Yale University

EliScholar – A Digital Platform for Scholarly Publishing at Yale

Cowles Foundation Discussion Papers | Cowles Foundation

10-1-2018

Adaptation Bounds for Confidence Bands under Self-Similarity

Timothy B. Armstrong

Follow this and additional works at: https://elischolar.library.yale.edu/cowles-discussion-paper-series

 Part of the Economics Commons

Recommended Citation

Armstrong, Timothy B., "Adaptation Bounds for Confidence Bands under Self-Similarity" (2018). *Cowles Foundation Discussion Papers*. 120.
https://elischolar.library.yale.edu/cowles-discussion-paper-series/120

This Discussion Paper is brought to you for free and open access by the Cowles Foundation at EliScholar – A Digital Platform for Scholarly Publishing at Yale. It has been accepted for inclusion in Cowles Foundation Discussion Papers by an authorized administrator of EliScholar – A Digital Platform for Scholarly Publishing at Yale. For more information, please contact elischolar@yale.edu.

ADAPTATION BOUNDS FOR CONFIDENCE BANDS UNDER SELF-SIMILARITY

By

Timothy B. Armstrong

October 2018

Revised July 2019

COWLES FOUNDATION DISCUSSION PAPER NO. 2146R



COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

Adaptation Bounds for Confidence Bands under Self-Similarity

Timothy B. Armstrong*
Yale University

July 18, 2019

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. Thanks to Richard Nickl helpful comments and discussion.

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 led 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 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 a more positive note, once ε is given, we 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 $\mathcal{F}(\tau)$ of function classes indexed by a regularity parameter $\tau \in \mathcal{T}$, 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 $\mathcal{F}(\tau)$, while also maintaining coverage $1-\alpha$ for each $\mathcal{F}(\tau)$ (so that $\mathcal{C}_n(\cdot) \in \mathcal{I}_{n,\alpha,\cup_{\tau\in\mathcal{T}}\mathcal{F}(\tau)}$). Suppose that a confidence band $\mathcal{C}_n(\cdot) \in \mathcal{I}_{n,\alpha,\cup_{\tau\in\mathcal{T}}\mathcal{F}(\tau)}$ achieves this goal up to a factor $A_n(\tau)$:

$$R_\beta(\mathcal{C}_n; \mathcal{F}(\tau)) \leq A_n(\tau) R^*_{n,\alpha,\beta}(\mathcal{F}(\tau), \mathcal{F}(\tau)) \text{ all } \tau \in \mathcal{T}.$$

We will call such a band *adaptive* to τ up to the *adaptation penalty* $A_n(\tau)$. 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 $R^*_{n,\alpha,\beta}(\mathcal{F}(\tau), \cup_{\tau\in\mathcal{T}}\mathcal{F}(\tau))/R^*_{n,\alpha,\beta}(\mathcal{F}(\tau), \mathcal{F}(\tau))$ 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) propose augmenting the Hölder condition with an auxiliary condition. Let $K : \mathbb{R}^2 \to \mathbb{R}$ be a function, called a kernel, such that $x \mapsto K(t, x)$ is of bounded variation for each t. Let $K_j(t, x) = 2^j K_j(2^j t, 2^j x)$ for any integer j, and let $\hat{f}(t, j) = \int K_j(t, x)\, dY(x)$. This allows for convolution kernels $K(t, x) = \tilde{K}(t - x)$ (in which case 2^{-j} is the bandwidth) and wavelet projection kernels $K(t, x) = \sum_k \phi(t - k)\phi(x - k)$ (in which case ϕ is the father wavelet and j is the resolution level). While the restriction to integer j is more natural for projection kernels than for convolution kernels, we maintain it throughout the paper in order to treat both cases with the same framework, following Giné and Nickl (2010). A previous version of this paper (Armstrong, 2018) focuses on convolution kernels and imposes the bound for all $h = \log_2 j$

on a set $(0, \overline{h}]$. The results are similar, although additional conditions are needed on the kernel and range of values of γ considered. Let $K_j f(t) = \int K(t,x) f(x)\, dx$. Note that $E_f \hat{f}(t,j) = K_j f(t)$, where E_f denotes expectation when $Y(x)$ is drawn according to f, so that the bias is given by $K_j f(t) - f(t)$. Under appropriate conditions on K, an upper bound on this bias for functions in $\mathcal{F}_{\text{Höl}}(\gamma, B)$ follows from standard calculations (see Appendix B):

$$\sup_{t \in [0,1]} |K_j f(t) - f(t)| \leq \tilde{C} B 2^{-j\gamma} \tag{2}$$

for some constant $\tilde{C}$. Giné and Nickl (2010) impose such a bound on bias directly, along with an analogous lower bound. For $\underline{j}, b_1, b_2 > 0$, let $\mathcal{F}_{\text{GN}}(\gamma, b_1, b_2) = \mathcal{F}_{\text{GN}}(\gamma, b_1, b_2; K, \underline{j})$ denote the set of functions f satisfying Condition 3 of Giné and Nickl (2010): for all integers $j \geq \underline{j}$,

$$b_1 2^{-j\gamma} \leq \sup_{t \in [0,1]} |K_j f(t) - f(t)| \leq b_2 2^{-j\gamma}. \tag{3}$$

Since we will also be imposing Hölder conditions, which, as noted above, satisfy the upper bound with $b_2 = \tilde{C} B$, 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 $\mathcal{F}_{\text{self-sim}}(\gamma, B, \varepsilon) = \mathcal{F}_{\text{self-sim}}(\gamma, B, \varepsilon; K, \underline{j})$ be the set of functions in $\mathcal{F}_{\text{Höl}}(\gamma, B)$ such that the lower bound in (3) holds with $b_1 = \varepsilon B$ for all integers $j \geq \underline{j}$. By the discussion above, this is equivalent to defining $\mathcal{F}_{\text{self-sim}}(\gamma, B, \varepsilon; K, \underline{j}) = \mathcal{F}_{\text{Höl}}(\gamma, B) \cap \mathcal{F}_{\text{GN}}(\gamma, \varepsilon B, CB; K, \underline{j})$ for any $C \geq \tilde{C}$. 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 $\mathcal{T}$ 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 $\mathcal{T}$ over which adaptation is required). For the lower bounds, we consider separately the cases of adaptation to B with γ known (i.e. $\mathcal{T} = \gamma \times [\underline{B}, \overline{B}]$) and adaptation to γ with B known

(i.e. $\mathcal{T} = [\underline{\gamma}, \overline{\gamma}] \times B$). 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 $\underline{\varepsilon}^{-1/(2\gamma+1)}$.

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 $f \in \mathcal{F}_{\text{self-sim}}(\gamma, B, \varepsilon_n)$ 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 with 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).

To our knowledge, this paper is the first to derive lower bounds on adaptation constants for 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. Our lower bounds apply to these weaker conditions as well. In addition, a large literature has considered adaptive confidence sets in related settings under conditions that are similar to the self-similarity condition used by Giné and Nickl (2010). In the Gaussian sequence setting, Szabó et al. (2015) propose a condition called a "polished tail" condition, and they show that this condition is weaker than a natural definition of self-similarity in that setting. They use this condition to show frequentist coverage of adaptive Bayesian credible sets (see also Sniekers and van der Vaart, 2015; van der Pas et al., 2017). Other applications of self-similarity type conditions include high dimensional sparse regression (Nickl and van de Geer, 2013), density estimation on the sphere (Kueh, 2012), locally adaptive confidence bands (Patschkowski and Rohde, 2019), binary regression (Mukherjee and Sen, 2018) and L_p confidence sets (Bull and Nickl, 2013; Carpentier, 2013; Nickl and Szabó, 2016). Self-similarity is also related to "signal strength" conditions used in other settings, such as "beta-min" conditions used to study variable selection in high dimensional regression (see Bühlmann and van de Geer, 2011, Section 7.4).

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 impose the following conditions on the kernel K:

$$\begin{aligned} &\text{there exists } C_K < \infty \text{ such that } K(y,x) = 0 \text{ for } |x-y| > C_K \\ &\text{and, for all } k \in \mathbb{Z} \text{ and } x, y \in \mathbb{R},\ K(y,x) = K(y-k, x-k). \end{aligned} \tag{4}$$

These conditions hold for convolution kernels with finite support, and for wavelet projection kernels for which the father wavelet has finite support.

We first consider adaptation to the constant B.

Theorem 2.1. *Let $\gamma > 0$ and let $0 < 2\alpha < \beta < 1$. Let K be a kernel satisfying (4). There exists $\underline{j}_{K,\gamma}$, $C_{K,\gamma,*} > 0$ and $\eta_{K,\gamma} > 0$ such that, for any $0 < \underline{B} \le B \le \overline{B}$, $\varepsilon \le \varepsilon' < \eta_{K,\gamma}$ and $\underline{\ell} \ge \underline{j}_{K,\gamma}$,*

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

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

Theorem 2.2. *Let $0 < \underline{\gamma} < \gamma \le \overline{\gamma}$ and let $0 < 2\alpha < \beta < 1$. Let K be a kernel that satisfies (4). There exist $C_{K,\overline{\gamma},*}$, $\underline{j}_{K,\overline{\gamma}}$ and $\eta_{K,\overline{\gamma}}$ depending only on K and $\overline{\gamma}$ such that, for all $\underline{\ell} \ge \underline{j}_{K,\overline{\gamma}}$ and $0 < \varepsilon \le \varepsilon' < \eta_{K,\overline{\gamma}}$,*

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

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 $\varepsilon' > \varepsilon$.

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.

We make some additional assumptions on the kernel:

$$\begin{array}{l} \sup_{t\in[0,1]} \int K(t,x)^2\,dx < \infty \text{ and there exists } \tau_K > 0 \\ \text{such that } \sup_{s,t\in[0,1]} \frac{\int [K(s,x)-K(t,x)]^2\,dx}{|s-t|^{\tau_K}} < \infty. \end{array} \tag{5}$$

Condition (5) is a mild continuity condition. For convolution kernels $K(y,x) = \tilde{K}(y-x)$ or wavelet projection kernels $K(y,x) = \sum_k \phi(y-k)\phi(x-k)$, it is sufficient for the kernel $\tilde{K}$ or father wavelet ϕ to be bounded with finite support and bounded first derivative (see Giné and Nickl, 2010, p. 1146 for the latter case).

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) and (5), such that, for some $\tilde{C}$, (2) holds for all $B \in [\underline{B}, \overline{B}]$ and all $\gamma \in [\underline{\gamma}, \overline{\gamma}]$. There exists a confidence band $\mathcal{C}_n(\cdot)$ and a constant $C^*_{K,\overline{\gamma},\tilde{C}}$ depending only on K, $\overline{\gamma}$ and $\tilde{C}$ such that, with probability approaching one uniformly over $\cup_{\gamma\in[\underline{\gamma},\overline{\gamma}]} \cup_{B\in[\underline{B},\overline{B}]} \mathcal{F}_{\text{self-sim}}(\gamma, B, \varepsilon)$,*

$$\sup_{x\in[0,1]} \text{length}\left(\mathcal{C}_n(x)\right) \leq C^*_{K,\overline{\gamma},\tilde{C}} \left(B\varepsilon^{-1}\right)^{1/(2\gamma+1)} \left(\sigma_n^2 \log(1/\sigma_n^2)\right)^{\gamma/(2\gamma+1)}$$

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

To prove this theorem, we construct a confidence band that has coverage for the class $\cup_{B\in[\underline{B},\overline{B}]} \cup_{\gamma\in[\underline{\gamma},\overline{\gamma}]} \mathcal{F}_{\text{GN}}(\gamma, \varepsilon B, B)$, such that the width is bounded by a constant times $(\varepsilon^{-1}B)^{1/(2\gamma+1)}(\sigma_n \log(1/\sigma_n))^{\gamma/(2\gamma+1)}$ with probability approaching one uniformly over the class $\mathcal{F}_{\text{GN}}(\gamma, \varepsilon B, B)$. Letting $\tilde{\varepsilon} = \varepsilon/\tilde{C}$ and $\tilde{B} = \tilde{C}B$, we have $\mathcal{F}_{\text{self-sim}}(\varepsilon, \gamma, B) \subseteq \mathcal{F}_{\text{GN}}(\gamma, \tilde{\varepsilon}\tilde{B}, \tilde{B})$ under (2), so that the conclusion of Theorem 2.3 holds for this confidence

band, constructed with $\tilde{\varepsilon} = \varepsilon/\tilde{C}$ in place of ε. We describe the confidence band here, with additional details in Appendix A.

Let $\Delta(j, j'; f) = \sup_{x\in[0,1]} |K_j f(x) - K_{j'} f(x)|$ and $\hat{\Delta}(j, j') = \sup_{x\in[0,1]} |\hat{f}(x, j) - \hat{f}(x, j')|$. Let $c(j)$ and $\tilde{c}(j, j')$ be critical values satisfying

$$|\hat{f}(x, j) - K_j f(x)| \leq c(j) \text{ all } x \in [0, 1],\ j \in \mathcal{J}_n \tag{6}$$

and

$$|\hat{\Delta}(j, j') - \Delta(j, j'; f)| \leq \tilde{c}(j, j') \text{ all } j, j' \in \mathcal{J}_n \tag{7}$$

with some prespecified probability for all $f \in \cup_{\gamma\in[\underline{\gamma},\overline{\gamma}]} \cup_{B\in[\underline{B},\overline{B}]} \mathcal{F}_{\mathrm{GN}}(\gamma, \varepsilon B, B)$, where $\mathcal{J}_n = \{\underline{\ell}_n, \underline{\ell}_n + 1, \ldots, \overline{\ell}_n\}$ for some $\underline{\ell}_n$, $\overline{\ell}_n$ (it suffices to set $c(j) = \bar{c}_K \sigma_n 2^{j/2}\sqrt{j}$ and $\tilde{c}(j, j') = c(j) + c(j')$ for a large enough constant $\bar{c}_K$ and to take $\underline{\ell}_n \to \infty$ with $\underline{\ell}_n/\log n \to 0$ and $\overline{\ell}_n/\log n \to \infty$; see Appendix A). We construct a confidence band that covers f for all $f \in \cup_{\gamma\in[\underline{\gamma},\overline{\gamma}]} \cup_{B\in[\underline{B},\overline{B}]} \mathcal{F}_{\mathrm{GN}}(\gamma, \varepsilon B, B; K, \underline{\ell}_n)$ on the event that (6) and (7) both hold.

To this end, we use $\Delta(j, j'; f)$ along with the self-similarity condition to bound the bias $|K_j f(x) - f(x)|$. This, along with the confidence bands $\hat{f}(x, j) \pm c(j)$ and $\hat{\Delta}(j, j') \pm \tilde{c}(j, j')$ for $K_j f(x)$ and $\Delta(j, j'; f)$ leads to a confidence band for f. First, note that, for $f \in \mathcal{F}_{\mathrm{GN}}(\gamma, \varepsilon B, B; K, \underline{\ell})$ and $j_1, j_2 \geq \underline{\ell}$,

$$\begin{aligned} B(\varepsilon 2^{-j_1\gamma} - 2^{-j_2\gamma}) &\leq \sup_{x\in[0,1]} |K_{j_1} f(x) - f(x)| - \sup_{x\in[0,1]} |K_{j_2} f(x) - f(x)| \\ &\leq \Delta(j_1, j_2; f) \leq \sup_{x\in[0,1]} |K_{j_1} f(x) - f(x)| + \sup_{x\in[0,1]} |K_{j_2} f(x) - f(x)| \leq B(2^{-j_1\gamma} + 2^{-j_2\gamma}) \end{aligned} \tag{8}$$

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

$$a(\varepsilon, j_1, j_2, j, \gamma_\ell, \gamma_u) = \max\left\{\varepsilon 2^{-\max\{(j_1-j)\gamma_u, (j_1-j)\gamma_\ell\}} - 2^{-\min\{(j_2-j)\gamma_u, (j_2-j)\gamma_\ell\}}, 0\right\}.$$

If $\gamma_\ell \leq \gamma \leq \gamma_u$ and $a(\varepsilon, j_1, j_2, j, \gamma_\ell, \gamma_u) > 0$, then $a(\varepsilon, j_1, j_2, j, \gamma_\ell, \gamma_u) \leq \frac{\varepsilon 2^{-j_1\gamma} - 2^{-j_2\gamma}}{2^{-j\gamma}}$ so that, for any $f \in \mathcal{F}_{\mathrm{GN}}(\gamma, \varepsilon B, B)$,

$$\sup_{x\in[0,1]} |K_j f(x) - f(x)| \leq B 2^{-j\gamma} \leq B \frac{\varepsilon 2^{-j_1\gamma} - 2^{-j_2\gamma}}{a(\varepsilon, j_1, j_2, j, \gamma_\ell, \gamma_u)} \leq \frac{\Delta(j_1, j_2; f)}{a(\varepsilon, j_1, j_2, j, \gamma_\ell, \gamma_u)} \tag{9}$$

where the last inequality uses (8).

In Appendix A.2, we provide an interval $[\hat{\gamma}_\ell, \hat{\gamma}_u]$ that contains γ on the event in (7). Letting $\hat{\jmath}$, $\hat{\jmath}_1$ and $\hat{\jmath}_2$ be data dependent values that are contained in $\mathcal{J}_n$ with probability one, it follows from (9) that, on the event that (6) and (7) both hold, the band

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

contains $f(x)$ for all $x \in [0, 1]$. Since $\hat{\jmath}_1$, $\hat{\jmath}_2$ and $\hat{\jmath}$ can be data dependent, we can simply choose them to minimize the length of this band. For concreteness, we will assume that $\mathcal{J}_n$ is finite for each n, so that a minimum is taken:

$$c(\hat{\jmath}) + \frac{\hat{\Delta}(\hat{\jmath}_1, \hat{\jmath}_2) + \tilde{c}(\hat{\jmath}_1, \hat{\jmath}_2)}{a(\varepsilon, \hat{\jmath}_1, \hat{\jmath}_2, \hat{\jmath}, \hat{\gamma}_\ell, \hat{\gamma}_u)} = \min_{j, j_1, j_2 \in \mathcal{J}_n} \left[c(j) + \frac{\hat{\Delta}(j_1, j_2) + \tilde{c}(j_1, j_2)}{a(\varepsilon, j_1, j_2, j, \hat{\gamma}_\ell, \hat{\gamma}_u)} \right],$$

where we use the convention that $\frac{\hat{\Delta}(j_1, j_2) + \tilde{c}(j_1, j_2)}{a(\varepsilon, j_1, j_2, j, \hat{\gamma}_\ell, \hat{\gamma}_u)}$ is equal to $+\infty$ if $a(\varepsilon, j_1, j_2, j, \hat{\gamma}_\ell, \hat{\gamma}_u) = 0$, so that the minimum is only over j, j_1, j_2 such that $a(\varepsilon, j_1, j_2, j, \hat{\gamma}_\ell, \hat{\gamma}_u) > 0$. The half-length of this band is then bounded by

$$\min_{j, j_1, j_2 \in \mathcal{J}_n} \left[c(j) + \frac{B(2^{-j_1\gamma} + 2^{-j_2\gamma}) + 2\tilde{c}(j_1, j_2)}{a(\varepsilon, j_1, j_2, j, \hat{\gamma}_\ell, \hat{\gamma}_u)} \right] \tag{10}$$

on the event that (6) and (7) both hold (here we use the upper bound in (8)). In Appendix A.3, we use this bound to show that this confidence band, constructed with $\tilde{\varepsilon} = \varepsilon / \tilde{C}$ in place of ε, satisfies the requirements of Theorem 2.3.

2.3 Alternative Definition of Self-Similarity Constant

We have defined $\mathcal{F}_{\text{self-sim}}(\gamma, B, \varepsilon)$ to be the class of functions in $\mathcal{F}_{\text{Höl}}(\gamma, B)$ such that the lower bound in (3) holds with $b_1 = \varepsilon B$. Under (2), this means that the self-similarity constant ε gives the ratio between the upper and lower bound on bias, up to the constant $\tilde{C}$. 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 $\overline{\mathcal{F}}_{\text{self-sim}}(\gamma, B, b_1) = \mathcal{F}_{\text{self-sim}}(\gamma, B, b_1/B)$. 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 $\cup_{B \in [\underline{B}, \overline{B}]} \overline{\mathcal{F}}_{\text{self-sim}}(\gamma, B, b_1) =$

$\overline{\mathcal{F}}_{\text{self-sim}}(\gamma, \overline{B}, b_1) = \mathcal{F}_{\text{self-sim}}(\gamma, \overline{B}, b_1/\overline{B})$. As the next theorem shows, this leads to a much more negative result: adaptation to the Hölder constant is completely impossible.

Theorem 2.4. *Let $\gamma > 0$ and let $0 < 2\alpha < \beta < 1$. Let K be a kernel satisfying (4). There exists $\underline{j}_{K,\gamma}$, $C_{K,\gamma,*} > 0$ and $\eta_{K,\gamma} > 0$ such that, for any $0 < B \le \overline{B}$, $b_1 \le \eta_{K,\gamma} B$ and $\underline{\ell} \ge \underline{j}_{K,\gamma}$,*

$$\begin{aligned} &R^*_{n,\alpha,\beta}(\overline{\mathcal{F}}_{\text{self-sim}}(\gamma, B, b_1; K, \underline{\ell}), \overline{\mathcal{F}}_{\text{self-sim}}(\gamma, \overline{B}, b_1; K, \underline{\ell})) \\ &\ge (1 + o(1)) C_{K,\gamma,*} \overline{B}^{1/(2\gamma+1)} \left(\sigma_n^2 \log(1/\sigma_n)\right)^{\gamma/(2\gamma+1)}. \end{aligned}$$

2.4 Discussion

The confidence band in Section 2.2 builds on the important work of 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). 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 (note, however, that 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.

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. We begin with bounds based on minimax testing (Section 3.1). We then construct self-similar functions that can be used along with these testing bounds to prove our results (Section 3.2). Finally, we combine these results to complete the proofs (Section 3.3).

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

$$P_{f_0}\left(\sup_{x \in [0,1]} \text{length}\left(\mathcal{C}_n(x)\right) > R\right) = 1 - P_{f_0}\left(\sup_{x \in [0,1]} \text{length}\left(\mathcal{C}_n(x)\right) \leq R\right) \leq 1 - \beta. \tag{11}$$

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

$$\inf_{f \in \mathcal{F} \cap \{f : \sup_{x \in [0,1]} |f(x) - f_0(x)| \geq \tilde{R}\}} E_f \phi \geq 1 - \alpha \tag{12}$$

(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 (11), 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 (12), 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 the first step, we use the following result, which is immediate from slight modifications of arguments in Lepski and Tsybakov (2000). 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 $[a,b]$. 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 3.2. *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 $\eta > 0$ 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 $C(\gamma, B, \kappa) = \left[\frac{4}{2\gamma+1} B^{1/\gamma}/\|\kappa\|^2\right]^{\frac{\gamma}{2\gamma+1}} \kappa(0)$. 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

$$\begin{aligned}
h_n &= \left(\frac{(1-\eta)C(\gamma, B, \kappa)}{B\kappa(0)}\right)^{1/\gamma} \left(\sigma_n^2 \log(1/\sigma_n)\right)^{1/(2\gamma+1)}, \\
M_n &= \left\lfloor \frac{b-a}{2A_\kappa h_n} \right\rfloor - 1, \quad x_{n,k} = a + (2k-1)A_\kappa h_n, \quad k = 1, \ldots, M_n \\
f_{k,n}(x) &= B h_n^\gamma \kappa\left(\frac{x - x_{n,k}}{h_n}\right).
\end{aligned}$$

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

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

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

Following arguments on pp. 35-36 of Lepski and Tsybakov (2000), it will then follow that $\sup_{\gamma\in[\underline{\gamma},\overline{\gamma}],B\in[\underline{B},\overline{B}]} d_{\text{test}}(\{0\},\{f_{n,1},f_{n,2},\ldots,f_{n,M_n}\}) \to 0$ so long as there exists $\delta > 0$ such that, for large enough n, $(s_n^2/\sigma_n^2)/(2\log M_n) \leq (1-\delta)$ for all $\gamma \in [\underline{\gamma},\overline{\gamma}]$ and $B \in [\underline{B},\overline{B}]$. Since each $f_{k,n}$ is contained in the set $\widetilde{\mathcal{F}}(\gamma,B,a,b) \cap \{f : \sup_{x\in[a,b]} |f(x)| = c_n(\gamma,B)\}$, this will complete the proof.

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

$$\begin{aligned}\frac{s_n^2}{\sigma_n^2} &= B^2\|\kappa\|^2 h_n^{2\gamma+1}\sigma_n^{-2} = B^2\|\kappa\|^2\left(\frac{(1-\eta)C(\gamma,B,\kappa)}{B\kappa(0)}\right)^{(2\gamma+1)/\gamma}\log(1/\sigma_n)\\ &= (1-\eta)^{(2\gamma+1)/\gamma}\frac{4}{2\gamma+1}\log(1/\sigma_n).\end{aligned}$$

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, $(s_n^2/\sigma_n^2)/(2\log M_n) \leq (1-\delta)$.

□

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 $\mathcal{F}_{\text{self-sim}}(\gamma,B,\varepsilon)$ for any $\varepsilon > 0$. 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.

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

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

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

$$d_{\text{test}}(\mathcal{F}, \mathcal{G} + \{g_0 - f_0\}) = \sup_{\phi} \inf_{f \in \mathcal{F},\, g \in \mathcal{G}} |E_{g+f_0-g_0}\phi(Y) - E_f\phi(Y)|$$

$$\leq \sup_{\phi} \inf_{f \in \mathcal{F},\, g \in \mathcal{G}} \left[|E_{g+f_0-g_0}\phi(Y) - E_g\phi(Y)| + |E_g\phi(Y) - E_f\phi(Y)| \right].$$

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 $1/\sqrt{2\pi} \leq 1$. □

3.2 Constructing Functions in Self-Similarity Classes

Let $\psi : \mathbb{R} \to \mathbb{R}$ be a function with $\|\psi\| = 1$ with support contained in $(-C_\psi, C_\psi)$ where $C_\psi < \infty$. Let $\psi_{\ell k}(x) = 2^{\ell/2}\psi(2^\ell x - k)$. Let k^* be a positive integer with $k^* > 3C_\psi$. Note that the lower endpoint of the support of $\psi_{\ell k^*}$ is $-2^{-\ell}C_\psi + 2^{-\ell}k^*$ and the upper endpoint of the support of $\psi_{(\ell+1)k^*}$ is $2^{-(\ell+1)}C_\psi + 2^{-(\ell+1)}k^*$, so that the supports of these functions do not overlap so long as $-2C_\psi + 2k^* \geq C_\psi + k^*$, which is guaranteed by the condition $k^* > 3C_\psi$. Furthermore, this guarantees that the support of each $\psi_{\ell k^*}$ is contained in $(0, \infty)$ and does not overlap with the support of $\psi_{\ell' k^*}$ for any $\ell' \neq \ell$. Given a positive integer $\underline{\ell}$ and a sequence $\left\{\tilde{\beta}_\ell\right\}_{\ell=\underline{\ell}}^{\infty}$ of real numbers, we will consider functions that take the form

$$f_{\{\tilde{\beta}\},\underline{\ell}}(x) = \sum_{\ell=\underline{\ell}}^{\infty} \tilde{\beta}_\ell \psi_{\ell k^*}(x). \tag{13}$$

Note that, since the support of each $\psi_{\ell k^*}$ is nonoverlapping and $\|\psi_{\ell k^*}\| = \|\psi\| = 1$, we have $\|f_{\{\tilde{\beta}\},\underline{\ell}}\|^2 = \sum_{\ell=\underline{\ell}}^{\infty} \tilde{\beta}_\ell^2$. If ψ is a mother wavelet for some wavelet basis, then f has ℓ, kth wavelet coefficient given by $\tilde{\beta}_\ell$ for $\ell \geq \underline{\ell}$ and $k = k^*$ and ℓ, kth wavelet coefficent 0 for all other ℓ, k. However, we do not require that ψ be a mother wavelet.

A Hölder condition for such functions can be obtained from the rate of decay of the

coefficients $\tilde{\beta}_\ell$. For a function $f : \mathbb{R} \to \mathbb{R}$, let $\|f\|_\infty = \sup_{t\in\mathbb{R}} |f(t)|$ denote the L_∞ norm of f.

Lemma 3.4. *Let $\gamma > 0$ and suppose that ψ is $\lfloor\gamma\rfloor + 1$ times differentiable. Let A be given and let $f(x) = f_{\{\tilde{\beta}\},\underline{\ell}}(x)$ be given by (13) where $|\tilde{\beta}_\ell| \le A2^{-\ell(\gamma+1/2)}$ for all ℓ. Then $f \in \mathcal{F}_{\text{Höl}}(\gamma, 2A\|\psi^{(\lfloor\gamma\rfloor+1)}\|_\infty(2C_\psi)^{1-(\gamma-\lfloor\gamma\rfloor)})$.*

Proof. By Lemma 3.5, it suffices to show that $x \mapsto \tilde{\beta}_\ell\psi_{\ell k^*}(x)$ is in $\mathcal{F}_{\text{Höl}}(\gamma, A\|\psi^{(\lfloor\gamma\rfloor)+1}\|_\infty(2C_\psi)^{1-(\gamma-\lfloor\gamma\rfloor)})$ for each ℓ. Given ℓ, let x and x' be in the support of $\psi_{\ell k^*}$ so that $x, x' \in [2^{-\ell}k^* - 2^{-\ell}C_\psi, 2^{-\ell}k^* + 2^{-\ell}C_\psi]$. Then

$$
\begin{aligned}
&\left|\tilde{\beta}_\ell\psi_{\ell k^*}^{(\lfloor\gamma\rfloor)}(x) - \tilde{\beta}_\ell\psi_{\ell k^*}^{(\lfloor\gamma\rfloor)}(x')\right| \\
&= |\tilde{\beta}_\ell|2^{\ell(\lfloor\gamma\rfloor+1/2)}\left|\psi^{(\lfloor\gamma\rfloor)}(2^\ell x + k) - \psi^{(\lfloor\gamma\rfloor)}(2^\ell x' + k)\right| \le \|\psi^{(\lfloor\gamma\rfloor+1)}\|_\infty \cdot |\tilde{\beta}_\ell|2^{\ell(\lfloor\gamma\rfloor+1/2)} \cdot 2^\ell|x - x'| \\
&= \|\psi^{(\lfloor\gamma\rfloor+1)}\|_\infty \cdot |\tilde{\beta}_\ell|2^{\ell(\lfloor\gamma\rfloor+1/2)} \cdot (2C_\psi) \cdot (2C_\psi)^{-1}2^\ell|x - x'| \\
&\le \|\psi^{(\lfloor\gamma\rfloor+1)}\|_\infty \cdot |\tilde{\beta}_\ell|2^{\ell(\lfloor\gamma\rfloor+1/2)} \cdot (2C_\psi) \cdot (2C_\psi)^{-(\gamma-\lfloor\gamma\rfloor)}2^{\ell(\gamma-\lfloor\gamma\rfloor)}|x - x'|^{\gamma-\lfloor\gamma\rfloor}
\end{aligned}
$$

where the last inequality uses the fact that $(2C_\psi)^{-1}2^\ell|x - x'| \le 1$ by the conditions on x, x'. If $|\tilde{\beta}_\ell| \le A2^{-\ell(\gamma+1/2)}$, then this is bounded by $A\|\psi^{(\lfloor\gamma\rfloor+1)}\|_\infty(2C_\psi)^{1-(\gamma-\lfloor\gamma\rfloor)}|x - x'|^{\gamma-\lfloor\gamma\rfloor}$ as required. □

We have used the following lemma.

Lemma 3.5. *Let $\{g_k\}_{k=1}^\infty$ be a sequence of functions with nonoveralapping support with $g_k \in \mathcal{F}_{\text{Höl}}(\gamma, B)$ for each k. Let $f = \sum_{k=1}^\infty g_k$. Then $f \in \mathcal{F}_{\text{Höl}}(\gamma, 2B)$.*

Proof. Let x, x' be given. We need to show that $|f^{\lfloor\gamma\rfloor}(x) - f^{\lfloor\gamma\rfloor}(x')| \le 2B|x - x'|^{\gamma-\lfloor\gamma\rfloor}$. If x and x' are both in the support of g_k for some k, or if x and x' are not in the support of g_k for any k, then this follows immediately. If x is in the support of g_k and x' is in the support of $g_{k'}$ for some $k' \ne k$, let $\overline{x}$ denote the upper endpoint of the support of g_k and let $\underline{x}'$ denote the lower endpoint of the support of $g_{k'}$, and assume without loss of generality that $\overline{x} \le \underline{x}'$. By the Hölder condition on g_k and $g_{k'}$, we have $g_k^{\lfloor\gamma\rfloor}(\overline{x}) = g_{k'}^{\lfloor\gamma\rfloor}(\underline{x}') = 0$, so that $|f^{\lfloor\gamma\rfloor}(x) - f^{\lfloor\gamma\rfloor}(x')| = |g_k^{\lfloor\gamma\rfloor}(x) - g_k^{\lfloor\gamma\rfloor}(\overline{x}) + g_{k'}^{\lfloor\gamma\rfloor}(x) - g_{k'}^{\lfloor\gamma\rfloor}(\underline{x}')| \le B|x - \overline{x}|^{\gamma-\lfloor\gamma\rfloor} + B|x' - \overline{x}'|^{\gamma-\lfloor\gamma\rfloor} \le 2B|x - x'|^{\gamma-\lfloor\gamma\rfloor}$. Finally, if x is in the support of some g_k and x' is not in the support of $g_{k'}$ for any k', then, letting $[\underline{x}, \overline{x}]$ denote the support of g_k, $|f^{\lfloor\gamma\rfloor}(x) - f^{\lfloor\gamma\rfloor}(x')| = |g_k^{\lfloor\gamma\rfloor}(x)| \le B\min\{|x - \underline{x}|^{\gamma-\lfloor\gamma\rfloor}, |x - \overline{x}|^{\gamma-\lfloor\gamma\rfloor}\} \le B|x - x'|^{\gamma-\lfloor\gamma\rfloor}$. □

Lemma 3.6. *Suppose that $K(y,x)$ satisfies (4), and let $f_{\{\tilde{\beta}\},\underline{\ell}}$ be defined as in (13), with $k^* > 4(C_\psi + C_K)$. Let f^* be a function supported on the set $(2^{-\underline{\ell}}(k^* + C_\psi + 2C_K), \infty)$, and let $f = f_{\{\tilde{\beta}\},\underline{\ell}} + f^*$. Then, for $j \geq \underline{\ell}$,*

$$\sup_{x \in [0, 2^{-j}(k^* + C_\psi + C_K)]} |K_j f(x) - f(x)| \geq |\tilde{\beta}_j| \cdot 2^{j/2} \sup_{x \in \mathbb{R}} |K_0 \psi(x) - \psi(x)|.$$

Proof. Note that $K_j(y,x) = 2^j K(2^j y, 2^j x) = 0$ whenever $|x - y| > 2^{-j} C_K$. Thus, for any function g with support contained in (a,b) for some $a < b$, the support of $K_j g$ is contained in $(a - 2^{-j}C_K, b + 2^{-j}C_K)$. In particular, the support of $K_j \psi_{\ell k^*}(x)$ is contained in $(2^{-\ell}k^* - 2^{-\ell}C_\psi - 2^{-j}C_K, 2^{-\ell}k^* + 2^{-\ell}C_\psi + 2^{-j}C_K)$. Let S_j denote this set with $\ell = j$. We will argue that S_j does not overlap with the support of $K_j f^* - f^*$ or $K_j \psi_{\ell k^*} - \psi_{\ell k^*}$ for $\ell \neq j$. This will imply that, for $x \in S_j$ and $j \geq \underline{\ell}$, $K_j f(x) - f(x) = \tilde{\beta}_j \left[K_j \psi_{j k^*}(x) - \psi_{j k^*}(x)\right]$. This gives the result since

$$\begin{aligned} &\sup_{x \in [0, 2^{-j}(k^* + C_\psi + C_K)]} |K_j f(x) - f(x)| \geq \sup_{x \in S_j} |K_j f(x) - f(x)| \\ &= |\tilde{\beta}_j| \sup_{x \in \mathbb{R}} |K_j \psi_{j k^*}(x) - \psi_{j k^*}(x)| = |\tilde{\beta}_j| \cdot 2^{j/2} \sup_{x \in \mathbb{R}} |K_0 \psi(x) - \psi(x)| \end{aligned}$$

where the last step follows by using a change of variables to note that $K_j \psi_{j k^*}(x) - \psi_{j k^*}(x) = 2^{j/2} \left[K_0 \psi(u - k^*) - \psi(u - k^*)\right]$.

To complete the proof, we need to show that S_j does not overlap with the support of $K_j f^* - f^*$ or $K_j \psi_{\ell k^*} - \psi_{\ell k^*}$ for $\ell \neq j$. For any $\ell \geq j + 1$, the upper support point of $K_j \psi_{\ell k^*} - \psi_{\ell k^*}$ is no greater than $2^{-\ell}k^* + 2^{-\ell}C_\psi + 2^{-j}C_K \leq 2^{-j-1}k^* + 2^{-j-1}C_\psi + 2^{-j}C_K$. Thus, the support of $K_j \psi_{\ell k^*}$ does not overlap with S_j so long as $2^{-j-1}k^* + 2^{-j-1}C_\psi + 2^{-j}C_K < 2^{-j}k^* - 2^{-j}C_\psi - 2^{-j}C_K$, which holds so long as $3C_\psi + 4C_K < k^*$. This is guaranteed by the condition $k^* > 4(C_\psi + C_K)$. For $\ell \leq j - 1$, the lower support point of $K_j \psi_{\ell k^*} - \psi_{\ell k^*}$ is no less than $2^{-\ell}k^* - 2^{-\ell}C_\psi - 2^{-j}C_K \geq 2^{-j+1}k^* - 2^{-j+1}C_\psi - 2^{-j}C_K$. Thus, the support of $K_j \psi_{\ell k^*}$ does not overlap with S_j so long as $2^{-j}k^* + 2^{-j}C_\psi + 2^{-j}C_K < 2^{-j+1}k^* - 2^{-j+1}C_\psi - 2^{-j}C_K$, which holds so long as $3C_\psi + 2C_K < k^*$. This is guaranteed by the condition $k^* > 4(C_\psi + C_K)$. Finally, the lower support point of $K_j f^* - f^*$ is bounded from below by $2^{-\underline{\ell}}(k^* + C_\psi + C_K)$, so that the support of $K_j f^* - f^*$ does not overlap with S_j.

□

Given $0 < \varepsilon < 1$ and $0 < \gamma - \delta \leq \gamma < \infty$, let $\tilde{f}_{\underline{\ell},\gamma,\delta,\varepsilon,1}$ be defined as in (13) with

$$\tilde{\beta}_\ell = \max\{2^{-\ell(\gamma+1/2)}, \varepsilon 2^{-\ell(\gamma-\delta+1/2)}\}.$$

Let $\tilde{g}_{\underline{\ell},\gamma,1}$ be defined as in (13) with

$$\tilde{\beta}_\ell = 2^{-\ell(\gamma+1/2)}.$$

Let $\tilde{f}_{\underline{\ell},\gamma,\delta,\varepsilon,A}(x) = A\tilde{f}_{\underline{\ell},\gamma,\delta,\varepsilon,1}(x)$ and let $\tilde{g}_{\underline{\ell},\gamma,A}(x) = A\tilde{g}_{\underline{\ell},\gamma,1}(x)$. The next two lemmas construct self-similar functions from $\tilde{f}_{\underline{\ell},\gamma,\delta,\varepsilon,A}$ and $\tilde{g}_{\underline{\ell},\gamma,A}$. Let ψ be given, and let $\underline{C}_{K,\psi} = \sup_{x\in[0,1]} |K_0\psi(x) - \psi(x)|$ and $\overline{C}_{K,\psi,\gamma} = 2\|\psi^{(\lfloor\gamma\rfloor+1)}\|_\infty (2C_\psi)^{1-(\gamma-\lfloor\gamma\rfloor)}$. Note that ψ can be chosen so that $\overline{C}_{K,\psi,\gamma}$ is bounded from above over $\gamma \leq \overline{\gamma}$, and so that $\underline{C}_{K,\psi} > 0$. Recall that $\widetilde{\mathcal{F}}(\gamma, B, a, b)$ denotes the class of functions in $\mathcal{F}_{\text{Höl}}(\gamma, B)$ with support in $[a, b]$.

Lemma 3.7. *Let $0 < a < b$, $A > 0$ and $\tilde{B} \geq 0$ be given, and let K be a kernel that satisfies (4). Let $k^* > 4(C_\psi + C_K)$, and let $\underline{\ell}$ be large enough so that $2^{-\underline{\ell}}(k^* + C_\psi + C_K) < a$. Then, for any $A^* \geq \overline{C}_{K,\psi,\gamma}A + \tilde{B}$ and $\varepsilon^* \leq \underline{C}_{K,\psi}A/A^*$,*

$$\widetilde{\mathcal{F}}(\gamma, \tilde{B}, a, b) + \{\tilde{g}_{\underline{\ell},\gamma,A}\} \subseteq \mathcal{F}_{\text{self-sim}}(\gamma, A^*, \varepsilon^*; K, \underline{\ell}).$$

Proof. Let $f^* \in \widetilde{\mathcal{F}}(\gamma, \tilde{B}, a, b)$ and let $f = f^* + \tilde{g}_{\underline{\ell},\gamma,A}$. It follows from Lemma 3.4 that $\tilde{g}_{\underline{\ell},\gamma,A} \in \mathcal{F}_{\text{Höl}}(\gamma, \overline{C}_{K,\psi,\gamma}A)$, so that $f \in \mathcal{F}_{\text{Höl}}(\gamma, \overline{C}_{K,\psi,\gamma}A + \tilde{B}) \subseteq \mathcal{F}_{\text{Höl}}(\gamma, A^*)$. From Lemma 3.6, it follows that, for $j \geq \underline{\ell}$, $\sup_{x\in[0,1]} |K_j f(x) - f(x)| \geq A2^{-j(\gamma+1/2)} \cdot 2^{j/2}\underline{C}_{K,\psi} = A2^{-j\gamma}\underline{C}_{K,\psi} = (\underline{C}_{K,\psi}A/A^*) \cdot A^* \cdot 2^{-j\gamma} \geq \varepsilon^* A^* \cdot 2^{-j\gamma}$. Thus, $f \in \mathcal{F}_{\text{self-sim}}(\gamma, A^*, \varepsilon^*; K, \underline{\ell})$.

□

Lemma 3.8. *Let $0 < a < b$, $\tilde{\varepsilon} > 0$, $A > 0$ and $\tilde{B} \geq 0$ be given, and let K be a kernel that satisfies (4). Let $k^* > 4(C_\psi + C_K)$, and let $\underline{\ell}$ be large enough so that $2^{-\underline{\ell}}(k^* + C_\psi + C_K) < a$. Then, for any $A^* \geq \overline{C}_{K,\psi,\gamma-\delta}A + \tilde{B}$ and $\varepsilon^* \leq \tilde{\varepsilon}\underline{C}_{K,\psi}A/A^*$,*

$$\widetilde{\mathcal{F}}(\gamma - \delta, \tilde{B}, a, b) + \{\tilde{f}_{\underline{\ell},\gamma,\delta,\tilde{\varepsilon},A}\} \subseteq \mathcal{F}_{\text{self-sim}}(\gamma - \delta, A^*, \varepsilon^*; K, \underline{\ell}).$$

Proof. Let $f^* \in \widetilde{\mathcal{F}}(\gamma - \delta, \tilde{B}, a, b)$ and let $f = f^* + \tilde{f}_{\underline{\ell},\gamma,\delta,\tilde{\varepsilon},A}$. It follows from Lemma 3.4 that $\tilde{f}_{\underline{\ell},\gamma,\delta,\tilde{\varepsilon},A} \in \mathcal{F}_{\text{Höl}}(\gamma - \delta, \overline{C}_{K,\psi,\gamma-\delta}A)$ so that $f \in \mathcal{F}_{\text{Höl}}(\gamma - \delta, \overline{C}_{K,\psi,\gamma-\delta}A + \tilde{B}) \subseteq \mathcal{F}_{\text{Höl}}(\gamma - \delta, A^*)$. From Lemma 3.6, it follows that, for $j \geq \underline{\ell}$, $\sup_{x\in[0,1]} |K_j f(x) - f(x)| \geq \tilde{\varepsilon}A2^{-j(\gamma-\delta+1/2)} \cdot 2^{j/2}\underline{C}_{K,\psi} = \tilde{\varepsilon}A2^{-j(\gamma-\delta)}\underline{C}_{K,\psi} = \tilde{\varepsilon}(\underline{C}_{K,\psi}A/A^*) \cdot A^* \cdot 2^{-j(\gamma-\delta)} \geq \varepsilon^* \cdot A^* \cdot 2^{-j(\gamma-\delta)}$. Thus, $f \in \mathcal{F}_{\text{self-sim}}(\gamma - \delta, A^*, \varepsilon^*; K, \underline{\ell})$.

□

3.3 Testing Bounds for Self-Similar Functions

According to Lemma 3.7, we can obtain bounds for adaptation to the Hölder constant subject to coverage over self-similarity classes using the classes $\widetilde{\mathcal{F}}(\gamma, B, a, b)+\{\tilde{g}_{\underline{\ell},\gamma,A}\}$, thereby completing the proofs of Theorems 2.1 and 2.2. Similarly, Lemma 3.8 allows us to obtain bounds for adaptation to the Hölder exponent using the classes $\widetilde{\mathcal{F}}(\gamma, B, a, b) + \{\tilde{f}_{\underline{\ell},\gamma,\delta,\varepsilon,A}\}$, thereby completing the proof of Theorem 2.4. To obtain these bounds, we use the results from Section 3.1. We begin with a bound that will be useful for adaptation to the constant.

Lemma 3.9. *Let $A > 0$, $B > 0$ and $0 < a < b$ be given, and let K be a kernel that satisfies (4). Let $k^* > 4(C_\psi + C_K)$, and let $\underline{\ell}$ be large enough so that $2^{-\underline{\ell}}(k^* + C_\psi + C_K) < a$. Let*

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

where κ_γ^* is a function in $\mathcal{F}_{\text{Höl}}(\gamma, 1)$ with compact support. Then, if $0 < 2\alpha < \beta < 1$,

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

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

We are now ready to prove Theorems 2.1 and 2.4.

Proof of Theorem 2.1. Let k^* and $\underline{\ell}$ be chosen so that $k^* > 4(C_\psi + C_K)$ and $2^{-\underline{\ell}}(k^* + C_\psi + C_K) < 1/2$. Let $A = B/(2\max\{\overline{C}_{K,\psi,\gamma}, 1\})$. Then, by Lemma 3.7, $\tilde{g}_{\underline{\ell},\gamma,A} \in \mathcal{F}_{\text{self-sim}}(\gamma, B, \varepsilon'; K, \underline{\ell})$ so long as $\varepsilon' \leq \underline{C}_{K,\psi}/(2\max\{\overline{C}_{K,\psi,\gamma}, 1\})$. Let $\tilde{B} = \min\{\tilde{\varepsilon}^{-1}B, \overline{B}\} - \overline{C}_{K,\psi,\gamma}A$ where $\tilde{\varepsilon} = 2\varepsilon\max\{\overline{C}_{K,\psi,\gamma}, 1\}/\underline{C}_{K,\psi}$. Applying Lemma 3.7 with $\min\{\tilde{\varepsilon}^{-1}B, \overline{B}\}$ playing the role of A^*, we have $\widetilde{\mathcal{F}}(\gamma, \tilde{B}, 1/2, 1) + \{\tilde{g}_{\underline{\ell},\gamma,A}\} \subseteq \mathcal{F}_{\text{self-sim}}(\gamma, \min\{\tilde{\varepsilon}^{-1}B, \overline{B}\}, \varepsilon; K, \underline{\ell})$, where we use the fact that the choice of $\tilde{\varepsilon}$ guarantees $\underline{C}_{K,\psi}A/A^* \geq \varepsilon$. If $\eta_{K,\gamma}$ is small enough, then we will have $\min\{\tilde{\varepsilon}^{-1}B, \overline{B}\} \in [\underline{B}, \overline{B}]$, so that this implies $\widetilde{\mathcal{F}}(\gamma, \tilde{B}, 1/2, 1)+\{\tilde{g}_{\underline{\ell},\gamma,A}\} \subseteq \cup_{B'\in[\underline{B},\overline{B}]}\mathcal{F}_{\text{self-sim}}(\gamma, B', \varepsilon; K, \underline{\ell})$. Applying Lemma 3.9, it follows that $R^*_{n,\alpha,\beta}(\mathcal{F}_{\text{self-sim}}(\gamma, B, \varepsilon'; K, \underline{\ell}), \cup_{B'\in[\underline{B},\overline{B}]}\mathcal{F}_{\text{self-sim}}(\gamma, B', \varepsilon; K, \underline{\ell}))$ is bounded from below by $(1+o(1))\tilde{B}^{1/(2\gamma+1)} \left(\sigma_n^2 \log(1/\sigma_n)\right)^{\gamma/(2\gamma+1)}$ times a term that depends only on γ. The result follows by noting that, if $\eta_{K,\gamma}$ is chosen small enough, then $\tilde{B}$ is bounded

from below by a constant times $\min\{\varepsilon^{-1}B, \overline{B}\}$, where the constant depends only on $\underline{C}_{K,\psi}$ and $\overline{C}_{K,\psi,\gamma}$.

□

Proof of Theorem 2.4. Let k^* and $\underline{\ell}$ be chosen so that $k^* > 4(C_\psi + C_K)$ and $2^{-\underline{\ell}}(k^* + C_\psi + C_K) < 1/2$. Let $A = B/(2\max\{\overline{C}_{K,\psi,\gamma}, 1\})$. Then, by Lemma 3.7, $\tilde{g}_{\underline{\ell},\gamma,A} \in \mathcal{F}_{\text{self-sim}}(\gamma, B, b_1/B; K, \underline{\ell}) = \overline{\mathcal{F}}_{\text{self-sim}}(\gamma, B, b_1; K, \underline{\ell})$ so long as $b_1/B \leq \underline{C}_{K,\psi}/(2\max\{\overline{C}_{K,\psi,\gamma}, 1\})$. Let $\tilde{B} = \overline{B} - \overline{C}_{K,\psi,\gamma}A = \overline{B} - B\overline{C}_{K,\psi,\gamma}/(2\max\{\overline{C}_{K,\psi,\gamma}, 1\})$. Applying Lemma 3.7 with $\overline{B}$ playing the role of A^*, we have $\widetilde{\mathcal{F}}(\gamma, \tilde{B}, 1/2, 1) + \{\tilde{g}_{\underline{\ell},\gamma,A}\} \subseteq \mathcal{F}_{\text{self-sim}}(\gamma, \overline{B}, b_1/\overline{B}; K, \underline{\ell}) = \overline{\mathcal{F}}_{\text{self-sim}}(\gamma, \overline{B}, b_1; K, \underline{\ell})$, so long as $b_1 \leq \underline{C}_{K,\psi}A = \underline{C}_{K,\psi}B/(2\max\{\overline{C}_{K,\psi,\gamma}, 1\})$. The result follows by applying Lemma 3.9 and noting that $\tilde{B} \geq \overline{B}/2$.

□

For adaptation to the exponent, we will use testing bounds for the classes $\{\tilde{f}_{\underline{\ell},\gamma,\delta_n,\varepsilon,A}\}$ and $\widetilde{\mathcal{F}}(\gamma, A, a, b) + \{\tilde{g}_{\underline{\ell},\gamma,A}\}$ 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 $\|\tilde{f}_{\underline{\ell},\gamma,\delta_n,\varepsilon,A} - \tilde{g}_{\underline{\ell},\gamma,A}\|/\sigma_n$, 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 3.10. *Let $\delta_n = C_n/\log n$ where $C_n = (1-b_n)(2\gamma+1)\log\varepsilon^{-1}$ with $b_n = 1/(\log n)^{1/2}$. Then*

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

Proof. It suffices to prove the result for $A = 1$. We have

$$\|\tilde{f}_{\underline{\ell},\gamma,\delta,\varepsilon,1} - \tilde{g}_{\underline{\ell},\gamma,1}\|^2 = \sum_{\ell=\tilde{\ell}}^{\infty} \left(\varepsilon 2^{-\ell(\gamma-\delta+1/2)} - 2^{-\ell(\gamma+1/2)}\right)^2 = \sum_{\ell=\tilde{\ell}}^{\infty} 2^{-\ell(2\gamma+1)} \left(\varepsilon 2^{\ell\delta} - 1\right)^2$$

where $\tilde{\ell} = \tilde{\ell}(\varepsilon, \delta)$ is the minimum value of $\ell \geq \underline{\ell}$ such that $\varepsilon 2^{\ell\delta} > 1$. The above display is bounded by

$$\varepsilon^2 \sum_{\ell=\tilde{\ell}}^{\infty} 2^{-\ell(2(\gamma-\delta)+1)} = \varepsilon^2 \sum_{\ell=0}^{\infty} 2^{-(\ell+\tilde{\ell})(2(\gamma-\delta)+1)} = \varepsilon^2 2^{-\tilde{\ell}(2(\gamma-\delta)+1)} \sum_{\ell=0}^{\infty} 2^{-\ell(2(\gamma-\delta)+1)}.$$

Note that $2^{-\tilde{\ell}} < \varepsilon^{1/\delta}$, so $2^{-\tilde{\ell}(2(\gamma-\delta)+1)} < \varepsilon^{(2(\gamma-\delta)+1)/\delta}$. From this and the bound $\sum_{\ell=0}^{\infty} 2^{-\ell(2(\gamma-\delta)+1)} \leq \sum_{\ell=0}^{\infty} 2^{-\ell} = 2$, it follows that the above display is bounded by $2\varepsilon^{2+(2(\gamma-\delta)+1)/\delta} = 2\varepsilon^{(2\gamma+1)/\delta}$.

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

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

which diverges to $-\infty$, so that exponentiating gives a sequence that converges to 0 as required.

□

Lemma 3.11. *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 3.10 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 3.12. *Let $A > 0$, $B > 0$ and $0 < a < b$ be given, and let K be a kernel that satisfies (4). Let $k^* > 4(C_\psi + C_K)$, and let $\underline{\ell}$ be large enough so that $2^{-\underline{\ell}}(k^* + C_\psi + C_K) < a$. 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 3.10. Let*

$$c_n = \varepsilon^{-1/(2\gamma+1)}\left[\frac{4}{2\gamma+1}\|\kappa^*\|^{-2}\right]^{\frac{\gamma}{2\gamma+1}} B^{\frac{1}{2\gamma+1}}\kappa^*(0)\left(\sigma_n^2\log(1/\sigma_n)\right)^{\gamma/(2\gamma+1)}$$

where κ^* *is a function with* $\kappa^* \in \mathcal{F}_{\text{Höl}}(1,\gamma-\delta)$ *for* $\delta \geq 0$ *small enough, with finite support. Then, if* $0 < 2\alpha < \beta < 1$*,*

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

Proof. First, note that, since $\tilde{f}_{\underline{\ell},\gamma,\delta_n,\varepsilon,A}(x) = 0$ for $x \in [a,b]$,

$$\begin{aligned}&\left\{\widetilde{\mathcal{F}}(\gamma-\delta_n,B,a,b) + \{\tilde{f}_{\underline{\ell},\gamma,\delta_n,\varepsilon,A}\}\right\} \cap \{f : \sup_{x\in[a,b]}|f(x)| \geq c_n(1-\eta)\}\\ &= \widetilde{\mathcal{F}}(\gamma-\delta_n,B,a,b) \cap \{f : \sup_{x\in[a,b]}|f(x)| \geq c_n(1-\eta)\} + \{\tilde{f}_{\underline{\ell},\gamma,\delta_n,\varepsilon,A}\}\end{aligned}$$

for any $\eta > 0$. By Lemma 3.3,

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

The second term converges to zero by Lemma 3.10. By Lemma 3.2, the first term will converge to zero so long as

$$\limsup_{n\to\infty} \frac{c_n(1-\eta)}{C(\gamma-\delta_n,B,\kappa^*)\left(\sigma_n^2\log(1/\sigma_n)\right)^{(\gamma-\delta_n)/[2(\gamma-\delta_n)+1]}} < 1,$$

which holds by Lemma 3.11 and the fact that $C(\gamma-\delta_n,B,\kappa^*) \to C(\gamma,B,\kappa^*)$. The result now follows by Lemma 3.1. □

We are now ready to prove Theorem 2.2.

Proof of Theorem 2.2. Let k^* and $\underline{\ell}$ be chosen so that $k^* > 4(C_\psi + C_K)$ and $2^{-\underline{\ell}}(k^* + C_\psi + C_K) < 1/2$. Let $C = \sup_{\gamma'\in(0,\overline{\gamma}]} \overline{C}_{K,\psi,\gamma'}$. By Lemma 3.7, $\tilde{g}_{\underline{\ell},\gamma,1/(2C)} \in \mathcal{F}_{\text{self-sim}}(\gamma,1,\varepsilon') \subseteq$

$\mathcal{F}_{\text{self-sim}}(\gamma,1,\varepsilon)$ for any $\varepsilon \le \varepsilon' \le \underline{C}_{K,\psi}/(2C)$. Applying Lemma 3.8 with $\tilde{\varepsilon} = 2\varepsilon C/\underline{C}_{K,\psi}$, $\tilde{B} = 1/2$, $A = 1/(2C)$, we have $\widetilde{\mathcal{F}}(\gamma-\delta,1/2,1/2,1)+\tilde{f}_{\underline{\ell},\gamma,\delta,\tilde{\varepsilon},1/(2C)} \subseteq \mathcal{F}_{\text{self-sim}}(\gamma-\delta,1,\varepsilon)$. Let δ_n be defined as in Lemma 3.12, with $\tilde{\varepsilon}$ in place of ε. Once n is large enough so that $\gamma-\delta_n > \underline{\gamma}$, we will have $\widetilde{\mathcal{F}}(\gamma-\delta,1/2,1/2,1)+\tilde{f}_{\underline{\ell},\gamma,\delta,\tilde{\varepsilon},1/(2C)} \subseteq \cup_{\gamma'\in[\underline{\gamma},\overline{\gamma}]}\mathcal{F}_{\text{self-sim}}(\gamma',1,\varepsilon)$. Using this and the fact that $\tilde{g}_{\underline{\ell},\gamma,1/(2C)} \in \mathcal{F}_{\text{self-sim}}(\gamma,1,\varepsilon')$, it follows that $R^*_{n,\alpha,\beta}(\mathcal{F}_{\text{self-sim}}(\gamma,1,\varepsilon';K,\underline{\ell}),\cup_{\gamma'\in[\underline{\gamma},\overline{\gamma}]}\mathcal{F}_{\text{self-sim}}(\gamma',1,\varepsilon';K,\underline{\ell}))$ is bounded from below by $R^*_{n,\alpha,\beta}(\{\tilde{g}_{\underline{\ell},\gamma,1/(2C)}\},\widetilde{\mathcal{F}}(\gamma-\delta,1/2,1/2,1)+\{\tilde{f}_{\underline{\ell},\gamma,\delta,\tilde{\varepsilon},1/(2C)}\})$. By Lemma 3.12, this is bounded from below by a positive constant that depends only on $\overline{\gamma}$ times $(1+o(1))\tilde{\varepsilon}^{-1/(2\gamma+1)}\left(\sigma_n^2\log(1/\sigma_n)\right)^{\gamma/(2\gamma+1)}$ (note that κ^* can be chosen to depend only on $\overline{\gamma}$). The result follows since $(\tilde{\varepsilon}/\varepsilon)^{-1/(2\gamma+1)}$ is bounded from below by a constant that depends only on $\overline{\gamma}$.

□

A Details for Section 2.2

This appendix provides details for the results in Section 2.2.

A.1 Critical Value

The critical value $c(j) = \bar{c}_K\sigma_n 2^{j/2}\sqrt{j}$ is justified by the following lemma.

Lemma A.1. *Let $c(j) = \bar{c}_K\sigma_n 2^{j/2}\sqrt{j}$ and suppose that (4) and (5) hold. Then, if $\bar{c}_K$ is larger than a constant that depends only on the kernel K, we will have, for any sequence $\underline{\ell}_n \to \infty$,*

$$P\left(|\hat{f}(t,j) - K_j f(t)| \le c(j) \text{ all } t\in[0,1], j \ge \underline{\ell}_n\right) \to 1.$$

Proof. Let $\mathbb{T}_n(t,j) = \sigma_n^{-1}2^{-j/2}\left[\hat{f}(t,j) - K_j f(t)\right] = \int 2^{j/2}K(2^j t, 2^j x)\,dW(x)$. Note that the distribution of the process $t \mapsto \mathbb{T}_n(2^{-j}(t+k))$ is the same for all j,k,n, since $cov(\mathbb{T}_n(2^{-j}(s+k),j),\mathbb{T}_n(2^{-j}(t+k),j)) = \int 2^j K(s+k,2^j x)K(t+k,2^j x)\,dx = \int K(s,u)K(t,u)\,du$, using change of variables $u = 2^j x - k$ and the fact that $K(t+k,u+k) = K(t,u)$. Thus,

$$\begin{aligned} P\left(\sup_{t\in[0,1]}|\mathbb{T}_n(t,j)| > \bar{c}_K\sqrt{j}\right) &\le \sum_{k=0}^{2^j-1} P\left(\sup_{s\in[0,1]}|\mathbb{T}_n(2^{-j}(s+k),j)| > \bar{c}_K\sqrt{j}\right) \\ &= 2^j P\left(\sup_{t\in[0,1]}|\mathbb{T}_n(t,1)| > \bar{c}_K\sqrt{j}\right). \end{aligned}$$

By (5), we can apply Theorem 8.1 in Piterbarg (1996) to the process $\mathbb{T}_n(t,1)$, which, along with the tail bound $\Phi(-x) \leq (x\sqrt{2\pi})^{-1}\exp(-x^2/2)$ where Φ is the standard normal cdf, gives the bound $P\left(\sup_{t\in[0,1]}|\mathbb{T}_n(t,1)| > \bar{c}_K\sqrt{j}\right) \leq Cj^{1/\tau_K-1}\exp(-j\bar{c}_K/C)$ for some constant C that depends only on the kernel K. Thus,

$$1 - P\left(|\hat{f}(t,j) - K_j f(t)| \leq c(j) \text{ all } t \in [0,1], j \geq \underline{\ell}_n\right) \leq \sum_{j=\underline{\ell}_n}^{\infty} 2^j P\left(\sup_{t\in[0,1]}|\mathbb{T}_n(t,1)| > \bar{c}_K\sqrt{j}\right)$$
$$\leq \sum_{j=\underline{\ell}_n}^{\infty} 2^j C j^{1/\tau_K-1}\exp(-j\bar{c}_K/C) = \sum_{j=\underline{\ell}_n}^{\infty} C j^{1/\tau_K-1}\exp(-j(\bar{c}_K/C - \log 2).$$

For $\bar{c}_K > C\log 2$, this converges to 0 as $n \to \infty$.

□

A.2 Confidence Interval for γ

We construct a confidence interval $[\hat{\gamma}_\ell, \hat{\gamma}_u]$ for γ, which can be used in the confidence band described in Section 2.2. The confidence interval covers γ on the event in (7), so that the resulting cofidence band for f contains f on the event that (6) and (7) both hold.

Let $\underline{G}(j_1,j_2) = \underline{G}(\varepsilon, \underline{B}, \overline{B}, \underline{\gamma}, \overline{\gamma}, j_1, j_2) = \min_{B\in[\underline{B},\overline{B}],\gamma\in[\underline{\gamma},\overline{\gamma}]} B(\varepsilon - 2^{-(j_2-j_1)\gamma})$ and $\overline{G}(j_1,j_2) = \overline{G}(\underline{B}, \overline{B}, \underline{\gamma}, \overline{\gamma}, j_1, j_2) = \max_{B\in[\underline{B},\overline{B}],\gamma\in[\underline{\gamma},\overline{\gamma}]} B(1 + 2^{-(j_2-j_1)\gamma})$. Let

$$\tilde{\gamma}_\ell(j_1,j_2) = \frac{\log_2 \underline{G}(j_1,j_2) - \log_2\left[\hat{\Delta}(j_2,j_2) + \tilde{c}(j_1,j_2)\right]}{j_1}$$

with the convention that $\tilde{\gamma}_\ell(j_1,j_2) = \underline{\gamma}$ when $\underline{G}(j_1,j_2) \leq 0$. Let

$$\tilde{\gamma}_u(j_1,j_2) = \frac{\log_2 \overline{G}(j_1,j_2) - \log_2\left[\hat{\Delta}(j_2,j_2) - \tilde{c}(j_1,j_2)\right]}{j_1}$$

with the convention that $\tilde{\gamma}_u(j_1,j_2) = \overline{\gamma}$ when $\log_2\left[\hat{\Delta}(j_2,j_2) - \tilde{c}(j_1,j_2)\right] \leq 0$. Let

$$\hat{\gamma}_\ell = \max_{j\in\mathcal{J}_n} \tilde{\gamma}_\ell(j_1,j_2) \text{ and } \hat{\gamma}_u = \min_{j\in\mathcal{J}_n} \tilde{\gamma}_u(j_1,j_2).$$

Then $\gamma \in [\hat{\gamma}_\ell, \hat{\gamma}_u]$ on the event in (7). To see this, note that, by (8), we have, for all $j_1, j_2 \in \mathcal{J}_n$

$$2^{-j_1\gamma}\underline{G}(j_1,j_2) \leq 2^{-j_1\gamma}B(\varepsilon - 2^{-(j_2-j_1)\gamma}) \leq \Delta(j_1,j_2;f) \leq \hat{\Delta}(j_1,j_2) + \tilde{c}(j_1,j_2), \tag{14}$$

and

$$\hat{\Delta}(j_2, j_2) - \tilde{c}(j_1, j_2) \leq \Delta(j_1, j_2; f) \leq 2^{-j_1\gamma} B(1 + 2^{-(j_2 - j_1)\gamma}) \leq 2^{-j_1\gamma} \overline{G}(j_1, j_2).$$

Taking logs and rearranging gives $\gamma \in [\tilde{\gamma}_\ell(j_1, j_2), \tilde{\gamma}_u(j_1, j_2)]$. Note also that

$$\begin{aligned}\tilde{\gamma}_u(j_1, j_2) - \tilde{\gamma}_\ell(j_1, j_2) &\leq \frac{\log_2 \overline{G}(j_1, j_2) - \log_2 \underline{G}(j_1, j_2)}{j_1} + \frac{2\tilde{c}(j_1, j_2)}{j_1(\hat{\Delta}(j_1, j_2) - \tilde{c}(j_1, j_2)) \log 2} \\ &\leq \frac{\log_2 \overline{G}(j_1, j_2) - \log_2 \underline{G}(j_1, j_2)}{j_1} + \frac{2\tilde{c}(j_1, j_2)}{j_1(2^{-j_1\overline{\gamma}} \underline{G}(j_1, j_2) - 2\tilde{c}(j_1, j_2)) \log 2}\end{aligned}$$

where the first inequality uses $|\log a - \log b| \leq |a - b| / \min\{a, b\}$ and the second inequality uses (14).

Let $\tilde{c}(j_1, j_2) = \bar{c}_K \sigma_n 2^{j_1/2} \sqrt{j_1} + \bar{c}_K \sigma_n 2^{j_2/2} \sqrt{j_2}$, so that Lemma A.1 applies. Let j_1, j_2 satisfy $j_1, j_2 \to \infty$, $j_2 - j_1 \to \infty$, and $j_2 / \log n \to 0$. Then the above display is bounded by a constant times j_1^{-1}. To see this, note that $\underline{G}(j_1, j_2)$ and $\overline{G}(j_1, j_2)$ converge to positive constants, and $2^{j_1\overline{\gamma}} \tilde{c}(j_1, j_2) \to 0$ by the conditions on j_1 and j_2.

We collect these results in a theorem.

Theorem A.1. *Let $\hat{\gamma}_\ell$ and $\hat{\gamma}_u$ be given above. Then, on the event in (7), we have $\gamma \in [\hat{\gamma}_\ell, \hat{\gamma}_u]$ for $f \in \mathcal{F}_{\text{self-sim}}(\gamma, B, \varepsilon)$ with $B \in [\underline{B}, \overline{B}]$ and $\gamma \in [\underline{\gamma}, \overline{\gamma}]$. Furthermore, if we take $\tilde{c}(j_1, j_2) = \bar{c}_K \sigma_n 2^{j_1/2} \sqrt{j_1} + \bar{c}_K \sigma_n 2^{j_2/2} \sqrt{j_2}$ and $\mathcal{J}_n$ contains sequences $j_1 = j_{1,n}$ and $j_2 = j_{2,n}$ which satisfy $j_1, j_2 \to \infty$, $j_2 - j_1 \to \infty$, and $j_2 / \log n \to 0$, then, for any sequence r_n with $r_n \to 0$ and $r_n / j_1 \to \infty$, we have*

$$\gamma - r_n \leq \hat{\gamma}_\ell \leq \gamma \leq \hat{\gamma}_u \leq \gamma + r_n$$

with probability approaching one uniformly over $\cup_{\gamma \in [\underline{\gamma}, \overline{\gamma}], B \in [\underline{B}, \overline{B}]} \mathcal{F}_{\text{GN}}(\varepsilon, \varepsilon B, B)$.

A.3 Length of the Confidence Band

We now bound the length of this confidence band. From (10), it follows that, on the event $\gamma - r_n \leq \hat{\gamma}_\ell \leq \gamma \leq \hat{\gamma}_u \leq \gamma + r_n$, the length of the confidence band is bounded by

$$\sup_{\gamma_u, \gamma_\ell \in [\gamma - r_n, \gamma + r_n]} \min_{j, j_1, j_2 \in \mathcal{J}_n} \left[c(j) + \frac{B(2^{-j_1\gamma} + 2^{-j_2\gamma}) + 2c(j_1) + 2c(j_2)}{a(\varepsilon, j_1, j_2, j, \gamma_\ell, \gamma_u)} \right]$$

where $c(j) = \bar{c}_K \sigma 2^{j/2} \sqrt{j/n}$.

It turns out that it will suffice to get an upper bound for the minimum in the above display by taking $j = j_{n,\gamma} = \lfloor \rho_\gamma + (2\gamma+1)^{-1}(\log_2(n/\log_2 n)) \rfloor$, $j_1 = j_{1,n,\gamma} = j_{n,\gamma} - m_{1,n}$ and $j_2 = j_{2,n,\gamma} = j_{n,\gamma} - m_{2,n}$ where $m_{1,n}$ and $m_{2,n}$ are sequences such that $m_{2,n} \to \infty$, $m_{1,n} - m_{2,n} \to \infty$, $r_n m_{1,n} \to 0$ and, for all $\gamma \in [\underline{\gamma}, \overline{\gamma}]$, $j_{1,n,\gamma} \to \infty$ and $j_{2,n,\gamma} \to \infty$. Applying the lemmas below gives the bound

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

where the $o(1)$ term is over $\gamma \in [\underline{\gamma}, \overline{\gamma}]$, $B \in [\underline{B}, \overline{B}]$. Setting $\rho_\gamma = \log_2 \left(\sigma^{-1} B \varepsilon^{-1}\right)^{2/(2\gamma+1)}$ so that $2^{\rho_\gamma/2} = \left(\sigma^{-1} B \varepsilon^{-1}\right)^{1/(2\gamma+1)} = \sigma^{2\gamma/(2\gamma+1)-1} \left(B\varepsilon^{-1}\right)^{1/(2\gamma+1)}$ gives

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

Since $\sigma_n^2 \log(1/\sigma_n) = (\sigma^2/n)\left((1/2)\log n - \log \sigma\right) = (1 + o(1))(\sigma^2/2)(\log n)/n$, this gives a bound of $(\sigma_n^2 \log(1/\sigma_n))^{\gamma/(2\gamma+1)}$ times a constant that is bounded uniformly over $\gamma \leq \overline{\gamma}$, as required.

Lemma A.2.

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

Proof. For n large enough, we have, for any $\gamma \in [\underline{\gamma}, \overline{\gamma}]$ and γ_ℓ, γ_u with $\gamma - r_n \leq \gamma_\ell \leq \gamma_u \leq \gamma + r_n$,

$$\varepsilon 2^{m_{1,n}(\gamma - r_n)} - 2^{m_{2,n}(\gamma + r_n)} \leq a(\varepsilon, j_{1,n,\gamma}, j_{2,n,\gamma}, j_{n,\gamma}, \gamma_\ell, \gamma_u) \leq \varepsilon 2^{m_{1,n}(\gamma + r_n)} - 2^{m_{2,n}(\gamma - r_n)}$$

and $a(\varepsilon, j_{1,n,\gamma}, j_{2,n,\gamma}, j_{n,\gamma}, \gamma, \gamma) = \varepsilon 2^{m_{1,n}\gamma} - 2^{m_{2,n}\gamma}$. Thus,

$$\frac{a(\varepsilon, j_{1,n,\gamma}, j_{2,n,\gamma}, j_{n,\gamma}, \gamma_\ell, \gamma_u)}{a(\varepsilon, j_{1,n,\gamma}, j_{2,n,\gamma}, j_{n,\gamma}, \gamma, \gamma)} \leq \frac{\varepsilon 2^{m_{1,n}(\gamma + r_n)} - 2^{m_{2,n}(\gamma - r_n)}}{\varepsilon 2^{m_{1,n}\gamma} - 2^{m_{2,n}\gamma}} = \frac{2^{m_{1,n} r_n} - \varepsilon^{-1} 2^{-m_{2,n} r_n + (m_{2,n} - m_{1,n})\gamma}}{1 - \varepsilon^{-1} 2^{(m_{2,n} - m_{1,n})\gamma}}$$

which converges to one uniformly over $\gamma \in [\underline{\gamma}, \overline{\gamma}]$ by the conditions on $m_{1,n}$ and $m_{2,n}$. The result follows from this and a similar argument with the lower bound. □

Lemma A.3.

$$\frac{2^{-\gamma j_{1,n,\gamma}} + 2^{-\gamma j_{2,n,\gamma}}}{a(\varepsilon, j_{1,n,\gamma}, j_{2,n,\gamma}, j_{n,\gamma}, \gamma, \gamma)} = 2^{-\gamma j_{n,\gamma}} \varepsilon^{-1}(1 + o(1))$$

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

Proof. We have

$$\frac{2^{-\gamma j_{1,n,\gamma}} + 2^{-\gamma j_{2,n,\gamma}}}{2^{-\gamma j_{n,\gamma}} \varepsilon^{-1} a(\varepsilon, j_{1,n,\gamma}, j_{2,n,\gamma}, j_{n,\gamma}, \gamma, \gamma)} = \frac{2^{-\gamma(j_{1,n,\gamma} - j_{n,\gamma})} + 2^{-\gamma(j_{2,n,\gamma} - j_{n,\gamma})}}{2^{m_{1,n}\gamma} - \varepsilon^{-1} 2^{m_{2,n}\gamma}} = \frac{1 + 2^{-(m_{1,n} - m_{2,n})\gamma}}{1 - \varepsilon^{-1} 2^{-(m_{1,n} - m_{2,n})\gamma}}$$

which converges to one uniformly over $\gamma \in [\underline{\gamma}, \overline{\gamma}]$ by the conditions on $m_{1,n}$ and $m_{2,n}$. □

Lemma A.4. *If* ρ_γ *is bounded over* $\gamma \in [\underline{\gamma}, \overline{\gamma}]$, *then* $c(j_{1,n,\gamma})/2^{-\gamma j_{1,n,\gamma}} \to 0$ *and* $c(j_{2,n,\gamma})/2^{-\gamma j_{2,n,\gamma}} \to 0$ *uniformly over* $\gamma \in [\underline{\gamma}, \overline{\gamma}]$. *Furthermore,* $c(j_{n,\gamma}) \leq \overline{c}_K \sigma 2^{\rho_\gamma/2} (2\gamma + 1)^{-1/2} (n/\log n)^{-\gamma/(2\gamma+1)}$ *and* $2^{-\gamma j_{n,\gamma}} \leq 2^{\gamma(1-\rho_\gamma)} (n/\log_2 n)^{-\gamma/(2\gamma+1)}$.

Proof. We have

$$c(j_{n,\gamma})^2/(\overline{c}_K \sigma)^2 = 2^{j_{n,\gamma}} j_{n,\gamma}/n = 2^{\lfloor \rho_\gamma + (2\gamma+1)^{-1}(\log_2(n/\log_2 n))\rfloor} \lfloor (2\gamma+1)^{-1}(\log_2 n - \log_2 \log_2 n)\rfloor/n$$
$$\leq 2^{\rho_\gamma} 2^{(2\gamma+1)^{-1}(\log_2(n/\log_2 n))} (2\gamma+1)^{-1}(\log_2 n)/n = 2^{\rho_\gamma}(2\gamma+1)^{-1}(n/\log_2 n)^{-2\gamma/(2\gamma+1)}.$$

and

$$2^{-\gamma j_{n,\gamma}} = 2^{-\gamma \lfloor \rho_\gamma + (2\gamma+1)^{-1} \log_2(n/\log_2 n)\rfloor} \leq 2^{\gamma(1-\rho_\gamma) - \gamma(2\gamma+1)^{-1} \log_2(n/\log_2 n)} = 2^{\gamma(1-\rho_\gamma)} (n/\log_2 n)^{\gamma/(2\gamma+1)}.$$

For any $m \geq \rho_\gamma$, we have

$$c(j_{n,\gamma} - m)^2/(2^{-\gamma(j_{n,\gamma}-m)} \overline{c}_K \sigma)^2 = 2^{(2\gamma+1)(j_{n,\gamma}-m)}(j_{n,\gamma} - m)/n$$
$$\leq 2^{\log_2(n/\log_2 n) - (m-\rho_\gamma)(2\gamma+1)} (2\gamma+1)^{-1}(\log_2 n)/n = 2^{-(m-\rho_\gamma)(2\gamma+1)}(2\gamma+1)^{-1}$$

Setting $m = m_{1,n} \to \infty$ it follows that $c(j_{1,n,\gamma})/2^{-\gamma j_{1,n,\gamma}} \to 0$ uniformly over $\gamma \in [\underline{\gamma}, \overline{\gamma}]$ and similarly for $j_{2,n,\gamma}$.

□

B Approximation Bounds

As noted by Giné and Nickl (2010), (2) holds under a Hölder condition under appropriate regularity conditions on the kernel K. In particular, for a bounded kernel satisfying (4), it suffices to assume that, for all $v \in \mathbb{R}$,

$$\int K(v, v+u)\,du = 1, \int K(v, v+u)u^\ell\,du = 0 \text{for } \ell = 1, \ldots, \lfloor\gamma\rfloor, \tag{15}$$

This is Condition 4.1.4, p. 301 in Giné and Nickl (2015), without the moment bound, which is implied by the support condition. For convolution kernels, this simply requires that the kernel be of order at least $\lfloor\gamma\rfloor + 1$. See Section 4.2.2 in Giné and Nickl (2015) for projection kernels.

Lemma B.1. *Let K be a bounded kernel satisfying (4) and (15). Then there exists a constant $\tilde{C}_K$, depending only on the kernel K, such that, for any $\gamma, B > 0$ and $f \in \mathcal{F}_{\text{Höl}}(\gamma, B)$, $|K_j f(t) - f(t)| \le \tilde{C}_K B 2^{-j\gamma}$.*

Proof. We have

$$K_j f(t) - f(t) = \int 2^j K(2^j t, 2^j x) f(x)\,dx - f(t) = \int K(2^j t, 2^j t + u) f(t + 2^{-j}u)\,du - f(t)$$
$$= \int K(2^j t, 2^j t + u)[f(t + 2^{-j}u) - f(t)]\,du.$$

By a Taylor approximation, we have, letting $r = \lfloor\gamma\rfloor$,

$$f(t+s) - f(t) = \sum_{\ell=1}^{r-1} s^\ell f^{(\ell)}(t)/\ell! + \frac{s^r}{(r-1)!}\int_0^1 (1-\tau)^{r-1} f^{(r)}(t+\tau s)\,d\tau.$$

Substituting this Taylor approximation and using the moment conditions on K gives

$$\int K(2^j t, 2^j t + u)\frac{2^{-jr}u^r}{(r-1)!}\int_0^1 (1-\tau)^{r-1}[f^{(r)}(t + \tau 2^{-j}u) - f^{(r)}(t)]\,d\tau\,du$$
$$\le B\int |K(2^j t, 2^j t + u)|\frac{2^{-j\gamma}|u|^\gamma}{(r-1)!}\int_0^1 (1-\tau)^{r-1}|\tau|^{\gamma - r}\,d\tau\,du$$
$$\le \frac{2^{-j\gamma}B}{(r-1)!}\sup_{v\in\mathbb{R}}\int K(v, v+u)|u|^\gamma\,du.$$

If K satisfies (4) and is bounded by some constant $\overline{K}$, this is bounded by

$$\frac{2^{-j\gamma}B}{(r-1)!}\sup_{v\in\mathbb{R}}\int K(v,v+u)|u|^{\gamma}\,du \leq \frac{2^{-j\gamma}B}{(r-1)!}\overline{K}C_K^{\gamma}$$

where C_K is the support bound in (4) and $\overline{K}$ is a bound on $|K(u,v)|$. The result follows by noting that $C_K^{\gamma}/(\lfloor\gamma\rfloor-1)!$ is bounded uniformly over γ.

□

References

Armstrong, T. B. (2018): "Adaptation Bounds for Confidence Bands under Self-Similarity," *arXiv:1810.09762v1 [math, stat].*

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.

Bull, A. D. and R. Nickl (2013): "Adaptive confidence sets in L^2," *Probability Theory and Related Fields*, 156, 889–919.

Bühlmann, P. and S. van de Geer (2011): *Statistics for High-Dimensional Data: Methods, Theory and Applications*, Heidelberg ; New York: Springer, 2011 edition ed.

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

Carpentier, A. (2013): "Honest and adaptive confidence sets in L_p," *Electronic Journal of Statistics*, 7, 2875–2923.

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.

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

——— (2015): *Mathematical Foundations of Infinite-Dimensional Statistical Models*, New York, NY: Cambridge University Press, 1 edition ed.

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.

Kueh, A. (2012): "Locally Adaptive Density Estimation on the Unit Sphere Using Needlets," *Constructive Approximation*, 36, 433–458.

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.

Mukherjee, R. and S. Sen (2018): "Optimal adaptive inference in random design binary regression," *Bernoulli*, 24, 699–739.

Nickl, R. and B. Szabó (2016): "A sharp adaptive confidence ball for self-similar functions," *Stochastic Processes and their Applications*, 126, 3913–3934.

Nickl, R. and S. van de Geer (2013): "Confidence sets in sparse regression," *The Annals of Statistics*, 41, 2852–2876.

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

Patschkowski, T. and A. Rohde (2019): "Locally adaptive confidence bands," *The Annals of Statistics*, 47, 349–381.

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

Piterbarg, V. I. (1996): *Asymptotic Methods in the Theory of Gaussian Processes and Fields*, American Mathematical Soc.

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.

Sniekers, S. and A. van der Vaart (2015): "Adaptive Bayesian credible sets in regression with a Gaussian process prior," *Electronic Journal of Statistics*, 9, 2475–2527.

Szabó, B., A. W. van der Vaart, and J. H. van Zanten (2015): "Frequentist coverage of adaptive nonparametric Bayesian credible sets," *The Annals of Statistics*, 43, 1391–1428.

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

van der Pas, S., B. Szabó, and A. van der Vaart (2017): "Uncertainty Quantification for the Horseshoe (with Discussion)," *Bayesian Analysis*, 12, 1221–1274.