SFB
823

Locally adaptive confidence bands

Tim Patschkowski, Angelika Rohde

Discussion Paper

Nr. 60/2016

SFB
823

LOCALLY ADAPTIVE CONFIDENCE BANDS*

By Tim Patschkowski and Angelika Rohde

Ruhr-Universität Bochum and Albert-Ludwigs-Universität Freiburg

We develop honest and locally adaptive confidence bands for probability densities. They provide substantially improved confidence statements in case of inhomogeneous smoothness, and are easily implemented and visualized. The article contributes conceptual work on locally adaptive inference as a straightforward modification of the global setting imposes severe obstacles for statistical purposes. Among others, we introduce a statistical notion of local Hölder regularity and prove a correspondingly strong version of local adaptivity. We substantially relax the straightforward localization of the self-similarity condition in order not to rule out prototypical densities. The set of densities permanently excluded from the consideration is shown to be pathological in a mathematically rigorous sense. On a technical level, the crucial component for the verification of honesty is the identification of an asymptotically least favorable stationary case by means of Slepian's comparison inequality.

1. Introduction. Let $X_1, \ldots, X_n$ be independent real-valued random variables which are identically distributed according to some unknown probability measure $\mathbb{P}_p$ with Lebesgue density p. Assume that p belongs to a nonparametric function class $\mathcal{P}$. For any interval $[a,b]$ and any significance level $\alpha \in (0,1)$, a confidence band for p, described by a family of random intervals $C_n(t,\alpha), t \in [a,b]$, is said to be (asymptotically) honest with respect to $\mathcal{P}$ if the coverage inequality

$$\liminf_n \inf_{p\in\mathcal{P}} \mathbb{P}_p^{\otimes n}\Big(p(t) \in C_n(t,\alpha) \text{ for all } \ t \in [a,b]\Big) \geq 1-\alpha$$

is satisfied. The aim of this article is to develop honest confidence bands $C_n(t,\alpha), t \in [a,b]$, with smallest possible width $|C_n(t,\alpha)|$ for every $t \in [a,b]$. Adaptive confidence sets maintain specific coverage probabilities over a large union of models while shrinking at the fastest possible nonparametric rate simultaneously over all submodels. If $\mathcal{P}$ is some class of densities within a union of Hölder balls $\mathcal{H}(\beta, L)$ with fixed radius $L > 0$, the confidence band is called globally adaptive, cf. Cai and Low (2004), if for every $\beta > 0$ and for every $\varepsilon > 0$ there exists some constant $c > 0$, such that

$$\limsup_n \sup_{p\in\mathcal{H}(\beta,L)\cap\mathcal{P}} \mathbb{P}_p^{\otimes n}\left(\sup_{t\in[a,b]} |C_n(t,\alpha)| \geq c \cdot r_n(\beta)\right) < \varepsilon.$$

*Supported by the DFG Collaborative Research Center 823, Subproject C1, and DFG Research Grant RO 3766/4-1.

Keywords and phrases: Local regularity and local adaptivity, honesty, confidence bands in density estimation.

Here, $r_n(\beta)$ denotes the minimax-optimal rate of convergence for estimation under supremum norm loss over $\mathcal{H}(\beta, L) \cap \mathcal{P}$, possibly inflated by additional logarithmic factors. However, if $\mathcal{P}$ equals the set of *all* densities contained in

$$\bigcup_{0<\beta\leq\beta^*} \mathcal{H}(\beta, L),$$

honest and adaptive confidence bands provably do not exist although adaptive estimation is possible. Indeed, Low (1997) shows that honest random-length intervals for a probability density at a fixed point cannot have smaller expected width than fixed-length confidence intervals with the size corresponding to the lowest regularity under consideration. Consequently, it is not even possible to construct a family of random intervals $C_n(t, \alpha), t \in [a, b]$, whose expected length shrinks at the fastest possible rate simultaneously over two distinct nested Hölder balls with fixed radius, and which is at the same time asymptotically honest for the union $\mathcal{P}$ of these Hölder balls. Numerous attempts have been made to tackle this adaptation problem in alternative formulations. Whereas Genovese and Wasserman (2008) relax the coverage property and do not require the confidence band to cover the function itself but a simpler surrogate function capturing the original function's significant features, most of the approaches are based on a restriction of the parameter space. Under qualitative shape constraints, Hengartner and Stark (1995), Dümbgen (1998, 2003), and Davies, Kovac and Meise (2009) achieve adaptive inference. Within the models of nonparametric regression and Gaussian white noise, Picard and Tribouley (2000) succeeded to construct pointwise adaptive confidence intervals under a self-similarity condition on the parameter space, see also Kueh (2012) for thresholded needlet estimators. Under a similar condition, Giné and Nickl (2010) even develop asymptotically honest confidence bands for probability densities whose width is adaptive to the global Hölder exponent. Bull (2012) proved that a slightly weakened version of the self-similarity condition is necessary and sufficient. Kerkyacharian, Nickl and Picard (2012) develop corresponding results in the context of needlet density estimators on compact homogeneous manifolds. Under the same type of self-similarity condition, adaptive confidence bands are developed under a considerably generalized Smirnov-Bickel-Rosenblatt assumption based on Gaussian multiplier bootstrap, see Chernozhukov, Chetverikov and Kato (2014a). Hoffmann and Nickl (2011) introduce a nonparametric distinguishability condition, under which adaptive confidence bands exist for finitely many models under consideration. Their condition is shown to be necessary and sufficient.

Similar important conclusions concerning adaptivity in terms of confidence statements are obtained under Hilbert space geometry with corresponding L^2-loss, see Juditsky and Lambert-Lacroix (2003), Baraud (2004), Genovese and Wasserman (2005), Cai and Low (2006), Robins and van der Vaart (2006), Bull and Nickl (2013), and Nickl and Szabó (2016). Concerning L^p-loss, we also draw attention to Carpentier (2013).

In this article, we develop locally adaptive confidence bands. They provide substantially improved confidence statements in case of inhomogeneous smoothness. Conceptual work on locally adaptive inference is contributed as a straightforward modification of the global setting imposes severe obstacles for statistical purposes. It is already delicate to specify what a "locally adaptive confidence band" should be. Disregarding any measurability issues, one possibility is to require a confidence band $C_{n,\alpha} = (C_{n,\alpha}(t))_{t\in[0,1]}$ to satisfy for every interval $U \subset [a,b]$ and for every β (possibly restricted to a prescribed range)

$$\limsup_{n\to\infty} \sup_{\substack{p\in\mathcal{P}:\\ p_{|U_\delta}\in\mathcal{H}_{U_\delta}(\beta,L^*)}} \mathbb{P}_p^{\otimes n}\left(|C_{n,\alpha}(t)| \geq \eta\, r_n(\beta) \text{ for some } t\in U\right) \to 0$$

as $\eta \to \infty$, where U_δ is the δ-enlargement of U. However, this definition reflects a weaker notion of local adaptivity than the statistician may have in mind. On the other hand, we prove that, uniformly over the function class $\mathcal{P}$ under consideration, adaptation to the local or pointwise regularity in the sense of Daoudi, Lévy Véhel and Meyer (1998), Seuret and Lévy Véhel (2002) or Jaffard (1995, 2006) is impossible. Indeed, not even adaptive estimation with respect to pointwise regularity at a fixed point is achievable. On this way, we introduce a statistically suitable notion of local regularity $\beta_{n,p}(t), t \in [a,b]$, depending in particular on the sample size n. We prove a corresponding strong version of local adaptivity, while we substantially relax the straightforward localization of the global self-similarity condition in order not to rule out prototypical densities. The set of functions which is excluded from our parameter space diminishes for growing sample size and the set of permanently excluded functions is shown to be pathological in a mathematically rigorous sense. Our new confidence band appealingly relies on a discretized evaluation of a modified Lepski-type density estimator, including an additional supremum in the empirical bias term in the bandwidth selection criterion. A suitable discretization and a locally constant approximation allow to piece the pointwise constructions together in order to obtain a continuum of confidence statements. The complex construction makes the asymptotic calibration of the confidence band to the level α non-trivial. Whereas the related globally adaptive procedure of Giné and Nickl (2010) reduces to the limiting distribution of the supremum of a stationary Gaussian process, our locally adaptive approach leads to a highly non-stationary situation. A crucial component is therefore the identification of a stationary process as a least favorable case by means of Slepian's comparison inequality, subsequent to a Gaussian reduction using recent non-asymptotic techniques of Chernozhukov, Chetverikov and Kato (2014b). Due to the discretization, the band is computable and feasible from a practical point of view without losing optimality between the mesh points. Our results are exemplarily formulated in the density estimation framework but can be mimicked in other nonparametric models. To keep the representation concise we restrict the theory to locally adaptive kernel density estimators. The ideas can be transferred to wavelet estimators to a large extent as has been done for globally adaptive confidence bands in Giné and Nickl (2010).

The article is organized as follows. Basic notations are introduced in Section 2. Section 3 presents the main contributions, that is a substantially relaxed localized self-similarity condition in Subsection 3.1, the construction and in particular the asymptotic calibration of the confidence band in Subsection 3.2 as well as its strong local adaptivity properties in Subsection 3.3. Important auxiliary results are postponed to Section 4, whereas Section 5 presents the proofs of the main results. Appendix A contains technical tools for the proofs of the main results.

2. Preliminaries and notation. Let $X_1, \ldots, X_n$, $n \geq 4$, be independent random variables identically distributed according to some unknown probability measure $\mathbb{P}_p$ on $\mathbb{R}$ with continuous Lebesgue density p. Subsequently, we consider kernel density estimators

$$\hat{p}_n(\cdot, h) = \frac{1}{n} \sum_{i=1}^{n} K_h \left(X_i - \cdot \right)$$

with bandwidth $h > 0$ and rescaled kernel $K_h(\cdot) = h^{-1} K(\cdot/h)$, where K is a measurable and symmetric kernel with support contained in $[-1, 1]$, integrating to one, and of bounded variation. Furthermore, K needs to be of order $l \in \mathbb{N}$, that is

$$\int x^j K(x) \, \mathrm{d}x = 0 \quad \text{for } 1 \leq j \leq l, \qquad \int x^{l+1} K(x) \, \mathrm{d}x = c \quad \text{with } c \neq 0.$$

For bandwidths of the form $h = 2^{-j}, j \in \mathbb{N}$, we abbreviate the notation writing $\hat{p}_n(\cdot, h) = \hat{p}_n(\cdot, j)$ and $K_h = K_j$. The open Euclidean ball of radius r around some point $t \in \mathbb{R}$ is referred to as $B(t, r)$. Subsequently, the sample is split into two subsamples. For simplicity, we divide the sample into two parts of equal size $\tilde{n} = \lfloor n/2 \rfloor$, leaving possibly out the last observation. Let

$$\chi_1 = \{X_1, \ldots, X_{\tilde{n}}\}, \quad \chi_2 = \{X_{\tilde{n}+1}, \ldots, X_{2\tilde{n}}\}$$

be the distinct subsamples and denote by $\hat{p}_n^{(1)}(\cdot, h)$ and $\hat{p}_n^{(2)}(\cdot, h)$ the kernel density estimators with bandwidth h based on χ_1 and χ_2, respectively. $\mathbb{E}_p^{\chi_1}$, $\mathbb{E}_p^{\chi_2}$, and $\mathbb{E}_p^{\chi}$ denote the expectations with respect to the product measures

$$\begin{aligned}
\mathbb{P}_p^{\chi_1} &= \text{joint distribution of } X_1, \ldots, X_{\tilde{n}}, \\
\mathbb{P}_p^{\chi_2} &= \text{joint distribution of } X_{\tilde{n}+1}, \ldots, X_{2\tilde{n}}, \\
\mathbb{P}_p^{\chi} &= \mathbb{P}_p^{\otimes 2\tilde{n}} = \mathbb{P}_p^{\chi_1} \otimes \mathbb{P}_p^{\chi_2},
\end{aligned}$$

respectively.

For some measure Q, we denote by $\| \cdot \|_{L^p(Q)}$ the L^p-norm with respect to Q. Is Q the Lebesgue measure, we just write $\| \cdot \|_p$, whereas $\| \cdot \|_{\sup}$ denotes the uniform norm. For any metric space (M, d), we define the covering number $N(M, d, \varepsilon)$ as the minimum number of closed balls with radius at most ε (with respect to d) needed to cover M. As has been shown by Nolan and Pollard (1987) (Section 4 and Lemma 22), the class

$$\mathcal{K} = \left\{ K \left(\frac{\cdot - t}{h} \right) : t \in \mathbb{R}, \, h > 0 \right\}$$

with constant envelope $\|K\|_{\sup}$ satisfies

$$N \left(\mathcal{K}, \| \cdot \|_{L^p(Q)}, \varepsilon \|K\|_{\sup} \right) \leq \left(\frac{A}{\varepsilon} \right)^{\nu}, \quad 0 < \varepsilon \leq 1, \quad p = 1, 2 \tag{2.1}$$

for all probability measures Q and for some finite and positive constants A and ν. For $k \in \mathbb{N}$ we denote the k-th order Taylor polynomial of the function p at point y by $P_{y,k}^{p}$. Denoting furthermore by $\lfloor \beta \rfloor = \max\{n \in \mathbb{N} \cup \{0\} : n < \beta\}$, the Hölder class $\mathcal{H}_U(\beta)$ to the parameter $\beta > 0$ on the open interval $U \subset \mathbb{R}$ is defined as the set of functions $p : U \to \mathbb{R}$ admitting derivatives up to the order $\lfloor \beta \rfloor$ and having finite Hölder norm

$$\|p\|_{\beta,U} = \sum_{k=0}^{\lfloor \beta \rfloor} \|p^{(k)}\|_U + \sup_{\substack{x,y \in U \\ x \neq y}} \frac{|p^{(\lfloor \beta \rfloor)}(x) - p^{(\lfloor \beta \rfloor)}(y)|}{|x-y|^{\beta - \lfloor \beta \rfloor}} < \infty.$$

The corresponding Hölder ball with radius $L > 0$ is denoted by $\mathcal{H}_U(\beta, L)$. With this definition of $\|\cdot\|_{\beta,U}$, the Hölder balls are nested, that is

$$\mathcal{H}_U(\beta_2, L) \subset \mathcal{H}_U(\beta_1, L).$$

for $0 < \beta_1 \leq \beta_2 < \infty$ and $|U| < 1$. Finally, $\mathcal{H}_U(\infty, L) = \bigcap_{\beta > 0} \mathcal{H}_U(\beta, L)$ and $\mathcal{H}_U(\infty) = \bigcap_{\beta > 0} \mathcal{H}_U(\beta)$. Subsequently, for any real function $f(\beta)$, the expression $f(\infty)$ is to be read as $\lim_{\beta \to \infty} f(\beta)$, provided that this limit exists. Additionally, the class of probability densities p, such that $p_{|U}$ is contained in the Hölder class $\mathcal{H}_U(\beta, L)$ is denoted by $\mathcal{P}_U(\beta, L)$. The indication of U is omitted when $U = \mathbb{R}$.

3. Main results.

In this section we pursue the new approach of locally adaptive confidence bands and present the main contribution of this article. A notion of local Hölder regularity tailored to statistical purposes, a corresponding condition of admissibility of a class of functions over which both asymptotic honesty and adaptivity (in a sense to be specified) can be achieved, as well as the construction of the new confidence band are presented. As compared to globally adaptive confidence bands, our confidence bands provide improved confidence statements for functions with inhomogeneous smoothness. Figure 1 illustrates the kind of adaptivity that the construction should reveal. The shaded area sketches the intended locally adaptive confidence band as compared to the globally adaptive band (dashed line) for the triangular density and for fixed sample size n. This density is not smoother than Lipschitz at its maximal point but infinitely smooth at both sides. The region where globally and locally adaptive confidence bands coincide up to logarithmic factors (light gray regime in Figure 1) should shrink as the sample size increases, resulting in a substantial benefit of the locally adaptive confidence band outside of a shrinking neighborhood of the maximal point.

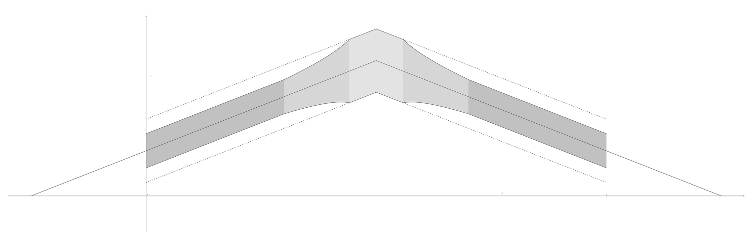

FIG 1. *Comparison of locally and globally adaptive confidence bands*

3.1. *Admissible functions.* As already pointed out in the introduction, no confidence band does exist which is simultaneously honest and adaptive. It is necessary to impose a condition which guarantees the possibility of recovering of the unknown smoothness parameter from the data. The subsequently introduced notion of admissibility aligns to the self-similarity condition as used in Picard and Tribouley (2000) and Giné and Nickl (2010) among others. Their self-similarity condition ensures that the data contains enough information to infer on the function's regularity. As also emphasized in Nickl (2015), self-similarity conditions turn out to be compatible with commonly used adaptive procedures and have been shown to be sufficient and necessary for adaptation to a continuum of smoothing parameters in Bull (2012) when measuring the performance by the L^∞-loss. Giné and Nickl (2010) consider globally adaptive confidence bands over the set

(3.1)

$$\bigcup_{\beta_* \leq \beta \leq \beta^*} \left\{ p \in \mathcal{P}(\beta, L) : p \geq \delta \text{ on } [-\varepsilon, 1+\varepsilon],\ \frac{c}{2^{j\beta}} \leq \|K_j * p - p\|_{\sup} \text{ for all } j \geq j_0 \right\}$$

for some constant $c > 0$ and $0 < \varepsilon < 1$, where $\beta^* = l + 1$ with l the order of the kernel. They work on the scale of Hölder-Zygmund rather than Hölder classes. For this reason they include the corresponding bias upper bound condition which is not automatically satisfied for $\beta = \beta^*$ in that case.

Remark 1. *As mentioned in Giné and Nickl (2010), if $K(\cdot) = \frac{1}{2}\mathbb{1}\{\cdot \in [-1,1]\}$ is the rectangular kernel, the set of all twice differentiable densities that are supported in a fixed compact interval $[a,b]$ satisfies (3.1) with a constant $c > 0$. The reason is that due to the constraint of being a probability density, $\|p''\|_{\sup}$ is bounded away from zero uniformly over this class, in particular p'' cannot vanish everywhere.*

A localized version of the self-similarity condition characterizing the above class reads as follows.

For any nondegenerate interval $(u,v) \subset [0,1]$, there exists some $\beta \in [\beta_*, \beta^*]$ with $p_{|(u,v)} \in \mathcal{P}_{(u,v)}(\beta, L^*)$ and

$$c \cdot 2^{-j\beta} \leq \sup_{s \in (u+2^{-j}, v-2^{-j})} |(K_j * p)(s) - p(s)| \tag{3.2}$$

for all $j \geq j_0 \vee \log_2(1/(v-u))$.

Remark 2. *Inequality (3.2) can be satisfied only for*

$$\tilde{\beta} = \tilde{\beta}_p(U) = \sup\left\{\beta \in (0,\infty] : p_{|U} \in \mathcal{H}_U(\beta)\right\}.$$

The converse is not true, however.

(i) There exist functions $p : U \to \mathbb{R}$, $U \subset \mathbb{R}$ some interval, which are not Hölder continuous to their exponent $\tilde{\beta}$. The Weierstraß function $W_1 : U \to \mathbb{R}$ with

$$W_1(\cdot) = \sum_{n=0}^{\infty} 2^{-n} \cos(2^n \pi \cdot)$$

is such an example. Indeed, Hardy (1916) proves that

$$W_1(x+h) - W_1(x) = O\left(|h| \log\left(\frac{1}{|h|}\right)\right),$$

which implies the Hölder continuity to any parameter $\beta < 1$, hence $\tilde{\beta} \geq 1$. Moreover, he shows in the same reference that W_1 is nowhere differentiable, meaning that it cannot be Lipschitz continuous, that is $\tilde{\beta} = 1$ but $W_1 \notin \mathcal{H}_U(\tilde{\beta})$.

(ii) It can also happen that $p_{|U} \in \mathcal{H}_U(\tilde{\beta})$ but

$$\limsup_{\delta \to 0} \sup_{\substack{|x-y| \leq \delta \\ x,y \in U}} \frac{|p^{(\lfloor \tilde{\beta} \rfloor)}(x) - p^{(\lfloor \tilde{\beta} \rfloor)}(y)|}{|x-y|^{\tilde{\beta} - \lfloor \tilde{\beta} \rfloor}} = 0, \tag{3.3}$$

meaning that the left-hand side of (3.2) is violated. In the analysis literature, the subset of functions in $\mathcal{H}_U(\tilde{\beta})$ satisfying (3.3) is called little Lipschitz (or little Hölder) space. As a complement of an open and dense set, it forms a nowhere dense subset of $\mathcal{H}_U(\tilde{\beta})$.

Due to the localization, a condition like (3.2) rules out examples which seem to be typical to statisticians. Assume that K is a kernel of order l with $l \geq 1$, and recall $\beta^* = l+1$. Then (3.2) excludes for instance the triangular density in Figure 1 because both sides are linear, in particular the second derivative exists and vanishes when restricted to an interval U which does not contain the maximal point. In contrast to the observation in Remark 1, $\|p''\|_U$ may vanish for subintervals $U \subset [a,b]$. For the same reason, densities with a constant piece are excluded. In general, if p restricted to the 2^{-j_0}-enlargement of U is a polynomial of order at most l, (3.2) is violated as the left-hand side is not equal to zero. In view of these deficiencies, a condition like (3.2) is insufficient for statistical purposes.

To circumvent this deficit, we introduce $\|\cdot\|_{\beta,\beta^*,U}$ by

$$\|p\|_{\beta,\beta^*,U} = \sum_{k=0}^{\lfloor \beta \wedge \beta^* \rfloor} \left\|p^{(k)}\right\|_U + \sup_{\substack{x,y \in U \\ x \neq y}} \frac{\left|p^{(\lfloor \beta \wedge \beta^* \rfloor)}(x) - p^{(\lfloor \beta \wedge \beta^* \rfloor)}(y)\right|}{|x-y|^{\beta - \lfloor \beta \wedge \beta^* \rfloor}} \tag{3.4}$$

for $\beta > 0$ and for some bounded open subinterval $U \subset \mathbb{R}$. As verified in Lemma A.4, $\|p\|_{\beta_1,\beta^*,U} \leq \|\cdot\|_{\beta_2,\beta^*,U}$ for $0 < \beta_1 \leq \beta_2 < \infty$ and $|U| \leq 1$. With the help of $\|\cdot\|_{\beta,\beta^*,U}$, we formulate a localized self-similarity type condition in the subsequent Assumption 3.1, which does not exclude these prototypical densities as mentioned above. For any bounded open interval $U \subset \mathbb{R}$, let $\mathcal{H}_{\beta^*,U}(\beta, L)$ be the set of functions $p : U \to \mathbb{R}$ admitting derivatives up to the order $\lfloor \beta \wedge \beta^* \rfloor$ with $\|p\|_{\beta,\beta^*,U} \leq L$. Moreover, $\mathcal{H}_{\beta^*,U}(\beta)$ is the set of functions $p : U \to \mathbb{R}$, such that $\|p\|_{\beta,\beta^*,U}$ is well-defined and finite. Correspondingly, $\mathcal{H}_{\beta^*,U}(\infty, L) = \bigcap_{\beta>0} \mathcal{H}_{\beta^*,U}(\beta, L)$ and $\mathcal{H}_{\beta^*,U}(\infty) = \bigcap_{\beta>0} \mathcal{H}_{\beta^*,U}(\beta)$. Define furthermore

$$\beta_p(U) = \sup\left\{\beta \in (0,\infty] : p_{|U} \in \mathcal{H}_{\beta^*,U}(\beta, L^*)\right\}. \tag{3.5}$$

Remark 3. *If for some open interval $U \subset [0,1]$ the derivative $p_{|U}^{(\beta^*)}$ exists and*

$$p_{|U}^{(\beta^*)} \equiv 0,$$

then $\|p\|_{\beta,\beta^*,U}$ is finite uniformly over all $\beta > 0$. If

$$p_{|U}^{(\beta^*)} \not\equiv 0,$$

then β^*, $\|p\|_{\beta,\beta^*,U}$ is finite if and only if $\beta \leq \beta^*$ as a consequence of the mean value theorem. That is, $\beta_p(U) \in (0, \beta^*] \cup \{\infty\}$.

Assumption 3.1. *For sample size $n \in \mathbb{N}$, some $0 < \varepsilon < 1$, $0 < \beta_* < 1$, and $L^* > 0$, a density p is said to be admissible if $p \in \mathcal{P}_{(-\varepsilon,1+\varepsilon)}(\beta_*, L^*)$ and the following holds true: for any $t \in [0,1]$ and for any $h \in \mathcal{G}_\infty$ with*

$$\mathcal{G}_\infty = \{2^{-j} : j \in \mathbb{N},\, j \geq j_{\min} = \lceil 2 \vee \log_2(2/\varepsilon) \rceil\},$$

there exists some $\beta \in [\beta_*, \beta^*] \cup \{\infty\}$ such that the following conditions are satisfied for $u = h$ or $u = 2h$:

$$p_{|B(t,u)} \in \mathcal{H}_{\beta^*, B(t,u)}(\beta, L^*) \tag{3.6}$$

and

$$\sup_{s \in B(t,u-g)} |(K_g * p)(s) - p(s)| \geq \frac{g^\beta}{\log n} \tag{3.7}$$

for all $g \in \mathcal{G}_\infty$ with $g \leq u/8$.

The set of admissible densities is denoted by $\mathscr{P}_n^{\mathrm{adm}} = \mathscr{P}_n^{\mathrm{adm}}(K, \beta_*, L^*, \varepsilon)$.

Lemma 3.2. *Any admissible density $p \in \mathscr{P}_n^{\mathrm{adm}}(K, \beta_*, L^*, \varepsilon)$ can satisfy (3.6) and (3.7) for $\beta = \beta_p(B(t,u))$ only.*

By construction, the collection of admissible densities is increasing with the number of observations, that is $\mathscr{P}_n^{\mathrm{adm}} \subset \mathscr{P}_{n+1}^{\mathrm{adm}}$, $n \in \mathbb{N}$. The logarithmic denominator even weakens the assumption for growing sample size, permitting smaller and smaller Lipschitz constants.

Remark 4. *Assumption 3.1 does not require an admissible function to be totally "unsmooth" everywhere. For instance, if K is the rectangular kernel and L^* is sufficiently large, the triangular density as depicted in Figure 1 is (eventually – for sufficiently large n) admissible. It is globally not smoother than Lipschitz, and the bias lower bound condition (3.7) is (eventually) satisfied for $\beta = 1$ and pairs (t,h) with $|t - 1/2| < (7/8)h$. Although the bias lower bound condition to the exponent $\beta^* = 2$ is not satisfied for any (t,h) with $t \in [0,1] \setminus (1/2 - h, 1/2 + h)$, these tuples (t,h) fulfill (3.6) and (3.7) for $\beta = \infty$, which is not excluded anymore by the new*

Assumption 3.1. Finally, if the conditions (3.6) and (3.7) are not simultaneously satisfied for some pair (t, h) with

$$\frac{1}{2} + \frac{7}{8}h < |t| < \frac{1}{2} + h,$$

then they are fulfilled for the pair $(t, 2h)$ and $\beta = 1$, because $|t - 1/2| < (7/8)2h$.

In view of this remark, it is crucial not to require (3.6) and (3.7) to hold for *every* pair (t, h). We now denote by

$$\mathscr{P}_n = \mathscr{P}_n(L^*, \beta_*, M, K, \varepsilon) = \left\{ p \in \mathscr{P}_n^{\mathrm{adm}}(K, \beta_*, L^*, \varepsilon) : \inf_{x \in [-\varepsilon, 1+\varepsilon]} p(x) \geq M \right\}$$

the set of admissible densities being bounded below by $M > 0$ on $[-\varepsilon, 1+\varepsilon]$. We restrict our considerations to combinations of parameters for which the class $\mathscr{P}_n$ is non-empty.

The remaining results of this subsection are about the massiveness of the function classes $\mathscr{P}_n$. They are stated for the particular case of the rectangular kernel. Other kernels may be treated with the same idea; verification of (3.7) however appears to require a case-by-case analysis for different kernels. The following proposition demonstrates that the pointwise minimax rate of convergence remains unchanged when passing from the class $\mathcal{H}(\beta, L^*)$ to $\mathscr{P}_n \cap \mathcal{H}(\beta, L^*)$.

PROPOSITION 3.3 (Lower pointwise risk bound). *For the rectangular kernel K_R there exists some constant $M > 0$, such that for any $t \in [0, 1]$, for any $\beta \in [\beta_*, 1]$, for any $0 < \varepsilon < 1$, and for any $k \geq k_0(\beta_*)$ there exists some $x > 0$ and some $L(\beta) > 0$ with*

$$\inf_{T_n} \sup_{\substack{p \in \mathscr{P}_k: \\ p_{|(-\varepsilon, 1+\varepsilon)} \in \mathcal{H}_{(-\varepsilon, 1+\varepsilon)}(\beta, L)}} \mathbb{P}_p^{\otimes n} \left(n^{\frac{\beta}{2\beta+1}} \left| T_n(t) - p(t) \right| \geq x \right) > 0$$

for all $L \geq L(\beta)$, for the class $\mathscr{P}_k = \mathscr{P}_k(L, \beta_*, M, K_R, \varepsilon)$, where the infimum is running over all estimators T_n based on $X_1, \ldots, X_n$.

Note that the classical construction for the sequence of hypotheses in order to prove minimax lower bounds consists of a smooth density distorted by small β-smooth perturbations, properly scaled with the sample size n. However, there does not exist a fixed constant $c > 0$, such that all of its members are contained in the class (3.1). Thus, the constructed hypotheses in our proof are substantially more complex, for which reason we restrict attention to $\beta \leq 1$.

Although Assumption 3.1 is getting weaker for growing sample size, some densities are permanently excluded from consideration. The following proposition states that the exceptional set of permanently excluded densities is pathological.

PROPOSITION 3.4. *For the rectangular kernel $K_R(\cdot) = \frac{1}{2}\mathbb{1}\{\cdot \in [-1,1]\}$, let*

$$\mathscr{R} = \bigcup_{n\in\mathbb{N}} \mathscr{P}_n^{\mathrm{adm}}(K_R, \beta_*, L^*, \varepsilon).$$

Then, for any $t \in [0,1]$, for any $h \in \mathcal{G}_\infty$ and for any $\beta \in [\beta_*, 1)$, the set

$$\mathcal{P}_{B(t,h)}(\beta, L^*) \setminus \mathscr{R}_{|B(t,h)}$$

is nowhere dense in $\mathcal{P}_{B(t,h)}(\beta, L^*)$ with respect to $\|\cdot\|_{\beta,B(t,h)}$.

Among more involved approximation steps, the proof reveals the existence of functions with the same regularity in the sense of Assumption 3.1 on *every* interval for $\beta \in (0,1)$. This property is closely related to but does not coincide with the concept of mono-Hölder continuity from the analysis literature, see for instance Barral et al. (2013). Hardy (1916) shows that the Weierstraß function is mono-Hölder continuous for $\beta \in (0,1)$. For any $\beta \in (0,1]$, the next lemma shows that Weierstraß' construction

$$W_\beta(t) = \sum_{n=0}^{\infty} 2^{-n\beta}\cos(2^n \pi t) \tag{3.8}$$

satisfies the bias condition (3.7) for the rectangular kernel to the exponent β on any subinterval $B(t,h)$, $t \in [0,1]$, $h \in \mathcal{G}_\infty$.

LEMMA 3.5. *For all $\beta \in (0,1)$, the Weierstraß function W_β as defined in (3.8) satisfies $W_{\beta|U} \in \mathcal{H}_U(\beta, L_W)$ with some Lipschitz constant $L_W = L_W(\beta)$ for every open interval U. For the rectangular kernel K_R and $\beta \in (0,1]$, the Weierstraß function fulfills the bias lower bound condition*

$$\sup_{s\in B(t,h-g)} |(K_{R,g} * W_\beta)(s) - W_\beta(s)| > \left(\frac{4}{\pi} - 1\right) g^\beta$$

for any $t \in \mathbb{R}$ and for any $g, h \in \mathcal{G}_\infty$ with $g \le h/2$.

The whole scale of parameters $\beta \in [\beta_*, 1]$ in Proposition 3.4 can be covered by passing over from Hölder classes to Hölder-Zygmund classes in the definition of $\mathscr{P}_n$. Although the Weierstraß function W_1 in (3.8) is not Lipschitz, a classical result, see Heurteaux (2005) or Mauldin and Williams (1986) and references therein, states that W_1 is indeed contained in the Zygmund class Λ_1. That is, it satisfies

$$|W_1(x+h) - W_1(x-h) - 2W_1(x)| \le C|h|$$

for some $C > 0$ and for all $x \in \mathbb{R}$ and for all $h > 0$. Due to the symmetry of the rectangular kernel K_R, it therefore fulfills the bias upper bound

$$\|K_{R,g} * W_1 - W_1\|_{\sup} \le C' g^\beta \quad \text{for all } g \in (0,1].$$

The local adaptivity theory can be likewise developed on the scale of Hölder-Zygmund rather than Hölder classes – here, we restrict attention to Hölder classes because they are commonly considered in the theory of kernel density estimation.

3.2. *Construction of the confidence band.* The new confidence band is based on a kernel density estimator with variable bandwidth incorporating a localized but not the fully pointwise Lepski (1990) bandwidth selection procedure. A suitable discretization and a locally constant approximation allow to piece the pointwise constructions together in order to obtain a continuum of confidence statements. The complex construction makes the asymptotic calibration of the confidence band to the level α non-trivial. Whereas the related globally adaptive procedure of Giné and Nickl (2010) reduces to the limiting distribution of the supremum of a stationary Gaussian process, our locally adaptive approach leads to a highly non-stationary situation. An essential component is therefore the identification of a stationary process as a least favorable case by means of Slepian's comparison inequality.

We now describe the procedure. The interval $[0,1]$ is discretized into equally spaced grid points, which serve as evaluation points for the locally adaptive estimator. We discretize by a mesh of width

$$\delta_n = \left\lceil 2^{j_{\min}} \left(\frac{\log \tilde{n}}{\tilde{n}}\right)^{-\kappa_1} (\log \tilde{n})^{\frac{2}{\beta_*}} \right\rceil^{-1}$$

with $\kappa_1 \geq 1/(2\beta_*)$ and set $\mathcal{H}_n = \{k\delta_n : k \in \mathbb{Z}\}$. Fix now constants

$$c_1 > \frac{2}{\beta_* \log 2} \quad \text{and} \quad \kappa_2 > c_1 \log 2 + 4. \tag{3.9}$$

Consider the set of bandwidth exponents

$$\mathcal{J}_n = \left\{ j \in \mathbb{N} \ : \ j_{\min} \leq j \leq j_{\max} = \left\lfloor \log_2 \left(\frac{\tilde{n}}{(\log \tilde{n})^{\kappa_2}} \right) \right\rfloor \right\}.$$

The bound $j_{\min}$ ensures that $2^{-j} \leq \varepsilon \wedge 1/4$ for all $j \in \mathcal{J}_n$, and therefore avoids that infinite smoothness in (3.15) and the corresponding local parametric rate is only attainable in trivial cases as the interval under consideration is $[0,1]$. The bound $j_{\max}$ is standard and particularly guarantees consistency of the kernel density estimator with minimal bandwidth within the dyadic grid of bandwidths

$$\mathcal{G}_n = \left\{ 2^{-j} : j \in \mathcal{J}_n \right\}.$$

We define the set of admissible bandwidths for $t \in [0,1]$ as

$$\begin{aligned} \mathcal{A}_n(t) = \Big\{ j \in \mathcal{J}_n : \max_{s \in B\left(t, \frac{7}{8}\cdot 2^{-j}\right) \cap \mathcal{H}_n} \left| \hat{p}_n^{(2)}(s,m) - \hat{p}_n^{(2)}(s,m') \right| \leq c_2 \sqrt{\frac{\log \tilde{n}}{\tilde{n} 2^{-m}}} \\ \text{for all } m, m' \in \mathcal{J}_n \text{ with } m > m' > j+2 \Big\}, \end{aligned} \tag{3.10}$$

with constant $c_2 = c_2(A, \nu, \beta_*, L^*, K, \varepsilon)$ specified in the proof of Proposition 4.1. Furthermore, let

$$\hat{j}_n(t) = \min \mathcal{A}_n(t), \quad t \in [0,1], \tag{3.11}$$

and $\hat{h}_n(t) = 2^{-\hat{j}_n(t)}$. Note that a slight difference to the classical Lepski procedure is the additional maximum in (3.10), which reflects the idea of adapting localized but not completely pointwise for fixed sample size n. The bandwidth (3.11) is determined for all mesh points $k\delta_n, k \in T_n = \{1, \dots, \delta_n^{-1}\}$ in $[0,1]$, and set piecewise constant in between. Accordingly, with

$$\hat{h}_{n,1}^{loc}(k) = 2^{-\hat{j}_n((k-1)\delta_n) - u_n}, \quad \hat{h}_{n,2}^{loc}(k) = 2^{-\hat{j}_n(k\delta_n) - u_n},$$

where $u_n = c_1 \log\log \tilde{n}$ is some sequence implementing the undersmoothing, the estimators are defined as

$$\begin{aligned} &\hat{h}_n^{loc}(t) = \hat{h}_{n,k}^{loc} = \min\left\{\hat{h}_{n,1}^{loc}(k), \hat{h}_{n,2}^{loc}(k)\right\} \qquad \text{and} \\ &\hat{p}_n^{loc}(t,h) = \hat{p}_n^{(1)}(k\delta_n, h) \end{aligned} \tag{3.12}$$

for $t \in I_k = [(k-1)\delta_n, k\delta_n)$, $k \in T_n \setminus \{\delta_n^{-1}\}$, $I_{\delta_n^{-1}} = [1-\delta_n, 1]$. Defining furthermore the width function of the confidence band

$$\hat{z}_n(\cdot) = \left(\tilde{n}\hat{h}_n^{loc}(\cdot)\right)^{-\frac{1}{2}}, \tag{3.13}$$

the following theorem lays the foundation for the construction of honest and locally adaptive confidence bands.

Theorem 3.6 (Least favorable case). *For the estimators defined in* (3.12) *and normalizing sequences*

$$a_n = c_3(-2\log\delta_n)^{1/2}, \quad b_n = \frac{3}{c_3}\left\{(-2\log\delta_n)^{1/2} - \frac{\log(-\log\delta_n) + \log 4\pi}{2(-2\log\delta_n)^{1/2}}\right\},$$

with $c_3 = \sqrt{2}/TV(K)$, *it holds*

$$\liminf_{n\to\infty} \inf_{p\in\mathscr{P}_n} \mathbb{P}_p^{\chi}\left(a_n\left(\sup_{t\in[0,1]} \frac{\left|\hat{p}_n^{loc}(t,\hat{h}_n^{loc}(t)) - p(t)\right|}{\hat{z}_n(t)} - b_n\right) \le x\right) \ge 2\,\mathbb{P}\left(\sqrt{L^*}G \le x\right) - 1$$

for some standard Gumbel distributed random variable G.

The proof of Theorem 3.6 is based on several completely non-asymptotic approximation techniques. The asymptotic Komlós-Major-Tusnády-approximation technique, used in Giné and Nickl (2010), has been evaded using non-asymptotic Gaussian approximation results recently developed in Chernozhukov, Chetverikov and Kato (2014b). The essential component of the proof of Theorem 3.6 is the application of Slepian's comparison inequality to reduce considerations from a non-stationary Gaussian process to the least favorable case of a maximum of δ_n^{-1} independent and identical standard normal random variables.

With $q_{1-\alpha/2}$ denoting the $(1-\alpha/2)$-quantile of the standard Gumbel distribution, we define the confidence band as the family of piecewise constant random intervals $C_{n,\alpha} = (C_{n,\alpha}(t))_{t\in[0,1]}$ with

$$C_{n,\alpha}(t) = \left[\hat{p}_n^{loc}(t, \hat{h}_n^{loc}(t)) - q_n(\alpha)\hat{z}_n(t),\quad \hat{p}_n^{loc}(t, \hat{h}_n^{loc}(t)) + q_n(\alpha)\hat{z}_n(t)\right] \tag{3.14}$$

and

$$q_n(\alpha) = \frac{\sqrt{L^*}\cdot q_{1-\alpha/2}}{a_n} + b_n.$$

For fixed $\alpha > 0$, $q_n(\alpha) = O(\sqrt{\log n})$ as n goes to infinity.

Corollary 3.7 (Honesty). *The confidence band as defined in* (3.14) *satisfies*

$$\liminf_{n\to\infty} \inf_{p\in\mathscr{P}_n} \mathbb{P}_p^{\chi}\Big(p(t) \in C_{n,\alpha}(t) \text{ for every } t \in [0,1]\Big) \geq 1-\alpha.$$

3.3. *Local Hölder regularity and local adaptivity.* In the style of global adaptivity in connection with confidence sets one may call a confidence band $C_{n,\alpha} = (C_{n,\alpha}(t))_{t\in[0,1]}$ locally adaptive if for every interval $U \subset [0,1]$,

$$\limsup_{n\to\infty} \sup_{p\in\mathscr{P}_{n|U_\delta}\cap\mathcal{H}_{\beta^*,U_\delta}(\beta,L^*)} \mathbb{P}_p^{\chi_2}\left(|C_{n,\alpha}(t)| \geq \eta\left(\frac{\log\tilde{n}}{\tilde{n}}\right)^{\frac{\beta}{2\beta+1}} \text{ for some } t\in U\right) \to 0$$

as $\eta \to \infty$, for every $\beta \in [\beta_*, \beta^*]$, where U_δ is the δ-enlargement of U. As a consequence of the subsequently formulated Theorem 3.12, our confidence band satisfies this notion of local adaptivity up to a logarithmic factor. However, in view of the imagination illustrated in Figure 1 the statistician aims at a stronger notion of adaptivity, where the asymptotic statement is not formulated for an arbitrary but fixed interval U only. Precisely, the goal would be to adapt even to some pointwise or local Hölder regularity, two well established notions from analysis.

Definition 3.8 (Pointwise Hölder exponent, Seuret and Lévy Véhel (2002)). *Let $p : \mathbb{R} \to \mathbb{R}$ be a function, $\beta > 0$, $\beta \notin \mathbb{N}$, and $t \in \mathbb{R}$. Then $p \in \mathcal{H}_t(\beta)$ if and only if there exists a real $R > 0$, a polynomial P with degree less than $\lfloor\beta\rfloor$, and a constant c such that*

$$|p(x) - P(x-t)| \leq c|x-t|^\beta$$

for all $x \in B(t,R)$. The pointwise Hölder exponent is denoted by

$$\beta_p(t) = \sup\{\beta : p \in \mathcal{H}_t(\beta)\}.$$

Definition 3.9 (Local Hölder exponent, Seuret and Lévy Véhel (2002)). *Let* $p : \Omega \to \mathbb{R}$ *be a function and* $\Omega \subset \mathbb{R}$ *an open set. One classically says that* $p \in \mathcal{H}_{loc}(\beta, \Omega)$*, where* $0 < \beta < 1$*, if there exists a constant* c *such that*

$$|p(x) - p(y)| \leq c|x - y|^{\beta}$$

for all $x, y \in \Omega$*. If* $m < \beta < m + 1$ *for some* $m \in \mathbb{N}$*, then* $p \in \mathcal{H}_{loc}(\beta, \Omega)$ *means that there exists a constant* c *such that*

$$|\partial^m p(x) - \partial^m p(y)| \leq c|x - y|^{\beta - m}$$

for all $x, y \in \Omega$*. Set now*

$$\beta_p(\Omega) = \sup\{\beta : p \in \mathcal{H}_{loc}(\beta, \Omega)\}.$$

Finally, the local Hölder exponent in t *is defined as*

$$\beta_p^{loc}(t) = \sup\{\beta_p(O_i) : i \in I\},$$

where $(O_i)_{i \in I}$ *is a decreasing family of open sets with* $\cap_{i \in I} O_i = \{t\}$*. [By Lemma 2.1 in Seuret and Lévy Véhel (2002), this notion is well defined, that is, it does not depend on the particular choice of the decreasing sequence of open sets.]*

The next proposition however shows that attaining the minimax rates of convergence corresponding to the pointwise or local Hölder exponent (possibly inflated by some logarithmic factor) uniformly over $\mathscr{P}_n$ is an unachievable goal.

Proposition 3.10. *For the rectangular kernel* K_R *there exists some constant* $M > 0$*, such that for any* $t \in [0, 1]$*, for any* $\beta \in [\beta_*, 1]$*, for any* $0 < \varepsilon < 1$*, and for any* $k \geq k_0(\beta_*)$ *there exists some* $x > 0$ *and constants* $L = L(\beta) > 0$ *and* $c_4 = c_4(\beta) > 0$ *with*

$$\inf_{T_n} \sup_{p \in \mathscr{S}_k(\beta)} \mathbb{P}_p^{\otimes n}\left(n^{\frac{\beta}{2\beta+1}} |T_n(t) - p(t)| \geq x\right) > 0 \quad \text{for all } k \geq k_0(\beta_*)$$

with

$$\begin{aligned}\mathscr{S}_k(\beta) &= \mathscr{S}_k(L, \beta, \beta_*, M, K_R, \varepsilon)\\ &= \Big\{p \in \mathscr{P}_k(L, \beta_*, M, K_R, \varepsilon) : \exists\, r \geq c_4\, n^{-\frac{1}{2\beta+1}}\\ &\qquad \text{such that } p_{|B(t,r)} \in \mathcal{H}_{B(t,r)}(\infty, L)\Big\} \cap \mathcal{H}_{(-\varepsilon, 1+\varepsilon)}(\beta, L),\end{aligned}$$

where the infimum is running over all estimators T_n *based on* $X_1, \ldots, X_n$*.*

The proposition furthermore reveals that if a density $p \in \mathscr{P}_k$ is Hölder smooth to some exponent $\eta > \beta$ on a ball around t with radius at least of the order $n^{-1/(2\beta+1)}$, then no estimator for $p(t)$ can achieve a better rate than $n^{-\beta/(2\beta+1)}$. We therefore introduce an n-dependent statistical notion of local regularity for any point t. Roughly speaking, we intend it to be the maximal β such that the density attains this Hölder exponent within $B(t, h_{\beta,n})$, where $h_{\beta,n}$ is of the optimal adaptive bandwidth order $(\log n/n)^{1/(2\beta+1)}$. We realize this idea with $\|\cdot\|_{\beta,\beta^*,U}$ as defined in (3.4) and used in Assumption 3.1.

DEFINITION 3.11 (n-dependent local Hölder exponent). *With the classical optimal bandwidth within the class* $\mathcal{H}(\beta)$

$$h_{\beta,n} = 2^{-j_{\min}} \cdot \left(\frac{\log \tilde{n}}{\tilde{n}}\right)^{\frac{1}{2\beta+1}},$$

define the class $\mathcal{H}_{\beta^*,n,t}(\beta, L)$ *as the set of functions* $p : B(t, h_{\beta,n}) \to \mathbb{R}$*, such that* p *admits derivatives up to the order* $\lfloor \beta \wedge \beta^* \rfloor$ *and* $\|p\|_{\beta,\beta^*,B(t,h_{\beta,n})} \leq L$*, and* $\mathcal{H}_{\beta^*,n,t}(\beta)$ *the class of functions* $p : B(t, h_{\beta,n}) \to \mathbb{R}$ *for which* $\|p\|_{\beta,\beta^*,B(t,h_{\beta,n})}$ *is well-defined and finite. The* n*-dependent local Hölder exponent for the function* p *at point* t *is defined as*

$$\beta_{n,p}(t) = \sup \Big\{ \beta > 0 : p_{|B(t,h_{\beta,n})} \in \mathcal{H}_{\beta^*,n,t}(\beta, L^*) \Big\}. \tag{3.15}$$

If the supremum is running over the empty set, we set $\beta_{n,p}(t) = 0$*.*

Finally, the next theorem shows that the confidence band adapts to the n-dependent local Hölder exponent.

THEOREM 3.12 (Strong local adaptivity). *There exists some universal constant* $c > 0$ *and some* $\gamma = \gamma(c_1)$*, such that for all* $\delta > 0$ *there exists some* $n_0(\delta) \in \mathbb{N}$ *with*

$$\sup_{p \in \mathscr{P}_n} \mathbb{P}_p^{\chi_2} \left(|C_{n,\alpha}(t)| \geq c \left(\frac{\log \tilde{n}}{\tilde{n}}\right)^{\frac{\beta_{n,p}(t)}{2\beta_{n,p}(t)+1}} q_n(\alpha)(\log \tilde{n})^{\gamma} \text{ for some } t \in [0,1] \right) \leq \delta$$

for all $n \geq n_0(\delta)$*.*

Note that the case $\beta_{n,p}(t) = \infty$ is not excluded in the formulation of Theorem 3.12. That is, if $p_{|U}$ can be represented as a polynomial of degree strictly less than β^*, the confidence band attains even adaptively the parametric width $n^{-1/2}$, up to logarithmic factors. In particular, the band can be tighter than $n^{-\beta^*/(2\beta^*+1)}$. In general,

$$\beta_{n,p}(t) \geq \beta_p(U_\delta) \quad \text{for all } t \in U$$

as long as $\varepsilon \leq \delta$.

COROLLARY 3.13 (Weak local adaptivity). *For every interval* $U \subset [0,1]$*,*

$$\limsup_{n\to\infty} \sup_{p \in \mathscr{P}_{n|U_\delta} \cap \mathcal{H}_{\beta^*,U_\delta}(\beta, L^*)} \mathbb{P}_p^{\chi_2} \left(|C_{n,\alpha}(t)| \geq \eta \left(\frac{\log \tilde{n}}{\tilde{n}}\right)^{\frac{\beta}{2\beta+1}} \text{ for some } t \in U \right) \to 0$$

as $\eta \to \infty$*, for every* $\beta \in [\beta_*, \beta^*]$*, where* U_δ *is the* δ*-enlargement of* U*.*

4. Auxiliary notation and results. The following auxiliary results are crucial ingredients in the proofs of Theorem 3.6 and Theorem 3.12.
Recalling the quantity $h_{\beta,n}$ in Definition 3.11, Proposition 4.1 shows that $2^{-\hat{j}_n(\cdot)}$ lies in a band around

$$\bar{h}_n(\cdot) = h_{\beta_{n,p}(\cdot),n} \tag{4.1}$$

uniformly over all admissible densities $p \in \mathscr{P}_n$. Proposition 4.1 furthermore reveals the necessity to undersmooth, which has been already discovered by Bickel and Rosenblatt (1973), leading to a bandwidth deflated by some logarithmic factor. Set now

$$\bar{j}_n(\cdot) = \left\lfloor \log_2 \left(\frac{1}{\bar{h}_n(\cdot)} \right) \right\rfloor + 1,$$

such that the bandwidth $2^{-\bar{j}_n(\cdot)}$ is an approximation of $\bar{h}_n(\cdot)$ by the next smaller bandwidth on the grid $\mathcal{G}_n$ with

$$\frac{1}{2}\bar{h}_n(\cdot) \leq 2^{-\bar{j}_n(\cdot)} \leq \bar{h}_n(\cdot).$$

The next proposition states that the procedure chooses a bandwidth which simultaneously in the location t is neither too large nor too small.

PROPOSITION 4.1. *The bandwidth $\hat{j}_n(\cdot)$ defined in (3.11) satisfies*

$$\lim_{n\to\infty} \sup_{p\in\mathscr{P}_n} \left\{ 1 - \mathbb{P}_p^{\chi_2} \left(\hat{j}_n(k\delta_n) \in \left[k_n(k\delta_n), \, \bar{j}_n(k\delta_n) + 1 \right] \text{ for all } k \in T_n \right) \right\} = 0$$

where $k_n(\cdot) = \bar{j}_n(\cdot) - m_n$, and $m_n = \frac{1}{2}c_1 \log\log \tilde{n}$.

LEMMA 4.2. *Let $s, t \in [0,1]$ be two points with $s < t$, and let $z \in (s,t)$. If*

$$|s - t| \leq \frac{1}{8}\, h_{\beta_*,n} \tag{4.2}$$

then

$$\frac{1}{3}\, \bar{h}_n(z) \leq \min\left\{ \bar{h}_n(s), \bar{h}_n(t) \right\} \leq 3\, \bar{h}_n(z).$$

LEMMA 4.3. *There exist positive and finite constants $c_5 = c_5(A,\nu,K)$ and $c_6 = c_6(A,\nu,L^*,K)$, and some $\eta_0 = \eta_0(A,\nu,L^*,K) > 0$, such that*

$$\sup_{p\in\mathscr{P}_n} \mathbb{P}_p^{\chi_i} \left(\sup_{s\in\mathcal{H}_n} \max_{h\in\mathcal{G}_n} \sqrt{\frac{\tilde{n}h}{\log \tilde{n}}} \left| \hat{p}_n^{(i)}(s,h) - \mathbb{E}_p^{\chi_i} \hat{p}_n^{(i)}(s,h) \right| > \eta \right) \leq c_5 \tilde{n}^{-c_6\eta}, \quad i = 1,2$$

for sufficiently large $n \geq n_0(A,\nu,L^*,K)$ and for all $\eta \geq \eta_0$.

The next lemma states extends the classical upper bound on the bias for the modified Hölder classes $\mathcal{H}_{\beta^*,B(t,U)}(\beta,L)$.

LEMMA 4.4. *Let $t\in\mathbb{R}$ and $g,h>0$. Any density $p:\mathbb{R}\to\mathbb{R}$ with $p_{|B(t,g+h)}\in\mathcal{H}_{\beta^*,B(t,g+h)}(\beta,L)$ for some $0<\beta\leq\infty$ and some $L>0$ satisfies*

$$\sup_{s\in B(t,g)}|(K_h*p)(s)-p(s)|\leq b_2h^\beta \tag{4.3}$$

for some positive and finite constant $b_2=b_2(L,K)$.

LEMMA 4.5. *For symmetric kernels K and $\beta=1$, the bias bound (4.3) continues to hold if the Lipschitz balls are replaced by the corresponding Zygmund balls.*

5. Proofs. We first prove the results of Section 3 in Subsection 5.1 and afterwards proceed with the proofs of the results Section 4 in Subsection 5.2.
For the subsequent proofs we recall the following notion of the theory of empirical processes.

DEFINITION 5.1. *A class of measurable functions $\mathscr{H}$ on a measure space $(S,\mathscr{S})$ is a Vapnik-Červonenkis class (VC class) of functions with respect to the envelope H if there exists a measurable function H which is everywhere finite with $\sup_{h\in\mathscr{H}}|h|\leq H$ and finite numbers A and v, such that*

$$\sup_Q N\Big(\mathscr{H},\|\cdot\|_{L^2(Q)},\varepsilon\|H\|_{L^2(Q)}\Big)\leq\left(\frac{A}{\varepsilon}\right)^v$$

for all $0<\varepsilon<1$, where the supremum is running over all probability measures Q on $(S,\mathscr{S})$ for which $\|H\|_{L^2(Q)}<\infty$.

Nolan and Pollard (1987) call a class *Euclidean* with respect to the envelope H and with characteristics A and ν if the same holds true with $L^1(Q)$ instead of $L^2(Q)$. The following auxiliary lemma is a direct consequence of the results in the same reference.

LEMMA 5.2. *If a class of measurable functions $\mathscr{H}$ is Euclidean with respect to a constant envelope H and with characteristics A and ν, then the class*

$$\tilde{\mathscr{H}}=\{h-\mathbb{E}_{\mathbb{P}}h:h\in\mathscr{H}\}$$

is a VC class with envelope $2H$ and characteristics $A'=4\sqrt{A}\vee 2A$ and $\nu'=3\nu$ for any probability measure $\mathbb{P}$.

PROOF. For any probability measure $\mathbb{P}$ and for any functions $\tilde{h}_1=h_1-\mathbb{E}_{\mathbb{P}}h_1$, $\tilde{h}_2=h_2-\mathbb{E}_{\mathbb{P}}h_2\in\tilde{\mathscr{H}}$ with $h_1,h_2\in\mathscr{H}$, we have

$$\|\tilde{h}_1-\tilde{h}_2\|_{L^2(Q)}\leq\|h_1-h_2\|_{L^2(Q)}+\|h_1-h_2\|_{L^1(\mathbb{P})}.$$

For any $0 < \varepsilon \le 1$, we obtain as a direct consequence of Lemma 14 in Nolan and Pollard (1987)

$$N\left(\tilde{\mathscr{H}}, L^2(Q), 2\varepsilon\|H\|_{L^2(Q)}\right) \le N\left(\mathscr{H}, L^2(Q), \frac{\varepsilon\|H\|_{L^2(Q)}}{2}\right) \cdot N\left(\mathscr{H}, L^1(\mathbb{P}), \frac{\varepsilon\|H\|_{L^1(\mathbb{P})}}{2}\right). \tag{5.1}$$

Nolan and Pollard (1987), page 789, furthermore state that the Euclidean class $\mathscr{H}$ is also a VC class with respect to the envelope H and with

$$N\left(\mathscr{H}, L^2(Q), \frac{\varepsilon\|H\|_{L^2(Q)}}{2}\right) \le \left(\frac{4\sqrt{A}}{\varepsilon}\right)^{2\nu},$$

whereas

$$N\left(\mathscr{H}, L^1(\mathbb{P}), \frac{\varepsilon\|H\|_{L^1(\mathbb{P})}}{2}\right) \le \left(\frac{2A}{\varepsilon}\right)^{\nu}.$$

Inequality (5.1) thus implies

$$N\left(\tilde{\mathscr{H}}, L^2(Q), 2\varepsilon\|H\|_{L^2(Q)}\right) \le \left(\frac{4\sqrt{A} \vee 2A}{\varepsilon}\right)^{3\nu}.$$

□

5.1. *Proofs of the results in Section 3.*

Proof of Lemma 3.2. Let $p \in \mathscr{P}_n^{\mathrm{adm}}(K, \beta_*, L^*, \varepsilon)$ be an admissible density. That is, for any $t \in [0,1]$ and for any $h \in \mathcal{G}_\infty$ there exists some $\beta \in [\beta_*, \beta^*] \cup \{\infty\}$, such that for $u = h$ or $u = 2h$ both

$$p_{|B(t,u)} \in \mathcal{H}_{\beta^*, B(t,u)}(\beta, L^*)$$

and

$$\sup_{s \in B(t,u-g)} |(K_g * p)(s) - p(s)| \ge \frac{g^\beta}{\log n} \quad \text{for all } g \in \mathcal{G}_\infty \text{ with } g \le u/8$$

hold. By definition of $\beta_p(B(t,u))$ in (3.5), we obtain $\beta_p(B(t,u)) \ge \beta$. We now prove by contradiction that also $\beta_p(B(t,u)) \le \beta$. If $\beta = \infty$, the proof is finished. Assume now that $\beta < \infty$ and that $\beta_p(B(t,u)) > \beta$. Then, by Lemma A.4, there exists some $\beta < \beta' < \beta_p(B(t,u))$ with $p_{|B(t,u)} \in \mathcal{H}_{\beta^*, B(t,u)}(\beta', L^*)$. By Lemma 4.4, there exists some constant $b_2 = b_2(L^*, K)$ with

$$b_2 g^{\beta'} \ge \sup_{s \in B(t,u-g)} |(K_g * p)(s) - p(s)| \ge \frac{g^\beta}{\log n}$$

for all $g \in \mathcal{G}_\infty$ with $g \le u/8$, which is a contradiction. □

Proof of Proposition 3.3. The proof is based on a reduction of the supremum over the class to a maximum over two distinct hypotheses.

Part 1. For $\beta \in [\beta_*, 1)$, the construction of the hypotheses is based on the Weierstraß function as defined in (3.8) and is depicted in Figure 2. Consider the function $p_0 : \mathbb{R} \to \mathbb{R}$ with

$$p_0(x) = \begin{cases} 0, & \text{if } \ |x-t| \geq \frac{10}{3} \\ \frac{1}{4} + \frac{3}{16}(x-t+2), & \text{if } \ -\frac{10}{3} < x-t < -2 \\ \frac{1}{6} + \frac{1-2^{-\beta}}{12} W_\beta(x-t), & \text{if } \ |x-t| \leq 2 \\ \frac{1}{4} - \frac{3}{16}(x-t-2), & \text{if } \ 2 < x-t < \frac{10}{3} \end{cases}$$

and the function $p_{1,n} : \mathbb{R} \to \mathbb{R}$ with

$$p_{1,n}(x) = p_0(x) + q_{t+\frac{9}{4},n}(x) - q_{t,n}(x), \quad x \in \mathbb{R},$$

where

$$q_{a,n}(x) = \begin{cases} 0, & \text{if } \ |x-a| > g_{\beta,n} \\ \frac{1-2^{-\beta}}{12}\Big(W_\beta(x-a) - W_\beta(g_{\beta,n})\Big), & \text{if } \ |x-a| \leq g_{\beta,n} \end{cases}$$

for $g_{\beta,n} = \frac{1}{4} n^{-1/(2\beta+1)}$ and $a \in \mathbb{R}$. Note that $p_{1,n|B(t,g_{\beta,n})}$ is constant with value

$$p_{1,n}(x) = \frac{1}{6} + \frac{1-2^{-\beta}}{12} W_\beta(g_{\beta,n}) \quad \text{ for all } x \in B(t, g_{\beta,n}).$$

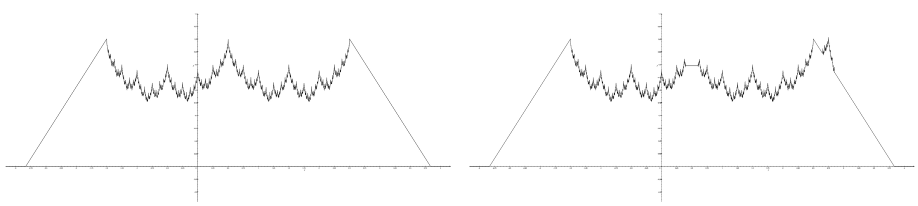

FIG 2. *Functions p_0 and $p_{1,n}$ for $t = 0.5$, $\beta = 0.5$ and $n = 100$*

We now show that both p_0 and $p_{1,n}$ are contained in the class $\mathscr{P}_k$ for sufficiently large $k \geq k_0(\beta_*)$ with

$$p_{0|(-\varepsilon,1+\varepsilon)},\, p_{1,n|(-\varepsilon,1+\varepsilon)} \in \mathcal{H}_{(-\varepsilon,1+\varepsilon)}(\beta, L^*).$$

(i) We first verify that p_0 integrates to one. Then, it follows directly that also $p_{1,n}$ integrates to one. We have

$$\begin{aligned}\int p_0(x)\,\mathrm{d}x = &\int_{t-\frac{10}{3}}^{t-2} \left(\frac{1}{4} + \frac{3}{16}(x-t+2)\right) \mathrm{d}x \\ &+ \int_{t-2}^{t+2} \left(\frac{1}{6} + \frac{1-2^{-\beta}}{12} W_\beta(x-t)\right) \mathrm{d}x \\ &+ \int_{t+2}^{t+\frac{10}{3}} \left(\frac{1}{4} - \frac{3}{16}(x-t-2)\right) \mathrm{d}x\end{aligned}$$

$$= \frac{1}{6} + \frac{2}{3} + \frac{1-2^{-\beta}}{12}\int_{-2}^{2} W_\beta(x)\,\mathrm{d}x + \frac{1}{6}$$
$$= 1,$$

where the last equality is due to

$$\int_{-2}^{2} W_\beta(x)\,\mathrm{d}x = \sum_{k=0}^{\infty} 2^{-k\beta}\int_{-2}^{2}\cos(2^k\pi x)\,\mathrm{d}x = 0.$$

(ii) Next, we check the non-negativity of p_0 and $p_{1,n}$ to show that they are probability density functions. We prove non-negativity for p_0, whereas non-negativity of $p_{1,n}$ is an easy implication. Since $p_0(-10/3) = 0$ and p_0 is linear on $(t-10/3, t-2)$ with positive derivative, p_0 is non-negative on $(t-10/3, t-2)$. Analogously, p_0 is non-negative on $(t+2, t+10/3)$. Note furthermore that

$$|W_\beta(x)| \le W_\beta(0) = \sum_{k=0}^{\infty} 2^{-k\beta} = \frac{1}{1-2^{-\beta}} \tag{5.2}$$

for all $x \in \mathbb{R}$. Thus, for any $x \in \mathbb{R}$ with $|x-t| \le 2$, we have

$$p_0(x) = \frac{1}{6} + \frac{1-2^{-\beta}}{12} W_\beta(x-t) \ge \frac{1}{6} - \frac{1}{12} = \frac{1}{12} > 0.$$

(iii) As p_0 and also $p_{1,n}$ are bounded from below by $M = 1/12$ on $B(t,2)$, we furthermore conclude that they are bounded from below by M on $(-1,2) \subset B(t,2)$, and therefore on any interval $[-\varepsilon, 1+\varepsilon]$ with $0 < \varepsilon < 1$.

(iv) We now verify that $p_{0|(-\varepsilon,1+\varepsilon)}, p_{1,n|(-\varepsilon,1+\varepsilon)} \in \mathcal{H}_{(-\varepsilon,1+\varepsilon)}(\beta, L(\beta))$ for some positive constant $L(\beta)$. Note again that for any $0 < \varepsilon < 1$ and any $t \in [0,1]$, the inclusion $(-\varepsilon, 1+\varepsilon) \subset B(t,2)$ holds. Thus,

$$\sup_{\substack{x,y \in (-\varepsilon,1+\varepsilon)\\ x \ne y}} \frac{|p_0(x)-p_0(y)|}{|x-y|^\beta} = \frac{1-2^{-\beta}}{12} \cdot \sup_{\substack{x,y \in (-\varepsilon,1+\varepsilon)\\ x \ne y}} \frac{|W_\beta(x-t) - W_\beta(y-t)|}{|(x-t)-(y-t)|^\beta},$$

which is bounded by some constant $c(\beta)$ according to Lemma 3.5. Together with (5.2) and with the triangle inequality, we obtain that

$$p_{0|(-\varepsilon,1+\varepsilon)} \in \mathcal{H}_{(-\varepsilon,1+\varepsilon)}(\beta, L)$$

for some Lipschitz constant $L = L(\beta)$. The Hölder continuity of $p_{1,n}$ is now a simple consequence. The function $p_{1,n}$ is constant on $B(t, g_{\beta,n})$ and coincides with p_0 on $(-\varepsilon, 1+\varepsilon) \setminus B(t, g_{\beta,n})$. Hence, it remains to investigate combinations of points $x \in (-\varepsilon, 1+\varepsilon) \setminus B(t, g_{\beta,n})$ and $y \in B(t, g_{\beta,n})$. Without loss of generality assume that $x \le t - g_{\beta,n}$. Then,

$$\frac{|p_{1,n}(x)-p_{1,n}(y)|}{|x-y|^\beta} = \frac{|p_{1,n}(x)-p_{1,n}(t-g_{\beta,n})|}{|x-y|^\beta} \le \frac{|p_{1,n}(x)-p_{1,n}(t-g_{\beta,n})|}{|x-(t-g_{\beta,n})|^\beta} \le L,$$

which proves that also

$$p_{1,n|(-\varepsilon,1+\varepsilon)} \in \mathcal{H}_{(-\varepsilon,1+\varepsilon)}(\beta, L).$$

(v) Finally, we address the verification of Assumption 3.1 for the hypotheses p_0 and $p_{1,n}$. Again, for any $t' \in [0,1]$ and any $h \in \mathcal{G}_\infty$ the inclusion $B(t',2h) \subset B(t,2)$ holds, such that in particular

$$p_{0|B(t',h)} \in \mathcal{H}_{\beta^*,B(t',h)}(\beta, L_W(\beta))$$

for any $t' \in [0,1]$ and for any $h \in \mathcal{G}_\infty$ by Lemma 3.5. Simultaneously, Lemma 3.5 implies

$$\sup_{s\in B(t',h-g)} |(K_{R,g} * p_0)(s) - p_0(s)| > \frac{1-2^{-\beta_*}}{12}\left(\frac{4}{\pi}-1\right) g^\beta \geq \frac{g^\beta}{\log k}$$

for all $g \leq h/2$ and for sufficiently large $k \geq k_0(\beta_*)$. That is, for any $t' \in [0,1]$, both (3.6) and (3.7) are satisfies for p_0 with $u = h$ for any $h \in \mathcal{G}_\infty$.
Concerning $p_{1,n}$ we distinguish between several combinations of pairs (t',h) with $t' \in [0,1]$ and $h \in \mathcal{G}_\infty$.

$(v.1)$ If $B(t',h) \cap B(t, g_{\beta,n}) = \emptyset$, the function $p_{1,n}$ coincides with p_0 on $B(t',h)$, for which Assumption 3.1 has been already verified.

$(v.2)$ If $B(t',h) \subset B(t, g_{\beta,n})$, the function $p_{1,n}$ is constant on $B(t',h)$, such that (3.6) and (3.7) trivially hold for $u = h$ and $\beta = \infty$.

$(v.3)$ If $B(t',h) \cap B(t, g_{\beta,n}) \neq \emptyset$ and $B(t',h) \not\subset B(t, g_{\beta,n})$, we have that $t' + h > t + g_{\beta,n}$ or $t' - h < t - g_{\beta,n}$. As $p_{1,n|B(t,2)}$ is symmetric around t we assume $t' + h > t + g_{\beta,n}$ without loss of generality. In this case,

$$(t' + 2h - g) - (t + g_{\beta,n}) > 2\left(\frac{h}{2} - g\right),$$

such that

$$B\left(t' + \frac{3}{2}h, \frac{h}{2} - g\right) \subset B(t', 2h - g) \setminus B(t, g_{\beta,n}).$$

Consequently, we obtain

$$\sup_{s\in B(t',2h-g)} |(K_{R,g} * p_{1,n})(s) - p_{1,n}(s)| \geq \sup_{s\in B\left(t'+\frac{3}{2}h, \frac{h}{2}-g\right)} |(K_{R,g} * p_{1,n})(s) - p_{1,n}(s)|.$$

If $2h \geq 8g$, we conclude that $h/2 \geq 2g$, so that Lemma 3.5 finally proves Assumption 3.1 for $u = 2h$ to the exponent β for sufficiently large $k \geq k_0(\beta_*)$.

Combining $(i) - (v)$, we conclude that p_0 and $p_{1,n}$ are contained in the class $\mathscr{P}_k$ with $p_{0|(-\varepsilon,1+\varepsilon)}, p_{1,n|(-\varepsilon,1+\varepsilon)} \in \mathcal{H}_{(-\varepsilon,1+\varepsilon)}(\beta, L^*)$ for sufficiently large $k \geq k_0(\beta_*)$. The absolute distance of the two hypotheses in t is at least

$$|p_0(t) - p_{1,n}(t)| = \frac{1-2^{-\beta}}{12}\left(W_\beta(0) - W_\beta(g_{\beta,n})\right)$$

$$
\begin{aligned}
&= \frac{1-2^{-\beta}}{12}\sum_{k=0}^{\infty} 2^{-k\beta}\left(1-\cos(2^k\pi g_{\beta,n})\right)\\
&\geq \frac{1-2^{-\beta_*}}{12} 2^{-\tilde{k}\beta}\left(1-\cos(2^{\tilde{k}}\pi g_{\beta,n})\right)\\
&\geq 2c_7\, g_{\beta,n}^{\beta} \qquad (5.3)
\end{aligned}
$$

where $\tilde{k} \in \mathbb{N}$ is chosen such that $2^{-(\tilde{k}+1)} < g_{\beta,n} \leq 2^{-\tilde{k}}$ and

$$
c_7 = c_7(\beta_*) = \frac{1-2^{-\beta_*}}{24}.
$$

It remains to bound the distance between the associated product probability measures $\mathbb{P}_0^{\otimes n}$ and $\mathbb{P}_{1,n}^{\otimes n}$. For this purpose, we analyze the Kullback-Leibler divergence between these probability measures, which can be bounded from above by

$$
\begin{aligned}
K(\mathbb{P}_{1,n}^{\otimes n}, \mathbb{P}_0^{\otimes n}) &= n\, K(\mathbb{P}_{1,n}, \mathbb{P}_0)\\
&= n \int p_{1,n}(x) \log\left(\frac{p_{1,n}(x)}{p_0(x)}\right) \mathbb{1}\left\{p_0(x) > 0\right\} \mathrm{d}x\\
&= n \int p_{1,n}(x) \log\left(1 + \frac{q_{t+\frac{9}{4},n}(x) - q_{t,n}(x)}{p_0(x)}\right) \mathbb{1}\left\{p_0(x) > 0\right\} \mathrm{d}x\\
&\leq n \int q_{t+\frac{9}{4},n}(x) - q_{t,n}(x) + \frac{\left(q_{t+\frac{9}{4},n}(x) - q_{t,n}(x)\right)^2}{p_0(x)} \mathbb{1}\left\{p_0(x) > 0\right\} \mathrm{d}x\\
&= n \int \frac{\left(q_{t+\frac{9}{4},n}(x) - q_{t,n}(x)\right)^2}{p_0(x)} \mathbb{1}\left\{p_0(x) > 0\right\} \mathrm{d}x\\
&\leq 12n \int \left(q_{t+\frac{9}{4},n}(x) - q_{t,n}(x)\right)^2 \mathrm{d}x\\
&= 24n \int q_{0,n}(x)^2\, \mathrm{d}x\\
&= 24n \left(\frac{1-2^{-\beta}}{12}\right)^2 \int_{-g_{\beta,n}}^{g_{\beta,n}} \left(W_\beta(x) - W_\beta(g_{\beta,n})\right)^2 \mathrm{d}x\\
&\leq 24L(\beta)^2 n \left(\frac{1-2^{-\beta}}{12}\right)^2 \int_{-g_{\beta,n}}^{g_{\beta,n}} (g_{\beta,n} - x)^{2\beta}\, \mathrm{d}x\\
&\leq c_8 n g_{\beta,n}^{2\beta+1}\\
&\leq c_8
\end{aligned}
$$

using the inequality $\log(1+x) \leq x$, $x > -1$, Lemma 3.5, and

$$
p_0(t + 5/2) = \frac{5}{32} > M = \frac{1}{12},
$$

where

$$c_8 = c_8(\beta) = 48L(\beta)^2 4^{-(2\beta+1)} 2^{2\beta} \left(\frac{1-2^{-\beta}}{12} \right)^2 .$$

Using now Theorem 2.2 in Tsybakov (2009),

$$\inf_{T_n} \sup_{\substack{p \in \mathscr{P}_k: \\ p_{|(-\varepsilon,1+\varepsilon)} \in \mathcal{H}_{(-\varepsilon,1+\varepsilon)}(\beta,L^*)}} \mathbb{P}_p^{\otimes n} \left(n^{\frac{\beta}{2\beta+1}} |T_n(t) - p(t)| \geq c_7 \right)$$
$$\geq \max \left\{ \frac{1}{4} \exp(-c_8), \, \frac{1-\sqrt{c_8/2}}{2} \right\} > 0.$$

Part 2. For $\beta = 1$, consider the function $p_0 : \mathbb{R} \to \mathbb{R}$ with

$$p_0(x) = \begin{cases} 0, & \text{if } |x-t| > 4 \\ \frac{1}{4} - \frac{1}{16}|x-t|, & \text{if } |x-t| \leq 4 \end{cases}$$

and the function $p_{1,n} : \mathbb{R} \to \mathbb{R}$ with

$$p_{1,n}(x) = p_0(x) + q_{t+\frac{9}{4},n}(x) - q_{t,n}(x), \quad x \in \mathbb{R},$$

where

$$q_{a,n}(x) = \begin{cases} 0, & \text{if } |x-a| > g_{1,n} \\ \frac{1}{16}(g_{1,n} - |x-a|), & \text{if } |x-a| \leq g_{1,n} \end{cases}$$

for $g_{1,n} = n^{-1/3}$ and $a \in \mathbb{R}$. The construction is depicted in Figure 3 below.

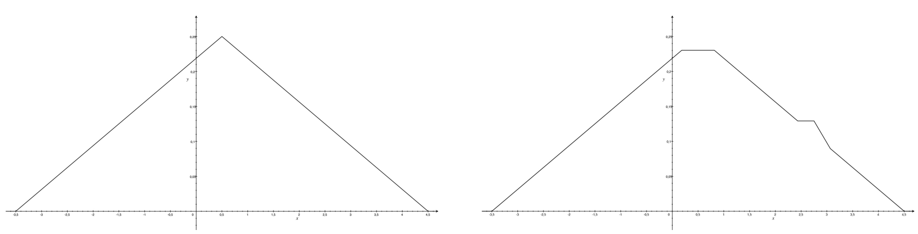

FIG 3. *Functions p_0 and $p_{1,n}$ for $t = 0.5$, $\beta = 0.5$ and $n = 10$*

$(i) - (iii)$ Easy calculations show that both p_0 and $p_{1,n}$ are probability densities, which are bounded from below by $M = 1/8$ on $B(t,2)$.

(iv) We now verify that $p_{0|(-\varepsilon,1+\varepsilon)}, p_{1,n|(-\varepsilon,1+\varepsilon)} \in \mathcal{H}_{(-\varepsilon,1+\varepsilon)}(1,L)$ for some Lipschitz constant $L > 0$. Note again that for any $0 < \varepsilon < 1$ and any $t \in [0,1]$, the inclusion $(-\varepsilon, 1+\varepsilon) \subset B(t,2)$ holds. Thus,

$$\sup_{\substack{x,y \in (-\varepsilon,1+\varepsilon) \\ x \neq y}} \frac{|p_0(x) - p_0(y)|}{|x-y|} = \frac{1}{16} \cdot \sup_{\substack{x,y \in (-\varepsilon,1+\varepsilon) \\ x \neq y}} \frac{||x-t| - |y-t||}{|x-y|} \leq \frac{1}{16}.$$

Since p_0 has maximal value $1/4$, we obtain that

$$p_{0|(-\varepsilon,1+\varepsilon)} \in \mathcal{H}_{(-\varepsilon,1+\varepsilon)}\left(1, \frac{5}{16}\right).$$

For the same reasons as before, we also obtain

$$p_{1,n|(-\varepsilon,1+\varepsilon)} \in \mathcal{H}_{(-\varepsilon,1+\varepsilon)}\left(1, \frac{5}{16}\right).$$

(v) Finally, we address the verification of Assumption 3.1 for the hypotheses p_0 and $p_{1,n}$. Again, for any $t' \in [0,1]$ and any $h \in \mathcal{G}_\infty$ the inclusion $B(t',2h) \subset B(t,2)$ holds, and we distinguish between several combinations of pairs (t',h) with $t' \in [0,1]$ and $h \in \mathcal{G}_\infty$. We start with p_0.

$(v.1)$ If $t \notin B(t',h)$, it holds that $\|p\|_{\beta,B(t',h)} \leq 5/16$ for all $\beta > 0$, such that (3.6) and (3.7) trivially hold for $u = h$ and $\beta = \infty$.

$(v.2)$ In case $t \in B(t',h)$, the function $p_{0|B(t',2h)}$ is not differentiable and

$$\|p_0\|_{1,B(t',2h)} \leq 5/16.$$

Furthermore, $t \in B(t', 2h - g)$ for any $g \in \mathcal{G}_\infty$ with $g < 2h/16$ and thus

$$\sup_{s \in B(t',2h-g)} |(K_{R,g} * p)(s) - p(s)| \geq |(K_{R,g} * p)(t) - p(t)| = \frac{1}{32} g.$$

That is, (3.6) and (3.7) are satisfied for $u = 2h$ and $\beta = 1$ for sufficiently large $n \geq n_0$.

The density $p_{1,n}$ can be treated in a similar way. It is constant on the interval $B(t, g_{\beta,n})$. If $B(t',h)$ does not intersect with $\{t - g_{\beta,n}, t + g_{\beta,n}\}$, Assumption 3.1 is satisfied for $u = h$ and $\beta = \infty$. If the two sets intersect, $t - g_{\beta,n}$ or $t + g_{\beta,n}$ is contained in $B(t', 2h - g)$ for any $g \in \mathcal{G}_\infty$ with $g < 2h/16$, and we proceed as before.

Again, combining $(i) - (v)$, it follows that p_0 and $p_{1,n}$ are contained in the class $\mathscr{P}_k$ with $p_{0|(-\varepsilon,1+\varepsilon)}, p_{1,n|(-\varepsilon,1+\varepsilon)} \in \mathcal{H}_{(-\varepsilon,1+\varepsilon)}(1, L)$ for sufficiently large $k \geq k_0$ and some universal constant $L > 0$. The absolute distance of the two hypotheses in t equals

$$|p_0(t) - p_{1,n}(t)| = \frac{1}{16} g_{1,n}.$$

To bound the Kullback-Leibler divergence between the associated product probability measures $\mathbb{P}_0^{\otimes n}$ and $\mathbb{P}_{1,n}^{\otimes n}$, we derive as before

$$\begin{aligned} K(\mathbb{P}_{1,n}^{\otimes n}, \mathbb{P}_0^{\otimes n}) &\leq n \int \frac{\left(q_{t+\frac{9}{4},n}(x) - q_{t,n}(x)\right)^2}{p_0(x)} \mathbb{1}\left\{p_0(x) > 0\right\} \mathrm{d}x \\ &\leq 16n \int \left(q_{t+\frac{9}{4},n}(x) - q_{t,n}(x)\right)^2 \mathrm{d}x \end{aligned}$$

$$= 32n \int q_{0,n}(x)^2 \, \mathrm{d}x$$
$$= \frac{1}{12},$$

using $p_0(t+5/2) > 1/16$. Using Theorem 2.2 in Tsybakov (2009) again,

$$\inf_{T_n} \sup_{\substack{p \in \mathscr{P}_k: \\ p_{|(-\varepsilon,1+\varepsilon)} \in \mathcal{H}_{(-\varepsilon,1+\varepsilon)}(1,L^*)}} \mathbb{P}_p^{\otimes n} \left(n^{\frac{1}{3}} \left| T_n(t) - p(t) \right| \geq \frac{1}{32} \right)$$
$$\geq \max \left\{ \frac{1}{4} \exp(-1/12), \, \frac{1 - \sqrt{1/24}}{2} \right\} > 0.$$

□

PROOF OF PROPOSITION 3.4. Define

$$\tilde{\mathscr{R}} = \bigcup_{n \in \mathbb{N}} \tilde{\mathscr{R}}_n$$

with

$$\tilde{\mathscr{R}}_n = \Bigg\{ p \in \mathcal{H}_{(-\varepsilon,1+\varepsilon)}(\beta_*) : \forall\, t \in [0,1] \;\forall\, h \in \mathcal{G}_\infty \;\exists\, \beta \in [\beta_*, \beta^*] \text{ with}$$
$$p_{|B(t,h)} \in \mathcal{H}_{B(t,h)}(\beta) \text{and } \|(K_{R,g} * p) - p\|_{B(t,h-g)} \geq \frac{g^\beta}{\log n}$$
$$\text{for all } g \in \mathcal{G}_\infty \text{ with } g \leq h/8 \Bigg\}.$$

Furthermore, let

$$E_n(\beta) = \Bigg\{ p \in \mathcal{H}_{(-\varepsilon,1+\varepsilon)}(\beta) : \|(K_{R,g} * p) - p\|_{B(t,h-g)} \geq \frac{2}{\log n} g^\beta \text{ for all } t \in [0,1],$$
$$\text{for all } h \in \mathcal{G}_\infty, \text{ and for all } g \in \mathcal{G}_\infty \text{ with } g \leq h/8 \Bigg\}.$$

Note that Lemma 3.5 shows that $E_n(\beta)$ is non-empty as soon as

$$\frac{2}{\log n} \leq 1 - \frac{4}{\pi}.$$

Note additionally that $E_n(\beta) \subset \tilde{\mathscr{R}}_n$ for any $\beta \in [\beta_*, \beta^*]$, and

$$\bigcup_{n \in \mathbb{N}} E_n(\beta) \subset \tilde{\mathscr{R}}.$$

With

$$A_n(\beta) = \left\{ \tilde{f} \in \mathcal{H}_{(-1,2)}(\beta) : \|\tilde{f} - f\|_{\beta,(-\varepsilon,1+\varepsilon)} < \frac{\|K_R\|_1^{-1}}{\log n} \text{ for some } f \in E_n(\beta) \right\},$$

we get for any $\tilde{f} \in A_n(\beta)$ and a corresponding $f \in E_n(\beta)$ with

$$\|\check{f}\|_{\beta,(-\varepsilon,1+\varepsilon)} < \|K_R\|_1^{-1} \frac{1}{\log n}$$

and $\check{f} = \tilde{f} - f$, the lower bound

$$\begin{aligned}
&\left\| (K_{R,g} * \tilde{f}) - \tilde{f} \right\|_{B(t,h-g)} \\
&\geq \left\| (K_{R,g} * f) - f \right\|_{B(t,h-g)} - \left\| \check{f} - (K_{R,g} * \check{f}) \right\|_{B(t,h-g)} \\
&= \frac{2}{\log n} g^\beta - \sup_{s \in B(t,h-g)} \left| \int K_R(x) \left\{ \check{f}(s+gx) - \check{f}(s) \right\} \mathrm{d}x \right| \\
&\geq \frac{2}{\log n} g^\beta - g^\beta \cdot \int |K_R(x)| \sup_{s \in B(t,h-g)} \sup_{\substack{s' \in B(s,g) \\ s' \neq s}} \frac{\left| \check{f}(s') - \check{f}(s) \right|}{|s-s'|^\beta} \, \mathrm{d}x \\
&\geq \frac{2}{\log n} g^\beta - g^\beta \cdot \|K_R\|_1 \cdot \|\check{f}\|_{\beta,(-\varepsilon,1+\varepsilon)} \\
&\geq \frac{1}{\log n} g^\beta
\end{aligned}$$

for all $g, h \in \mathcal{G}_\infty$ with $g \leq h/8$ and for all $t \in [0,1]$, and therefore

$$A = \bigcup_{n \in \mathbb{N}} A_n(\beta) \subset \tilde{\mathscr{R}}.$$

Clearly, $A_n(\beta)$ is open in $\mathcal{H}_{(-\varepsilon,1+\varepsilon)}(\beta)$. Hence, the same holds true for A. Next, we verify that A is dense in $\mathcal{H}_{(-\varepsilon,1+\varepsilon)}(\beta)$. Let $p \in \mathcal{H}_{(-\varepsilon,1+\varepsilon)}(\beta)$ and let $\delta > 0$. We now show that there exists some function $\tilde{p}_\delta \in A$ with $\|p - \tilde{p}_\delta\|_{\beta,(-\varepsilon,1+\varepsilon)} \leq \delta$. For the construction of the function $\tilde{p}_\delta$, set the grid points

$$t_{j,1}(k) = (4j+1)2^{-k}, \quad t_{j,2}(k) = (4j+3)2^{-k}$$

for $j \in \{-2^{k-2}, -2^{k-2}+1, \ldots, 2^{k-1}-1\}$ and $k \geq 2$. The function $\tilde{p}_\delta$ shall be defined as the limit of a recursively constructed sequence. The idea is to recursively add appropriately rescaled sine waves at those locations where the bias condition is violated. Let $p_{1,\delta} = p$, and denote

$$\begin{aligned}
J_k = \Bigg\{ j \in \{-2^{k-2}, \ldots, 2^{k-1}-1\} : \max_{i=1,2} \left| (K_{R,2^{-k}} * p_{k-1,\delta})(t_{j,i}(k)) - p_{k-1,\delta}(t_{j,i}(k)) \right| \\
< \frac{1}{2} c_9 \, \delta \left(1 - \frac{2}{\pi}\right) 2^{-k\beta} \Bigg\}
\end{aligned}$$

for $k \geq 2$, where

$$c_9 = c_9(\beta) = \left(\frac{3\pi}{2} \cdot \frac{1}{1 - 2^{\beta - 1}} + \frac{7}{1 - 2^{-\beta}} \right)^{-1}.$$

For any $k \geq 2$ set

$$p_{k,\delta}(x) = p_{k-1,\delta}(x) + c_9\, \delta \sum_{j \in J_k} S_{k,\beta,j}(x)$$

with functions

$$S_{k,\beta,j}(x) = 2^{-k\beta} \sin\left(2^{k-1}\pi x\right) \mathbb{1}\left\{ |(4j+2)2^{-k} - x| \leq 2^{-k+1} \right\}$$

exemplified in Figure 4. That is,

$$p_{k,\delta}(x) = p(x) + c_9\, \delta \sum_{l=2}^{k} \sum_{j \in J_l} S_{l,\beta,j}(x),$$

and we define $\tilde{p}_\delta$ as the limit

$$\begin{aligned} \tilde{p}_\delta(x) &= p(x) + c_9\, \delta \sum_{l=2}^{\infty} \sum_{j \in J_l} S_{l,\beta,j}(x) \\ &= p_{k,\delta}(x) + c_9\, \delta \sum_{l=k+1}^{\infty} \sum_{j \in J_l} S_{l,\beta,j}(x). \end{aligned}$$

The function $\tilde{p}_\delta$ is well-defined as the series on the right-hand side converges: for fixed $l \in \mathbb{N}$, the indicator functions

$$\mathbb{1}\left\{ |(4j+2)2^{-k} - x| \leq 2^{-k+1} \right\}, \quad j \in \{-2^{l-2}, -2^{l-2}+1, \ldots, 2^{l-1}-1\}$$

have disjoint supports, such that

$$\left\| \sum_{j \in J_l} S_{l,\beta,j} \right\|_{(-\varepsilon, 1+\varepsilon)} \leq 2^{-l\beta}.$$

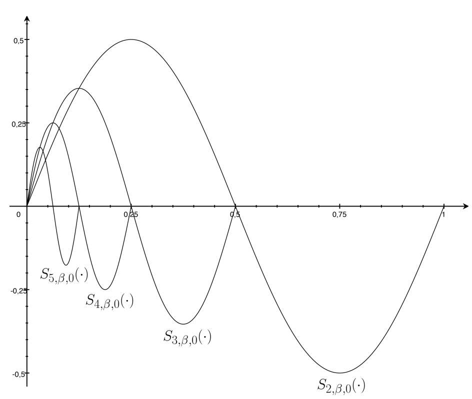

FIG 4. *Functions $S_{k,\beta,0}$ for $k = 2, \ldots, 5$ and $\beta = 0.5$*

It remains to verify that $\tilde{p}_\delta \in \bigcup_{n\in\mathbb{N}} E_n(\beta) \subset A$ and also $\|p-\tilde{p}_\delta\|_{\beta,(-\varepsilon,1+\varepsilon)} \le \delta$. As concerns the inequality $\|p-\tilde{p}_\delta\|_{\beta,(-\varepsilon,1+\varepsilon)} \le \delta$, it remains to show that

$$\left\| \sum_{l=2}^{\infty} \sum_{j\in J_l} S_{l,\beta,j} \right\|_{\beta,(-\varepsilon,1+\varepsilon)} \le \frac{1}{c_9}.$$

For $s,t \in (-\varepsilon, 1+\varepsilon)$ with $|s-t| \le 1$, we obtain

$$\begin{aligned}
&\left| \sum_{l=2}^{\infty} \sum_{j\in J_l} S_{l,\beta,j}(s) - \sum_{l=2}^{\infty} \sum_{j\in J_l} S_{l,\beta,j}(t) \right| \\
&\quad \le \sum_{l=2}^{\infty} 2^{-l\beta} \Bigg| \sin(2^{l-1}\pi s) \sum_{j\in J_l} \mathbb{1}\{|(4j+2)2^{-l} - s| \le 2^{-l+1}\} \\
&\qquad\qquad - \sin(2^{l-1}\pi t) \sum_{j\in J_l} \mathbb{1}\{|(4j+2)2^{-l} - t| \le 2^{-l+1}\} \Bigg|. \qquad (5.4)
\end{aligned}$$

Choose now $k' \in \mathbb{N}$ maximal, such that both

$$(4j+2)2^{-k'} - 2^{-k'+1} \le s \le (4j+2)2^{-k'} + 2^{-k'+1}$$

and

$$(4j+2)2^{-k'} - 2^{-k'+1} \le t \le (4j+2)2^{-k'} + 2^{-k'+1}$$

for some $j \in \{-2^{k'-2}, \dots, 2^{k'-1}-1\}$. For $2 \le l \le k'$, we have

$$\begin{aligned}
&\Bigg| \sin(2^{l-1}\pi s) \sum_{j\in J_l} \mathbb{1}\{|(4j+2)2^{-l} - s| \le 2^{-l+1}\} \\
&\qquad - \sin(2^{l-1}\pi t) \sum_{j\in J_l} \mathbb{1}\{|(4j+2)2^{-l} - t| \le 2^{-l+1}\} \Bigg| \\
&\quad \le \left| \sin(2^{l-1}\pi s) - \sin(2^{l-1}\pi t) \right| \\
&\quad \le \min\left\{ 2^{l-1}\pi|s-t|, 2 \right\} \qquad (5.5)
\end{aligned}$$

by the mean value theorem. For $l \ge k'+1$,

$$\begin{aligned}
&\Bigg| \sin(2^{l-1}\pi s) \sum_{j\in J_l} \mathbb{1}\{|(4j+2)2^{-l} - s| \le 2^{-l+1}\} \\
&\qquad - \sin(2^{l-1}\pi t) \sum_{j\in J_l} \mathbb{1}\{|(4j+2)2^{-l} - t| \le 2^{-l+1}\} \Bigg| \\
&\quad \le \max\left\{ \left|\sin(2^{l-1}\pi s)\right|, \left|\sin(2^{l-1}\pi t)\right| \right\}.
\end{aligned}$$

Furthermore, due to the choice of k', there exists some $z \in [s,t]$ with

$$\sin(2^{l-1}\pi z) = 0$$

for all $l \geq k'+1$. Thus, for any $l \geq k'+1$, by the mean value theorem,

$$\begin{aligned}\left|\sin(2^{l-1}\pi s)\right| &= \left|\sin(2^{l-1}\pi s) - \sin(2^{l-1}\pi z)\right| \\ &\leq \min\left\{2^{l-1}\pi|s-z|, 1\right\} \\ &\leq \min\left\{2^{l-1}\pi|s-t|, 1\right\}.\end{aligned}$$

Analogously, we obtain

$$\left|\sin(2^{l-1}\pi t)\right| \leq \min\left\{2^{l-1}\pi|s-t|, 1\right\}.$$

Consequently, together with inequality (5.4) and (5.5),

$$\left|\sum_{l=2}^{\infty}\sum_{j\in J_l} S_{l,\beta,j}(s) - \sum_{l=2}^{\infty}\sum_{j\in J_l} S_{l,\beta,j}(t)\right| \leq \sum_{l=2}^{\infty} 2^{-l\beta}\min\left\{2^{l-1}\pi|s-t|,\, 2\right\}.$$

Choose now $k \in \mathbb{N}\cup\{0\}$, such that $2^{-(k+1)} < |s-t| \leq 2^{-k}$. If $k \leq 1$,

$$\sum_{l=2}^{\infty} 2^{-l\beta}\min\left\{2^{l-1}\pi|s-t|,\, 2\right\} \leq 2\frac{2^{-2\beta}}{1-2^{-\beta}} \leq \frac{2}{1-2^{-\beta}}|s-t|^{\beta}.$$

If $k \geq 2$, we decompose

$$\begin{aligned}\sum_{l=2}^{\infty} 2^{-l\beta}\min\left\{2^{l-1}\pi|s-t|,\, 2\right\} &\leq \frac{\pi}{2}|s-t|\sum_{l=0}^{k} 2^{l(1-\beta)} + 2\sum_{l=k+1}^{\infty} 2^{-l\beta} \\ &= \frac{\pi}{2}|s-t|\,\frac{2^{k(1-\beta)}-2^{\beta-1}}{1-2^{\beta-1}} + 2\cdot\frac{2^{-(k+1)\beta}}{1-2^{-\beta}} \\ &\leq |s-t|^{\beta}\cdot\left(\frac{\pi}{2}\cdot\frac{1}{1-2^{\beta-1}} + \frac{2}{1-2^{-\beta}}\right).\end{aligned}$$

Since furthermore

$$\left\|\sum_{l=2}^{\infty}\sum_{j\in J_l} S_{l,\beta,j}\right\|_{\sup} \leq \frac{1}{1-2^{-\beta}},$$

we have

$$\left\|\sum_{l=2}^{\infty}\sum_{j\in J_l} S_{l,\beta,j}\right\|_{\beta,(-\varepsilon,1+\varepsilon)} \leq 3\left(\frac{\pi}{2}\cdot\frac{1}{1-2^{\beta-1}} + \frac{2}{1-2^{-\beta}}\right) + \frac{1}{1-2^{-\beta}} = \frac{1}{c_9}$$

and finally $\|p - \tilde{p}_\varepsilon\|_{\beta,(-\varepsilon,1+\varepsilon)} \leq \delta$. In particular $\tilde{p}_\delta \in \mathcal{H}_{(-\varepsilon,1+\varepsilon)}(\beta)$.

We now show that the function $\tilde{p}_\delta$ is contained in $\bigcup_{n\in\mathbb{N}} E_n(\beta) \subset A$. For any bandwidths $g, h \in \mathcal{G}_\infty$ with $g \leq h/8$, it holds that $h - g \geq 4g$. Thus, for any $g = 2^{-k}$ with $k \geq 2$ and for any $t \in (-\varepsilon, 1+\varepsilon)$, there exists some $j = j(t,h,g) \in \{-2^{k-2}, \ldots, 2^{k-1}-1\}$ such that both $t_{j,1}(k)$ and $t_{j,2}(k)$ are contained in $B(t, h-g)$, which implies

$$\sup_{s\in B(t,h-g)} |(K_{R,g} * \tilde{p}_\delta)(s) - \tilde{p}_\delta(s)| \geq \max_{i=1,2} |(K_{R,g} * \tilde{p}_\delta)(t_{j,i}(k)) - \tilde{p}_\delta(t_{j,i}(k))| . \tag{5.6}$$

By linearity of the convolution and the theorem of dominated convergence,

$$\begin{aligned}
&(K_{R,g} * \tilde{p}_\delta)(t_{j,i}(k)) - \tilde{p}_\delta(t_{j,i}(k)) \\
&\quad = (K_{R,g} * p_{k,\delta})(t_{j,i}(k)) - p_{k,\delta}(t_{j,i}(k)) \\
&\qquad + c_9\,\delta \sum_{l=k+1}^{\infty} \sum_{j\in J_l} \Big((K_{R,g} * S_{l,\beta,j})(t_{j,i}(k)) - S_{l,\beta,j}(t_{j,i}(k)) \Big).
\end{aligned} \tag{5.7}$$

We analyze the convolution $K_{R,g} * S_{l,\beta,j}$ for $l \geq k+1$. Here,

$$\sin\big(2^{l-1}\pi\, t_{j,1}(k)\big) = \sin\big(2^{l-k-1}\pi\,(4j+1)\big) = 0$$

and

$$\sin\big(2^{l-1}\pi\, t_{j,2}(k)\big) = \sin\big(2^{l-k-1}\pi\,(4j+3)\big) = 0.$$

Hence,

$$\sum_{j\in J_l} S_{l,\beta,j}(t_{j,i}(k)) = 0, \quad i = 1,2$$

for any $l \geq k+1$. Furthermore,

$$\begin{aligned}
(K_{R,g} * S_{l,\beta,j})(t_{j,i}(k)) &= \frac{1}{2g}\int_{-g}^{g} S_{l,\beta,j}(t_{j,i}(k) - x)\,\mathrm{d}x \\
&= \frac{1}{2g}\int_{t_{j,i}(k)-g}^{t_{j,i}(k)+g} S_{l,\beta,j}(x)\,\mathrm{d}x, \quad i = 1,2.
\end{aligned}$$

Due to the identities

$$\begin{aligned}
(4j+2)2^{-k} - 2^{-k+1} &= t_{j,1}(k) - g \\
(4j+2)2^{-k} + 2^{-k+1} &= t_{j,2}(k) + g,
\end{aligned}$$

we have either

$$\big[(4j+2)2^{-l} - 2^{-l+1}, (4j+2)2^{-l} + 2^{-l+1}\big] \subset [t_{j,1}(k) - g, t_{j,2}(k) + g]$$

or

$$\big[(4j+2)2^{-l} - 2^{-l+1}, (4j+2)2^{-l} + 2^{-l+1}\big] \cap [t_{j,1}(k) - g, t_{j,2}(k) + g] = \emptyset$$

for any $l \geq k+1$. Therefore, for $i = 1, 2$,

$$\begin{aligned}
&\sum_{j \in J_l} (K_{R,g} * S_{l,\beta,j})(t_{j,i}(k)) \\
&\quad = \sum_{j \in J_l} \frac{1}{2g} \int_{t_{j,i}(k)-g}^{t_{j,i}(k)+g} 2^{-l\beta} \sin\left(2^{l-1}\pi x\right) \mathbb{1}\left\{|(4j+2)2^{-l} - x| \leq 2^{-l+1}\right\} \mathrm{d}x \\
&\quad = 0
\end{aligned}$$

such that equation (5.7) then simplifies to

$$(K_{R,g} * \tilde{p}_\delta)(t_{j,i}(k)) - \tilde{p}_\delta(t_{j,i}(k)) = (K_{R,g} * p_{k,\delta})(t_{j,i}(k)) - p_{k,\delta}(t_{j,i}(k)), \quad i = 1, 2.$$

Together with (5.6), we obtain

$$\sup_{s \in B(t,h-g)} |(K_{R,g} * \tilde{p}_\delta)(s) - \tilde{p}_\delta(s)| \geq \max_{i=1,2} |(K_{R,g} * p_{k,\delta})(t_{j,i}(k)) - p_{k,\delta}(t_{j,i}(k))|$$

for some $j \in \{-2^{k-2}, -2^{k-2}+1, \ldots, 2^{k-2}-1\}$. If $j \notin J_k$, then

$$\begin{aligned}
&\max_{i=1,2} |(K_{R,g} * p_{k,\delta})(t_{j,i}(k)) - p_{k,\delta}(t_{j,i}(k))| \\
&\qquad = \max_{i=1,2} |(K_{R,g} * p_{k-1,\delta})(t_{j,i}(k)) - p_{k-1,\delta}(t_{j,i}(k))| \\
&\qquad \geq \frac{1}{2} c_9\, \delta \left(1 - \frac{2}{\pi}\right) g^\beta.
\end{aligned}$$

If $j \in J_k$, then

$$\begin{aligned}
&\max_{i=1,2} |(K_{R,g} * p_{k,\delta})(t_{j,i}(k)) - p_{k,\delta}(t_{j,i}(k))| \\
&\quad \geq c_9\, \delta \max_{i=1,2} |(K_{R,g} * S_{k,\beta,j})(t_{j,i}(k)) - S_{k,\beta,j}(t_{j,i}(k))| \\
&\qquad - \max_{i=1,2} |(K_{R,g} * p_{k-1,\delta})(t_{j,i}(k)) - p_{k-1,\delta}(t_{j,i}(k))| \\
&\quad \geq c_9\, \delta \max_{i=1,2} |(K_{R,g} * S_{k,\beta,j})(t_{j,i}(k)) - S_{k,\beta,j}(t_{j,i}(k))| - \frac{1}{2} c_9\, \delta \left(1 - \frac{2}{\pi}\right) g^\beta.
\end{aligned}$$

Similar as above we obtain

$$\begin{aligned}
&(K_{R,g} * S_{k,\beta,j})(t_{j,1}(k)) - S_{k,\beta,j}(t_{j,1}(k)) \\
&\qquad = \frac{1}{2g} \int_{t_{j,1}(k)-g}^{t_{j,1}(k)+g} 2^{-k\beta} \sin\left(2^{k-1}\pi x\right) \mathrm{d}x - 2^{-k\beta} \\
&\qquad = \frac{1}{2g} 2^{-k\beta} \int_0^{2^{-k+1}} \sin\left(2^{k-1}\pi x\right) \mathrm{d}x - 2^{-k\beta} \\
&\qquad = g^\beta \left(\frac{2}{\pi} - 1\right)
\end{aligned}$$

as well as

$$(K_{R,g} * S_{k,\beta,j})(t_{j,2}(k)) - S_{k,\beta,j}(t_{j,2}(k)) = g^{\beta}\left(1 - \frac{2}{\pi}\right),$$

such that

$$\max_{i=1,2} |(K_{R,g} * p_{k,\delta})(t_{j,i}(k)) - p_{k,\delta}(t_{j,i}(k))| \geq \frac{1}{2} c_9\, \delta \left(1 - \frac{2}{\pi}\right) g^{\beta}.$$

Combining the two cases finally gives

$$\sup_{s \in B(t,h-g)} |(K_{R,g} * \tilde{p}_{\delta})(s) - \tilde{p}_{\delta}(s)| \geq \frac{1}{2} c_9\, \delta \left(1 - \frac{2}{\pi}\right) g^{\beta}.$$

In particular, $\tilde{p}_{\delta} \in E_n(\beta)$ for sufficiently large $n \geq n_0(\beta, \delta)$, and thus $\tilde{p}_{\delta} \in A$.

Since A is open and dense in the class $\mathcal{H}_{(-\varepsilon,1+\varepsilon)}(\beta)$ and $A \subset \tilde{\mathscr{R}}$, the complement $\mathcal{H}_{(-\varepsilon,1+\varepsilon)}(\beta) \setminus \tilde{\mathscr{R}}$ is nowhere dense in $H_{(-\varepsilon,1+\varepsilon)}(\beta)$. Thus, because of

$$\mathcal{H}_{(-\varepsilon,1+\varepsilon)}(\beta)_{|B(t,h)} = \mathcal{H}_{B(t,h)}(\beta),$$

and the fact that for any $x \in \mathcal{H}_{(-\varepsilon,1+\varepsilon)}(\beta)$ and any $z' \in \mathcal{H}_{B(t,h)}(\beta)$ with

$$\|x_{|B(t,h)} - z'\|_{\beta,B(t,h)} < \delta$$

there exists an extension $z \in \mathcal{H}_{(-\varepsilon,1+\varepsilon)}(\beta)$ of z' with

$$\|x - z\|_{\beta,(-\varepsilon,1+\varepsilon)} < \delta,$$

the set $\mathcal{H}_{B(t,h)}(\beta) \setminus \tilde{\mathscr{R}}_{|B(t,h)}$ is nowhere dense in $\mathcal{H}_{B(t,h)}(\beta)$. Since the property "nowhere dense" is stable when passing over to intersections and the corresponding relative topology, we conclude that

$$\mathcal{P}_{B(t,h)}(\beta, L^*) \setminus \mathscr{R}_{|B(t,h)}$$

is nowhere dense in $\mathcal{P}_{B(t,h)}(\beta, L^*)$ with respect to $\|\cdot\|_{\beta,B(t,h)}$. □

Proof of Lemma 3.5. As it has been proven in Hardy (1916) the Weierstraß function W_{β} is β-Hölder continuous everywhere. For the sake of completeness, we state the proof here. Because the Weierstraß function is 2-periodic, it suffices to consider points $s, t \in \mathbb{R}$ with $|s-t| \leq 1$. Note first that

$$\begin{aligned} |W_{\beta}(s) - W_{\beta}(t)| &\leq 2 \sum_{n=0}^{\infty} 2^{-n\beta} \left|\sin\left(\frac{1}{2} 2^n \pi (s+t)\right)\right| \cdot \left|\sin\left(\frac{1}{2} 2^n \pi (s-t)\right)\right| \\ &\leq 2 \sum_{n=0}^{\infty} 2^{-n\beta} \left|\sin\left(\frac{1}{2} 2^n \pi (s-t)\right)\right|. \end{aligned}$$

Choose $k \in \mathbb{N} \cup \{0\}$ such that $2^{-(k+1)} < |s-t| \leq 2^{-k}$. For all summands with index $n \leq k$, use the inequality $|\sin(x)| \leq |x|$ and for all summands with index $n > k$ use $|\sin(x)| \leq 1$, such that

$$|W_\beta(s) - W_\beta(t)| \leq 2 \sum_{n=0}^{k} 2^{-n\beta} \left| \frac{1}{2} 2^n \pi (s-t) \right| + 2 \sum_{n=k+1}^{\infty} 2^{-n\beta}$$
$$= \pi \, |s-t| \sum_{n=0}^{k} 2^{n(1-\beta)} + 2 \sum_{n=k+1}^{\infty} 2^{-n\beta}.$$

Note that,

$$\sum_{n=0}^{k} 2^{n(1-\beta)} = \frac{2^{(k+1)(1-\beta)} - 1}{2^{1-\beta} - 1} = \frac{2^{k(1-\beta)} - 2^{\beta-1}}{1 - 2^{\beta-1}} \leq \frac{2^{k(1-\beta)}}{1 - 2^{\beta-1}},$$

and, as $2^{-\beta} < 1$,

$$\sum_{n=k+1}^{\infty} 2^{-n\beta} = \frac{2^{-(k+1)\beta}}{1 - 2^{-\beta}}.$$

Consequently, we have

$$|W_\beta(s) - W_\beta(t)| \leq \pi \, |s-t| \, \frac{2^{k(1-\beta)}}{1 - 2^{\beta-1}} + 2 \frac{2^{-(k+1)\beta}}{1 - 2^{-\beta}}$$
$$\leq |s-t|^\beta \left(\frac{\pi}{1 - 2^{\beta-1}} + \frac{2}{1 - 2^{-\beta}} \right).$$

Furthermore

$$\|W_\beta\|_{\sup} \leq \sum_{n=0}^{\infty} 2^{-n\beta} = \frac{1}{1 - 2^{-\beta}},$$

so that for any interval $U \subset \mathbb{R}$,

$$\|W_\beta\|_{\beta,U} \leq \frac{\pi}{1 - 2^{\beta-1}} + \frac{3}{1 - 2^{-\beta}}.$$

We now turn to the proof of bias lower bound condition. For any $0 < \beta \leq 1$, for any $h \in \mathcal{G}_\infty$, for any $g = 2^{-k} \in \mathcal{G}_\infty$ with $g \leq h/2$, and for any $t \in \mathbb{R}$, there exists some $s_0 \in [t-(h-g), t+(h-g)]$ with $\cos\left(2^k \pi s_0\right) = 1$, since the function $x \mapsto \cos(2^k \pi x)$ is 2^{1-k}-periodic. Note that in this case also

$$\cos\left(2^n \pi s_0\right) = 1 \quad \text{for all } n \geq k. \tag{5.8}$$

The following supremum is now lower bounded by

$$\sup_{s \in B(t,h-g)} \left| \int K_{R,g}(x-s) W_\beta(x) \, \mathrm{d}x - W_\beta(s) \right|$$
$$\geq \left| \int_{-1}^{1} K_R(x) W_\beta(s_0 + gx) \, \mathrm{d}x - W_\beta(s_0) \right|.$$

As furthermore

$$\sup_{x\in\mathbb{R}} \left| K_R(x) 2^{-n\beta} \cos\left(2^n \pi (s_0 + gx)\right) \right| \le \|K_R\|_{\sup} \cdot 2^{-n\beta}$$

and

$$\sum_{n=0}^{\infty} \|K_R\|_{\sup} \cdot 2^{-n\beta} = \frac{\|K_R\|_{\sup}}{1-2^{-\beta}} < \infty,$$

the dominated convergence theorem implies

$$\left| \int_{-1}^{1} K_R(x) W_\beta(s_0 + gx)\, \mathrm{d}x - W_\beta(s_0) \right| = \left| \sum_{n=0}^{\infty} 2^{-n\beta} I_n(s_0, g) \right|$$

with

$$I_n(s_0, g) = \int_{-1}^{1} K_R(x) \cos\left(2^n \pi (s_0 + gx)\right) \mathrm{d}x - \cos\left(2^n \pi s_0\right).$$

Recalling (5.8), it holds for any index $n \ge k$

$$\begin{aligned} I_n(s_0, g) &= \frac{1}{2} \cdot \frac{\sin(2^n \pi (s_0 + g)) - \sin(2^n \pi (s_0 - g))}{2^n \pi g} - 1 \\ &= \frac{\sin(2^n \pi g)}{2^n \pi g} - 1 \\ &= -1. \end{aligned} \tag{5.9}$$

Furthermore, for any index $0 \le n \le k-1$ holds

$$\begin{aligned} I_n(s_0, g) &= \frac{1}{2} \cdot \frac{\sin(2^n \pi (s_0 + g)) - \sin(2^n \pi (s_0 - g))}{2^n \pi g} - \cos\left(2^n \pi s_0\right) \\ &= \cos(2^n \pi s_0) \left(\frac{\sin(2^n \pi g)}{2^n \pi g} - 1 \right). \end{aligned} \tag{5.10}$$

Using this representation, the inequality $\sin(x) \le x$ for $x \ge 0$, and Lemma A.3, we obtain

$$\begin{aligned} 2^{-n\beta} I_n(s_0, g) &\le 2^{-n\beta} \left(1 - \frac{\sin(2^n \pi g)}{2^n \pi g} \right) \\ &\le 2^{-n\beta} \cdot \frac{(2^n \pi g)^2}{6} \\ &\le 2^{-n\beta + 2(n-k) + 1}. \end{aligned}$$

Since $k - n - 1 \ge 0$ and $\beta \le 1$, this is in turn bounded by

$$\begin{aligned} 2^{-n\beta} I_n(s_0, g) &\le 2^{-(2k-n-2)\beta} \cdot 2^{2(n-k)+1+2(k-n-1)\beta} \\ &\le 2^{-(2k-n-2)\beta} \cdot 2^{2(n-k)+1+2(k-n-1)} \\ &\le 2^{-(2k-n-2)\beta}. \end{aligned} \tag{5.11}$$

Taking together (5.9) and (5.11), we arrive at

$$\sum_{n=0}^{k-3} 2^{-n\beta} I_n(s_0, g) + \sum_{n=k+1}^{2k-2} 2^{-n\beta} I_n(s_0, g) \leq \sum_{n=0}^{k-3} 2^{-(2k-n-2)\beta} - \sum_{n=k+1}^{2k-2} 2^{-n\beta} = 0.$$

Since by (5.9) also

$$\sum_{n=2k-1}^{\infty} 2^{-n\beta} I_n(s_0, g) = - \sum_{n=2k-1}^{\infty} 2^{-n\beta} < 0,$$

it remains to investigate

$$\sum_{n=k-2}^{k} 2^{-n\beta} I_n(s_0, g).$$

For this purpose, we distinguish between the three cases

$$\begin{aligned}
(i) &\quad \cos(2^{k-1}\pi s_0) = \cos(2^{k-2}\pi s_0) = 1 \\
(ii) &\quad \cos(2^{k-1}\pi s_0) = -1,\ \cos(2^{k-2}\pi s_0) = 0 \\
(iii) &\quad \cos(2^{k-1}\pi s_0) = 1,\ \cos(2^{k-2}\pi s_0) = -1
\end{aligned}$$

and subsequently use the representation in (5.10). In case (i), obviously

$$\sum_{n=k-2}^{k} 2^{-n\beta} I_n(s_0, g) \leq -2^{-k\beta} < 0.$$

using $\sin(x) \leq x$ for $x \geq 0$ again. In case (ii), we obtain for $\beta \leq 1$

$$\sum_{n=k-2}^{k} 2^{-n\beta} I_n(s_0, g) = 2^{-k\beta} 2^{\beta} \left(1 - \frac{\sin(\pi/2)}{\pi/2}\right) - 2^{-k\beta} \leq 2^{-k\beta} \left(1 - \frac{4}{\pi}\right) < 0.$$

Finally, in case (iii), for $\beta \leq 1$,

$$\begin{aligned}
&\sum_{n=k-2}^{k} 2^{-n\beta} I_n(s_0, g) \\
&\quad = 2^{-(k-1)\beta} \left(\frac{\sin(\pi/2)}{\pi/2} - 1\right) - 2^{-(k-2)\beta} \left(\frac{\sin(\pi/4)}{\pi/4} - 1\right) - 2^{-k\beta} \\
&\quad = 2^{-(k-1)\beta} \left(\left(\frac{2}{\pi} - 1\right) + 2^{\beta} \left(1 - \frac{\sin(\pi/4)}{\pi/4}\right)\right) - 2^{-k\beta} \\
&\quad < 2^{-(k-1)\beta} \left(\frac{2}{\pi} + 1 - 8\,\frac{\sin(\pi/4)}{\pi}\right) - 2^{-k\beta} \\
&\quad < -2^{-k\beta} \\
&\quad < 0.
\end{aligned}$$

That is,

$$\sup_{s\in B(t,h-g)}\left|\int K_{R,g}(x-s)W_\beta(x)\,\mathrm{d}x-W_\beta(s)\right|$$

$$\geq\left|\int_{-1}^{1}K_R(x)W_\beta(s_0+gx)\,\mathrm{d}x-W_\beta(s_0)\right|$$

$$=-\sum_{n=0}^{\infty}2^{-n\beta}I_n(s_0,g)$$

$$\geq-\sum_{n=k-2}^{k}2^{-n\beta}I_n(s_0,g)$$

$$>\left(\frac{4}{\pi}-1\right)g^\beta.$$

□

Proof of Theorem 3.6. The proof is structured as follows. First, we show that the bias term is negligible. Then, we conduct several reduction steps to non-stationary Gaussian processes. We pass over to the supremum over a stationary Gaussian process by means of Slepian's comparison inequality, and finally, we employ extreme value theory for its asymptotic distribution.

Step 1 (Negligibility of the bias). For any $t\in[0,1]$, there exists some $k_t\in T_n$ with $t\in I_{k_t}$. Hence,

$$\frac{\left|\mathbb{E}_p^{\chi_1}\hat{p}_n^{loc}(t,\hat{h}_n^{loc}(t))-p(t)\right|}{\hat{z}_n(t)}$$

$$=\sqrt{\tilde{n}\hat{h}_{n,k_t}^{loc}}\left|\mathbb{E}_p^{\chi_1}\hat{p}_n^{(1)}(k_t\delta_n,\hat{h}_{n,k_t}^{loc})-p(t)\right|$$

$$\leq\sqrt{\tilde{n}\hat{h}_{n,k_t}^{loc}}\left|\mathbb{E}_p^{\chi_1}\hat{p}_n^{(1)}(k_t\delta_n,\hat{h}_{n,k_t}^{loc})-p(k_t\delta_n)\right|+\sqrt{\tilde{n}\hat{h}_{n,k_t}^{loc}}\left|p(k_t\delta_n)-p(t)\right|.$$

Assume $\hat{j}_n(k\delta_n)\geq k_n(k\delta_n)=\bar{j}_n(k\delta_n)-m_n$ for all $k\in T_n$. Since $\delta_n\leq\frac{1}{8}h_{\beta_*,n}$ for sufficiently large $n\geq n_0(\beta_*,\varepsilon)$,

$$\begin{aligned}\hat{h}_{n,k_t}^{loc}&=2^{m_n-u_n}\cdot\min\left\{2^{-\hat{j}_n((k_t-1)\delta_n)-m_n},2^{-\hat{j}_n(k_t\delta_n)-m_n}\right\}\\&\leq 2^{m_n-u_n}\cdot\min\left\{\bar{h}_n((k_t-1)\delta_n),\bar{h}_n(k_t\delta_n)\right\}\\&\leq 3\cdot 2^{m_n-u_n}\cdot\bar{h}_n(t)\end{aligned}$$

by Lemma 4.2. In particular, $\delta_n+\hat{h}_{n,k_t}^{loc}\leq 2^{-(\bar{j}_n(t)+1)}$ holds for sufficiently large $n\geq n_0(c_1)$, so that Assumption 3.1, Lemma 3.2, and Lemma 4.4 yield

$$\sup_{p\in\mathscr{P}_n}\sqrt{\tilde{n}\hat{h}_{n,k_t}^{loc}}\left|\mathbb{E}_p^{\chi_1}\hat{p}_n^{(1)}(k_t\delta_n,\hat{h}_{n,k_t}^{loc})-p(k_t\delta_n)\right|$$

$$\leq\sup_{p\in\mathscr{P}_n}\sqrt{\tilde{n}\hat{h}_{n,k_t}^{loc}}\sup_{s\in B_{\delta_n}(t)}\left|\mathbb{E}_p^{\chi_1}\hat{p}_n^{(1)}(s,\hat{h}_{n,k_t}^{loc})-p(s)\right|$$

$$
\begin{aligned}
&\le \sup_{p\in\mathscr{P}_n} b_2\sqrt{\tilde{n}\hat{h}^{loc}_{n,k_t}}\left(\hat{h}^{loc}_{n,k_t}\right)^{\beta_p(B(t,2^{-\bar{j}_n(t)}))}\\
&\le \sup_{p\in\mathscr{P}_n} b_2\sqrt{\tilde{n}\hat{h}^{loc}_{n,k_t}}\left(\hat{h}^{loc}_{n,k_t}\right)^{\beta_p(B(t,\bar{h}_n(t)))}\\
&\le \sup_{p\in\mathscr{P}_n} b_2\left(3\cdot 2^{m_n-u_n}\right)^{\frac{2\beta_*+1}{2}}\sqrt{\frac{\tilde{n}\bar{h}_n(t)}{\log\tilde{n}}}\bar{h}_n(t)^{\beta_{n,p}(t)}\\
&\le c_{10}\cdot(\log\tilde{n})^{-\frac{1}{4}c_1(2\beta_*+1)\log 2} \qquad (5.12)
\end{aligned}
$$

for some constant $c_{10} = c_{10}(\beta_*, L^*, K)$, on the event

$$
\left\{\hat{j}_n(k\delta_n) \ge k_n(k\delta_n) \text{ for all } k \in T_n\right\}.
$$

Furthermore, for $t \in I_k$ and for $n \ge n_0$,

$$
\delta_n^{\beta_*} \le 2^{-j_{\min}}\left(\frac{\log\tilde{n}}{\tilde{n}}\right)^{\kappa_1\beta_*} \le 2^{-j_{\min}}\left(\frac{\log\tilde{n}}{\tilde{n}}\right)^{\frac{1}{2}} \le \bar{h}_n(t)^{\beta_{n,p}(t)},
$$

such that on the same event

$$
\begin{aligned}
\sup_{p\in\mathscr{P}_n}\sqrt{\tilde{n}\hat{h}^{loc}_{n,k_t}}\,|p(k_t\delta_n)-p(t)| &\le \sup_{p\in\mathscr{P}_n}\sqrt{3}L^*\cdot 2^{\frac{1}{2}(m_n-u_n)}\sqrt{\frac{\tilde{n}\bar{h}_n(t)}{\log\tilde{n}}}\cdot\delta_n^{\beta_*}\\
&\le c_{11}\cdot(\log\tilde{n})^{-\frac{1}{4}c_1\log 2} \qquad (5.13)
\end{aligned}
$$

for some constant $c_{11} = c_{11}(\beta_*, L^*)$. Taking (5.12) and (5.13) together,

$$
\sup_{p\in\mathscr{P}_n}\sup_{t\in[0,1]} a_n\frac{\left|\mathbb{E}_p^{\chi_1}\hat{p}^{loc}_n(t,\hat{h}^{loc}_n(t))-p(t)\right|}{\hat{z}_n(t)}\mathbb{1}\left\{\hat{j}_n(k\delta_n)\ge k_n(k\delta_n)\forall\, k\in T_n\right\}\le\varepsilon_{1,n},
$$

with

$$
\varepsilon_{1,n} = c_{10}\cdot a_n(\log n)^{-\frac{1}{4}c_1(2\beta_*+1)\log 2} + c_{11}\cdot a_n(\log n)^{-\frac{1}{4}c_1\log 2}.
$$

According to the definition of c_1 in (3.9), $\varepsilon_{1,n}$ converges to zero. Observe furthermore that

$$
\sup_{t\in[0,1]}\frac{\left|\hat{p}^{loc}_n(t,\hat{h}^{loc}_n(t))-p(t)\right|}{\hat{z}_n(t)} \qquad (5.14)
$$

can be written as

$$
\begin{aligned}
&\max_{k\in T_n}\sup_{t\in I_k}\frac{\left|\hat{p}^{loc}_n(t,\hat{h}^{loc}_n(t))-p(t)\right|}{\hat{z}_n(t)}\\
&\quad=\max_{k\in T_n}\sqrt{\tilde{n}\hat{h}^{loc}_{n,k}}\max\left\{\hat{p}^{(1)}_n(k\delta_n,\hat{h}^{loc}_{n,k})-\inf_{t\in I_k}p(t),\ \ \sup_{t\in I_k}p(t)-\hat{p}^{(1)}_n(k\delta_n,\hat{h}^{loc}_{n,k})\right\}
\end{aligned}
$$

with the definitions in (3.12) and (3.13). That is, the supremum in (5.14) is measurable. Then, by means of Proposition 4.1, with $x_{1,n} = x - \varepsilon_{1,n}$,

$$\sup_{p\in\mathscr{P}_n} \mathbb{P}_p^{\chi}\left(a_n\left\{\sup_{t\in[0,1]}\frac{\left|\hat{p}_n^{loc}(t,\hat{h}_n^{loc}(t)) - p(t)\right|}{\hat{z}_n(t)} - b_n\right\} \le x\right)$$

$$\ge \sup_{p\in\mathscr{P}_n} \mathbb{P}_p^{\chi}\left(a_n\left\{\sup_{t\in[0,1]}\frac{\left|\hat{p}_n^{loc}(t,\hat{h}_n^{loc}(t)) - p(t)\right|}{\hat{z}_n(t)} - b_n\right\} \le x,\right.$$

$$\left.\hat{j}_n(k\delta_n) \ge k_n(k\delta_n) \text{ for all } k\in T_n\right)$$

$$\ge \sup_{p\in\mathscr{P}_n} \mathbb{P}_p^{\chi}\left(a_n\left\{\sup_{t\in[0,1]}\frac{\left|\hat{p}_n^{loc}(t,\hat{h}_n^{loc}(t)) - \mathbb{E}_p^{\chi_1}\hat{p}_n^{loc}(t,\hat{h}_n^{loc}(t))\right|}{\hat{z}_n(t)} - b_n\right\} \le x_{1,n},\right.$$

$$\left.\hat{j}_n(k\delta_n) \ge k_n(k\delta_n) \text{ for all } k\in T_n\right)$$

$$\ge \sup_{p\in\mathscr{P}_n} \mathbb{P}_p^{\chi}\left(a_n\left\{\sup_{t\in[0,1]}\frac{\left|\hat{p}_n^{loc}(t,\hat{h}_n^{loc}(t)) - \mathbb{E}_p^{\chi_1}\hat{p}_n^{loc}(t,\hat{h}_n^{loc}(t))\right|}{\hat{z}_n(t)} - b_n\right\} \le x_{1,n}\right)$$

$$- \sup_{p\in\mathscr{P}_n} P_p^{\chi_2}\left(\hat{j}_n(k\delta_n) < k_n(k\delta_n) \text{ for some } k\in T_n\right)$$

$$= \sup_{p\in\mathscr{P}_n} \mathbb{E}_p^{\chi}\left[\mathbb{P}_p^{\chi}\left(a_n\left\{\max_{k\in T_n}\sqrt{\tilde{n}\hat{h}_{n,k}^{loc}}\left|\hat{p}_n^{(1)}(k\delta_n,\hat{h}_{n,k}^{loc}) - \mathbb{E}_p^{\chi_1}\hat{p}_n^{(1)}(k\delta_n,\hat{h}_{n,k}^{loc})\right|\right.\right.\right.$$

$$\left.\left.\left. - b_n\right\} \le x_{1,n}\,\middle|\,\chi_2\right)\right] + o(1) \tag{5.15}$$

for $n\to\infty$.

Step 2 (Reduction to the supremum over a non-stationary Gaussian process). In order to bound (5.15) from below note first that

$$\mathbb{P}_p^{\chi}\left(a_n\left\{\max_{k\in T_n}\sqrt{\tilde{n}\hat{h}_{n,k}^{loc}}\left|\hat{p}_n^{(1)}(k\delta_n,\hat{h}_{n,k}^{loc}) - \mathbb{E}_p^{\chi_1}\hat{p}_n^{(1)}(k\delta_n,\hat{h}_{n,k}^{loc})\right| - b_n\right\} \le x_{1,n}\,\middle|\,\chi_2\right)$$

$$\ge \mathbb{P}_p^{\chi}\left(a_n\left\{\max_{k\in T_n}\sqrt{\frac{\tilde{n}\hat{h}_{n,k}^{loc}}{p(k\delta_n)}}\left|\hat{p}_n^{(1)}(k\delta_n,\hat{h}_{n,k}^{loc}) - \mathbb{E}_p^{\chi_1}\hat{p}_n^{(1)}(k\delta_n,\hat{h}_{n,k}^{loc})\right| - b_n\right\} \le \frac{x_{1,n}}{\sqrt{L^*}}\,\middle|\,\chi_2\right).$$

Using the identity $|x| = \max\{x, -x\}$, we arrive at

$$\mathbb{P}_p^{\chi}\left(a_n\left\{\max_{k\in T_n}\sqrt{\tilde{n}\hat{h}_{n,k}^{loc}}\left|\hat{p}_n^{(1)}(k\delta_n,\hat{h}_{n,k}^{loc}) - \mathbb{E}_p^{\chi_1}\hat{p}_n^{(1)}(k\delta_n,\hat{h}_{n,k}^{loc})\right| - b_n\right\} \le x_{1,n}\middle|\chi_2\right) \ge 1 - P_{1,p} - P_{2,p}$$

with

$$P_{1,p} = \mathbb{P}_p^{\chi}\left(a_n\left\{\max_{k\in T_n}\sqrt{\frac{\tilde{n}\hat{h}_{n,k}^{loc}}{p(k\delta_n)}}\Big(\hat{p}_n^{(1)}(k\delta_n,\hat{h}_{n,k}^{loc}) - \mathbb{E}_p^{\chi_1}\hat{p}_n^{(1)}(k\delta_n,\hat{h}_{n,k}^{loc})\Big) - b_n\right\} > \frac{x_{1,n}}{\sqrt{L^*}}\middle|\chi_2\right)$$

$$P_{2,p} = \mathbb{P}_p^{\chi}\left(a_n\left\{\max_{k\in T_n}\sqrt{\frac{\tilde{n}\hat{h}_{n,k}^{loc}}{p(k\delta_n)}}\Big(\mathbb{E}_p^{\chi_1}\hat{p}_n^{(1)}(k\delta_n,\hat{h}_{n,k}^{loc}) - \hat{p}_n^{(1)}(k\delta_n,\hat{h}_{n,k}^{loc})\Big) - b_n\right\} > \frac{x_{1,n}}{\sqrt{L^*}}\middle|\chi_2\right).$$

In order to approximate the maxima in $P_{1,p}$ and $P_{2,p}$ by a supremum over a Gaussian process, we verify the conditions in Corollary 2.2 developed recently in Chernozhukov, Chetverikov and Kato (2014b). For this purpose, consider the empirical process

$$\mathbb{G}_n^p f = \frac{1}{\sqrt{\tilde{n}}}\sum_{i=1}^{\tilde{n}}\Big(f(X_i) - \mathbb{E}_p f(X_i)\Big), \quad f \in \mathcal{F}_n$$

indexed by

$$\mathcal{F}_n^p = \{f_{n,k} \ : \ k \in T_n\}$$

with

$$\begin{aligned} f_{n,k} &: \mathbb{R} \to \mathbb{R} \\ x &\mapsto \left(\tilde{n}\hat{h}_{n,k}^{loc}\, p(k\delta_n)\right)^{-\frac{1}{2}} K\left(\frac{k\delta_n - x}{\hat{h}_{n,k}^{loc}}\right). \end{aligned}$$

Note that Chernozhukov, Chetverikov and Kato (2014b) require the class of functions to be centered. We subsequently show that the class $\mathcal{F}_n^p$ is Euclidean, which implies by Lemma 5.2 that the corresponding centered class is VC. It therefore suffices to consider the uncentered class $\mathcal{F}_n^p$. Note furthermore that $f_{n,k}$ are random functions but depend on the second sample χ_2 only. Conditionally on χ_2, any function $f_{n,k} \in \mathcal{F}_n^p$ is measurable as K is continuous. Due to the choice of κ_2 and due to

$$\hat{h}_{n,k}^{loc} \ge 2^{-u_n}\cdot\frac{(\log\tilde{n})^{\kappa_2}}{\tilde{n}} \ge \frac{(\log\tilde{n})^{\kappa_2 - c_1\log 2}}{\tilde{n}}$$

the factor

$$\left(\tilde{n}\hat{h}^{loc}_{n,k}\,p(k\delta_n)\right)^{-\frac{1}{2}} \leq \frac{1}{\sqrt{M}}(\log\tilde{n})^{\frac{1}{2}(c_1\log 2-\kappa_2)} \tag{5.16}$$

tends to zero logarithmically. We now show that $\mathcal{F}_n^p$ is Euclidean with envelope

$$F_n = \frac{\|K\|_{\sup}}{\sqrt{M}}(\log\tilde{n})^{\frac{1}{2}(c_1\log 2-\kappa_2)}.$$

Note first that

$$\mathcal{F}_n^p \subset \mathscr{F} = \left\{ f_{u,h,t} : t\in\mathbb{R},\ 0<u\leq \frac{1}{\sqrt{M}}(\log\tilde{n})^{\frac{1}{2}(c_1\log 2-\kappa_2)},\ 0<h\leq 1\right\}$$

with

$$f_{u,h,t}(\cdot) = u\cdot K\left(\frac{t-\cdot}{h}\right).$$

Hence,

$$N\left(\mathcal{F}_n^p, \|\cdot\|_{L^1(Q)}, \varepsilon F_n\right) \leq N\left(\mathscr{F}, \frac{\|\cdot\|_{L^1(Q)}}{F_n}, \varepsilon\right)$$

for all probability measures Q and it therefore suffices to show that $\mathscr{F}$ is Euclidean. To this aim, note that for any $f_{u,h,t}, f_{v,g,s}\in\mathscr{F}$ and for any probability measure Q,

$$\begin{aligned}
&\frac{\|f_{u,h,t}-f_{v,g,s}\|_{L^1(Q)}}{F_n}\\
&\qquad\leq \frac{\|f_{u,h,t}-f_{v,h,t}\|_{L^1(Q)}}{F_n} + \frac{\|f_{v,h,t}-f_{v,g,s}\|_{L^1(Q)}}{F_n}\\
&\qquad\leq |u-v|\cdot\frac{\|K\|_{\sup}}{F_n} + \frac{1}{\|K\|_{\sup}}\left\|K\left(\frac{t-\cdot}{h}\right) - K\left(\frac{s-\cdot}{g}\right)\right\|_{L^1(Q)}.
\end{aligned}$$

Thus, using the estimate of the covering numbers in (2.1) and Lemma 14 in Nolan and Pollard (1987), there exist constants $A' = A'(A,K)$ and $\nu' = \nu+1$ with

$$\sup_Q N\left(\mathscr{F}, \frac{\|\cdot\|_{L^1(Q)}}{F_n}, \varepsilon\right) \leq \left(\frac{A'}{\varepsilon}\right)^{\nu'}$$

for all $0<\varepsilon\leq 1$. That is, $\mathscr{F}$ is Euclidean with the constant function F_n as envelope, and in particular

$$\limsup_{n\to\infty}\sup_Q N\left(\mathcal{F}_n^p, \|\cdot\|_{L^1(Q)}, \varepsilon F_n\right) \leq \left(\frac{A'}{\varepsilon}\right)^{\nu'}. \tag{5.17}$$

Hence, by Lemma 5.2, the $\mathbb{P}_p$-centered class $\mathcal{F}_n^{p,0}$ corresponding to $\mathcal{F}_n^p$ is VC with envelope $2F_n$ and

$$N\left(\mathcal{F}_n^{p,0}, \|\cdot\|_{L^2(Q)}, 2\varepsilon F_n\right) \leq \left(\frac{A''}{\varepsilon}\right)^{\nu''}$$

and VC characteristics $A'' = A''(A, K)$ and $\nu'' = \nu''(\nu)$. Next, we verify the Bernstein condition

$$\sup_{p\in\mathscr{P}_n}\sup_{f\in\mathcal{F}_n}\int |f(y)|^l p(y)\,\mathrm{d}y \le \sigma_n^2 B_n^{l-2}$$

for some $B_n \ge \sigma_n > 0$ and $B_n \ge 2F_n$ and $l = 2, 3$. First,

$$\begin{aligned}
&\max_{k\in T_n}\int |f_{n,k}(y)|^2\, p(y)\,\mathrm{d}y\\
&= \max_{k\in T_n}\Big(\tilde{n}\, p(k\delta_n)\Big)^{-1}\int_{-1}^{1} K(x)^2 p\left(k\delta_n + \hat{h}_{n,k}^{loc}x\right)\mathrm{d}x\\
&\le \sigma_n^2
\end{aligned}$$

with

$$\sigma_n^2 = \frac{2L^*\|K\|_{\sup}^2}{M\tilde{n}}.$$

Also, using (5.16),

$$\begin{aligned}
&\max_{k\in T_n}\int |f_{n,k}(y)|^3\, p(y)\,\mathrm{d}y\\
&= \max_{k\in T_n}\Big(\tilde{n}\hat{h}_{n,k}^{loc}\, p(k\delta_n)\Big)^{-3/2}\hat{h}_{n,k}^{loc}\int_{-1}^{1} K\left(x\right)^3 p\left(k\delta_n + \hat{h}_{n,k}^{loc}x\right)\mathrm{d}x\\
&\le \sigma_n^2\|K\|_{\sup}\cdot\max_{k\in T_n}\Big(\tilde{n}\hat{h}_{n,k}^{loc}\, p(k\delta_n)\Big)^{-1/2}\\
&\le \sigma_n^2\cdot B_n
\end{aligned}$$

with $B_n = 2F_n$. Furthermore, it holds that $\|2F_n\|_{\sup} = B_n$. According to Corollary 2.2 in Chernozhukov, Chetverikov and Kato (2014b), for sufficiently large $n \ge n_0(c_1, \kappa_2, L^*, K)$ such that $B_n \ge \sigma_n$, there exists a random variable

$$Z_{n,p} \overset{\mathcal{D}}{=} \max_{f\in\mathcal{F}_n^p} G_{\mathbb{P}_p} f,$$

and universal constants c_{12} and c_{13}, such that for $\eta = \frac{1}{4}(\kappa_2 - c_1\log 2 - 4) > 0$

$$\sup_{p\in\mathscr{P}_n}\mathbb{P}\left(a_n\sqrt{\tilde{n}}\left|\max_{f\in\mathcal{F}_n^p}\mathbb{G}_n f - Z_{n,p}\right| > \varepsilon_{2,n}\middle|\chi_2\right) \le c_{12}\left((\log\tilde{n})^{-\eta} + \frac{\log\tilde{n}}{\tilde{n}}\right),$$

where

$$\varepsilon_{2,n} = a_n\left(\frac{B_nK_n}{(\log\tilde{n})^{-\eta/2}} + \frac{\tilde{n}^{1/4}\sqrt{B_n\sigma_n}K_n^{3/4}}{(\log\tilde{n})^{-\eta/2}} + \frac{\tilde{n}^{1/3}(B_n\sigma_n^2K_n^2)^{1/3}}{(\log\tilde{n})^{-\eta/3}}\right)$$

with $K_n = c_{13}\nu''(\log\tilde{n}\vee\log(A''B_n/\sigma_n))$, and $G_{\mathbb{P}_p}$ is a version of the $\mathbb{P}_p$-Brownian motion. That is, it is centered and has the covariance structure

$$\mathbb{E}_p^{\chi_1} f(X_1)g(X_1)$$

for all $f, g \in \mathcal{F}_n^p$. As can be seen from an application of the Itô isometry, it possesses in particular the distributional representation

$$(G_{\mathbb{P}_p} f)_{f\in\mathcal{F}_n^p} \stackrel{\mathcal{D}}{=} \left(\int f(x)\sqrt{p(x)}\,\mathrm{d}W(x)\right)_{f\in\mathcal{F}_n^p}, \tag{5.18}$$

where W is a standard Brownian motion independent of χ_2. An easy calculation furthermore shows that $\varepsilon_{2,n}$ tends to zero for $n \to \infty$ logarithmically due to the choice of η. Finally,

$$\begin{aligned}
&\sup_{p\in\mathscr{P}_n} P_{1,p} \\
&\leq \sup_{p\in\mathscr{P}_n} \mathbb{P}_p^{\chi}\left(a_n\left(\sqrt{\tilde{n}}\max_{f\in\mathcal{F}_n^p}\mathbb{G}_n^p f - b_n\right) > \frac{x_{1,n}}{\sqrt{L^*}},\ a_n\sqrt{\tilde{n}}\left|\max_{f\in\mathcal{F}_n^p}\mathbb{G}_n^p f - Z_{n,p}\right| \leq \varepsilon_{2,n}\middle|\chi_2\right) \\
&\qquad + \sup_{p\in\mathscr{P}_n} \mathbb{P}_p^{\chi}\left(a_n\sqrt{\tilde{n}}\left|\max_{f\in\mathcal{F}_n^p}\mathbb{G}_n^p f - Z_{n,p}\right| > \varepsilon_{2,n}\middle|\chi_2\right) \\
&\leq \sup_{p\in\mathscr{P}_n} \mathbb{P}_p^{\chi}\left(a_n\left(\sqrt{\tilde{n}}\,Z_{n,p} - b_n\right) > x_{2,n}\middle|\chi_2\right) + o(1)
\end{aligned}$$

for $n \to \infty$, with

$$x_{2,n} = \frac{x_{1,n}}{\sqrt{L^*}} - \varepsilon_{2,n} = \frac{x - \varepsilon_{1,n}}{\sqrt{L^*}} - \varepsilon_{2,n} = \frac{x}{\sqrt{L^*}} + o(1).$$

The probability $P_{2,p}$ is bounded in the same way, leading to

$$\begin{aligned}
&\sup_{p\in\mathscr{P}_n} \mathbb{P}_p^{\chi}\left(a_n\left\{\max_{k\in T_n}\sqrt{\tilde{n}\hat{h}_{n,k}^{loc}}\left|\hat{p}_n^{(1)}(k\delta_n, \hat{h}_{n,k}^{loc}) - \mathbb{E}_p^{\chi_1}\hat{p}_n^{(1)}(k\delta_n, \hat{h}_{n,k}^{loc})\right| - b_n\right\} \leq x_{1,n}\middle|\chi_2\right) \\
&\qquad\qquad \geq 2\sup_{p\in\mathscr{P}_n} \mathbb{P}_p^{\chi}\left(a_n\left(\sqrt{\tilde{n}}\,Z_{n,p} - b_n\right) \leq x_{2,n}\middle|\chi_2\right) - 1 + o(1).
\end{aligned}$$

Finally we conduct a further approximation

$$Y_{n,p}(k) \stackrel{\mathcal{D}}{=} \frac{1}{\sqrt{\hat{h}_{n,k}^{loc}}}\int K\left(\frac{k\delta_n - x}{\hat{h}_{n,k}^{loc}}\right)\mathrm{d}W(x)$$

for the right-hand side in (5.18) in order to obtain to a suitable intermediate process for the next step. With

$$\begin{aligned}
V_{n,p}(k) &= \sqrt{\tilde{n}}\,W(f_{n,k}\sqrt{p}) - Y_{n,p}(k) \\
&\stackrel{\mathcal{D}}{=} \sqrt{\tilde{n}}\int f_{n,k}(x)\left(\sqrt{p(x)} - \sqrt{p(k\delta_n)}\right)\mathrm{d}W(x),
\end{aligned}$$

it remains to show that

$$\lim_{n\to\infty}\sup_{p\in\mathscr{P}_n} \mathbb{P}^W\left(a_n\max_{k=1,\ldots,\delta_n}|V_{n,p}(k)| > \varepsilon_{3,n}\right) = 0$$

for some sequence $(\varepsilon_{3,n})_{n\in\mathbb{N}}$ converging to zero. Note that $V_{n,p}(k), k \in T_n$ is a centered Gaussian process and in particular subgaussian with respect to the metric

$$d_{V_{n,p}}(k,l) = \left(\mathbb{E}_W \left(V_{n,p}(k) - V_{n,p}(l)\right)^2\right)^{1/2}.$$

Hence, by Dudley's entropy bound in the form of Koltchinskii (2011), Theorem 3.1, there exists a numerical constant $c_{14} > 0$, such that

$$\mathbb{E}_W \max_{k\in T_n} |Z_{n,p}(k) - Z_{n,p}(k_0)| \leq c_{14} \int_0^{D(T_n)} \sqrt{\log N(T_n, d_{Z_{n,p}}, \varepsilon)}\, \mathrm{d}\varepsilon \tag{5.19}$$

for all $k_0 \in T_n$, where

$$D(T_n) = D(T_n, d_{V_{n,p}}) = \sup_{k,l\,\in\,T_n} d_{V_{n,p}}(k,l)$$

denotes the diameter of the space T_n with respect to $d_{V_{n,p}}$. It remains to bound the entropy integral. Note that by Itô's isometry and by the Minkowski inequality

$$\begin{aligned} d_{V_{n,p}}(k,l) \leq & \left(\tilde{n} \int f_{n,k}(x)^2 \left(\sqrt{p(x)} - \sqrt{p(k\delta_n)}\right)^2 \mathrm{d}x\right)^{1/2} \\ & + \left(\tilde{n} \int f_{n,l}(x)^2 \left(\sqrt{p(x)} - \sqrt{p(l\delta_n)}\right)^2 \mathrm{d}x\right)^{1/2}. \end{aligned}$$

Furthermore, as $p \in \mathscr{P}_n$,

$$\begin{aligned} & \tilde{n} \int f_{n,k}(x)^2 \left(\sqrt{p(x)} - \sqrt{p(k\delta_n)}\right)^2 \mathrm{d}x \\ & = \frac{1}{p(k\delta_n)} \int K(x)^2 \left(\sqrt{p(k\delta_n + \hat{h}_{n,k}^{loc} x)} - \sqrt{p(k\delta_n)}\right)^2 \mathrm{d}x \\ & \leq \frac{1}{p(k\delta_n)} \int K(x)^2 \left|p(k\delta_n + \hat{h}_{n,k}^{loc} x) - p(k\delta_n)\right| \mathrm{d}x \\ & \leq \frac{L^*\|K\|_2^2}{p(k\delta_n)} \left(\hat{h}_{n,k}^{loc}\right)^{\beta_*} \\ & \leq \frac{L^*\|K\|_2^2}{M} (\log \tilde{n})^{-c_1\beta_* \log 2}, \end{aligned} \tag{5.20}$$

we get

$$D(T_n) \leq \Delta_n = 2\frac{\sqrt{L^*}\|K\|_2}{\sqrt{M}} (\log \tilde{n})^{-\frac{1}{2}c_1\beta_* \log 2}.$$

Because of

$$N(T_n, d_{V_{n,p}}, \varepsilon) = N\left(\mathcal{G}_n^p, \|\cdot\|_{L^2(\mathbb{P}_p)}, \varepsilon\right)$$

with

$$\mathcal{G}_n^p = \left\{ \tilde{f}_t(\cdot) = \sqrt{\tilde{n}} f_{n,k}(\cdot) \left(1 - \sqrt{\frac{p(k\delta_n)}{p(\cdot)}} \right) : k \in T_n \right\},$$

it remains to bound the covering numbers $N\left(\mathcal{G}_n^p, \|\cdot\|_{L^2(\mathbb{P}_p)}, \varepsilon\right)$. Moreover,

$$\mathcal{G}_n^p \subset \mathcal{H}_n^p = \left\{ g_{n,k,u}(\cdot) = u \cdot K\left(\frac{k\delta_n - \cdot}{\hat{h}_{n,k}^{loc}} \right) : k \in T_n, \ |u| \le U_n \right\}$$

for any $p \in \mathscr{P}_n$, with

$$U_n = \frac{\sqrt{L^*}}{M} \left(\frac{\tilde{n}}{(\log \tilde{n})^{\kappa_2 - c_1 \log 2}} \right)^{\frac{1-\beta_*}{2}},$$

since for $k \in T_n$ and $|k\delta_n - x| \le \hat{h}_{n,k}^{loc}$

$$\begin{aligned} \left| \left(\hat{h}_{n,k}^{loc}\, p(k\delta_n) \right)^{-1/2} \left(1 - \sqrt{\frac{p(k\delta_n)}{p(x)}} \right) \right| &\le \left(\hat{h}_{n,k}^{loc} \right)^{-1/2} \left(\frac{|p(x) - p(k\delta_n)|}{p(x)\, p(k\delta_n)} \right)^{1/2} \\ &\le \frac{\sqrt{L^*}}{M} (\hat{h}_{n,k}^{loc})^{\frac{\beta_*-1}{2}} \\ &\le U_n. \end{aligned}$$

As before, for any $k, l \in T_n$ and for any u, v with $|u| \le U_n$ and $|v| \le U_n$ and for any probability measure Q, by Minkowski inequality,

$$\begin{aligned} &\|g_{n,k,u} - g_{n,l,v}\|_{L^2(Q)} \\ &\quad \le \|g_{n,k,u} - g_{n,k,v}\|_{L^2(Q)} + \|g_{n,k,v} - g_{n,l,v}\|_{L^2(Q)} \\ &\quad \le |u - v| \cdot \|K\|_{\sup} + U_n \cdot \left\| K\left(\frac{k\delta_n - \cdot}{\hat{h}_{n,k}^{loc}} \right) - K\left(\frac{l\delta_n - \cdot}{\hat{h}_{n,l}^{loc}} \right) \right\|_{L^2(Q)} \end{aligned}$$

which together with (2.1) implies that

$$N\left(\mathcal{H}_n^p, \|\cdot\|_{L^2(Q)}, \varepsilon\right) \le \left(\frac{A' U_n}{\varepsilon} \right)^{\nu+1}$$

for $0 < \varepsilon \le \|K\|_{\sup}$, and for some universal constant $A' = A'(A, K)$, see Nolan and Pollard (1987), Lemma 14. Consequently,

$$\sup_{p \in \mathscr{P}_n} N(T_n, d_{V_{n,p}}, \varepsilon) \le \left(\frac{A' U_n}{\varepsilon} \right)^{\nu+1},$$

and therefore, recalling (5.19),

$$\begin{aligned} &\mathbb{E}_W \max_{k \in T_n} |V_{n,p}(k) - V_{n,p}(k_0)| \\ &\qquad \le c_{14} \sqrt{\nu + 1} \int_0^{\Delta_n} \left\{ \log(A') + \log(U_n) + \log(1/\varepsilon) \right\}^{1/2} \mathrm{d}\varepsilon \end{aligned}$$

for all $k_0 \in T_n$ and for sufficiently large $n \geq n_0(\beta_*, L^*, K, M)$. Finally, as Δ_n tends to zero and U_n goes to infinity for $n \to \infty$,

$$\begin{aligned}
&\mathbb{E}_W \max_{k \in T_n} |V_{n,p}(k) - V_{n,p}(k_0)| \\
&\qquad \leq c_{14}\sqrt{\nu+1}\left(2\Delta_n \sqrt{\log(U_n)} + \int_0^{\Delta_n} \sqrt{\log(1/\varepsilon)}\, \mathrm{d}\varepsilon\right)
\end{aligned}$$

for sufficiently large $n \geq n_0(\beta_*, L^*, K, M, A)$ independent of p. Additionally, with a change of variable and integration by parts,

$$\begin{aligned}
\int_0^{\Delta_n} \sqrt{\log(1/\varepsilon)}\, \mathrm{d}\varepsilon &= \int_{\Delta_n^{-1}}^{\infty} \frac{\sqrt{\log(\varepsilon)}}{\varepsilon^2}\, \mathrm{d}\varepsilon \\
&= \Delta_n \sqrt{-\log(\Delta_n)} + \frac{1}{2}\int_{\Delta_n^{-1}}^{\infty} \frac{1}{\varepsilon^2 \sqrt{\log(\varepsilon)}}\, \mathrm{d}\varepsilon \\
&\leq \Delta_n \sqrt{-\log(\Delta_n)} + \frac{1}{2}\int_{\Delta_n^{-1}}^{\infty} \frac{\sqrt{\log(\varepsilon)}}{\varepsilon^2}\, \mathrm{d}\varepsilon
\end{aligned}$$

for sufficiently large $n \geq n_0(\beta_*, L^*, K, M)$ independent of p, and thus

$$\int_0^{\Delta_n} \sqrt{\log(1/\varepsilon)}\, \mathrm{d}\varepsilon \leq 2\Delta_n \sqrt{-\log(\Delta_n)}.$$

Finally,

$$\begin{aligned}
&\mathbb{E}_W \max_{k \in T_n} |V_{n,p}(k) - V_{n,p}(k_0)| \\
&\qquad \leq 2c_{14}\sqrt{\nu+1}\Delta_n \left(\sqrt{\log(U_n)} + \sqrt{-\log(\Delta_n)}\right),
\end{aligned}$$

and consequently, recalling (5.20) and using Jensen's inequality,

$$\begin{aligned}
&\mathbb{E}_W \max_{k \in T_n} |V_{n,p}(k)| \\
&\quad \leq 2c_{14}\sqrt{\nu+1}\Delta_n \left(\sqrt{\log(U_n)} + \sqrt{-\log(\Delta_n)}\right) + \mathbb{E}_W |V_{n,p}(k_0)| \\
&\quad \leq 2c_{14}\sqrt{\nu+1}\Delta_n \left(\sqrt{\log(U_n)} + \sqrt{-\log(\Delta_n)}\right) + \left(\mathbb{E}_W |V_{n,p}(k_0)|^2\right)^{1/2} \\
&\quad \leq 2c_{14}\sqrt{\nu+1}\Delta_n \left(\sqrt{\log(U_n)} + \sqrt{-\log(\Delta_n)}\right) + \frac{\Delta_n}{2} \\
&\quad \leq (2c_{14}\sqrt{\nu+1} + 1/2)\Delta_n \left(\sqrt{\log(U_n)} + \sqrt{-\log(\Delta_n)}\right).
\end{aligned}$$

For the sequence

$$\varepsilon_{3,n} = (\log \tilde{n})^{-\frac{1}{4}(c_1 \beta_* \log 2 - 2)},$$

Markov's inequality gives

$$\begin{aligned}
&\sup_{p \in \mathscr{P}_n} \mathbb{P}_W\left(a_n \max_{k \in T_n} |V_{n,p}(k)| > \varepsilon_{3,n}\right) \\
&\qquad \leq (2c_{14}\sqrt{\nu+1} + 1/2)\frac{a_n \Delta_n}{\varepsilon_{3,n}} \left(\sqrt{\log(U_n)} + \sqrt{-\log(\Delta_n)}\right), \qquad (5.21)
\end{aligned}$$

where both $\varepsilon_{3,n}$ and the expression in (5.21) converge to 0 due to the choice of c_1 in (3.9). Following the same steps as before, we obtain

$$\sup_{p\in\mathscr{P}_n} \mathbb{P}_p^{\chi}\left(a_n\left(\sqrt{\tilde n}\max_{k\in T_n} G_{\mathbb{P}_p} f_{n,k} - b_n\right) \leq x_{2,n}\middle|\chi_2\right) \geq \sup_{p\in\mathscr{P}_n} \mathbb{P}^W\left(a_n\left(\max_{k\in T_n} Y_{n,p}(k) - b_n\right) \leq x_{3,n}\right) + o(1)$$

for $n\to\infty$, with $x_{3,n} = x_{2,n} - \varepsilon_{3,n}$.

Step 3 (Reduction to the supremum over a stationary Gaussian process). We now use Slepian's comparison inequality in order to pass over to the least favorable case. Since K is symmetric and of bounded variation, it possesses a representation

$$K(x) = \int_{-1}^{x} g\,\mathrm{d}P$$

for all but at most countably many $x\in[-1,1]$, where P is some symmetric probability measure on $[-1,1]$ and g is some measurable odd function with $|g|\leq TV(K)$. Using this representation, and denoting by

$$\begin{aligned} W_{k,l}(z) &= \sqrt{\frac{1}{\hat h^{loc}_{n,k}}}\left\{W(k\delta_n + \hat h^{loc}_{n,k}) - W(k\delta_n + z\hat h^{loc}_{n,k})\right\} \\ &\quad - \sqrt{\frac{1}{\hat h^{loc}_{n,l}}}\left\{W(l\delta_n + \hat h^{loc}_{n,l}) - W(l\delta_n + z\hat h^{loc}_{n,l})\right\} \\ \tilde W_{k,l}(z) &= \sqrt{\frac{1}{\hat h^{loc}_{n,k}}}\left\{W(k\delta_n - z\hat h^{loc}_{n,k}) - W(k\delta_n + z\hat h^{loc}_{n,k})\right\} \\ &\quad + \sqrt{\frac{1}{\hat h^{loc}_{n,l}}}\left\{W(l\delta_n + z\hat h^{loc}_{n,l}) - W(l\delta_n - z\hat h^{loc}_{n,l})\right\}, \end{aligned} \tag{5.22}$$

Fubini's theorem and the Cauchy-Schwarz inequality yield for any $k,l\in T_n$

$$\begin{aligned} &\mathbb{E}_W\left(Y_{n,p}(k) - Y_{n,p}(l)\right)^2 \\ &= \mathbb{E}_W\left(\sqrt{\frac{1}{\hat h^{loc}_{n,k}}}\int\int_{-1}^{\frac{x-k\delta_n}{\hat h^{loc}_{n,k}}} g(z)\,\mathrm{d}P(z)\mathbb{1}\left\{|x-k\delta_n|\leq \hat h^{loc}_{n,k}\right\}\mathrm{d}W(x)\right. \\ &\qquad \left. -\sqrt{\frac{1}{\hat h^{loc}_{n,l}}}\int\int_{-1}^{\frac{x-l\delta_n}{\hat h^{loc}_{n,l}}} g(z)\,\mathrm{d}P(z)\mathbb{1}\left\{|x-l\delta_n|\leq \hat h^{loc}_{n,l}\right\}\mathrm{d}W(x)\right)^2 \\ &= \mathbb{E}_W\left(\int_{-1}^{1} g(z)\left\{\sqrt{\frac{1}{\hat h^{loc}_{n,k}}}\int \mathbb{1}\left\{k\delta_n + z\hat h^{loc}_{n,k}\leq x\leq k\delta_n + \hat h^{loc}_{n,k}\right\}\mathrm{d}W(x)\right.\right. \end{aligned}$$

$$
\begin{aligned}
&\qquad -\sqrt{\frac{1}{\hat{h}^{loc}_{n,l}}}\int \mathbb{1}\left\{l\delta_n + z\hat{h}^{loc}_{n,l} \le x \le l\delta_n + \hat{h}^{loc}_{n,l}\right\}\,\mathrm{d}W(x)\Bigg\}\,dP(z)\Bigg)^2 \\
&= \mathbb{E}_W\left(\int_{-1}^{1} g(z)W_{k,l}(z)\,\mathrm{d}P(z)\right)^2 \\
&= \mathbb{E}_W\left(\int_{0}^{1} g(z)\Big(W_{k,l}(z) - W_{k,l}(-z)\Big)\,\mathrm{d}P(z)\right)^2 \\
&= \mathbb{E}_W\int_0^1\int_0^1 g(z)g(z')\tilde{W}_{k,l}(z)\tilde{W}_{k,l}(z')\,\mathrm{d}P(z)\,\mathrm{d}P(z') \\
&\le \int_0^1\int_0^1 |g(z)g(z')|\left\{\mathbb{E}_W\tilde{W}_{k,l}(z)^2\mathbb{E}_W\tilde{W}_{k,l}(z')^2\right\}^{1/2}\mathrm{d}P(z)\,\mathrm{d}P(z').
\end{aligned}
$$

We verify in Lemma A.1 that

$$\mathbb{E}_W\tilde{W}_{k,l}(z)^2 \le 4$$

for $z\in[0,1]$, so that

$$\mathbb{E}_W(Y_{n,p}(k) - Y_{n,p}(l))^2 \le 4\left(\int_0^1 |g(z)|\,\mathrm{d}P(z)\right)^2 \le TV(K)^2$$

for all $k,l\in T_n$. Consider now the Gaussian process

$$Y_{n,\min}(k) = \frac{c_{15}}{\sqrt{\delta_n}}\int K\left(\frac{k\delta_n - x}{\delta_n/2}\right)\mathrm{d}W(x), \quad k\in T_n,$$

with

$$c_{15} = \frac{TV(K)}{\|K\|_2}.$$

Furthermore,

$$\mathbb{E}_W\left(Y_{n,\min}(k) - Y_{n,\min}(l)\right)^2 = \mathbb{E}_W Y_{n,\min}(k)^2 + \mathbb{E}_W Y_{n,\min}(l)^2 = TV(K)^2$$

for all $k,l\in T_n$ with $k\neq l$, so that

$$\mathbb{E}_W(Y_{n,p}(k) - Y_{n,p}(l))^2 \le \mathbb{E}_W\left(Y_{n,\min}(k) - Y_{n,\min}(l)\right)^2 \tag{5.23}$$

for all $k,l\in T_n$. In order to apply Slepian's comparison inequality we additionally need coinciding second moments. For this aim, we analyze the modified Gaussian processes

$$
\begin{aligned}
\bar{Y}_{n,p}(k) &= Y_{n,p}(k) + c_{16}Z \\
\bar{Y}_{n,\min}(k) &= Y_{n,\min}(k) + c_{17}Z
\end{aligned}
$$

with

$$c_{16} = c_{16}(K) = \frac{TV(K)}{\sqrt{2}}, \qquad c_{17} = c_{17}(K) = \|K\|_2,$$

and for some standard normally distributed random variable Z independent of $(Y_{n,p}(k))_{k\in T_n}$ and $(Y_{n,\min}(k))_{k\in T_n}$. For sufficiently large $n \geq n_0(K)$, these processes have the same increments as the processes before but additionally coinciding second moments. Obviously,

$$\mathbb{E}_W \bar{Y}_{n,p}(k)^2 = \mathbb{E}_W \bar{Y}_{n,\min}(k)^2 = \frac{TV(K)^2}{2} + \|K\|_2^2$$

for all $k \in T_n$, and

$$\begin{aligned}\mathbb{E}_W \left(\bar{Y}_{n,p}(k) - \bar{Y}_{n,p}(l)\right)^2 &= \mathbb{E}_W \left(Y_{n,p}(k) - Y_{n,p}(l)\right)^2 \\ &\leq \mathbb{E}_W \left(Y_{n,\min}(k) - Y_{n,\min}(l)\right)^2 \\ &= \mathbb{E}_W \left(\bar{Y}_{n,\min}(k) - \bar{Y}_{n,\min}(l)\right)^2\end{aligned}$$

for all $k, l \in T_n$ by inequality (5.23). Then,

$$\begin{aligned}&\sup_{p\in\mathscr{P}_n} \mathbb{P}^W \left(a_n \left(\max_{k\in T_n} Y_{n,p}(k) - b_n\right) \leq x_{3,n}\right) \\ &\quad= \sup_{p\in\mathscr{P}_n} \mathbb{P}^W \left(a_n \left(\max_{k\in T_n} \bar{Y}_{n,p}(k) - c_{16}Z - b_n\right) \leq x_{3,n}\right) \\ &\quad\geq \sup_{p\in\mathscr{P}_n} \mathbb{P}^W \left(a_n \left(\max_{k\in T_n} \bar{Y}_{n,p}(k) - c_{16}Z - b_n\right) \leq x_{3,n},\ -Z \leq \frac{1}{3c_{16}} b_n\right) \\ &\quad\geq \sup_{p\in\mathscr{P}_n} \mathbb{P}^W \left(a_n \left(\max_{k\in T_n} \bar{Y}_{n,p}(k) - \frac{2}{3} b_n\right) \leq x_{3,n}\right) - \mathbb{P}\left(-Z > \frac{1}{3c_{16}} b_n\right) \\ &\quad\geq \sup_{p\in\mathscr{P}_n} \mathbb{P}^W \left(a_n \left(\max_{k\in T_n} \bar{Y}_{n,p}(k) - \frac{2}{3} b_n\right) \leq x_{3,n}\right) + o(1)\end{aligned}$$

for $n \to \infty$. Slepian's inequality in the form of Corollary 3.12 in Ledoux and Talagrand (1991) yields

$$\begin{aligned}&\sup_{p\in\mathscr{P}_n} \mathbb{P}^W \left(a_n \left(\max_{k\in T_n} \bar{Y}_{n,p}(k) - \frac{2}{3} b_n\right) \leq x_{3,n}\right) \\ &\qquad\geq \mathbb{P}^W \left(a_n \left(\max_{k\in T_n} \bar{Y}_{n,\min}(k) - \frac{2}{3} b_n\right) \leq x_{3,n}\right).\end{aligned}$$

Step 4 (Limiting distribution theory). Finally, we pass over to an iid sequence and apply extreme value theory. Together with

$$\begin{aligned}&\mathbb{P}^W \left(a_n \left(\max_{k\in T_n} \bar{Y}_{n,\min}(k) - \frac{2}{3} b_n\right) \leq x_{3,n}\right) \\ &\quad\geq \mathbb{P}^W \left(a_n \left(\max_{k\in T_n} Y_{n,\min}(k) + c_{17}Z - \frac{2}{3} b_n\right) \leq x_{3,n},\ Z \leq \frac{1}{3c_{17}} b_n\right) \\ &\quad\geq \mathbb{P}^W \left(a_n \left(\max_{k\in T_n} Y_{n,\min}(k) - \frac{1}{3} b_n\right) \leq x_{3,n}\right) - \mathbb{P}\left(Z > \frac{1}{3c_{17}} b_n\right)\end{aligned}$$

$$= \mathbb{P}^W \left(a_n \left(\max_{k \in T_n} Y_{n,\min}(k) - \frac{1}{3} b_n \right) \leq x_{3,n} \right) + o(1)$$

as $n \to \infty$, we finally obtain

$$\sup_{p \in \mathscr{P}_n} \mathbb{P}_p^\chi \left(a_n \left(\sup_{t \in [0,1]} \frac{\left| \hat{p}_n^{loc}(t, \hat{h}_n^{loc}(t)) - p(t) \right|}{\hat{z}_n(t)} - b_n \right) \leq x \right)$$
$$\geq 2\, \mathbb{P} \left(a_n \left(\max_{k \in T_n} Y_{n,\min}(k) - \frac{1}{3} b_n \right) \leq x_{3,n} \right) - 1 + o(1).$$

Theorem 1.5.3 in Leadbetter, Lindgren and Rootzén (1983) yields now

(5.24)
$$F_n(x) = \mathbb{P}^W \left(a_n \left(\max_{k \in T_n} Y_{n,\min}(k) - \frac{1}{3} b_n \right) \leq x \right) \longrightarrow F(x) = \exp(-\exp(-x))$$

for any $x \in \mathbb{R}$. It remains to show, that $F_n(x_n) \to F(x)$ for some sequence $x_n \to x$ as $n \to \infty$. Because F is continuous in x, for any $\varepsilon > 0$ there exists some $\delta = \delta(\varepsilon) > 0$ such that for all $y \in \mathbb{R}$ with $|y - x| \leq \delta$ holds $|F(x) - F(y)| \leq \varepsilon/2$. In particular, for $y = x \pm \delta$,

$$|F(x) - F(x+\delta)| \leq \frac{\varepsilon}{2} \quad \text{and} \quad |F(x) - F(x-\delta)| \leq \frac{\varepsilon}{2}. \tag{5.25}$$

Since $x_n \to x$, there exists some $N_1 = N_1(\varepsilon)$, such that $|x_n - x| \leq \delta$ for all $n \geq N_1$. Therefore, employing the monotonicity of F_n,

$$|F_n(x_n) - F(x)| \leq |F_n(x+\delta) - F(x)| \vee |F_n(x-\delta) - F(x)|$$

for $n \geq N_1$, where

$$|F_n(x \pm \delta) - F(x)| \leq |F_n(x \pm \delta) - F(x \pm \delta)| + |F(x \pm \delta) - F(x)| \leq \varepsilon$$

for $n \geq N_2 = N_2(\varepsilon)$ due to (5.24) and (5.25). Consequently,

$$\lim_{n \to \infty} \inf_{p \in \mathscr{P}_n} \mathbb{P}_p^\chi \left(a_n \left(\sup_{t \in [0,1]} \frac{\left| \hat{p}_n^{loc}(t, \hat{h}_n^{loc}(t)) - p(t) \right|}{\hat{z}_n(t)} - b_n \right) \leq x \right)$$
$$\geq 2 \lim_{n \to \infty} \mathbb{P} \left(a_n \left(\max_{k \in T_n} Y_{n,\min}(k) - \frac{1}{3} b_n \right) \leq x_{3,n} \right) - 1 + o(1)$$
$$= 2\, \mathbb{P} \left(\sqrt{L^*} G \leq x \right) - 1 + o(1), \quad n \to \infty,$$

for some standard Gumbel distributed random variable G. □

PROOF OF PROPOSITION 3.10. The proof is based on a reduction of the supremum over the class to a maximum over two distinct hypotheses.

Part 1. For $\beta \in [\beta_*, 1)$, the construction of the hypotheses is based on the Weierstraß function as defined in (3.8). As in the proof of Proposition 3.3 consider the function $p_0 : \mathbb{R} \to \mathbb{R}$ with

$$p_0(x) = \begin{cases} 0, & \text{if } |x-t| \geq \frac{10}{3} \\ \frac{1}{4} + \frac{3}{16}(x-t+2), & \text{if } -\frac{10}{3} < x-t < -2 \\ \frac{1}{6} + \frac{1-2^{-\beta}}{12} W_\beta(x-t), & \text{if } |x-t| \leq 2 \\ \frac{1}{4} - \frac{3}{16}(x-t-2), & \text{if } 2 < x-t < \frac{10}{3} \end{cases}$$

and the functions $p_{1,n}$, $p_{2,n} : \mathbb{R} \to \mathbb{R}$ with

$$p_{1,n}(x) = p_0(x) + q_{t+\frac{9}{4},n}(x; g_{\beta,n}) - q_{t,n}(x; g_{\beta,n}), \quad x \in \mathbb{R}$$
$$p_{2,n}(x) = p_0(x) + q_{t+\frac{9}{4},n}(x; c_{18} \cdot g_{\beta,n}) - q_{t,n}(x; c_{18} \cdot g_{\beta,n}), \quad x \in \mathbb{R}$$

for $g_{\beta,n} = \frac{1}{4} n^{-1/(2\beta+1)}$ and $c_{18} = c_{18}(\beta) = (2L_W(\beta))^{-1/\beta}$, where

$$q_{a,n}(x; g) = \begin{cases} 0, & \text{if } |x-a| > g \\ \frac{1-2^{-\beta}}{12}\Big(W_\beta(x-a) - W_\beta(g)\Big), & \text{if } |x-a| \leq g \end{cases}$$

for $a \in \mathbb{R}$ and $g > 0$.

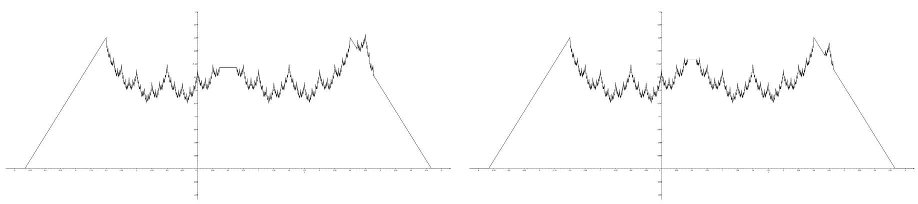

FIG 5. *Functions $p_{1,n}$ and $p_{2,n}$ for $t = 0.5$, $\beta = 0.5$ and $n = 50$*

Following the lines of the proof of Proposition 3.3, both $p_{1,n}$ and $p_{2,n}$ are contained in the class $\mathscr{P}_k(L, \beta_*, M, K_R, \varepsilon)$ for sufficiently large $k \geq k_0(\beta_*)$. Moreover, both $p_{1,n}$ and $p_{2,n}$ are constant on $B(t, c_{18} \cdot g_{\beta,n})$, so that

$$p_{1,n|B(t,c_{18}\cdot g_{\beta,n})},\ p_{2,n|B(t,c_{18}\cdot g_{\beta,n})} \in \mathcal{H}_{B(t,c_{18}\cdot g_{\beta,n})}(\infty, L)$$

for some constant $L = L(\beta)$. Using Lemma 3.5 and (5.3), the absolute distance of the two hypotheses in t is at least

$$\begin{aligned} |p_{1,n}(t) - p_{2,n}(t)| &= |q_{t,n}(t; g_{\beta,n}) - q_{t,n}(t; c_{18} \cdot g_{\beta,n})| \\ &= \frac{1-2^{-\beta}}{12} |W_\beta(g_{\beta,n}) - W_\beta(c_{18} \cdot g_{\beta,n})| \\ &\geq \frac{1-2^{-\beta_*}}{12} \Big(|W_\beta(g_{\beta,n}) - W_\beta(0)| - |W_\beta(c_{18} \cdot g_{\beta,n}) - W_\beta(0)| \Big) \\ &\geq \frac{1-2^{-\beta_*}}{12} \Big(g_{\beta,n}^\beta - L_W(\beta) \, (c_{18} \cdot g_{\beta,n})^\beta \Big) \\ &\geq 2c_{19} g_{\beta,n}^\beta \end{aligned}$$

where

$$c_{19} = c_{19}(\beta_*) = \frac{1 - 2^{-\beta_*}}{48}.$$

Since furthermore

$$\int (p_{2,n}(x) - p_{1,n}(x))\,\mathrm{d}x = 0,$$

and $\log(1+x) \leq x$ for $x > -1$, the Kullback-Leibler divergence between the associated product probability measures $\mathbb{P}_{1,n}^{\otimes n}$ and $\mathbb{P}_{2,n}^{\otimes n}$ is bounded from above by

$$\begin{aligned}
K(\mathbb{P}_{2,n}^{\otimes n}, \mathbb{P}_{1,n}^{\otimes n}) &\leq n \int \frac{(p_{2,n}(x) - p_{1,n}(x))^2}{p_{1,n}(x)}\,\mathrm{d}x \\
&\leq 12\,n \int (p_{2,n}(x) - p_{1,n}(x))^2\,\mathrm{d}x \\
&= 24\,n \int (q_{0,n}(x; g_{\beta,n}) - q_{0,n}(x, c_{18}\cdot g_{\beta,n}))^2\,\mathrm{d}x \\
&= 24\,n \left(\frac{1-2^{-\beta}}{12}\right)^2 \left(2\int_{c_{18}\cdot g_{\beta,n}}^{g_{\beta,n}} \Big(W_\beta(x) - W_\beta(g_{\beta,n})\Big)^2\,\mathrm{d}x \right.\\
&\qquad\qquad \left. + \int_{-c_{18}\cdot g_{\beta,n}}^{c_{18}\cdot g_{\beta,n}} \Big(W_\beta(c_{18}\cdot g_{\beta,n}) - W_\beta(g_{\beta,n})\Big)^2\,\mathrm{d}x\right) \\
&\leq 24\,n\,L_W(\beta)^2 \left(\frac{1-2^{-\beta}}{12}\right)^2 \left(2\int_{c_{18}\cdot g_{\beta,n}}^{g_{\beta,n}} (g_{\beta,n} - x)^{2\beta}\,\mathrm{d}x \right.\\
&\qquad\qquad \left. + 2(1-c_{18})^2 c_{18} g_{\beta,n}^{2\beta+1}\right) \\
&= c_{20}
\end{aligned}$$

with

$$c_{20} = c_{20}(\beta) = 48\,L_W(\beta)^2 4^{-(2\beta+1)} \left(\frac{1-2^{-\beta}}{12}\right)^2 \left(\frac{(1-c_{18})^{2\beta+1}}{2\beta+1} + (1-c_{18})^2 c_{18}\right),$$

where we used Lemma 3.5 in the last inequality. Theorem 2.2 in Tsybakov (2009) then yields

$$\begin{aligned}
\inf_{T_n} \sup_{p\in\mathscr{S}_k(\beta)} \mathbb{P}_p^{\otimes n} &\left(n^{\frac{\beta}{2\beta+1}}\,|T_n(t) - p(t)| \geq c_{19}\right) \\
&\geq \max\left\{\frac{1}{4}\exp(-c_{20}),\ \frac{1-\sqrt{c_{20}/2}}{2}\right\} > 0.
\end{aligned}$$

Part 2. For $\beta = 1$, consider the function $p_0 : \mathbb{R} \to \mathbb{R}$ with

$$p_0(x) = \begin{cases} 0, & \text{if } \; |x-t| > 4 \\ \frac{1}{4} - \frac{1}{16}|x-t|, & \text{if } \; |x-t| \le 4 \end{cases}$$

and the functions $p_{1,n}, p_{2,n} : \mathbb{R} \to \mathbb{R}$ with

$$\begin{aligned} p_{1,n}(x) &= p_0(x) + q_{t+\frac{9}{4},n}(x; g_{1,n}) - q_{t,n}(x; g_{1,n}) \\ p_{2,n}(x) &= p_0(x) + q_{t+\frac{9}{4},n}(x; g_{1,n}/2) - q_{t,n}(x; g_{1,n}/2) \end{aligned}$$

for $g_{1,n} = \frac{1}{4} n^{-1/3}$, where

$$q_{a,n}(x; g) = \begin{cases} 0, & \text{if } \; |x-a| > g \\ \frac{1}{16}(g - |x-a|), & \text{if } \; |x-a| \le g \end{cases}$$

for $a \in \mathbb{R}$ and $g > 0$. Following the lines of the proof of Proposition 3.3, both $p_{1,n}$ and $p_{2,n}$ are contained in the class $\mathscr{P}_k$ for sufficiently large $k \ge k_0(\beta_*)$. Moreover, both $p_{1,n}$ and $p_{2,n}$ are constant on $B(t, g_{1,n}/2)$, so that

$$p_{1,n|B(t,g_{1,n}/2)},\, p_{2,n|B(t,g_{1,n}/2)} \in \mathcal{H}_{B(t,g_{1,n}/2)}(\infty, 1/4).$$

The absolute distance of $p_{1,n}$ and $p_{2,n}$ in t is given by

$$|p_{1,n}(t) - p_{2,n}(t)| = \frac{1}{32} g_{1,n},$$

whereas the Kullback-Leibler divergence between the associated product probability measures $\mathbb{P}_{1,n}^{\otimes n}$ and $\mathbb{P}_{2,n}^{\otimes n}$ is upper bounded by

$$\begin{aligned} K(\mathbb{P}_{2,n}^{\otimes n}, \mathbb{P}_{1,n}^{\otimes n}) &\le n \int \frac{(p_{2,n}(x) - p_{1,n}(x))^2}{p_{1,n}(x)} \, \mathrm{d}x \\ &\le 16\, n \int (p_{2,n}(x) - p_{1,n}(x))^2 \, \mathrm{d}x \\ &= 32\, n \int (q_{0,n}(x; g_{1,n}) - q_{0,n}(x, g_{1,n}/2))^2 \, \mathrm{d}x \\ &= 32\, n \left(2 \int_{g_{1,n}/2}^{g_{1,n}} \left(\frac{1}{16}(g_{1,n} - x) \right)^2 \mathrm{d}x + \int_{-g_{1,n}/2}^{g_{1,n}/2} \left(\frac{g_{1,n}}{32} \right)^2 \mathrm{d}x \right) \\ &= \frac{2}{3 \cdot 32^2} + \frac{1}{32}. \end{aligned}$$

Together with Theorem 2.2 in Tsybakov (2009) the result follows. □

Proof of Theorem 3.12. Recall the notation of Subsection 3.2 and set the exponent $\gamma = \gamma(c_1) = \frac{1}{2}(c_1 \log 2 - 1)$. To show that the confidence band is adaptive, note that according to Proposition 4.1 and Lemma 4.2 for any $\delta > 0$ there exists some $n_0(\delta)$, such that

$$\begin{aligned}
&\sup_{p\in\mathscr{P}_n} \mathbb{P}_p^{\chi_2}\left(\exists\, t\in[0,1] : |C_{n,\alpha}(t)| \geq \sqrt{6}\cdot 2^{\frac{j_{\min}}{2}+1} q_n(\alpha)\left(\frac{\log\tilde{n}}{\tilde{n}}\right)^{\frac{\beta_{n,p(t)}}{2\beta_{n,p(t)}+1}} (\log\tilde{n})^{\gamma}\right)\\
&\quad= \sup_{p\in\mathscr{P}_n} \mathbb{P}_p^{\chi_2}\left(\exists\, t\in[0,1] : \hat{z}_n(t) \geq \sqrt{6}\cdot\sqrt{\frac{\log\tilde{n}}{\tilde{n}\bar{h}_n(t)}}\cdot\sqrt{\frac{2^{u_n}}{\log\tilde{n}}}\right)\\
&\quad= \sup_{p\in\mathscr{P}_n} \mathbb{P}_p^{\chi_2}\left(\exists\, t\in[0,1] : \hat{h}_n^{loc}(t) \leq \frac{1}{6}\cdot\bar{h}_n(t)\cdot 2^{-u_n}\right)\\
&\quad= \sup_{p\in\mathscr{P}_n} \mathbb{P}_p^{\chi_2}\left(\sup_{t\in[0,1]} \frac{\bar{h}_n(t)}{\hat{h}_n^{loc}(t)}\cdot 2^{-u_n} \geq 6\right)\\
&\quad= \sup_{p\in\mathscr{P}_n} \mathbb{P}_p^{\chi_2}\left(\max_{k\in T_n}\sup_{t\in I_k} \frac{\bar{h}_n(t)}{\min\left\{2^{-\hat{j}_n((k-1)\delta_n)}, 2^{-\hat{j}_n(k\delta_n)}\right\}} \geq 6\right)\\
&\quad\leq \sup_{p\in\mathscr{P}_n} \mathbb{P}_p^{\chi_2}\left(\max_{k\in T_n} \frac{\min\left\{\bar{h}_n((k-1)\delta_n), \bar{h}_n(k\delta_n)\right\}}{\min\left\{2^{-\hat{j}_n((k-1)\delta_n)}, 2^{-\hat{j}_n(k\delta_n)}\right\}} \geq 2\right)\\
&\quad\leq \sup_{p\in\mathscr{P}_n} \mathbb{P}_p^{\chi_2}\left(\exists\, k\in T_n : \frac{\min\left\{2^{-\bar{j}_n((k-1)\delta_n)}, 2^{-\bar{j}_n(k\delta_n)}\right\}}{\min\left\{2^{-\hat{j}_n((k-1)\delta_n)}, 2^{-\hat{j}_n(k\delta_n)}\right\}} \geq 1\right)\\
&\quad= \sup_{p\in\mathscr{P}_n}\left\{1-\mathbb{P}_p^{\chi_2}\left(\forall\, k\in T_n : \frac{\min\left\{2^{-\bar{j}_n((k-1)\delta_n)}, 2^{-\bar{j}_n(k\delta_n)}\right\}}{\min\left\{2^{-\hat{j}_n((k-1)\delta_n)}, 2^{-\hat{j}_n(k\delta_n)}\right\}} < 1\right)\right\}\\
&\quad\leq \sup_{p\in\mathscr{P}_n}\left\{1-\mathbb{P}_p^{\chi_2}\left(\hat{j}_n(k\delta_n) < \bar{j}_n(k\delta_n) \text{ for all } k\in T_n\right)\right\}\\
&\quad\leq \delta
\end{aligned}$$

for all $n \geq n_0(\delta)$. □

5.2. *Proofs of the results in Section 4.*

Proof of Proposition 4.1. We prove first that

$$\lim_{n\to\infty}\sup_{p\in\mathscr{P}_n} \mathbb{P}_p^{\chi_2}\left(\hat{j}_n(k\delta_n) > \bar{j}_n(k\delta_n)+1 \text{ for some } k\in T_n\right) = 0. \tag{5.26}$$

Note first that if $\hat{j}_n(k\delta_n) > \bar{j}_n(k\delta_n)+1$ for some $k\in T_n$, then $\bar{j}_n(k\delta_n)+1$ cannot be an admissible exponent according to the construction of the bandwidth selection scheme in (3.10), that is, $\bar{j}_n(k\delta_n)+1 \notin \mathcal{A}_n(k\delta_n)$. By definition of $\mathcal{A}_n(k\delta_n)$ there exist exponents $m_{n,k}, m'_{n,k} \in \mathcal{J}_n$ with $m_{n,k} > m'_{n,k} \geq \bar{j}_n(k\delta_n)+4$ such that

$$\max_{s\in B\left(k\delta_n, \frac{7}{8}\cdot 2^{-(\bar{j}_n(k\delta_n)+1)}\right)\cap\mathcal{H}_n} |\hat{p}_n^{(2)}(s, m_{n,k}) - \hat{p}_n^{(2)}(s, m'_{n,k})| > c_2\sqrt{\frac{\log\tilde{n}}{\tilde{n}2^{-m_{n,k}}}}.$$

Consequently,

$$
\begin{aligned}
&\mathbb{P}_p^{\chi_2}\left(\hat{j}_n(k\delta_n) > \bar{j}_n(k\delta_n)+1 \text{ for some } k\in T_n\right)\\
&\le \mathbb{P}_p^{\chi_2}\Bigg(\exists k\in T_n \text{ and } \exists\, m_{n,k}, m'_{n,k}\in\mathcal{J}_n \text{ with } m_{n,k}>m'_{n,k}\ge \bar{j}_n(k\delta_n)+4 \text{ such that}\\
&\max_{s\in B\left(k\delta_n,\frac{7}{8}\cdot 2^{-(\bar{j}_n(k\delta_n)+1)}\right)\cap\mathcal{H}_n} |\hat{p}_n^{(2)}(s,m_{n,k})-\hat{p}_n^{(2)}(s,m'_{n,k})| > c_2\sqrt{\frac{\log\tilde{n}}{\tilde{n}2^{-m_{n,k}}}}\Bigg)\\
&\le \sum_{m\in\mathcal{J}_n}\sum_{m'\in\mathcal{J}_n}\mathbb{P}_p^{\chi_2}\Bigg(m>m'\ge\bar{j}_n(k\delta_n)+4 \text{ and}\\
&\qquad \max_{s\in B\left(k\delta_n,\frac{7}{8}\cdot 2^{-(\bar{j}_n(k\delta_n)+1)}\right)\cap\mathcal{H}_n} |\hat{p}_n^{(2)}(s,m)-\hat{p}_n^{(2)}(s,m')|\\
&\qquad\qquad > c_2\sqrt{\frac{\log\tilde{n}}{\tilde{n}2^{-m}}} \text{ for some } k\in T_n\Bigg).
\end{aligned}
$$

We furthermore use the following decomposition into two stochastic terms and two bias terms

$$
\begin{aligned}
&\max_{s\in B\left(k\delta_n,\frac{7}{8}\cdot 2^{-(\bar{j}_n(k\delta_n)+1)}\right)\cap\mathcal{H}_n} \left|\hat{p}_n^{(2)}(s,m)-\hat{p}_n^{(2)}(s,m')\right|\\
&\le \max_{s\in B\left(k\delta_n,\frac{7}{8}\cdot 2^{-(\bar{j}_n(k\delta_n)+1)}\right)\cap\mathcal{H}_n} \left|\hat{p}_n^{(2)}(s,m)-\mathbb{E}_p^{\chi_2}\hat{p}_n^{(2)}(s,m)\right|\\
&\qquad + \max_{s\in B\left(k\delta_n,\frac{7}{8}\cdot 2^{-(\bar{j}_n(k\delta_n)+1)}\right)\cap\mathcal{H}_n} \left|\hat{p}_n^{(2)}(s,m')-\mathbb{E}_p^{\chi_2}\hat{p}_n^{(2)}(s,m')\right|\\
&\quad + \sup_{s\in B\left(k\delta_n,\frac{7}{8}\cdot 2^{-(\bar{j}_n(k\delta_n)+1)}\right)} \left|\mathbb{E}_p^{\chi_2}\hat{p}_n^{(2)}(s,m)-p(s)\right|\\
&\quad + \sup_{s\in B\left(k\delta_n,\frac{7}{8}\cdot 2^{-(\bar{j}_n(k\delta_n)+1)}\right)} \left|\mathbb{E}_p^{\chi_2}\hat{p}_n^{(2)}(s,m')-p(s)\right|.
\end{aligned}
$$

In order to bound the two bias terms, note first that for any $m>m'\ge\bar{j}_n(k\delta_n)+4$ both

$$
\frac{7}{8}\cdot 2^{-(\bar{j}_n(k\delta_n)+1)} = 2^{-(\bar{j}_n(k\delta_n)+1)} - \frac{1}{8}\cdot 2^{-(\bar{j}_n(k\delta_n)+1)} \le 2^{-(\bar{j}_n(k\delta_n)+1)} - 2^{-m}
$$

and

$$
\frac{7}{8}\cdot 2^{-(\bar{j}_n(k\delta_n)+1)} = 2^{-(\bar{j}_n(k\delta_n)+1)} - \frac{1}{8}\cdot 2^{-(\bar{j}_n(k\delta_n)+1)} \le 2^{-(\bar{j}_n(k\delta_n)+1)} - 2^{-m'}.
$$

According to Assumption 3.1 and Lemma 3.2,

$$
p_{|B\left(k\delta_n,2^{-(\bar{j}_n(k\delta_n)+1)}\right)} \in \mathcal{H}_{\beta^*,B\left(k\delta_n,2^{-(\bar{j}_n(k\delta_n)+1)}\right)}\left(\beta_p\left(B\left(k\delta_n,2^{-\bar{j}_n(k\delta_n)}\right)\right),L^*\right),
$$

so that Lemma 4.4 yields,

$$
\begin{aligned}
&\sup_{s\in B\left(k\delta_n,\frac{7}{8}\cdot 2^{-(\bar{j}_n(k\delta_n)+1)}\right)} \left|\mathbb{E}_p^{\chi_2}\hat{p}_n^{(2)}(s,m)-p(s)\right| \\
&\qquad\qquad \le \sup_{s\in B(k\delta_n,2^{-(\bar{j}_n(k\delta_n)+1)}-2^{-m})} \left|\mathbb{E}_p^{\chi_2}\hat{p}_n^{(2)}(s,m)-p(s)\right| \\
&\qquad\qquad \le b_2 2^{-m\beta_p\left(B\left(k\delta_n,2^{-\bar{j}_n(k\delta_n)}\right)\right)} \\
&\qquad\qquad \le b_2 2^{-m\beta_p\left(B\left(k\delta_n,\bar{h}_n(k\delta_n)\right)\right)} \\
&\qquad\qquad \le b_2 2^{-m\beta_{n,p}(k\delta_n)},
\end{aligned}
$$

with the bandwidth $\bar{h}_n(\cdot)$ as defined in (4.1), and analogously

$$
\sup_{s\in B\left(k\delta_n,\frac{7}{8}\cdot 2^{-(\bar{j}_n(k\delta_n)+1)}\right)} \left|\mathbb{E}_p^{\chi_2}\hat{p}_n^{(2)}(s,m')-p(s)\right| \le b_2 2^{-m'\beta_{n,p}(k\delta_n)}.
$$

Thus, the sum of the two bias terms is bounded from above by $2b_2\bar{h}_n(k\delta_n)^{\beta_{n,p}(k\delta_n)}$, such that

$$
\begin{aligned}
&\sqrt{\frac{\tilde{n}2^{-m}}{\log\tilde{n}}}\Bigg(\sup_{s\in B\left(k\delta_n,\frac{7}{8}\cdot 2^{-(\bar{j}_n(k\delta_n)+1)}\right)} \left|\mathbb{E}_p^{\chi_2}\hat{p}_n^{(2)}(s,m)-p(s)\right| \\
&\qquad\qquad + \sup_{s\in B\left(k\delta_n,\frac{7}{8}\cdot 2^{-(\bar{j}_n(k\delta_n)+1)}\right)} \left|\mathbb{E}_p^{\chi_2}\hat{p}_n^{(2)}(s,m')-p(s)\right|\Bigg) \\
&\quad \le \sqrt{\frac{\tilde{n}\bar{h}_n(k\delta_n)}{\log\tilde{n}}}\cdot 2b_2\bar{h}_n(k\delta_n)^{\beta_{n,p}(k\delta_n)} \\
&\quad \le c_{21},
\end{aligned}
$$

where $c_{21}=c_{21}(\beta_*,L^*,\varepsilon)=2b_2\cdot 2^{-j_{\min}(2\beta_*+1)/2}$. Thus, it holds

$$
\begin{aligned}
&\mathbb{P}_p^{\chi_2}\left(\hat{j}_n(k\delta_n)>\bar{j}_n(k\delta_n)+1 \text{ for some } k\in T_n\right) \\
&\quad \le \sum_{m\in\mathcal{J}_n}\sum_{m'\in\mathcal{J}_n}\Bigg\{\mathbb{P}_p^{\chi_2}\Bigg(\max_{k\in T_n}\max_{s\in B\left(k\delta_n,\frac{7}{8}\cdot 2^{-(\bar{j}_n(k\delta_n)+1)}\right)\cap\mathcal{H}_n}\left|\hat{p}_n^{(2)}(s,m)-\mathbb{E}_p^{\chi_2}\hat{p}_n^{(2)}(s,m)\right| \\
&\qquad\qquad\qquad\qquad > \frac{c_2-c_{21}}{2}\sqrt{\frac{\log\tilde{n}}{\tilde{n}2^{-m}}}\Bigg) \\
&\qquad + \mathbb{P}_p^{\chi_2}\Bigg(\max_{k\in T_n}\max_{s\in B\left(k\delta_n,\frac{7}{8}\cdot 2^{-(\bar{j}_n(k\delta_n)+1)}\right)\cap\mathcal{H}_n}\left|\hat{p}_n^{(2)}(s,m')-\mathbb{E}_p^{\chi_2}\hat{p}_n^{(2)}(s,m')\right| \\
&\qquad\qquad\qquad\qquad > \frac{c_2-c_{21}}{2}\sqrt{\frac{\log\tilde{n}}{\tilde{n}2^{-m'}}}\Bigg)\Bigg\} \\
&\quad \le 2\left|\mathcal{J}_n\right|^2\cdot\mathbb{P}_p^{\chi_2}\left(\sup_{s\in\mathcal{H}_n}\max_{h\in\mathcal{G}_n}\sqrt{\frac{\tilde{n}h}{\log\tilde{n}}}\left|\hat{p}_n^{(2)}(s,h)-\mathbb{E}_p^{\chi_2}\hat{p}_n^{(2)}(s,h)\right|>\frac{c_2-c_{21}}{2}\right).
\end{aligned}
$$

Choose $c_2 = c_2(A, \nu, \beta_*, L^*, K, \varepsilon)$ sufficiently large such that $c_2 \geq c_{21} + 2\eta_0$, where η_0 is given in Lemma 4.3. Then, Lemma 4.3 and the logarithmic cardinality of $\mathcal{J}_n$ yield (5.26). In addition, we show that

$$\lim_{n\to\infty} \sup_{p\in\mathscr{P}_n} \mathbb{P}_p^{\chi_2}\left(\hat{j}_n(k\delta_n) < k_n(k\delta_n) \text{ for some } k \in T_n\right) = 0. \tag{5.27}$$

For $t \in [0, 1]$, due to the sequential definition of the set of admissible bandwidths $\mathcal{A}_n(t)$ in (3.10), if $\hat{j}_n(t) < j_{\max}$, then both $\hat{j}_n(t)$ and $\hat{j}_n(t)+1$ are contained in $\mathcal{A}_n(t)$. Note furthermore, that $k_n(t) < j_{\max}$ for any $t \in [0, 1]$. Thus, if $\hat{j}_n(k\delta_n) < k_n(k\delta_n)$ for some $k \in T_n$, there exists some index $j < k_n(k\delta_n) + 1$ with $j \in \mathcal{A}_n(k\delta_n)$ and satisfying (3.6) and (3.7) for $u = 2^{-j}$ and $t = k\delta_n$. In particular,

$$\max_{s\in B\left(k\delta_n,\frac{7}{8}\cdot 2^{-j}\right)\cap\mathcal{H}_n} \left|\hat{p}_n^{(2)}(s, j+3) - \hat{p}_n^{(2)}(s, \bar{j}_n(k\delta_n))\right| \leq c_2\sqrt{\frac{\log\tilde{n}}{\tilde{n}2^{-\bar{j}_n(k\delta_n)}}}$$

for sufficiently large $n \geq n_0(c_1)$, using that $\bar{j}_n(k\delta_n) \in \mathcal{J}_n$ for any $k \in T_n$. Consequently

$$\begin{aligned}
&\mathbb{P}_p^{\chi_2}\left(\hat{j}_n(k\delta_n) < k_n(k\delta_n) \text{ for some } k \in T_n\right) \\
&\leq \sum_{j\in\mathcal{J}_n} \mathbb{P}_p^{\chi_2}\Bigg(\exists\, k \in T_n : j < k_n(k\delta_n) + 1 \text{ and } p_{|B(k\delta_n,2^{-j})} \in \mathcal{H}_{\beta^*,B(k\delta_n,2^{-j})}(\beta, L^*) \\
&\qquad \text{and} \sup_{s\in B(k\delta_n,2^{-j}-g)} |(K_g * p)(s) - p(s)| \geq \frac{g^\beta}{\log n} \text{ for all } g \in \mathcal{G}_\infty \text{ with} \\
&\qquad g \leq 2^{-(j+3)} \text{ and } \max_{s\in B\left(k\delta_n,\frac{7}{8}\cdot 2^{-j}\right)\cap\mathcal{H}_n} \left|\hat{p}_n^{(2)}(s, j+3) - \hat{p}_n^{(2)}(s, \bar{j}_n(k\delta_n))\right| \\
&\qquad\qquad \leq c_2\sqrt{\frac{\log\tilde{n}}{\tilde{n}2^{-\bar{j}_n(k\delta_n)}}}\Bigg).
\end{aligned} \tag{5.28}$$

The triangle inequality yields

$$\begin{aligned}
&\max_{s\in B\left(k\delta_n,\frac{7}{8}\cdot 2^{-j}\right)\cap\mathcal{H}_n} \left|\hat{p}_n^{(2)}(s, j+3) - \hat{p}_n^{(2)}(s, \bar{j}_n(k\delta_n))\right| \\
&\quad\geq \max_{s\in B\left(k\delta_n,\frac{7}{8}\cdot 2^{-j}\right)\cap\mathcal{H}_n} \left|\mathbb{E}_p^{\chi_2}\hat{p}_n^{(2)}(s, j+3) - \mathbb{E}_p^{\chi_2}\hat{p}_n^{(2)}(s, \bar{j}_n(k\delta_n))\right| \\
&\qquad - \max_{s\in B\left(k\delta_n,\frac{7}{8}\cdot 2^{-j}\right)\cap\mathcal{H}_n} \left|\hat{p}_n^{(2)}(s, j+3) - \mathbb{E}_p^{\chi_2}\hat{p}_n^{(2)}(s, j+3)\right| \\
&\qquad - \max_{s\in B\left(k\delta_n,\frac{7}{8}\cdot 2^{-j}\right)\cap\mathcal{H}_n} \left|\hat{p}_n^{(2)}(s, \bar{j}_n(k\delta_n)) - \mathbb{E}_p^{\chi_2}\hat{p}_n^{(2)}(s, \bar{j}_n(k\delta_n))\right|.
\end{aligned}$$

We further decompose

$$\max_{s\in B\left(k\delta_n,\frac{7}{8}\cdot 2^{-j}\right)\cap\mathcal{H}_n}\left|\mathbb{E}_p^{\chi_2}\hat{p}_n^{(2)}(s,j+3)-\mathbb{E}_p^{\chi_2}\hat{p}_n^{(2)}(s,\bar{j}_n(k\delta_n))\right|$$
$$\geq \max_{s\in B\left(k\delta_n,\frac{7}{8}\cdot 2^{-j}\right)\cap\mathcal{H}_n}\left|\mathbb{E}_p^{\chi_2}\hat{p}_n^{(2)}(s,j+3)-p(s)\right|$$
$$-\sup_{s\in B\left(k\delta_n,\frac{7}{8}\cdot 2^{-j}\right)}\left|\mathbb{E}_p^{\chi_2}\hat{p}_n^{(2)}(s,\bar{j}_n(k\delta_n))-p(s)\right|.$$

As Assumption 3.1 is satisfied for $u=2^{-j}$ and $t=k\delta_n$, together with Lemma 3.2 we both have

$$p_{|B(k\delta_n,2^{-j})}\in\mathcal{H}_{\beta^*,B(k\delta_n,2^{-j})}\Big(\beta_p\left(B\left(k\delta_n,2^{-j}\right)\right),L^*\Big) \tag{5.29}$$

and

$$\sup_{s\in B(k\delta_n,2^{-j}-g)}\left|(K_g*p)(s)-p(s)\right|\geq\frac{g^{\beta_p\left(B\left(k\delta_n,2^{-j}\right)\right)}}{\log n} \tag{5.30}$$

for all $g\in\mathcal{G}_\infty$ with $g\leq 2^{-(j+3)}$. In particular, (5.29) together with Lemma 4.4 gives the upper bias bound

$$\sup_{s\in B\left(k\delta_n,\frac{7}{8}\cdot 2^{-j}\right)}\left|\mathbb{E}_p^{\chi_2}\hat{p}_n^{(2)}(s,\bar{j}_n(k\delta_n))-p(s)\right|\leq b_2\cdot 2^{-\bar{j}_n(k\delta_n)\beta_p\left(B\left(k\delta_n,2^{-j}\right)\right)}$$

for sufficiently large $n\geq n_0(c_1)$, whereas (5.30) yields the bias lower bound

$$\sup_{s\in B\left(k\delta_n,\frac{7}{8}\cdot 2^{-j}\right)}\left|\mathbb{E}_p^{\chi_2}\hat{p}_n^{(2)}(s,j+3)-p(s)\right|$$
$$=\sup_{s\in B(k\delta_n,2^{-j}-2^{-(j+3)})}\left|\mathbb{E}_p^{\chi_2}\hat{p}_n^{(2)}(s,j+3)-p(s)\right|$$
$$\geq\frac{2^{-(j+3)\beta_p\left(B\left(k\delta_n,2^{-j}\right)\right)}}{\log n}. \tag{5.31}$$

To show that the above lower bound even holds for the maximum over the set $B\left(k\delta_n,\frac{7}{8}\cdot 2^{-j}\right)\cap\mathcal{H}_n$, note that for any point $k\delta_n-\frac{7}{8}2^{-j}\leq\tilde{t}\leq k\delta_n+\frac{7}{8}2^{-j}$ there exists some $t\in\mathcal{H}_n$ with $|t-\tilde{t}|\leq\delta_n$, and

$$\left|\mathbb{E}_p^{\chi_2}\hat{p}_n^{(2)}(t,j+3)-p(t)\right|$$
$$=\left|\int K(x)\Big\{p(t+2^{-(j+3)}x)-p(t)\Big\}\,\mathrm{d}x\right|$$
$$\geq\left|\int K(x)\Big\{p\left(\tilde{t}+2^{-(j+3)}x\right)-p\left(\tilde{t}\right)\Big\}\,\mathrm{d}x\right|$$
$$-\int|K(x)|\cdot\left|p(t+2^{-(j+3)}x)-p\left(\tilde{t}+2^{-(j+3)}x\right)\right|\mathrm{d}x$$
$$-\int|K(x)|\cdot\left|p(t)-p\left(\tilde{t}\right)\right|\mathrm{d}x$$

(5.32)

$$\geq \left| \int K(x) \Big\{ p\left(\tilde{t} + 2^{-(j+3)}x\right) - p\left(\tilde{t}\right) \Big\} \, \mathrm{d}x \right| - 2\|K\|_1 L^* \cdot |t - \tilde{t}|^{\beta_*},$$

where

$$\begin{aligned} |t-\tilde{t}|^{\beta_*} &\leq \delta_n^{\beta_*} \\ &\leq 2^{-j_{\min}} \left(\frac{\log \tilde{n}}{\tilde{n}} \right)^{\frac{1}{2}} (\log \tilde{n})^{-2} \\ &\leq \frac{\bar{h}_n(k\delta_n)^{\beta_{n,p}(k\delta_n)}}{(\log \tilde{n})^2} \\ &\leq \frac{2^{-(\bar{j}_n(k\delta_n)-1)\beta_{n,p}(k\delta_n)}}{(\log \tilde{n})^2} \\ &\leq \frac{2^{-(j+3)\beta_{n,p}(k\delta_n)}}{(\log \tilde{n})^2} \end{aligned}$$

for sufficiently large $n \geq n_0(c_1)$. For $n \geq n_0(c_1)$ and $j \in \mathcal{J}_n$ with $j < k_n(k\delta_n) + 1$,

$$2^{-j} > 2^{m_n - 1} \cdot 2^{-\bar{j}_n(k\delta_n)} > \bar{h}_n(k\delta_n).$$

Together with (5.29), this implies

$$\beta_p(B(k\delta_n, 2^{-j})) \leq \beta_{n,p}(k\delta_n) \tag{5.33}$$

since otherwise p would be β-Hölder smooth with $\beta > \beta_{n,p}(k\delta_n)$ on a ball $B(k\delta_n, r)$ with radius $r > \bar{h}_n(t)$, which would contradict the definition of $\beta_{n,p}(k\delta_n)$ together with Lemma A.4. This implies

$$|t - \tilde{t}|^{\beta_*} \leq \frac{2^{-(j+3)\beta_p(B(k\delta_n, 2^{-j}))}}{(\log \tilde{n})^2}.$$

Together with inequalities (5.31) and (5.32),

$$\begin{aligned} &\max_{s \in B\left(k\delta_n, \frac{7}{8} \cdot 2^{-j}\right) \cap \mathcal{H}_n} \left| \mathbb{E}_p^{\chi_2} \hat{p}_n^{(2)}(s, j+3) - p(s) \right| \\ &\qquad \geq \sup_{s \in B\left(k\delta_n, \frac{7}{8} \cdot 2^{-j}\right)} \left| \mathbb{E}_p^{\chi_2} \hat{p}_n^{(2)}(s, j+3) - p(s) \right| - 2\|K\|_1 L^* \frac{2^{-(j+3)\beta_p(B(k\delta_n, 2^{-j}))}}{(\log \tilde{n})^2} \\ &\qquad \geq \frac{1}{2} \cdot \frac{2^{-(j+3)\beta_p(B(k\delta_n, 2^{-j}))}}{\log \tilde{n}} \end{aligned}$$

for sufficiently large $n \geq n_0(L^*, K, c_1)$. Altogether, we get for $j < k_n(k\delta_n) + 1$,

$$\begin{aligned} &\sqrt{\frac{\tilde{n} 2^{-\bar{j}_n(k\delta_n)}}{\log \tilde{n}}} \max_{s \in B\left(k\delta_n, \frac{7}{8} \cdot 2^{-j}\right) \cap \mathcal{H}_n} \left| \mathbb{E}_p^{\chi_2} \hat{p}_n^{(2)}(s, j+3) - \mathbb{E}_p^{\chi_2} \hat{p}_n^{(2)}(s, \bar{j}_n(k\delta_n)) \right| \\ &\geq \sqrt{\frac{\tilde{n} 2^{-\bar{j}_n(k\delta_n)}}{\log \tilde{n}}} \left(\max_{s \in B\left(k\delta_n, \frac{7}{8} \cdot 2^{-j}\right) \cap \mathcal{H}_n} \left| \mathbb{E}_p^{\chi_2} \hat{p}_n^{(2)}(s, j+3) - p(s) \right| \right. \end{aligned}$$

$$- \max_{s \in B\left(k\delta_n, \frac{7}{8}\cdot 2^{-j}\right)\cap \mathcal{H}_n} \left| \mathbb{E}_p^{\chi_2} \hat{p}_n^{(2)}(s, \bar{j}_n(k\delta_n)) - p(s) \right| \Bigg)$$

$$\geq \sqrt{\frac{\tilde{n}\bar{h}_n(k\delta_n)}{2\log\tilde{n}}} \left(\frac{1}{2} \cdot \frac{2^{-(j+3)\beta_p(B(k\delta_n,2^{-j}))}}{\log\tilde{n}} - b_2 \cdot 2^{-\bar{j}_n(k\delta_n)\beta_p\left(B\left(k\delta_n,2^{-j}\right)\right)} \right)$$

$$\geq \sqrt{\frac{\tilde{n}\bar{h}_n(k\delta_n)}{2\log\tilde{n}}} 2^{-(\bar{j}_n(k\delta_n)-1)\beta_p(B(k\delta_n,2^{-j}))} \left(\frac{1}{2} \cdot \frac{2^{(\bar{j}_n(k\delta_n)-j-4)\beta_p(B(k\delta_n,2^{-j}))}}{\log\tilde{n}} - b_2 2^{-\beta_*} \right)$$

$$> \sqrt{\frac{\tilde{n}\bar{h}_n(k\delta_n)}{2\log\tilde{n}}} 2^{-(\bar{j}_n(k\delta_n)-1)\beta_p(B(k\delta_n,2^{-j}))} \left(\frac{2^{(m_n-5)\beta_*}}{2\,\log\tilde{n}} - b_2 2^{-\beta_*} \right).$$

We now show that for $j \in \mathcal{J}_n$ with $j < k_n(k\delta_n) + 1$, we have that

$$\beta_p(B(k\delta_n, 2^{-j})) \leq \beta^*. \tag{5.34}$$

According to (5.33), it remains to show that $\beta_{n,p}(k\delta_n) \leq \beta^*$. If $\beta_{n,p}(k\delta_n) = \infty$, then $\bar{j}_n(k\delta_n) = j_{\min}$. Since furthermore $j \in \mathcal{J}_n$ and therefore $j \geq j_{\min}$, this immediately contradicts $j < k_n(k\delta_n) + 1$. That is, $j < k_n(k\delta_n) + 1$ implies that $\beta_{n,p}(k\delta_n) < \infty$, which in turn implies $\beta_{n,p}(k\delta_n) \leq \beta^*$ according to Remark 3. Due to (3.9) and (5.34), the last expression is again lower bounded by

$$3c_2 \sqrt{\frac{\tilde{n}\bar{h}_n(k\delta_n)}{\log\tilde{n}}} \bar{h}_n(k\delta_n)^{\beta_p(B(k\delta_n,2^{-j}))} 2^{j_{\min}\frac{2\beta_p(B(k\delta_n,2^{-j}))+1}{2}}$$

for sufficiently large $n \geq n_0(L^*, K, \beta_*, \beta^*, c_1, c_2)$. Recalling (5.33), we obtain

$$\sqrt{\frac{\tilde{n}2^{-\bar{j}_n(k\delta_n)}}{\log\tilde{n}}} \max_{s \in B\left(k\delta_n, \frac{7}{8}\cdot 2^{-j}\right)\cap \mathcal{H}_n} \left| \mathbb{E}_p^{\chi_2} \hat{p}_n^{(2)}(s, j+3) - \mathbb{E}_p^{\chi_2} \hat{p}_n^{(2)}(s, \bar{j}_n(k\delta_n)) \right|$$

$$\geq 3c_2 \sqrt{\frac{\tilde{n}\bar{h}_n(k\delta_n)}{\log\tilde{n}}} \bar{h}_n(k\delta_n)^{\beta_{n,p}(k\delta_n)} 2^{j_{\min}\frac{2\beta_p(B(k\delta_n,2^{-j}))+1}{2}}$$

$$= 3c_2.$$

Thus, by the above consideration and (5.28),

$$\mathbb{P}_p^{\chi_2}\left(\hat{j}_n(k\delta_n) < k_n(k\delta_n) \text{ for some } k \in T_n \right) \leq \sum_{j \in \mathcal{J}_n} (P_{j,1} + P_{j,2})$$

for sufficiently large $n \geq n_0(L^*, K, \beta_*, \beta^*, c_1, c_2)$, with

$$P_{j,1} = \mathbb{P}_p^{\chi_2}\Bigg(\exists k \in T_n : j < k_n(k\delta_n) + 1 \text{ and } \sqrt{\frac{\tilde{n}2^{-(j+3)}}{\log\tilde{n}}}$$

$$\cdot \max_{s \in B\left(k\delta_n, \frac{7}{8}\cdot 2^{-j}\right)\cap \mathcal{H}_n} \left| \hat{p}_n^{(2)}(s, j+3) - \mathbb{E}_p^{\chi_2} \hat{p}_n^{(2)}(s, j+3) \right| \geq c_2 \Bigg)$$

$$P_{j,2} = \mathbb{P}_p^{\chi_2}\Bigg(\exists k \in T_n : j < k_n(k\delta_n) + 1 \text{ and } \sqrt{\frac{\tilde{n}2^{-\bar{j}_n(k\delta_n)}}{\log \tilde{n}}} \cdot \max_{s \in B\left(k\delta_n, \frac{7}{8}\cdot 2^{-j}\right)\cap \mathcal{H}_n} \left|\hat{p}_n^{(2)}(s, \bar{j}_n(k\delta_n)) - \mathbb{E}_p^{\chi_2}\hat{p}_n^{(2)}(s, \bar{j}_n(k\delta_n))\right| \geq c_2\Bigg).$$

Both $P_{j,1}$ and $P_{j,2}$ are bounded by

$$P_{j,i} \leq \mathbb{P}_p^{\chi_2}\left(\sup_{s\in\mathcal{H}_n}\max_{h\in\mathcal{G}_n}\sqrt{\frac{\tilde{n}h}{\log\tilde{n}}}\left|\hat{p}_n^{(2)}(s,h) - \mathbb{E}_p^{\chi_2}\hat{p}_n^{(2)}(s,h)\right| \geq c_2\right), \quad i = 1, 2.$$

For sufficiently large $c_2 \geq \eta_0$, Lemma 4.3 and the logarithmic cardinality of $\mathcal{J}_n$ yield (5.27). □

Proof of Lemma 4.2. We prove both inequalities separately.

Part (i). First, we show that the density p cannot be substantially unsmoother at $z \in (s,t)$ than at the boundary points s and t. Precisely, we shall prove that $\min\{\bar{h}_n(s), \bar{h}_n(t)\} \leq 2\bar{h}_n(z)$. In case

$$\beta_{n,p}(s) = \beta_{n,p}(t) = \infty,$$

that is $\bar{h}_n(s) = \bar{h}_n(t) = 2^{-j_{\min}}$, we immediately obtain $\bar{h}_n(z) \geq \frac{1}{2}2^{-j_{\min}}$ since

$$B\left(z, \frac{1}{2}2^{-j_{\min}}\right) \subset B(s, \bar{h}_n(s)) \cap B(t, \bar{h}_n(t)).$$

Hence, we subsequently assume that

$$\min\{\beta_{n,p}(s), \beta_{n,p}(t)\} < \infty.$$

Note furthermore that

$$\min\left\{\bar{h}_n(s), \bar{h}_n(t)\right\} = h_{\min\{\beta_{n,p}(s),\beta_{n,p}(t)\},n}. \tag{5.35}$$

In a first step, we subsequently conclude that

$$z + \frac{1}{2}h_{\min\{\beta_{n,p}(s),\beta_{n,p}(t)\},n} < s + \bar{h}_n(s) \tag{5.36}$$

or

$$z - \frac{1}{2}h_{\min\{\beta_{n,p}(s),\beta_{n,p}(t)\},n} > t - \bar{h}_n(t). \tag{5.37}$$

Note first that $|s-t| < h_{\beta,n}$ for all $\beta \geq \beta_*$ by condition (4.2). Assume now that (5.36) does not hold. Then, inequality (5.37) directly follows as

$$z - \frac{1}{2}\min\{\bar{h}_n(s), \bar{h}_n(t)\} = z + \frac{1}{2}\min\{\bar{h}_n(s), \bar{h}_n(t)\} - \min\{\bar{h}_n(s), \bar{h}_n(t)\}$$

$$
\begin{aligned}
&\geq s + \bar{h}_n(s) - \min\{\bar{h}_n(s), \bar{h}_n(t)\} \\
&\geq t - (t - s) \\
&> t - \bar{h}_n(t).
\end{aligned}
$$

Vice versa, if (5.37) does not hold, then a similar calculation as above shows that (5.36) is true. Subsequently, we assume without loss of generality that (5.36) holds. That is,

$$
\begin{aligned}
s - \bar{h}_n(s) &< z - \frac{1}{2} h_{\min\{\beta_{n,p}(s), \beta_{n,p}(t)\},n} \\
&< z + \frac{1}{2} h_{\min\{\beta_{n,p}(s), \beta_{n,p}(t)\},n} \qquad (5.38)\\
&< s + \bar{h}_n(s).
\end{aligned}
$$

There exists some $\tilde{\beta} > 0$ with

$$
h_{\tilde{\beta},n} = \frac{1}{2} \min\{\bar{h}_n(t), \bar{h}_n(s)\}. \qquad (5.39)
$$

for sufficiently large $n \geq n_0(\beta_*)$. Equation (5.39) implies that

$$
\tilde{\beta} < \min\{\beta_{n,p}(s), \beta_{n,p}(t)\} \leq \beta_{n,p}(s). \qquad (5.40)
$$

Finally, we verify that

$$
\beta_{n,p}(z) \geq \tilde{\beta}. \qquad (5.41)
$$

Using Lemma A.4 as well as (5.38), (5.39), and (5.40) we obtain

$$
\begin{aligned}
&\|p\|_{\tilde{\beta},\beta^*,B(z,h_{\tilde{\beta},n})} \\
&\quad = \sum_{k=0}^{\lfloor \tilde{\beta}\wedge\beta^* \rfloor} \|p^{(k)}\|_{B\left(z,\frac{1}{2}\min\{\bar{h}_n(t),\bar{h}_n(s)\}\right)} \\
&\qquad + \sup_{\substack{x,y \in B\left(z,\frac{1}{2}\min\{\bar{h}_n(t),\bar{h}_n(s)\}\right) \\ x \neq y}} \frac{|p^{(\lfloor \tilde{\beta}\wedge\beta^* \rfloor)}(x) - p^{(\lfloor \tilde{\beta}\wedge\beta^* \rfloor)}(y)|}{|x-y|^{\tilde{\beta} - \lfloor \tilde{\beta}\wedge\beta^* \rfloor}} \\
&\quad \leq L^*.
\end{aligned}
$$

Consequently, we conclude (5.41). With (5.35) and (5.39), this in turn implies

$$
\min\left\{\bar{h}_n(s), \bar{h}_n(t)\right\} = 2h_{\tilde{\beta},n} \leq 2h_{\beta_{n,p}(z),n} = 2\bar{h}_n(z).
$$

Part (ii). Now, we show that the density p cannot be substantially smoother at $z \in (s,t)$ than at the boundary points s and t. Without loss of generality, let $\beta_{n,p}(t) \leq \beta_{n,p}(s)$. We prove the result by contradiction: assume that

$$
\min\left\{\bar{h}_n(s), \bar{h}_n(t)\right\} < \frac{8}{17} \cdot \bar{h}_n(z). \qquad (5.42)
$$

Since $t - z \le h_{\beta,n}/8$ for all $\beta \ge \beta_*$ by condition (4.2), so that in particular $t - z \le \bar{h}_n(t)/8$, we obtain together with (5.42) that

$$\frac{1}{2}\left(z - t + \bar{h}_n(z)\right) > \frac{1}{2}\left(-\frac{1}{8}\bar{h}_n(t) + \frac{17}{8}\bar{h}_n(t)\right) = \bar{h}_n(t) > 0. \tag{5.43}$$

Because furthermore $\frac{1}{2}(z - t + \bar{h}_n(z)) < 1$, there exists some $\beta' = \beta'(n) > 0$ with

$$h_{\beta',n} = \frac{1}{2}\left(z - t + \bar{h}_n(z)\right).$$

This equation in particular implies that $h_{\beta',n} < \frac{1}{2}\bar{h}_n(z)$ and thus $\beta' < \beta_{n,p}(z)$. Since furthermore $t - z < \bar{h}_n(z)$ by condition (4.2) and therefore also

$$z - \bar{h}_n(z) < t - h_{\beta',n} < t + h_{\beta',n} < z + \bar{h}_n(z),$$

we immediately obtain

$$\|p\|_{\beta',\beta^*,B(t,h_{\beta',n})} \le L^*,$$

so that

$$\beta_{n,p}(t) \ge \beta'.$$

This contradicts inequality (5.43). □

Proof of Lemma 4.3. Without loss of generality, we prove the inequality for the estimator $\hat{p}_n^{(1)}(\cdot, h)$ based on χ_1. Note first, that

$$\sup_{s\in\mathcal{H}_n}\sup_{h\in\mathcal{G}_n}\sqrt{\frac{\tilde{n}h}{\log\tilde{n}}}\left|\hat{p}_n^{(1)}(s,h) - \mathbb{E}_p^{\chi_1}\hat{p}_n^{(1)}(s,h)\right| = \sup_{f\in\mathscr{E}_n}\left|\sum_{i=1}^{\tilde{n}}\left(f(X_i) - \mathbb{E}_p f(X_i)\right)\right|$$

with

$$\mathscr{E}_n = \left\{f_{n,s,h}(\cdot) = (\tilde{n}h\log\tilde{n})^{-\frac{1}{2}}K\left(\frac{\cdot - s}{h}\right) : s \in \mathcal{H}_n,\ h \in \mathcal{G}_n\right\}.$$

Observe first that

$$\begin{aligned}
\sup_{p\in\mathscr{P}_n}\operatorname{Var}_p(f_{n,s,h}(X_1)) &\le \sup_{p\in\mathscr{P}_n}\mathbb{E}_p f_{n,s,h}(X_1)^2 \\
&= \sup_{p\in\mathscr{P}_n}\frac{1}{\tilde{n}h\log\tilde{n}}\int K\left(\frac{x-s}{h}\right)^2 p(x)\,\mathrm{d}x \\
&\le \frac{L^*\|K\|_2^2}{\tilde{n}\log\tilde{n}} \\
&=: \sigma_n^2
\end{aligned}$$

uniformly over all $f_{n,s,h} \in \mathscr{E}_n$, and

$$\begin{aligned}\sup_{s\in\mathcal{H}_n}\max_{h\in\mathcal{G}_n}\|f_{n,s,h}\|_{\sup} &\leq \max_{h\in\mathcal{G}_n}\frac{\|K\|_{\sup}}{\sqrt{\tilde{n}h\log\tilde{n}}}\\ &= \|K\|_{\sup}(\log\tilde{n})^{-\frac{\kappa_2+1}{2}}\\ &\leq \frac{\|K\|_{\sup}}{(\log\tilde{n})^{3/2}} =: U_n,\end{aligned}$$

where the last inequality holds true because by definition of $\kappa_2 \geq 2$ in (3.9). In particular $\sigma_n \leq U_n$ for sufficiently large $n \geq n_0(L^*, K)$. Since $(\tilde{n}h\log\tilde{n})^{-1/2} \leq 1$ for all $h \in \mathcal{G}_n$ and for all $n \geq n_0$, the class $\mathscr{E}_n$ satisfies the VC property

$$\limsup_{n\to\infty}\sup_Q N\left(\mathcal{E}_n, \|\cdot\|_{L^2(Q)}, \varepsilon\|K\|_{\sup}\right) \leq \left(\frac{A''}{\varepsilon}\right)^{\nu''}$$

for some VC characteristics $A'' = A''(A, K)$ and $\nu' = \nu + 1$, by the same arguments as in (5.17). According to Proposition 2.2 in Giné and Guillou (2001), there exist constants $c_{22} = c_{22}(A'', \nu'')$ and $c_5 = c_5(A'', \nu'')$, such that

$$\begin{aligned}&\mathbb{P}_p^{\chi_1}\left(\sup_{s\in\mathcal{H}_n}\max_{h\in\mathcal{G}_n}\sqrt{\frac{\tilde{n}h}{\log\tilde{n}}}\left|\hat{p}_n^{(1)}(s,h) - \mathbb{E}_p^{\chi_1}\hat{p}_n^{(1)}(s,h)\right| > \eta\right)\\ &\quad\leq c_5\exp\left(-\frac{\eta}{c_5U_n}\log\left(1+\frac{\eta U_n}{c_5\left(\sqrt{\tilde{n}\sigma_n^2}+U_n\sqrt{\log(A''U_n/\sigma_n)}\right)^2}\right)\right) \qquad (5.44)\\ &\quad\leq c_5\exp\left(-\frac{\eta}{c_5U_n}\log\left(1+c_{23}\eta U_n\log\tilde{n}\right)\right)\end{aligned}$$

uniformly over all $p \in \mathscr{P}_n$, for all $n \geq n_0(A'', K, L^*)$ with $c_{23} = c_{23}(A'', \nu'', L^*, K)$, whenever

$$\eta \geq c_{22}\left(U_n\log\left(\frac{A''U_n}{\sigma_n}\right) + \sqrt{\tilde{n}\sigma_n^2}\sqrt{\log\left(\frac{A''U_n}{\sigma_n}\right)}\right). \qquad (5.45)$$

Since the right hand side in (5.45) is bounded from above by some positive constant $\eta_0 = \eta_0(A'', \nu'', L^*, K)$ for sufficiently large $n \geq n_0(A'', \nu'', L^*, K)$, inequality (5.44) holds in particular for all $n \geq n_0(A'', \nu, K, L^*)$ and for all $\eta \geq \eta_0$. Finally, using the inequality $\log(1+x) \geq \frac{x}{2}$ for $0 \leq x \leq 2$ (Lemma A.2), we obtain for all $\eta \geq \eta_0$

$$\begin{aligned}&\mathbb{P}_p^{\chi_1}\left(\sup_{s\in\mathcal{H}_n}\max_{h\in\mathcal{G}_n}\sqrt{\frac{\tilde{n}h}{\log\tilde{n}}}\left|\hat{p}_n^{(1)}(s,h) - \mathbb{E}_p^{\chi_1}\hat{p}_n^{(1)}(s,h)\right| > \eta\right)\\ &\qquad\leq c_5\exp\left(-c_{24}\eta(\log\tilde{n})^{3/2}\log\left(1+c_{25}\frac{\eta_0}{\sqrt{\log\tilde{n}}}\right)\right)\\ &\qquad\leq c_5\exp\left(-\frac{1}{2}c_{24}c_{25}\eta_0\eta\log\tilde{n}\right)\end{aligned}$$

uniformly over all $p \in \mathscr{P}_n$, for all $n \geq n_0(A'', \nu'', K, L^*)$ and positive constants $c_{24} = c_{24}(A'', \nu'', K)$ and $c_{25} = c_{25}(A'', \nu'', L^*, K)$, which do not depend on n or η. □

Proof of Lemma 4.4. Let $t \in \mathbb{R}$, $g, h > 0$, and

$$p_{|B(t,g+h)} \in \mathcal{H}_{\beta^*, B(t,g+h)}(\beta, L).$$

The three cases $\beta \leq 1$, $1 < \beta < \infty$, and $\beta = \infty$ are analyzed separately. In case $\beta \leq 1$, we obtain

$$\sup_{s \in B(t,g)} |(K_h * p)(s) - p(s)| \leq \int |K(x)| \sup_{s \in B(t,g)} |p(s+hx) - p(s)| \, \mathrm{d}x,$$

where

$$\sup_{s \in B(t,g)} |p(s+hx) - p(s)| \leq h^\beta \cdot \sup_{\substack{s,s' \in B(t,g+h) \\ s \neq s'}} \frac{|p(s') - p(s)|}{|s' - s|^\beta} \leq L h^\beta.$$

In case $1 < \beta < \infty$, we use the Peano form for the remainder of the Taylor polynomial approximation. Note that $\beta^* \geq 2$ because K is symmetric by assumption, and K is a kernel of order $\lfloor \beta^* \rfloor = \beta^* - 1$ in general, such that

$$\begin{aligned}
&\sup_{s \in B(t,g)} |(K_h * p)(s) - p(s)| \\
&= \sup_{s \in B(t,g)} \left| \int K(x) \Big\{ p(s+hx) - p(s) \Big\} \, \mathrm{d}x \right| \\
&= \sup_{s \in B(t,g)} \left| \int K(x) \left\{ p(s+hx) - P^p_{s, \lfloor \beta \wedge \beta^* \rfloor}(s+hx) + \sum_{k=1}^{\lfloor \beta \wedge \beta^* \rfloor} \frac{p^{(k)}(s)}{k!} \cdot (hx)^k \right\} \mathrm{d}x \right| \\
&\leq \int |K(x)| \sup_{s \in B(t,g)} \left| p(s+hx) - P^p_{s, \lfloor \beta \wedge \beta^* \rfloor}(s+hx) \right| \mathrm{d}x \\
&\leq \int |K(x)| \sup_{s \in B(t,g)} \sup_{s' \in B(s,h)} \left| \frac{p^{(\lfloor \beta \wedge \beta^* \rfloor)}(s') - p^{(\lfloor \beta \wedge \beta^* \rfloor)}(s)}{\lfloor \beta \wedge \beta^* \rfloor !} (hx)^{\lfloor \beta \wedge \beta^* \rfloor} \right| \mathrm{d}x \\
&\leq \frac{h^{\lfloor \beta \wedge \beta^* \rfloor} h^{\beta - \lfloor \beta \wedge \beta^* \rfloor}}{\lfloor \beta \wedge \beta^* \rfloor !} \cdot \int |K(x)| \sup_{s \in B(t,g)} \sup_{\substack{s' \in B(s,h) \\ s' \neq s}} \frac{\left| p^{(\lfloor \beta \wedge \beta^* \rfloor)}(s') - p^{(\lfloor \beta \wedge \beta^* \rfloor)}(s) \right|}{|s - s'|^{\beta - \lfloor \beta \wedge \beta^* \rfloor}} \, \mathrm{d}x \\
&\leq L \|K\|_1 h^\beta. \qquad (5.46)
\end{aligned}$$

In case $\beta = \infty$, the density p satisfies $p_{|B(t,g+h)} \in \mathcal{H}_{\beta^*, B(t,g+h)}(\beta, L^*)$ for all $\beta > 0$. That is, the upper bound (5.46) on the bias holds for any $\beta > 0$, implying that

$$\sup_{s \in B(t,g)} |(K_h * p)(s) - p(s)| = 0.$$

This completes the proof. □

PROOF OF LEMMA 4.5. Note that by symmetry of K

$$(K_h * p)(s) - p(s) = \frac{1}{2}\int_{-1}^{1} K(x)\Big(p(s+hx) + p(s-hx) - 2p(s)\Big)\,\mathrm{d}x.$$

The upper bound can thus be deduced exactly as in the proof of Lemma 4.4. □

APPENDIX A: AUXILIARY RESULTS

LEMMA A.1. *For $z \in [0,1]$, the second moments of $\tilde{W}_{k,l}(z), k, l \in T_n$ as defined in (5.22) are bounded by*

$$\mathbb{E}_W \tilde{W}_{k,l}(z)^2 \leq 4.$$

PROOF OF LEMMA A.1. As $\tilde{W}_{k,l}(\cdot) = -\tilde{W}_{l,k}(\cdot)$, we assume $k \leq l$ without loss of generality. For any $k, l \in T_n$

$$\mathbb{E}_W \tilde{W}_{k,l}(z)^2 = \sum_{i=1}^{10} E_i$$

with

$$\begin{aligned}
E_1 &= \frac{1}{\hat{h}_{n,k}^{loc}} \mathbb{E}_W W(k\delta_n - z\hat{h}_{n,k}^{loc})^2 \\
&= \frac{k\delta_n - z\hat{h}_{n,k}^{loc}}{\hat{h}_{n,k}^{loc}} \\
E_2 &= -\frac{2}{\hat{h}_{n,k}^{loc}} \mathbb{E}_W W(k\delta_n - z\hat{h}_{n,k}^{loc}) W(k\delta_n + z\hat{h}_{n,k}^{loc}) \\
&= -2\frac{k\delta_n - z\hat{h}_{n,k}^{loc}}{\hat{h}_{n,k}^{loc}} \\
E_3 &= \frac{2}{\sqrt{\hat{h}_{n,k}^{loc}\hat{h}_{n,l}^{loc}}} \mathbb{E}_W W(k\delta_n - z\hat{h}_{n,k}^{loc}) W(l\delta_n + z\hat{h}_{n,l}^{loc}) \\
&= 2\frac{k\delta_n - z\hat{h}_{n,k}^{loc}}{\sqrt{\hat{h}_{n,k}^{loc}\hat{h}_{n,l}^{loc}}} \\
E_4 &= -\frac{2}{\sqrt{\hat{h}_{n,k}^{loc}\hat{h}_{n,l}^{loc}}} \mathbb{E}_W W(k\delta_n - z\hat{h}_{n,k}^{loc}) W(l\delta_n - z\hat{h}_{n,l}^{loc}) \\
&= -2\frac{\min\{k\delta_n - z\hat{h}_{n,k}^{loc},\ l\delta_n - z\hat{h}_{n,l}^{loc}\}}{\sqrt{\hat{h}_{n,k}^{loc}\hat{h}_{n,l}^{loc}}} \\
E_5 &= \frac{1}{\hat{h}_{n,k}^{loc}} \mathbb{E}_W W(k\delta_n + z\hat{h}_{n,k}^{loc})^2 \\
&= \frac{k\delta_n + z\hat{h}_{n,k}^{loc}}{\hat{h}_{n,k}^{loc}}
\end{aligned}$$

$$\begin{aligned}
E_6 &= -\frac{2}{\sqrt{\hat{h}^{loc}_{n,k}\hat{h}^{loc}_{n,l}}}\mathbb{E}_W W(k\delta_n + z\hat{h}^{loc}_{n,k})W(l\delta_n + z\hat{h}^{loc}_{n,l}) \\
&= -2\frac{\min\{k\delta_n + z\hat{h}^{loc}_{n,k},\ l\delta_n + z\hat{h}^{loc}_{n,l}\}}{\sqrt{\hat{h}^{loc}_{n,k}\hat{h}^{loc}_{n,l}}} \\
E_7 &= \frac{2}{\sqrt{\hat{h}^{loc}_{n,k}\hat{h}^{loc}_{n,l}}}\mathbb{E}_W W(k\delta_n + z\hat{h}^{loc}_{n,k})W(l\delta_n - z\hat{h}^{loc}_{n,l}) \\
&= 2\frac{\min\{k\delta_n + z\hat{h}^{loc}_{n,k}, l\delta_n - z\hat{h}^{loc}_{n,l}\}}{\sqrt{\hat{h}^{loc}_{n,k}\hat{h}^{loc}_{n,l}}} \\
E_8 &= \frac{1}{\hat{h}^{loc}_{n,l}}\mathbb{E}_W W(l\delta_n + z\hat{h}^{loc}_{n,l})^2 \\
&= \frac{l\delta_n + z\hat{h}^{loc}_{n,l}}{\hat{h}^{loc}_{n,l}} \\
E_9 &= -\frac{2}{\hat{h}^{loc}_{n,l}}\mathbb{E}_W W(l\delta_n + z\hat{h}^{loc}_{n,l})W(l\delta_n - z\hat{h}^{loc}_{n,l}) \\
&= -2\frac{l\delta_n - z\hat{h}^{loc}_{n,l}}{\hat{h}^{loc}_{n,l}} \\
E_{10} &= \frac{1}{\hat{h}^{loc}_{n,l}}\mathbb{E}_W W(l\delta_n - z\hat{h}^{loc}_{n,l})^2 \\
&= \frac{l\delta_n - z\hat{h}^{loc}_{n,l}}{\hat{h}^{loc}_{n,l}}.
\end{aligned}$$

Altogether,

$$\begin{aligned}
\mathbb{E}_W\tilde{W}_{k,l}(z)^2 = 4z + \frac{2}{\sqrt{\hat{h}^{loc}_{n,k}\hat{h}^{loc}_{n,l}}}\Bigg(& k\delta_n - z\hat{h}^{loc}_{n,k} - \min\left\{k\delta_n - z\hat{h}^{loc}_{n,k}, l\delta_n - z\hat{h}^{loc}_{n,l}\right\} \\
& - \min\left\{k\delta_n + z\hat{h}^{loc}_{n,k}, l\delta_n + z\hat{h}^{loc}_{n,l}\right\} + \min\left\{k\delta_n + z\hat{h}^{loc}_{n,k}, l\delta_n - z\hat{h}^{loc}_{n,l}\right\}\Bigg).
\end{aligned}$$

We distinguish between the two cases

$$(i)\ k\delta_n - z\hat{h}^{loc}_{n,k} \le l\delta_n - z\hat{h}^{loc}_{n,l} \qquad \text{and} \qquad (ii)\ k\delta_n - z\hat{h}^{loc}_{n,k} > l\delta_n - z\hat{h}^{loc}_{n,l}.$$

In case (i), we obtain

$$\mathbb{E}_W \tilde{W}_{k,l}(z)^2 = 4z + \frac{2}{\sqrt{\hat{h}_{n,k}^{loc}\hat{h}_{n,l}^{loc}}}\Bigg(\min\left\{k\delta_n + z\hat{h}_{n,k}^{loc}, l\delta_n - z\hat{h}_{n,l}^{loc}\right\} - \min\left\{k\delta_n + z\hat{h}_{n,k}^{loc}, l\delta_n + z\hat{h}_{n,l}^{loc}\right\}\Bigg)$$

$$\leq 4.$$

In case (ii), we remain with

$$\mathbb{E}_W \tilde{W}_{k,l}(z)^2 = 4z + \frac{2}{\sqrt{\hat{h}_{n,k}^{loc}\hat{h}_{n,l}^{loc}}}\left(k\delta_n - z\hat{h}_{n,k}^{loc} - \min\left\{k\delta_n + z\hat{h}_{n,k}^{loc}, l\delta_n + z\hat{h}_{n,l}^{loc}\right\}\right).$$

If in the latter expression $k\delta_n + z\hat{h}_{n,k}^{loc} \leq l\delta_n + z\hat{h}_{n,l}^{loc}$, then

$$\mathbb{E}_W \tilde{W}_{k,l}(z)^2 = 4z - \frac{4z\hat{h}_{n,k}^{loc}}{\sqrt{\hat{h}_{n,k}^{loc}\hat{h}_{n,l}^{loc}}} \leq 4.$$

Otherwise, if $k\delta_n + z\hat{h}_{n,k}^{loc} > l\delta_n + z\hat{h}_{n,l}^{loc}$, we arrive at

$$\mathbb{E}_W \tilde{W}_{k,l}(z)^2 = 4z + \frac{2}{\sqrt{\hat{h}_{n,k}^{loc}\hat{h}_{n,l}^{loc}}}\left((k-l)\delta_n - z\left(\hat{h}_{n,k}^{loc} + \hat{h}_{n,l}^{loc}\right)\right) \leq 4$$

because $k \leq l$ and $z \in [0,1]$. Summarizing,

$$\mathbb{E}_W \tilde{W}_{k,l}(z)^2 \leq 4.$$

□

LEMMA A.2. *For any $x \in [0,1]$, we have*

$$e^x - 1 \leq 2x.$$

PROOF. Equality holds for $x = 0$, while $e - 1 \leq 2$. Hence, the result follows by convexity of the exponential function. □

LEMMA A.3. *For any $x \in \mathbb{R} \setminus \{0\}$, we have*

$$1 - \frac{\sin(x)}{x} \leq \frac{x^2}{6}.$$

PROOF. Since both sides of the inequality are symmetric in zero, we restrict our considerations to $x > 0$. For positive x, it is equivalent to

$$f(x) = \sin(x) - x + \frac{x^3}{6} \geq 0.$$

As $f(0) = 0$, it suffices to show that

$$f'(x) = \cos(x) - 1 + \frac{x^2}{2} \geq 0$$

for all $x > 0$. Since furthermore $f'(0) = 0$ and

$$f''(x) = -\sin(x) + x \geq 0$$

for all $x > 0$, the inequality follows. □

The next lemma shows that the monotonicity of the Hölder norms $\|\cdot\|_{\beta_1,U} \leq \|\cdot\|_{\beta_2,U}$ with $0 < \beta_1 \leq \beta_2$ stays valid for the modification $\|\cdot\|_{\beta,\beta^*,U}$.

LEMMA A.4. *For $0 < \beta_1 \leq \beta_2 < \infty$ and $p \in \mathcal{H}_{\beta^*,U}(\beta_2)$,*

$$\|p\|_{\beta_1,\beta^*,U} \leq \|p\|_{\beta_2,\beta^*,U}$$

for any open interval $U \subset \mathbb{R}$ with length less or equal than 1.

PROOF. If $\beta_1 \leq \beta_2$, but $\lfloor \beta_1 \wedge \beta^* \rfloor = \lfloor \beta_2 \wedge \beta^* \rfloor$, the statement follows directly with

$$\|p\|_{\beta_1,\beta^*,U} = \sum_{k=0}^{\lfloor \beta_2 \wedge \beta^* \rfloor} \|p^{(k)}\|_U + \sup_{\substack{x,y \in U \\ x \neq y}} \frac{|p^{(\lfloor \beta_2 \wedge \beta^* \rfloor)}(x) - p^{(\lfloor \beta_2 \wedge \beta^* \rfloor)}(y)|}{|x-y|^{\beta_1 - \lfloor \beta_2 \wedge \beta^* \rfloor}} \leq \|p\|_{\beta_2,\beta^*,U}.$$

If $\beta_1 < \beta_2$ and also $\lfloor \beta_1 \wedge \beta^* \rfloor < \lfloor \beta_2 \wedge \beta^* \rfloor$, we deduce that $\beta_1 < \beta^*$ and $\lfloor \beta_1 \rfloor + 1 \leq \lfloor \beta_2 \wedge \beta^* \rfloor$. Then, the mean value theorem yields

$$\begin{aligned}
\|p\|_{\beta_1,\beta^*,U} &= \sum_{k=0}^{\lfloor \beta_1 \rfloor} \|p^{(k)}\|_U + \sup_{\substack{x,y \in U \\ x \neq y}} \frac{|p^{(\lfloor \beta_1 \rfloor)}(x) - p^{(\lfloor \beta_1 \rfloor)}(y)|}{|x-y|^{\beta_1 - \lfloor \beta_1 \rfloor}} \\
&\leq \sum_{k=0}^{\lfloor \beta_1 \rfloor} \|p^{(k)}\|_U + \|p^{(\lfloor \beta_1 \rfloor + 1)}\|_U \sup_{\substack{x,y \in U \\ x \neq y}} |x-y|^{1-(\beta_1 - \lfloor \beta_1 \rfloor)} \\
&\leq \sum_{k=0}^{\lfloor \beta_1 \rfloor + 1} \|p^{(k)}\|_U \\
&\leq \sum_{k=0}^{\lfloor \beta_2 \wedge \beta^* \rfloor} \|p^{(k)}\|_U \\
&\leq \|p\|_{\beta_2,\beta^*,U}.
\end{aligned}$$

□

References.

Baraud, Y. (2004). Confidence balls in Gaussian regression. *Ann. Stat.* **32** 528-551.

Barral, J., Durand, S., Jaffard, S. and Seuret, S. (2013). Local multifractal analysis. *Fractal Geometry and Dynamical Systems in Pure and Applied Mathematics II: Fractals in Applied Mathematics, Amer. Math. Soc.* **601** 31-64.

Bickel, P. J. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.* **1** 1071-1095.

Bull, A. D. (2012). Honest adaptive confidence bands and self-similar functions. *Electron. J. Stat.* **6** 1490-1516.

Bull, A. D. and Nickl, R. (2013). Adaptive confidence sets in L^2. *Prob. Theory Relat. Fields* **156** 889-919.

Cai, T. T. and Low, M. G. (2004). An adaptation theory for nonparametric confidence intervals. *Ann. Statist.* **32** 1805-1840.

Cai, T. T. and Low, M. G. (2006). Adaptive confidence balls. *Ann. Statist.* **34** 202-228.

Carpentier, A. (2013). Adaptive confidence sets in L_p. *Elect. J. Stat.* **7** 2875-2923.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2014a). Anti-concentration and honest, adaptive confidence bands. *Ann. Statist.* **42** 1787-1818.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2014b). Gaussian approximation of suprema of empirical processes. *Ann. Statist.* **42** 1564-1597.

Daoudi, K., Lévy Véhel, J. and Meyer, Y. (1998). Construction of Continuous Functions with Prescribed Local Regularity. *Constr. Approx.* **14** 349-385.

Davies, L., Kovac, A. and Meise, M. (2009). Nonparametric regression, confidence regions and regularization. *Ann. Stat.* **37** 2597-2625.

Dümbgen, L. (1998). New goodness-of-fit tests and their application to nonparametric confidence sets. *Ann. Statist.* **26** 288-314.

Dümbgen, L. (2003). Optimal confidence bands for shape-restricted curves. *Bernoulli* **9** 423-449.

Genovese, C. and Wasserman, L. (2005). Confidence sets for nonparametric wavelet regression. *Ann. Statist.* **33** 698-729.

Genovese, C. and Wasserman, L. (2008). Adaptive confidence bands. *Ann. Statist.* **36** 875-905.

Giné, E. and Guillou, A. (2001). On consistency of kernel density estimators for randomly censored data: rates holding uniformly over adaptive intervals. *Ann. Inst. H. Poincaré Probab. Statist.* **37** 503-522.

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

Hardy, G. H. (1916). Weierstrass's non-differentiable function. *Trans. Amer. Math. Soc.* **17** 301-325.

Hengartner, N. W. and Stark, P. B. (1995). Finite-sample confidence envelopes for shape-restricted densities. *Ann. Statist.* **23** 525-550.

Heurteaux, Y. (2005). Weierstrass functions in Zygmund's class. *Proc. Amer. Math. Soc.* **133** 2711-2720.

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

Jaffard, S. (1995). Functions with prescribed Hölder exponent. *Appl. Comput. Harmon. Anal.* **2** 400-4001.

Jaffard, S. (2006). Wavelet techniques for pointwise regularity. *Ann. Fac. Sci. Toulouse* **15** 3-33.

Juditsky, A. and Lambert-Lacroix, S. (2003). Nonparametric confidence set estimation. *Math. Methods Statist.* **12** 410-428.

KERKYACHARIAN, G., NICKL, R. and PICARD, D. (2012). Concentration inequalities and confidence bands for needlet density estimators on compact homogeneous manifolds. *Probab. Theory Relat. Fields* **153** 363-404.

KOLTCHINSKII, V. (2011). *Oracle Inequalities in Empirical Risk Minimization and Sparse Recovery Problems. Lecture Notes in Math.* Springer, Heidelberg.

KUEH, A. (2012). Locally adaptive density estimation on the unit sphere using needlets. *Constr. Approx.* **36** 433-458.

LEADBETTER, M., LINDGREN, G. and ROOTZÉN, H. (1983). *Extremes and Related Properties of Random Sequences and Processes.* Springer, New York.

LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach spaces. Isoperimetry and processes.* Springer, New York.

LEPSKI, O. V. (1990). One problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.* **35** 459-470.

LOW, M. G. (1997). On nonparametric confidence intervals. *Ann. Statist.* **25** 2547-2554.

MAULDIN, A. D. and WILLIAMS, S. C. (1986). On the Hausdorff dimension of some graphs. *Trans. Amer. Math. Soc.* **298** 793-804.

NICKL, R. (2015). Discussion of "Frequentist coverage of adaptive nonparametric bayesian credible sets". *Ann. Statist.* **43** 1429-1436.

NICKL, R. and SZABÓ, B. (2016). A sharp adaptive confidence ball for self-similar functions. *Stoch. Proc. Appl.* to appear.

NOLAN, D. and POLLARD, D. (1987). U-processes: rates of convergence. *Ann. Statist.* **15** 780-799.

PICARD, D. and TRIBOULEY, K. (2000). Adaptive confidence interval for pointwise curve estimation. *Ann. Statist.* **28** 298-335.

ROBINS, J. and VAN DER VAART, A. W. (2006). Adaptive nonparametric confidence sets. *Ann. Statist.* **34** 229-253.

SEURET, S. and LÉVY VÉHEL, J. (2002). The local Hölder function of a continuous function. *Appl. Comput. Harmon. Anal.* **13** 263-276.

TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation.* Springer, New York.

FAKULTÄT FÜR MATHEMATIK
RUHR-UNIVERSITÄT BOCHUM
44780 BOCHUM
GERMANY
E-MAIL: tim.patschkowski@ruhr-uni-bochum.de

AND

MATHEMATISCHES INSTITUT
ECKERSTRASSE 1
79104 FREIBURG IM BREISGAU
GERMANY
E-MAIL: angelika.rohde@stochastik.uni-freiburg.de