tilde{f}_{t_0,\gamma,\delta,\varepsilon,A}\}$. To obtain these bounds, we can use the results from Section 3.1. We begin with a bound that will be useful for adaptation to the constant.

Lemma 3.7. *Let K be a kernel that satisfies (2) and let $\gamma > 0$, $A > 0$, $B > 0$ and $0 < a < b < t_0 < 1$. Let*

$$c_n = \left(\frac{4}{2\gamma + 1}\right)^{\frac{\gamma}{2\gamma+1}} B^{\frac{1}{2\gamma+1}} \kappa_\gamma^*(0) \left(\sigma_n^2 \log(1/\sigma_n)\right)^{\gamma/(2\gamma+1)}$$

where κ_γ^* is the function that solves the optimal recovery problem (4) in Lepski and Tsybakov

(2000). Then, for any $\eta > 0$,

$$\lim_{n\to\infty} d_{\text{test}}\left(\{\tilde{g}_{t_0,\gamma,A}\}, \left\{\widetilde{\mathcal{F}}(\gamma, B, a, b) + \{\tilde{g}_{t_0,\gamma,A}\}\right\} \cap \{f : \sup_{x\in[a,b]} |f(x)| \geq c_n(1-\eta)\}\right) = 0.$$

Furthermore, if $\beta > 2\alpha$,

$$R^*_{n,\alpha,\beta}\left(\{\tilde{g}_{t_0,\gamma,A}\}, \widetilde{\mathcal{F}}(\gamma, B, a, b) + \{\tilde{g}_{t_0,\gamma,A}\}\right) \geq (1 + o(1))c_n.$$

Proof. The first statement is immediate from Lemma 3.2 and Lemma 3.3, along with the fact that $\left\{\widetilde{\mathcal{F}}(\gamma, B, a, b) + \{\tilde{g}_{t_0,\gamma,A}\}\right\} \cap \{f : \sup_{x\in[a,b]} |f(x)| \geq c_n(1-\eta)\} = \widetilde{\mathcal{F}}(\gamma, B, a, b) \cap \{f : \sup_{x\in[a,b]} |f(x)| \geq c_n(1-\eta)\} + \{\tilde{g}_{t_0,\gamma,A}\}$ (since $\tilde{g}_{t_0,\gamma,A}(x) = 0$ for $x \in [a, b]$). The second statement is immediate from the first statement and Lemma 3.1. □

For adaptation to the exponent, we will use testing bounds for the classes $\{\tilde{f}_{t_0,\gamma,\delta_n,\varepsilon,A}\}$ and $\widetilde{\mathcal{F}}(\gamma, A, a, b) + \{\tilde{g}_{t_0,\gamma,A}\}$ where δ_n is a sequence converging to zero. To obtain these bounds using Lemma 3.2 and Lemma 3.3, we need to bound $\|\tilde{f}_{t_0,\gamma,\delta_n,\varepsilon,A} - \tilde{g}_{t_0,\gamma,A}\|/\sigma_n$, and to compute the limit of $(\sigma_n^2 \log(1/\sigma_n))^{(\gamma-\delta_n)/(2(\gamma-\delta_n)+1)}$. It turns out that setting δ_n to decrease at rate $1/\log n$ gives bounds for both terms.

Lemma 3.8. *Let $\delta_n = C_n/\log n$ where $C_n = (1 - b_n)(2\gamma + 1)\log \varepsilon^{-1}$ with $b_n = 1/(\log n)^{1/2}$ and let $\gamma \in (0, 1]$. Then*

$$\|\tilde{f}_{t_0,\gamma,\delta_n,\varepsilon,A} - \tilde{g}_{t_0,\gamma,A}\|^2/\sigma_n^2 \to 0.$$

Proof. For any $\delta \in [0, \gamma)$,

$$\begin{aligned}
&\|\tilde{f}_{t_0,\gamma,\delta,\varepsilon,A} - \tilde{g}_{t_0,\gamma,A}\|^2 = A^2 \int_{t=0}^{\varepsilon^{1/\delta}} (\varepsilon t^{\gamma-\delta} - t^\gamma)^2\, dt = A^2 \int_{t=0}^{\varepsilon^{1/\delta}} (\varepsilon^2 t^{2(\gamma-\delta)} + t^{2\gamma} - 2\varepsilon t^{2\gamma-\delta})\, dt \\
&= A^2 \left[\frac{\varepsilon^2}{2(\gamma-\delta)+1} t^{2(\gamma-\delta)+1} + \frac{1}{2\gamma+1} t^{2\gamma+1} - \frac{2\varepsilon}{2\gamma-\delta+1} t^{2\gamma-\delta+1}\right]_{t=0}^{\varepsilon^{1/\delta}} \\
&= A^2 \left[\frac{\varepsilon^2}{2(\gamma-\delta)+1} \varepsilon^{[2(\gamma-\delta)+1]/\delta} + \frac{1}{2\gamma+1} \varepsilon^{(2\gamma+1)/\delta} - \frac{2\varepsilon}{2\gamma-\delta+1} \varepsilon^{(2\gamma-\delta+1)/\delta}\right] \\
&= A^2 \left[\frac{1}{2(\gamma-\delta)+1} \varepsilon^{(2\gamma+1)/\delta} + \frac{1}{2\gamma+1} \varepsilon^{(2\gamma+1)/\delta} - \frac{2}{2\gamma-\delta+1} \varepsilon^{(2\gamma+1)/\delta}\right] \leq 4A^2 \varepsilon^{(2\gamma+1)/\delta}.
\end{aligned}$$

Plugging in $\delta_n = C_n/\log n$, dividing by σ_n^2 and taking logs gives

$$\log\left[\|\tilde{f}_{t_0,\gamma,\delta_n,\varepsilon,A} - \tilde{g}_{t_0,\gamma,A}\|^2/\sigma_n^2\right] \leq \frac{2\gamma+1}{\delta_n}\log\varepsilon + \log(4A^2) - \log(\sigma^2/n)$$
$$= \left(\frac{(2\gamma+1)\log\varepsilon}{C_n}+1\right)\log n + \log(4A^2/\sigma^2) = \frac{-b_n}{(1-b_n)}\log n + \log(4A^2/\sigma^2),$$

which diverges to $-\infty$, so that exponentiating gives a sequence that converges to zero, as required.

□

Lemma 3.9. *Let $C > 0$ and let $\delta_n = C_n/\log n$ where $C_n \to C$. Then*

$$\lim_{n\to\infty}\frac{(\sigma_n^2\log(1/\sigma_n))^{(\gamma-\delta_n)/(2(\gamma-\delta_n)+1)}}{(\sigma_n^2\log(1/\sigma_n))^{\gamma/(2\gamma+1)}} = \exp\left(\frac{C}{(2\gamma+1)^2}\right)$$

Proof. First, note that

$$\frac{\gamma-\delta_n}{2(\gamma-\delta_n)+1} - \frac{\gamma}{2\gamma+1} = -\frac{\delta_n}{[2(\gamma-\delta_n)+1](2\gamma+1)} = -\frac{\delta_n}{(2\gamma+1)^2}(1+o(1)).$$

Thus,

$$(\sigma_n^2)^{\frac{\gamma-\delta_n}{2(\gamma-\delta_n)+1}-\frac{\gamma}{2\gamma+1}} = (\sigma_n^2)^{-\frac{\delta_n}{(2\gamma+1)^2}(1+o(1))} = (1+o(1))n^{\frac{\delta_n}{(2\gamma+1)^2}(1+o(1))}$$
$$= \exp\left(\frac{\delta_n}{(2\gamma+1)^2}(1+o(1))\log n\right) = \exp\left(\frac{C}{(2\gamma+1)^2}(1+o(1))\right).$$

For the other term, we have

$$[\log(1/\sigma_n)]^{\frac{\gamma-\delta_n}{2(\gamma-\delta_n)+1}-\frac{\gamma}{2\gamma+1}} = [\log\sigma^{-1} + (1/2)\log n]^{\mathcal{O}(1/\log n)}$$
$$= \exp\left(\mathcal{O}(1/\log n)\log[\log\sigma^{-1} + (1/2)\log n]\right)$$

which converges to one as $n \to \infty$. □

Plugging in the constant $C = (2\gamma+1)\log\varepsilon^{-1}$ used in Lemma 3.8 gives $\exp\left(\frac{C}{(2\gamma+1)^2}\right) = \varepsilon^{-1/(2\gamma+1)}$. With these results in hand, we can state a lemma that bounds the scope for adaptation to the Hölder exponent.

Lemma 3.10. *Let K be a kernel that satisfies (2) and let $0 < \gamma \leq 1$, $\varepsilon \in (0,1)$, $A > 0$, $B > 0$ and $0 < a < b < t_0 < 1$. Let $\delta_n = C_n/\log n$ where $C_n = (1-b_n)(2\gamma+1)\log\varepsilon^{-1}$ with*

$b_n = 1/(\log n)^{1/2}$, *as in Lemma 3.8. Let*

$$c_n = \varepsilon^{-1/(2\gamma+1)} \left[\frac{4}{2\gamma+1} \left(\int \kappa^*(u)^2 \, du \right)^{-1} \right]^{\frac{\gamma}{2\gamma+1}} A^{\frac{1}{2\gamma+1}} \kappa^*(0) \left(\sigma_n^2 \log(1/\sigma_n) \right)^{\gamma/(2\gamma+1)}$$

where κ^* *is a function in* $\mathcal{F}_{\text{Höl}}(1,1)$ *with support contained in* $(-1/2, 1/2)$*. Then, for any* $\eta > 0$,

$$\lim_{n\to\infty} d_{\text{test}} \left(\{\tilde{g}_{t_0,\gamma,A}\}, \left\{ \widetilde{\mathcal{F}}(\gamma - \delta_n, A, a, b) + \{\tilde{f}_{t_0,\gamma,\delta_n,\varepsilon,A}\} \right\} \cap \{f : \sup_{x\in[a,b]} |f(x)| \geq c_n(1-\eta)\} \right) = 0.$$

Furthermore, if $0 < 2\alpha < \beta < 1$,

$$R^*_{n,\alpha,\beta} \left(\{\tilde{g}_{t_0,\gamma,A}\}, \left\{ \widetilde{\mathcal{F}}(\gamma - \delta_n, A, a, b) + \{\tilde{f}_{t_0,\gamma,\delta_n,\varepsilon,A}\} \right\} \cup \{\tilde{g}_{t_0,\gamma,A}\} \right) \geq (1 + o(1)) c_n.$$

Proof. First, note that, since $\tilde{f}_{t_0,\gamma,\delta_n,\varepsilon,A}(x) = 0$ for $x \in [a,b]$, $\left\{ \widetilde{\mathcal{F}}(\gamma - \delta_n, A, a, b) + \{\tilde{f}_{t_0,\gamma,\delta_n,\varepsilon,A}\} \right\} \cap \{f : \sup_{x\in[a,b]} |f(x)| \geq c_n(1-\eta)\} = \widetilde{\mathcal{F}}(\gamma - \delta_n, A, a, b) \cap \{f : \sup_{x\in[a,b]} |f(x)| \geq c_n(1-\eta)\} + \{\tilde{f}_{t_0,\gamma,\delta_n,\varepsilon,A}\}$. By Lemma 3.3,

$$\begin{aligned} &d_{\text{test}} \left(\{\tilde{g}_{t_0,\gamma,A}\}, \widetilde{\mathcal{F}}(\gamma - \delta_n, A, a, b) \cap \{f : \sup_{x\in[a,b]} |f(x)| \geq c_n(1-\eta)\} + \{\tilde{f}_{t_0,\gamma,\delta_n,\varepsilon,A}\} \right) \\ &\leq d_{\text{test}} \left(\{0\}, \widetilde{\mathcal{F}}(\gamma - \delta_n, A, a, b) \cap \{f : \sup_{x\in[a,b]} |f(x)| \geq c_n(1-\eta)\} \right) + \|\tilde{f}_{t_0,\gamma,\delta_n,\varepsilon,A} - \tilde{g}_{t_0,\gamma,A}\| / \sigma_n. \end{aligned}$$

The second term converges to zero by Lemma 3.8. By Lemma 3.2, the first term will converge to zero so long as $\limsup_{n\to\infty} \frac{c_n(1-\eta)}{C(\gamma-\delta_n,A,\kappa^*)(\sigma_n^2 \log(1/\sigma_n))^{(\gamma-\delta_n)/[2(\gamma-\delta_n)+1]}} < 1$, which holds by Lemma 3.9 and the fact that $C(\gamma - \delta_n, A, \kappa^*) \to C(\gamma, A, \kappa^*)$ (note that $\kappa^* \in \mathcal{F}_{\text{Höl}}(1,\gamma)$ for all $\gamma \in (0,1]$). This proves the first statement of the lemma, which immediately gives the second statement by Lemma 3.1. □

3.3 Adaptation to the Hölder Constant

We first state a general result, which we then use to prove Theorems 2.1 and 2.4.

Theorem 3.1. *Let* $\gamma > 0$*. Let* K *be a kernel that satisfies* $\int K(u)|u|^\gamma \neq 0$ *as well as (2). There exists a* $\overline{h}_K$ *depending only on the kernel* K*, as well as constants* $\overline{C}_{K,\gamma}$ *and* $\underline{C}_{K,\gamma}$

depending only on K and γ such that for any $A > 0$, $B > 0$ and $0 < 2\alpha < \beta < 1$,

$$\begin{aligned}
&R^*_{n,\alpha,\beta}\big(\mathcal{F}_{\text{Höl}}(\gamma, A) \cap \mathcal{F}_{\text{GN}}(\gamma, \underline{C}_{K,\gamma}A, \overline{C}_{K,\gamma}(A+B); K, \overline{h}_K),\\
&\qquad \mathcal{F}_{\text{Höl}}(\gamma, B+A) \cap \mathcal{F}_{\text{GN}}(\gamma, \underline{C}_{K,\gamma}A, \overline{C}_{K,\gamma}(A+B); K, \overline{h}_K)\big)\\
&\qquad \geq (1+o(1))\left(\frac{4}{2\gamma+1}\right)^{\frac{\gamma}{2\gamma+1}} B^{\frac{1}{2\gamma+1}}\kappa^*_\gamma(0)\left(\sigma_n^2\log(1/\sigma_n)\right)^{\gamma/(2\gamma+1)}.
\end{aligned}$$

Proof. Let a, b and t_0 be any constants that satisfy the conditions of Lemma 3.5 (say, $a = 1/4$, $b = 1/2$ and $t_0 = 3/4$). By Lemma 3.5, we can choose $\underline{C}_{K,\gamma}$ and $\overline{C}_{K,\gamma}$ such that

$$\widetilde{\mathcal{F}}(\gamma, B, a, b) + \{\tilde{g}_{t_0,\gamma,A}\} \subseteq \mathcal{F}_{\text{Höl}}(\gamma, B+A) \cap \mathcal{F}_{\text{GN}}(\gamma, \underline{C}_{K,\gamma}A, \overline{C}_{K,\gamma}(A+B); K, \overline{h}_{K,t_0,b})$$

and

$$\tilde{g}_{t_0,\gamma,A} \in \mathcal{F}_{\text{Höl}}(\gamma, A) \cap \mathcal{F}_{\text{GN}}(\gamma, \underline{C}_{K,\gamma}A, \overline{C}_{K,\gamma}(A+B); K, \overline{h}_{K,t_0,b}).$$

The result now follows from Lemma 3.7.

□

3.3.1 Proof of Theorem 2.1

Let $\overline{C}_{K,\gamma}$ and $\underline{C}_{K,\gamma}$ be as given in Theorem 3.1. We apply Theorem 3.1 with $\tilde{A} = B/2$ playing the role of A in Theorem 3.1 and $\tilde{B} = \min\{B\varepsilon^{-1}\underline{C}_{K,\gamma}/2, \overline{B}\} - B/2$ playing the role of B in Theorem 3.1. This gives a lower bound so long as $\varepsilon \leq \varepsilon' \leq \underline{C}_{K,\gamma}/2$. To see this, note that

$$\begin{aligned}
&\mathcal{F}_{\text{Höl}}(\gamma, \tilde{A}) \cap \mathcal{F}_{\text{GN}}(\gamma, \underline{C}_{K,\gamma}\tilde{A}, \overline{C}_{K,\gamma}(\tilde{A}+\tilde{B}); K, \overline{h}_K)\\
&\subseteq \mathcal{F}_{\text{self-sim}}(\gamma, B, \underline{C}_{K,\gamma}/2; K, \overline{h}_K) \subseteq \mathcal{F}_{\text{self-sim}}(\gamma, B, \varepsilon; K, \overline{h}_K)
\end{aligned}$$

where the last set inclusion holds so long as $\varepsilon' \leq \underline{C}_{K,\gamma}/2$. Also,

$$\begin{aligned}
&\mathcal{F}_{\text{Höl}}(\gamma, \tilde{A}+\tilde{B}) \cap \mathcal{F}_{\text{GN}}(\gamma, \underline{C}_{K,\gamma}\tilde{A}, \overline{C}_{K,\gamma}(\tilde{A}+\tilde{B}); K, \overline{h}_K)\\
&\subseteq \mathcal{F}_{\text{self-sim}}(\gamma, \tilde{A}+\tilde{B}, \underline{C}_{K,\gamma}\tilde{A}/(\tilde{A}+\tilde{B}); K, \overline{h}_K) \subseteq \cup_{B' \in [\underline{B}, \overline{B}]}\mathcal{F}_{\text{self-sim}}(\gamma, B', \varepsilon; K, \overline{h}_K)
\end{aligned}$$

where the last set inclusion follows so long as $\varepsilon \le \underline{C}_{K,\gamma}/2$ since $\tilde{A}+\tilde{B} = \min\{B\varepsilon^{-1}\underline{C}_{K,\gamma}/2, \overline{B}\} \in [\underline{B}, \overline{B}]$ where the last step uses $\varepsilon \le \underline{C}_{K,\gamma}/2$, and

$$\frac{\underline{C}_{K,\gamma}\tilde{A}}{\tilde{A}+\tilde{B}} = \frac{\underline{C}_{K,\gamma}B/2}{\min\{B\varepsilon^{-1}\underline{C}_{K,\gamma}/2, \overline{B}\}} \ge \frac{\underline{C}_{K,\gamma}B/2}{B\varepsilon^{-1}\underline{C}_{K,\gamma}/2} = \varepsilon.$$

Applying Theorem 3.1 gives the lower bound

$$\begin{aligned}
&R^*_{n,\alpha,\beta}\left(\mathcal{F}_{\text{self-sim}}(\gamma, B, \varepsilon'; K, \overline{h}_K), \cup_{B' \in [\underline{B},\overline{B}]}\mathcal{F}_{\text{self-sim}}(\gamma, B', \varepsilon; K, \overline{h}_K)\right) \\
&\ge (1+o(1))\left(\frac{4}{2\gamma+1}\right)^{\frac{\gamma}{2\gamma+1}} \tilde{B}^{\frac{1}{2\gamma+1}}\kappa^*_\gamma(0)\left(\sigma_n^2\log(1/\sigma_n)\right)^{\gamma/(2\gamma+1)}
\end{aligned}$$

The result follows since, so long as $\varepsilon \le \underline{C}_{K,\gamma}/2$, we have $\tilde{B} = \min\{B\varepsilon^{-1}(\underline{C}_{K,\gamma} - \varepsilon)/2, \overline{B} - B/2\} \ge \min\{B\varepsilon^{-1}\underline{C}_{K,\gamma}/4, \overline{B}/2\} \ge \min\{B\varepsilon^{-1}, \overline{B}\} \cdot \min\{\underline{C}_{K,\gamma}/4, 1/2\}$.

3.3.2 Proof of Theorem 2.4

We apply Theorem 3.1 with $\tilde{A} = b_1/\underline{C}_{K,\gamma}$ playing the role of A in Theorem 3.1 and $\tilde{B} = \overline{B} - b_1/\underline{C}_{K,\gamma}$ playing the role of B in Theorem 3.1. For b_1 small enough so that $b_1/\underline{C}_{K,\gamma} \le B$, applying Theorem 3.1 with $\tilde{A}$ and $\tilde{B}$ gives a lower bound for the quantity in the display in the theorem. If b_1 is small enough that $b_1 \le \underline{C}_{K,\gamma}\overline{B}/2$, then $\tilde{B} \ge \overline{B}/2$, and plugging this into the lower bound obtained using Theorem 3.1 gives the result.

3.4 Adaptation to the Hölder Exponent

We now prove Theorem 2.2. Let a, b and t_0 be any constants that satisfy the conditions of Lemma 3.6 (say, $a = 1/4$, $b = 1/2$ and $t_0 = 3/4$). Let δ_n be the sequence in Lemma 3.10. By Lemma 3.6 with $A = B = 1/2$, we have $\tilde{g}_{t_0,\gamma,1/2} \in \mathcal{F}_{\text{Höl}}(\gamma, 1) \cap \mathcal{F}_{\text{GN}}(\gamma, \underline{C}_K/2, \overline{C}_K; K, \overline{h}_{K,t_0,b})$. For $\varepsilon' \le \underline{C}_K/2$, this set is contained in $\mathcal{F}_{\text{self-sim}}(\gamma, 1, \varepsilon'; K, \overline{h}_{K,t_0,b})$ (and also in $\mathcal{F}_{\text{self-sim}}(\gamma, 1, \varepsilon; K, \overline{h}_{K,t_0,b})$, since $\varepsilon \le \varepsilon'$). By Lemma 3.6, we also have $\tilde{\mathcal{F}}(\gamma, 1/2, a, b) + \{\tilde{f}_{t_0,\gamma,\delta_n,2\varepsilon/\underline{C}_K,1/2}\} \subseteq \mathcal{F}_{\text{Höl}}(\gamma - \delta_n, 1) \cap \mathcal{F}_{\text{GN}}(\gamma - \delta_n, \varepsilon, \overline{C}_K; K, \overline{h}_{K,t_0,b})$. For large enough n, we will have $\gamma - \delta_n \in [\underline{\gamma}, \overline{\gamma}]$, so that this set is contained in $\cup_{\tilde{\gamma}\in[\underline{\gamma},\overline{\gamma}]}\left[\mathcal{F}_{\text{Höl}}(\tilde{\gamma}, 1) \cap \mathcal{F}_{\text{GN}}(\tilde{\gamma}, \varepsilon, \overline{C}_K; K, \overline{h})\right] = \cup_{\tilde{\gamma}\in[\underline{\gamma},\overline{\gamma}]}\left[\mathcal{F}_{\text{self-sim}}(\tilde{\gamma}, 1, \varepsilon; K, \overline{h})\right]$.

Applying Lemma 3.10 gives the lower bound

$$R^*_{n,\alpha,\beta}\left(\mathcal{F}_{\text{self-sim}}(\gamma,1,\varepsilon';K,\overline{h}_K),\cup_{\tilde{\gamma}\in[\underline{\gamma},\overline{\gamma}]}\left[\mathcal{F}_{\text{self-sim}}(\tilde{\gamma},1,\varepsilon;K,\overline{h}_K)\right]\right)$$
$$\geq (2\varepsilon/\underline{C}_K)^{-1/(2\gamma+1)}\left[\frac{4}{2\gamma+1}\left(\int\kappa^*(u)^2\,du\right)^{-1}\right]^{\frac{\gamma}{2\gamma+1}}(1/2)^{\frac{1}{2\gamma+1}}\kappa^*(0)\left(\sigma_n^2\log(1/\sigma_n)\right)^{\gamma/(2\gamma+1)}(1+o(1)).$$

which gives the result after noting that the terms other than $\varepsilon^{-1/(2\gamma+1)}\left(\sigma_n^2\log(1/\sigma_n)\right)^{\gamma/(2\gamma+1)}$ are bounded away from zero over $\gamma\in(0,1]$.

A Details for Section 2.2

This appendix provides details for the results in Section 2.2 concerning the confidence band constructed in that section. In particular, we prove that the conclusions of Theorem 2.3 hold for this confidence band. To this end, we need to show that the upper bound in Theorem 2.3 holds and $[\hat{\gamma}_\ell,\hat{\gamma}_u]$ contains γ with probability approaching one uniformly over $f\in\cup_{B\in[\underline{B},\overline{B}],\gamma\in[\underline{\gamma},\overline{\gamma}]}\mathcal{F}_{\text{GN}}(\gamma,\varepsilon B,\overline{C}_K B)$.

A.1 Length of the Confidence Band

This section derives the upper bound on the length of the confidence band. We proceed by bounding (9). By Theorem A.1 below, we will have $\max\{|\hat{\gamma}_u-\gamma|,|\hat{\gamma}_\ell-\gamma|\}\leq r_n$ with probability approaching one uniformly over $\cup_{B\in[\underline{B},\overline{B}],\gamma\in[\underline{\gamma},\overline{\gamma}]}\mathcal{F}_{\text{GN}}(\gamma,\varepsilon B,\overline{C}_K B)$ for some sequence $r_n\to 0$. Using this and substituting $\tilde{c}(h,h')=c(h)+c(h')$ gives the bound (up to the term $(1+\eta_n)\to 1$) of two times

$$\sup_{\gamma_\ell,\gamma_u\in[\gamma-r_n,\gamma+r_n]}\ \inf_{h,h_1,h_2\in(0,\overline{h}]}\left[c(h)+\frac{\overline{C}_K B(h_1^\gamma+h_2^\gamma)+2c(h_1)+2c(h_2)}{a(\varepsilon,h_1,h_2,h,\gamma_\ell,\gamma_u)}\right]$$

where

$$c(h)=\left\{q_n(\log e/(2Rh))^{-1/2}[\log\log(e^e/(2Rh))]+(2\log(1/(2Rh)))^{1/2}\right\}/\left[\sqrt{nh}/(\sigma\|K\|)\right]$$

for q_n a slowly increasing sequence, as defined in Section 2.2.

It turns out that it will suffice to get an upper bound for the infimum by taking $h=h_{n,\gamma}=\left[\lambda_\gamma\left(\log n\right)/n\right]^{1/(2\gamma+1)}$, $h_1=h_{1,n,\gamma}=d_{1,n}h_{n,\gamma}$ and $h_2=h_{2,n,\gamma}=d_{2,n}h_{n,\gamma}$ where λ_γ is a constant depending on γ and $d_{1,n}$ and $d_{2,n}$ are sequences satisfying $d_{2,n}\to\infty$, $d_{1,n}/d_{2,n}\to\infty$

and $r_n \log d_{1,n} \to 0$. For these choices of h, h_1 and h_2, it can be checked that, letting

$$\bar{c}_n(h) = \sigma \|K\| (2 \log h^{-1})^{1/2} / \sqrt{nh},$$

we have $c_n(h_{n,\gamma}) = \bar{c}_n(h_{n,\gamma})(1 + o(1))$ where the $o(1)$ term is uniform over $\gamma \in [\underline{\gamma}, \overline{\gamma}]$, and similarly for $h_{1,n,\gamma}$ and $h_{2,n,\gamma}$, so long as q_n, $d_{1,n}$ and $d_{2,n}$ increase slowly enough. In addition, we have

$$\sup_{\gamma \in [\underline{\gamma}, \overline{\gamma}]} \sup_{\gamma_\ell, \gamma_u \in [\gamma - r_n, \gamma + r_n]} \left| \frac{a(\varepsilon, h_{1,n,\gamma}, h_{2,n,\gamma}, h_{n,\gamma}, \gamma_\ell, \gamma_u)}{a(\varepsilon, h_{1,n,\gamma}, h_{2,n,\gamma}, h_{n,\gamma}, \gamma, \gamma)} - 1 \right| \to 0.$$

(see Lemma A.1 below) and

$$\frac{h_{1,n,\gamma}^{\gamma} + h_{2,n,\gamma}^{\gamma}}{a(\varepsilon, h_{1,n,\gamma}, h_{2,n,\gamma}, h_{n,\gamma}, \gamma, \gamma)} = h_{n,\gamma}^{\gamma} (C_K/\varepsilon)(1 + o(1))$$

(see Lemma A.2 below). This gives the upper bound

$$\left[\bar{c}_n(h_{n,\gamma}) + (C_K^2 B/\varepsilon) h_{n,\gamma}^{\gamma} + (C_K/\varepsilon) h_{n,\gamma}^{\gamma} \frac{2\bar{c}_n(h_{1,n,\gamma}) + 2\bar{c}_n(h_{2,n,\gamma})}{h_{1,n,\gamma}^{\gamma} + h_{2,n,\gamma}^{\gamma}} \right] (1 + o(1))$$

where the $o(1)$ term is uniform over $\gamma \in [\underline{\gamma}, \overline{\gamma}]$ and $B \in [\underline{B}, \overline{B}]$. Some calculation (Lemma A.3 below) shows that this is equal to $[(\log n)/n]^{\gamma/(2\gamma+1)}$ times

$$\sigma \|K\| \left(\frac{2}{2\gamma + 1} \right)^{1/2} \lambda_{\gamma}^{-1/[2(2\gamma+1)]} + (C_K^2 B/\varepsilon) \lambda_{\gamma}^{\gamma/(2\gamma+1)}$$

times a term that converges to one uniformly over $\gamma \in [\underline{\gamma}, \overline{\gamma}]$ and $B \in [\underline{B}, \overline{B}]$. Choosing λ_γ to minimize this expression for each γ (and noting that the resulting choice of λ_γ is bounded away from zero and ∞ as required by the lemmas below) gives the result.

Lemma A.1.

$$\sup_{\gamma \in [\underline{\gamma}, \overline{\gamma}]} \sup_{\gamma_\ell, \gamma_u \in [\gamma - r_n, \gamma + r_n]} \left| \frac{a(\varepsilon, h_{1,n,\gamma}, h_{2,n,\gamma}, h_{n,\gamma}, \gamma_\ell, \gamma_u)}{a(\varepsilon, h_{1,n,\gamma}, h_{2,n,\gamma}, h_{n,\gamma}, \gamma, \gamma)} - 1 \right| \to 0.$$

Proof. Note that, once n is large enough so that $d_{1,n} \geq 1$ and $d_{2,n} \geq 1$, we will have, for any $\gamma \in [\underline{\gamma}, \overline{\gamma}]$ and $\gamma_\ell, \gamma_u \in [\gamma - r_n, \gamma + r_n]$ with $\gamma_\ell \leq \gamma_u$,

$$(\varepsilon/\overline{C}_K) d_{1,n}^{\gamma - r_n} - d_{2,n}^{\gamma + r_n} \leq a(\varepsilon, h_{1,n,\gamma}, h_{2,n,\gamma}, h_{n,\gamma}, \gamma_\ell, \gamma_u) \leq (\varepsilon/\overline{C}_K) d_{1,n}^{\gamma + r_n} - d_{2,n}^{\gamma - r_n}$$

and $a(\varepsilon, h_{1,n,\gamma}, h_{2,n,\gamma}, h_{n,\gamma}, \gamma, \gamma) = (\varepsilon/\overline{C}_K)d_{1,n}^{\gamma} - d_{2,n}^{\gamma}$. Thus, uniformly over $\gamma \in [\underline{\gamma}, \overline{\gamma}]$,

$$\frac{a(\varepsilon, h_{1,n,\gamma}, h_{2,n,\gamma}, h_{n,\gamma}, \gamma_\ell, \gamma_u)}{a(\varepsilon, h_{1,n,\gamma}, h_{2,n,\gamma}, h_{n,\gamma}, \gamma, \gamma)} \leq \frac{(\varepsilon/\overline{C}_K)d_{1,n}^{\gamma+r_n} - d_{2,n}^{\gamma-r_n}}{(\varepsilon/\overline{C}_K)d_{1,n}^{\gamma} - d_{2,n}^{\gamma}} = \frac{d_{1,n}^{r_n} - (\varepsilon/\overline{C}_K)^{-1}d_{2,n}^{-r_n}(d_{2,n}/d_{1,n})^{\gamma}}{1 - (\varepsilon/\overline{C}_K)^{-1}(d_{2,n}/d_{1,n})^{\gamma}} \to 1,$$

using the fact that $d_{1,n}^{r_n} = \exp(r_n \log d_{1,n}) \to 1$ and $d_{2,n}^{r_n} = \exp(r_n \log d_{2,n}) \to 1$. The result follows from this and a similar argument for the lower bound. $\square$

Lemma A.2.

$$\frac{h_{1,n,\gamma}^{\gamma} + h_{2,n,\gamma}^{\gamma}}{a(\varepsilon, h_{1,n,\gamma}, h_{2,n,\gamma}, h_{n,\gamma}, \gamma, \gamma)} = h_{n,\gamma}^{\gamma}(\overline{C}_K/\varepsilon)(1 + o(1))$$

where the $o(1)$ term is uniform over all $\gamma \in [\underline{\gamma}, \overline{\gamma}]$.

Proof. Once $h_{2,n,\gamma}/h_{1,n,\gamma} = d_{2,n}/d_{1,n}$ is small enough,

$$\frac{h_{1,n,\gamma}^{\gamma} + h_{2,n,\gamma}^{\gamma}}{a(\varepsilon, h_{1,n,\gamma}, h_{2,n,\gamma}, h, \gamma, \gamma)} = \frac{(d_{1,n}h_{n,\gamma})^{\gamma} + (d_{2,n}h_{n,\gamma})^{\gamma}}{(\varepsilon/\overline{C}_K)d_{1,n}^{\gamma} - d_{2,n}^{\gamma}} = h_{n,\gamma}^{\gamma}(\overline{C}_K/\varepsilon)\frac{1 + (d_{2,n}/d_{1,n})^{\gamma}}{1 - (\overline{C}_K/\varepsilon)(d_{2,n}/d_{1,n})^{\gamma}}$$

and the last term converges to one uniformly over $\gamma \in [\underline{\gamma}, \overline{\gamma}]$ as required. $\square$

Lemma A.3. *Let $h_{n,\gamma}$ be given above with λ_γ bounded away from zero and infinity uniformly over $\gamma \in [\underline{\gamma}, \overline{\gamma}]$. Then*

$$\bar{c}_n(h_{n,\gamma}) = (1 + o(1))\sigma\|K\| \left(\frac{2}{2\gamma + 1}\right)^{1/2} \left[\frac{\log n}{n}\right]^{\gamma/(2\gamma+1)} \lambda_\gamma^{-1/[2(2\gamma+1)]}$$

where the $o(1)$ term is uniform over $\gamma \in [\underline{\gamma}, \overline{\gamma}]$. In addition, $\bar{c}_n(h_{1,n,\gamma})/h_{1,n,\gamma}^{\gamma} \to 0$ and $\bar{c}_n(h_{2,n,\gamma})/h_{2,n,\gamma}^{\gamma} \to 0$ uniformly over $\gamma \in [\underline{\gamma}, \overline{\gamma}]$.

Proof. We have

$$\bar{c}_n(h_{n,\gamma}) = \sigma\|K\|\left(\frac{2\log h_{n,\gamma}^{-1}}{nh_{n,\gamma}}\right)^{1/2} = \sigma\|K\|\left(\frac{2\log\left[\lambda_\gamma^{-1}n/\log n\right]^{1/(2\gamma+1)}}{n\left[\lambda_\gamma\left(\log n\right)/n\right]^{1/(2\gamma+1)}}\right)^{1/2}$$
$$= \sigma\|K\|\lambda_\gamma^{-1/[2(2\gamma+1)]}\left(\frac{\frac{2}{2\gamma+1}\left[\log n - \log\lambda_\gamma - \log\log n\right]}{n\left[\left(\log n\right)/n\right]^{1/(2\gamma+1)}}\right)^{1/2}$$
$$= \sigma\|K\|\lambda_\gamma^{-1/[2(2\gamma+1)]}\left(\frac{2}{2\gamma+1}\right)^{1/2}\left(\frac{\log n}{n\left[\left(\log n\right)/n\right]^{1/(2\gamma+1)}}\right)^{1/2}\left(\frac{\left[\log n - \log\lambda_\gamma - \log\log n\right]}{\log n}\right)^{1/2}$$
$$= \sigma\|K\|\lambda_\gamma^{-1/[2(2\gamma+1)]}\left(\frac{2}{2\gamma+1}\right)^{1/2}\left(\frac{\log n}{n}\right)^{\gamma/(2\gamma+1)}\left(\frac{\left[\log n - \log\lambda_\gamma - \log\log n\right]}{\log n}\right)^{1/2}.$$

The last term converges to one uniformly over λ_γ bounded away from zero and infinity. For the second part of the lemma, it follows from similar calculations that $\bar{c}_n(h_{1,n,\gamma})/h_{1,n,\gamma}^{\gamma}$ is equal to

$$\sigma\|K\|(d_{1,n}\lambda_\gamma)^{-(1+2\gamma)/[2(2\gamma+1)]}\left(\frac{2}{2\gamma+1}\right)^{1/2}\left(\frac{\left[\log n - \log(d_{1,n}\lambda_\gamma) - \log\log n\right]}{\log n}\right)^{1/2},$$

which converges to zero uniformly over $\gamma \in [\underline{\gamma}, \overline{\gamma}]$, and similarly for $\bar{c}_n(h_{2,n,\gamma})/h_{2,n,\gamma}^{\gamma}$. □

A.2 Estimating γ

The following theorem shows that the estimator $\hat{\gamma}$ given in Section 2.2 converges at the required rate uniformly over $\cup_{\gamma\in[\underline{\gamma},\overline{\gamma}]}\cup_{B\in[\underline{B},\overline{B}]}\mathcal{F}_{\text{self-sim}}(\varepsilon,\gamma,B)$.

Theorem A.1. *Let $\tilde{h}_{1,n}$ and $\tilde{h}_{2,n}$ be sequences satisfying $\tilde{h}_{1,n} \to 0$, $\tilde{h}_{2,n} \to 0$, $\tilde{h}_{1,n}/\tilde{h}_{2,n} \to \infty$ and such that $\tilde{h}_{2,n}n^{\delta} \to \infty$ for all $\delta > 0$. Let $\hat{\gamma} = -\frac{\log\hat{\Delta}(\tilde{h}_{1,n},\tilde{h}_{2,n})}{\log\tilde{h}_{1,n}^{-1}}$. Then, for any $\varepsilon > 0$, $\overline{h} > 0$, $0 < \underline{\gamma} < \overline{\gamma} < \infty$ and $0 < \underline{B} < \overline{B} < \infty$ and any sequence $\tilde{b}_n \to \infty$, we have $|\hat{\gamma} - \gamma| \leq \tilde{b}_n/\log\tilde{h}_{1,n}^{-1}$ with probability approaching one uniformly over $\cup_{\gamma\in[\underline{\gamma},\overline{\gamma}]}\cup_{B\in[\underline{B},\overline{B}]}\mathcal{F}_{\text{self-sim}}(\varepsilon,\gamma,B)$.*

Proof. Note that, by (5), we have, letting $\tilde{d}_n = \tilde{h}_{1,n}/\tilde{h}_{2,n}$,

$$\tilde{h}_{1,n}^{\gamma}\overline{C}_K B[\varepsilon/\overline{C}_K - \tilde{d}_n^{-\gamma}] \leq \Delta(\tilde{h}_{1,n},\tilde{h}_{2,n};f) \leq \tilde{h}_{1,n}^{\gamma}\overline{C}_K B[1 + \tilde{d}_n^{-\gamma}]$$

which implies

$$
\begin{aligned}
&-\gamma\log\tilde{h}_{1,n}^{-1}+\log\{\overline{C}_K B[\varepsilon/\overline{C}_K-\tilde{d}_n^{-\gamma}]\}\leq\log\Delta(\tilde{h}_{1,n},\tilde{h}_{2,n};f)\leq-\gamma\log\tilde{h}_{1,n}^{-1}+\log\{\overline{C}_K B[1+\tilde{d}_n^{-\gamma}]\}\\
&\Longrightarrow -\gamma+\frac{\log\{\overline{C}_K B[\varepsilon/\overline{C}_K-\tilde{d}_n^{-\gamma}]\}}{\log\tilde{h}_{1,n}^{-1}}\leq\frac{\log\Delta(\tilde{h}_{1,n},\tilde{h}_{2,n};f)}{\log\tilde{h}_{1,n}^{-1}}\leq-\gamma+\frac{\log\{\overline{C}_K B[1+\tilde{d}_n^{-\gamma}]\}}{\log\tilde{h}_{1,n}^{-1}}\\
&\Longrightarrow -\frac{\log\{\overline{C}_K B[\varepsilon/\overline{C}_K-\tilde{d}_n^{-\gamma}]\}}{\log\tilde{h}_{1,n}^{-1}}\geq-\frac{\log\Delta(\tilde{h}_{1,n},\tilde{h}_{2,n};f)}{\log\tilde{h}_{1,n}^{-1}}-\gamma\geq-\frac{\log\{\overline{C}_K B[1+\tilde{d}_n^{-\gamma}]\}}{\log\tilde{h}_{1,n}^{-1}}.
\end{aligned}
$$

For large enough n, we will have, for all $\gamma\in[\underline{\gamma},\overline{\gamma}]$, $B\in[\underline{B},\overline{B}]$,

$$\max\left\{|\log\{\overline{C}_K B[1+\tilde{d}_n^{-\gamma}]\}|,|\log\{\overline{C}_K B[\varepsilon/\overline{C}_K-\tilde{d}_n^{-\gamma}]\}|\right\}\leq\tilde{b}_n-1$$

so that

$$\left|-\frac{\log\Delta(\tilde{h}_{1,n},\tilde{h}_{2,n};f)}{\log\tilde{h}_{1,n}^{-1}}-\gamma\right|\leq\frac{\tilde{b}_n-1}{\log\tilde{h}_{1,n}^{-1}}.$$

By Theorem 2.1 of Dumbgen and Spokoiny (2001),

$$
\begin{aligned}
&|\hat{\Delta}(\tilde{h}_{1,n},\tilde{h}_{2,n})-\Delta(\tilde{h}_{1,n},\tilde{h}_{2,n};f)|\leq C\left[(\log\tilde{h}_{1,n}^{-1})^{1/2}/(n\tilde{h}_{1,n})^{1/2}+(\log\tilde{h}_{2,n}^{-1})^{1/2}/(n\tilde{h}_{2,n})^{1/2}\right]\\
&\leq 2C(\log\tilde{h}_{2,n}^{-1})^{1/2}/(n\tilde{h}_{2,n})^{1/2}
\end{aligned}
$$

with probability approaching one, where C is a constant that depends only on σ and $\|K\|$. Using the fact that $|\log a-\log b|\leq|a-b|/\min\{a,b\}$ along with the lower bound $\Delta(\tilde{h}_{1,n},\tilde{h}_{2,n};f)\geq\tilde{h}_{1,n}^{\gamma}\overline{C}_K B[\varepsilon/\overline{C}_K-\tilde{d}_n^{-\gamma}]\geq\tilde{h}_{1,n}^{\overline{\gamma}}\overline{C}_K B[\varepsilon/\overline{C}_K-\tilde{d}_n^{-\underline{\gamma}}]\geq\tilde{h}_{2,n}^{\overline{\gamma}}\overline{C}_K B[\varepsilon/\overline{C}_K-\tilde{d}_n^{-\underline{\gamma}}]$, this gives

$$|\log\hat{\Delta}(\tilde{h}_{1,n},\tilde{h}_{2,n})-\log\Delta(\tilde{h}_{1,n},\tilde{h}_{2,n};f)|\leq\frac{2C(\log\tilde{h}_{2,n}^{-1})^{1/2}/(n\tilde{h}_{2,n})^{1/2}}{\tilde{h}_{2,n}^{\overline{\gamma}}\overline{C}_K\underline{B}[\varepsilon/\overline{C}_K-\tilde{d}_n^{-\underline{\gamma}}]-2C(\log\tilde{h}_{2,n}^{-1})^{1/2}/(n\tilde{h}_{2,n})^{1/2}}$$

so long as the denominator on the right hand side is positive. For large enough n, we will have $\tilde{h}_{2,n}^{\overline{\gamma}}\overline{C}_K\underline{B}[\varepsilon/\overline{C}_K-\tilde{d}_n^{-\underline{\gamma}}]/2\geq 2C(\log\tilde{h}_{2,n}^{-1})^{1/2}/(n\tilde{h}_{2,n})^{1/2}$ (using the fact that $\tilde{h}_{2,n}$ decreases more slowly than any power of n), so that the right hand side will be bounded by 1. Thus, with probability approaching one uniformly over the relevant set of functions f, we have $|\hat{\gamma}-\gamma|\leq\tilde{b}_n/\log\tilde{h}_{1,n}^{-1}$ as required. □

References

Bahadur, R. R. and L. J. Savage (1956): "The Nonexistence of Certain Statistical Procedures in Nonparametric Problems," *The Annals of Mathematical Statistics*, 27, 1115–1122.

Brown, L. D. and M. G. Low (1996): "Asymptotic equivalence of nonparametric regression and white noise," *The Annals of Statistics*, 24, 2384–2398.

Bull, A. D. (2012): "Honest adaptive confidence bands and self-similar functions," *Electronic Journal of Statistics*, 6, 1490–1516.

Cai, T. T. and M. G. Low (2004): "An Adaptation Theory for Nonparametric Confidence Intervals," *The Annals of Statistics*, 32, 1805–1840.

Chernozhukov, V., D. Chetverikov, and K. Kato (2014): "Anti-concentration and honest, adaptive confidence bands," *The Annals of Statistics*, 42, 1787–1818.

Donoho, D. L. (1994): "Statistical Estimation and Optimal Recovery," *The Annals of Statistics*, 22, 238–270.

Dumbgen, L. and V. G. Spokoiny (2001): "Multiscale Testing of Qualitative Hypotheses," *The Annals of Statistics*, 29, 124–152.

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

Hoffmann, M. and R. Nickl (2011): "On adaptive inference and confidence bands," *The Annals of Statistics*, 39, 2383–2409.

Ingster, Y. and I. A. Suslina (2003): *Nonparametric Goodness-of-Fit Testing Under Gaussian Models*, Springer.

Lepski, O. and A. Tsybakov (2000): "Asymptotically exact nonparametric hypothesis testing in sup-norm and at a fixed point," *Probability Theory and Related Fields*, 117, 17–48.

Low, M. G. (1997): "On nonparametric confidence intervals," *The Annals of Statistics*, 25, 2547–2554.

Nussbaum, M. (1996): "Asymptotic equivalence of density estimation and Gaussian white noise," *The Annals of Statistics*, 24, 2399–2430.

Picard, D. and K. Tribouley (2000): "Adaptive confidence interval for pointwise curve estimation," *The Annals of Statistics*, 28, 298–335.

Robins, J. and A. van der Vaart (2006): "Adaptive nonparametric confidence sets," *The Annals of Statistics*, 34, 229–253.

Schennach, S. M. (2015): "A bias bound approach to nonparametric inference," Tech. rep., cemmap working paper, Centre for Microdata Methods and Practice.

Tsybakov, A. B. (1998): "Pointwise and sup-norm sharp adaptive estimation of functions on the Sobolev classes," *The Annals of Statistics*, 26, 2420–2469.

——— (2009): *Introduction to Nonparametric Estimation*, New York: Springer.