Ideal Bayesian Spatial Adaptation: Supplement

January 30, 2023

This supplementary material contains the proofs of Theorems in the main text and details of the simulation study. In particular, the proofs of Theorems 5 and 6 are in Section A. The proofs for Theorems 1 and 2 under the white noise model are presented in Section B, the proof of the non-spatial adaptation for common classes of hierarchical Gaussian process prior is presented in Section C and some of the technical lemmas used in the proof of Theorem 5 are presented in Section D. In Section E, we provide the proof of Lemma A.3 used in the proof of Theorem 6 and Section F contains some auxiliary results. Details on the simulation study are in Section G.

Contents

A Proofs of Theorems 5 and 6

A.1 Proof of Theorem 5

We write $L_{max} = \lfloor \log_2 n \rfloor$ and denote with $\mathbb{T}$ the set of binary trees whose deepest internal node depth is smaller than L_{max}. Recall the notation from Section 3.1 where we denoted the set of internal tree nodes with $\mathcal{T}_{int}$ and the set of external tree nodes with $\mathcal{T}_{ext}$. Using the definition of M_{lk}, η_{lk} and $k_l(x)$ in Lemma 1 we first define, for some $\bar{\gamma} > 0$,

$$d_l(x) = \left\lfloor \log_2 \left[C_l(x) \left(\frac{n}{\log n} \right)^{\frac{1}{2t(x)+1}} \right] \right\rfloor \quad \text{where} \quad C_l(x) = (2M_{lk_l(x)}/\bar{\gamma})^{\frac{1}{t(x)+1/2}}. \tag{A.1}$$

It turns out that when[1]

$$l \geq \widetilde{d}_l(x) \equiv \max\{\log_2(1/2\eta_{lk_l(x)}), d_l(x)\}, \tag{A.2}$$

the multiscale coefficient satisfies (from Lemma 1)

$$|\beta^0_{lk_l(x)}| \leq \bar{\gamma}\sqrt{\frac{\log n}{n}}. \tag{A.3}$$

Moreover, (A.2) implies that $|\beta^0_{l'k_{l'}(x)}| \leq \bar{\gamma}\sqrt{\frac{\log n}{n}}$ for *all* $(l', k_{l'}(x))$ where $l' > l$. Indeed, since $\mathrm{I}_{l'k_{l'}(x)} \subset \mathrm{I}_{lk_l(x)}$ we have $M_{l'k_{l'}(x)} \leq M_{lk_l(x)}$ and thereby

$$|\beta^0_{l'k_{l'}(x)}| \leq 2M_{l'k_{l'}(x)}2^{-l'(t(x)+1/2)} \leq 2M_{lk_l(x)}2^{-l(t(x)+1/2)} \leq \bar{\gamma}\sqrt{\log n/n}.$$

[1]Note that when η is bounded away from zero, we have $\widetilde{d}_l(x) = d_l(x)$ when n is large enough.

For a tree $\mathcal{T}$, we denote with $\widetilde{\mathcal{T}}_{int}$ a set of *pre-terminal nodes* such that both children are external nodes, i.e.

$$\widetilde{\mathcal{T}}_{int} = \{(l,k) \in \mathcal{T}_{int} \quad s.t. \quad \{(l+1,2k),(l+1,2k+1)\} \in \mathcal{T}_{ext}\}\,. \tag{A.4}$$

Note that for all $x \in [0,1]$ we have $\widetilde{d}_l(x) \geq \widetilde{d}_{l+1}(x)$.

The main difference between the regression case and the white noise model is the dependence of parameters in the posterior distribution due to the fact that the design is not necessarily regular. Let

$$A_n = \{\sup_{x\in\mathcal{X}} \zeta_n(x)|f(x) - f_0(x)| > M_n\}$$

and let T denote a set of trees $\mathcal{T} \in \mathbb{T}$ that (a) *capture signal* and (b) *that are suitably small locally*. Formally, we define the set T as

$$T = \left\{\mathcal{T} \in \mathbb{T} : l \leq \min_{x\in \mathrm{I}_{lk}} \widetilde{d}_l(x) \quad \forall (l,k) \in \widetilde{\mathcal{T}}_{int} \quad \text{and} \quad S(f_0, A; \upsilon) \subseteq \mathcal{T}_{int}\right\} \tag{A.5}$$

for some $A > 0$ where

$$S(f_0; A; \upsilon) \equiv \{(l,k) : |\beta^0_{lk}| > A \log^{1+\upsilon+1\vee\upsilon} n/\sqrt{n}\}$$

where $a \vee b = \max\{a,b\}$. Going further, with $\mathcal{E}(\mathcal{T})$ we denote the set of functions $f = \sum_{(l,k)\in\mathcal{T}_{int}} \psi_{lk}\beta_{lk}$ that live on the tree skeleton $\mathcal{T}$ and

$$\mathcal{E} = \bigcup_{\mathcal{T}\in T} \mathcal{E}(\mathcal{T}) = \{f : \mathcal{T} \in T\}. \tag{A.6}$$

With $\mathcal{E}$ introduced in (A.6), we show in Section D.1 and Section D.2 that $E_{f_0}\Pi(\mathcal{E}^c\,|Y) \to 0$. We can write, for $\mathcal{A}$ defined in (D.4) with $P_{f_0}(\mathcal{A}^c) \leq 2/p \to 0$ where $p = 2^{L_{max}}$,

$$E_{f_0}\Pi\left[f \in A_n \,\middle|\, Y\right] \leq P_{f_0}[\mathcal{A}^c] + E_{f_0}\Pi[\mathcal{E}^c\,|\,Y] + E_{f_0}\Pi\left[f \in A_n \cap \mathcal{E} \,\middle|\, Y\right]\mathbb{I}_{\mathcal{A}}$$

Using the Markov's inequality, one can bound the last display above with (denoting $\mathcal{X} = \{x_i : 1 \leq i \leq n\}$)

$$\Pi\left[f \in A_n \cap \mathcal{E} \,\middle|\, Y\right] \leq M_n^{-1} \int_{\mathcal{E}} \sup_{x\in\mathcal{X}} \zeta_n(x)|f(x) - f_0^d(x)| d\Pi(f\,|\,Y) + M_n^{-1}B, \tag{A.7}$$

where B is the bias term defined in (B.15) and is shown to be $\mathcal{O}(1)$ in Lemma B.2 and where

$$f_0^d(x) = \sum_{l\leq L_{max}} \sum_{k=0}^{2^l-1} \mathbb{I}[l \leq \widetilde{d}_l(x)]\, \psi_{lk}(x)\beta^0_{lk}.$$

Since trees $\mathcal{T} \in T$ catch large signals (according to the definition of T above) we have $|\beta^0_{lk}| < A \log^{1+\upsilon+1\vee\upsilon} n/\sqrt{n}$ for $(l,k) \notin \mathcal{T}_{int}$ and

$$\sup_{x\in\mathcal{X}}\left[\zeta_n(x)\sum_{(l,k)\notin\mathcal{T}_{int};l\leq\widetilde{d}_l(x)} 2^{l/2}\mathbb{I}_{x\in\mathrm{I}_{lk}}|\beta^0_{lk}|\right] \lesssim \frac{\log^{1+\upsilon+1\vee\upsilon} n}{\sqrt{n}}\sup_{x\in\mathcal{X}}\zeta_n(x)2^{\widetilde{d}_l(x)/2} \lesssim \log^{\upsilon+1\vee\upsilon+1/2} n.$$

It thereby suffices to focus on the active coordinates inside $\mathcal{T}_{int}$. We now show that on the event $\mathcal{A}$

$$\int \max_{(l,k)\in\mathcal{T}_{int}} |\beta_{lk} - \beta^0_{lk}| d\Pi(\boldsymbol{\beta} \mid \mathcal{T}, Y) \lesssim \frac{\log^{\upsilon+1\vee\upsilon} n}{\sqrt{n}}.$$

Set $\Sigma_{\mathcal{T}} = c_n(X'_{\mathcal{T}}X_{\mathcal{T}})^{-1}$ with $c_n = g_n/(1+g_n)$ and $\mu_{\mathcal{T}} = \Sigma_{\mathcal{T}}X'_{\mathcal{T}}[X\boldsymbol{\beta}_0 + \boldsymbol{\nu}]$ we have $\boldsymbol{\beta}_{\mathcal{T}} \mid Y \sim \mathcal{N}(\mu_{\mathcal{T}}, \Sigma_{\mathcal{T}})$ and we use Lemma 8 in [18] which yields for $\bar{\sigma} = \max \mathrm{diag}(\Sigma_{\mathcal{T}})$

$$E\|\boldsymbol{\beta}_{\mathcal{T}} - \boldsymbol{\beta}^0_{\mathcal{T}}\|_\infty \leq \|\mu_{\mathcal{T}} - \boldsymbol{\beta}^0_{\mathcal{T}}\|_\infty + \sqrt{2\bar{\sigma}^2 \log|\mathcal{T}_{int}|} + 2\sqrt{2\pi\bar{\sigma}}. \tag{A.8}$$

For the first term, we note (denoting $\|A\|_\infty = \max_i \sum_j |a_{ij}|$)

$$\|\mu_{\mathcal{T}} - \boldsymbol{\beta}^0_{\mathcal{T}}\|_\infty \leq (1-c_n)\|\boldsymbol{\beta}^0_{\mathcal{T}}\|_\infty + \|\Sigma_{\mathcal{T}}\|_\infty\|X'_{\mathcal{T}}(X_{\backslash\mathcal{T}}\boldsymbol{\beta}^0_{\backslash\mathcal{T}} + F_0 - X\boldsymbol{\beta}_0 + \boldsymbol{\varepsilon})\|_\infty$$

From Lemma F.3 and (D.9) we have on the event $\mathcal{A}$

$$\|X'_{\mathcal{T}}(X_{\backslash\mathcal{T}}\boldsymbol{\beta}^0_{\backslash\mathcal{T}} + F_0 - X\boldsymbol{\beta}_0 + \boldsymbol{\varepsilon})\|_\infty \lesssim \sqrt{n}\log^{\upsilon+1\vee\upsilon} n.$$

Denoting with $a(i,\mathcal{T})$ (resp. $a(\backslash i,\mathcal{T})$) the i^{th} diagonal (resp. off-diagonal) entry in the matrix $X'_{\mathcal{T}}X_{\mathcal{T}}$, we can write using the Gershgorin theorem (see e.g. [**?**]) and Lemma F.4

$$\|\Sigma_{\mathcal{T}}\|_\infty \leq \frac{c_n}{\min_i[a(i,\mathcal{T}) - a(\backslash i,\mathcal{T})]} \leq \frac{1}{\underline{\lambda} n}.$$

Next, $\bar{\sigma} \leq \|\Sigma_{\mathcal{T}}\|_\infty \leq 1/(\underline{\lambda} n)$ and from (A.8) we obtain $E\|\boldsymbol{\beta}_{\mathcal{T}} - \boldsymbol{\beta}^0_{\mathcal{T}}\|_\infty \lesssim \log^{\upsilon+1\vee\upsilon} n/\sqrt{n}$. Therefore, on the event $\mathcal{A}$

$$\begin{aligned}
A(\mathcal{T}) &\equiv \int \sup_{x\in\mathcal{X}}\left[\zeta_n(x)\sum_{(l,k)\in\mathcal{T}_{int}}\mathbb{I}_{x\in\mathrm{I}_{lk}}2^{l/2}|\beta_{lk} - \beta^0_{lk}|\right] d\,\Pi(\boldsymbol{\beta} \mid \mathcal{T}, Y)\\
&\lesssim \int \max_{(l,k)\in\mathcal{T}_{int}} |\beta_{lk} - \beta^0_{lk}| \sup_{x\in\mathcal{X}}\left[\zeta_n(x)2^{\widetilde{d}_l(x)/2}\right] d\,\Pi(\boldsymbol{\beta} \mid \mathcal{T}, Y)\\
&\lesssim \sqrt{\frac{n}{\log n}}\int \max_{(l,k)\in\mathcal{T}_{int}} |\beta_{lk} - \beta^0_{lk}| d\,\Pi(\boldsymbol{\beta} \mid \mathcal{T}, Y) \leq B_A \times \log^{\upsilon+1\vee\upsilon+1/2} n,
\end{aligned}$$

uniformly for all $\mathcal{T} \in T$ for some $B_A > 0$. We now put the pieces together. From the considerations above, we continue the calculations in (A.7) to obtain, on the event $\mathcal{A}$,

$$\Pi\left[f \in A_n \cap \mathcal{E} \,\middle|\, Y\right] \le M_n^{-1} \sum_{\mathcal{T} \in T} \Pi[\mathcal{T} \mid Y] \int_{\mathcal{E}(\mathcal{T})} \sup_{x \in \mathcal{X}} \zeta_n(x) \left|f(x) - f_0^d(x)\right| d\Pi(f \mid Y, \mathcal{T}) + o(1)$$
$$\le M_n^{-1} \mathcal{O}(\log^{\upsilon+1\vee\upsilon+1/2} n) + o(1).$$

The upper bound goes to zero as long as M_n is strictly faster than $\log^{\upsilon+1\vee\upsilon+1/2} n$.

A.2 Proof of Theorem 6

First, we show that $A^c_{\varepsilon_n}(\widetilde{M})$ is contained in

$$\bigcup_{l=1}^{N_n} \left\{ \frac{|f_\beta^S(z_l) - f_0(z_l)|}{\varepsilon_n(z_l)} > \widetilde{M}/2 \right\}$$

so that it suffices to focus on the discretization $\mathcal{I}_n$ of $[0,1]$. To show this, we note that for all $x \in [z_l, z_{l+1})$ we have $f_\beta^S(x) = f_\beta^S(z_l)$. Next, from the Assumption[2] 1 where $M(\cdot) \le \bar{M}$ and $\eta(\cdot) \ge \underline{\eta} > 0$, and by using (4.10) and the Assumption 5 we obtain for a sufficiently large n and a suitable $\alpha_l > 0$

$$|f_0(x) - f_0(z_l)| \le \bar{M} \left(\frac{C_2 \log n}{n}\right)^{t(z_l)} \le \bar{M} \left(\frac{C_2 \log n}{n}\right)^{t(x) - L_0 C_2^{\alpha_l} \left(\frac{\log n}{n}\right)^{\alpha_l}} = O(\varepsilon_n(x)^{2t(x)+1}) = o(\varepsilon_n(x))$$

Hence since $\varepsilon_n(z_l) = \varepsilon_n(x)(1 + o(1))$,

$$\sup_{x \in (0,1)} \frac{|f_\beta^S(x) - f_0(x)|}{\varepsilon_n(x)} \le \max_{l \le N_n} \frac{|f_\beta^S(z_l) - f_0(z_l)|}{\varepsilon_n(z_l)} + o(1)$$

and thereby

$$\Pi(A^c_{\varepsilon_n}(\widetilde{M}) \mid D_n) \le \sum_{l \le N_n} \Pi\left(\frac{|f_\beta^S(z_l) - f_0(z_l)|}{\varepsilon_n(z_l)} > \widetilde{M}/2\right).$$

We now focus on one particular knot value $x = z_l$ for some l. For the sake of simplicity we write hereafter (in this proof) ε_n in place of $\varepsilon_n(x)$ when there is no ambiguity.

For a given partition S, recall that I_x^S denotes the interval in S which contains x. We consider two types of partitions S ('small-bias' versus 'large-bias'), i.e. for some $M_1 > 0$ we distinguish between partitions S satisfying $\{|\bar{y}_{I_x^S} - f_0(x)| \le M_1 \varepsilon_n\}$ and $\{|\bar{y}_{I_x^S} - f_0(x)| > M_1 \varepsilon_n\}$, where

$$\bar{y}_I = \sum_{i: x_i \in I} \frac{Y_i}{n_I} \quad \text{and} \quad n_I = \sum_{i=1}^n \mathbb{I}(x_i \in I).$$

[2] Since in Assumption 1, $M(\cdot)$ and $\eta(\cdot)$ are bounded, they could be regarded as constants.

We further split the 'small-bias' partitions $\{|\bar{y}_{I_x^S} - f_0(x)| \leq M_1\varepsilon_n\}$ into two types (a 'small cell' I_x^S versus a 'large cell' I_x^S), i.e. for some small $\delta > 0$ we distinguish between

$$\{n_{I_x^S} > s_n(\delta)\} \quad \text{and} \quad \{n_{I_x^S} \leq s_n(\delta)\}, \quad s_n(\delta) = \frac{\delta \log n}{\varepsilon_n^2}. \tag{A.9}$$

We first prove that if S is a favorable partition, i.e. if it belongs to

$$B_n = \Big\{S : \Big\{|\bar{y}_{I_x^S} - f_0(x)| \leq M_1\varepsilon_n\Big\} \cap \{n_{I_x^S} > s_n(\delta)\Big\}\Big\}$$

then the conditional posterior distribution given S concentrates on $\{|f_0(x) - f_\beta^S(x)| \leq 2M_1\varepsilon_n\}$. We then prove that the posterior probability of the set of non-favorable partitions, i.e. B_n^c, goes to zero as n goes to infinity.

Recall the definition of $\mathcal{I}(x)$ in (4.14) as the set of intervals which either contain x or are neighboring intervals to the one which contains x. We now define the following events for $u_0, u_2 > 0$ and $\bar{\epsilon}_I = \sum_{i:x_i \in I} \epsilon_i / n_I$

$$\Omega_{n,y}(u_0) = \left\{\forall S \in \mathbb{S} : |\bar{\epsilon}_{I_x^S}| \leq u_0 \sqrt{\frac{\log n}{n_{I_x^S}}}\right\}, \ \Omega_{n,y,2}(u_2) = \left\{\forall I \in \mathcal{I}(x) : |\bar{\epsilon}_I| \leq u_2 \sqrt{\frac{\log n}{n_I}}\right\}.$$

Since for a given I_x^S and $X = (x_1, \dots, x_n)'$ the standard Hoeffding Gaussian tail bound (see e.g. (2.10) in [**?**]) yields

$$P\left(|\bar{\epsilon}_{I_x^S}| > u_0 \sqrt{\frac{\log n}{n_{I_x^S}}} \,\middle|\, X\right) \leq 2 \exp\left\{-\frac{u_0^2 \log n}{2}\right\}$$

and since the number of possible intervals I_x^S in the definition of $\Omega_{n,y}(u_0)$ is of the order $O((n/\log n)^2)$, we have

$$P(\Omega_{n,y}(u_0)^c) = o(\log n/n) \qquad \text{as soon as } u_0^2 \geq 6.$$

Similarly, note that the number of intervals involved in the definition of $\Omega_{n,y,2}$ is of order $O((n/\log n)^4)$. By choosing $u_2^2 > 8$ we thus obtain that $P(\Omega_{n,y,2}(u_2)^c) = o(1/n)$. In the following lemmata, we will thus condition on the high-probability events $\Omega_{n,y}(u_0)$ and $\Omega_{n,y,2}(u_2)$ and we set $\Omega_n = \Omega_{n,x}(u_1) \cap \Omega_{n,y}(u_0) \cap \Omega_{n,y,2}(u_2)$.

Given the structure of the prior, for a given partition S the marginal likelihood density has a product form and is proportional to

$$m(S) = \prod_j m(I_j^S), \quad m(I_j^S) = \mathrm{e}^{-\sum_{i \in I_j^S} (Y_i - \bar{y}_{I_j^S})^2/2} \int_{\mathbb{R}} \mathrm{e}^{-n_{I_j^S}(\beta - \bar{y}_{I_j^S})^2/2} g_j(\beta) d\beta.$$

We will use repeatedly the following inequality on $|\bar{y}_{I_j^S}| < B_0 - \epsilon$ for some arbitrarily small but fixed ϵ

$$\frac{c_0(1+o(1))\mathrm{e}^{-\sum_{i\in I_j^S}(Y_i-\bar{y}_{I_j^S})^2/2}\sqrt{2\pi}}{\sqrt{n_{I_j^S}}} \leq m(I_j^S) \leq \frac{c_1\mathrm{e}^{-\sum_{i\in I_j^S}(Y_i-\bar{y}_{I_j^S})^2/2}\sqrt{2\pi}}{\sqrt{n_{I_j^S}}}. \tag{A.10}$$

Lemma A.1. *Assume the prior (4.9) with (4.11) and (4.12). For any $a > 0$ and if $M_1 \geq \max(2a/\sqrt{\delta}, 1/\sqrt{2\delta})$ then*

$$E\left[\mathbb{I}_{\Omega_n}\Pi(\{|f_0(x) - f_\beta^S(x)| > 2M_1\varepsilon_n\} \cap B_n | D_n)\right] \lesssim n^{-a}.$$

Proof of Lemma A.1. If $S \in B_n$ and $|f_0(x) - f_\beta^S(x)| > 2M_1\varepsilon_n$ then we have

$$|\bar{y}_{I_x^S} - f_\beta^S(x)| \geq |f_0(x) - f_\beta^S(x)| - |\bar{y}_{I_x^S} - f_0(x)| \geq M_1\varepsilon_n.$$

Using (A.10), we then have

$$\begin{aligned}
\Pi\left(|f_0(x) - f_\beta^S(x)| > 2M_1\varepsilon_n | D_n, S\right) &\leq \frac{2\sqrt{n_{I_x^S}}}{c_0\sqrt{2\pi}(1+o(1))} \int\limits_{|\beta-\bar{y}_{I_x^S}|>M_1\varepsilon_n} \exp\{-\frac{n_{I_x^S}}{2}(\beta - \bar{y}_{I_x^S})^2\}g(\beta)d\beta \\
&\leq \frac{2c_1\sqrt{n_{I_x^S}}}{c_0\sqrt{2\pi}(1+o(1))} \exp\left\{-\frac{n_{I_x^S}M_1^2\varepsilon_n^2}{2}\right\} \\
&\lesssim \exp\left\{-\delta M_1^2 \log n/4\right\} = o(n^{-a})
\end{aligned}$$

if $\delta M_1^2 > \max(4a, 1/2)$. $\square$

We now prove that the unfavorable partitions have posterior probability going to 0. Using Lemma A.2 (below), with $a > 1$ we obtain on Ω_n that

$$\Pi\left(S : \{|\bar{y}_{I_x^S} - f_0(x)| > M_1\varepsilon_n\} \cap \{n_{I_x^S} > s_n(\delta)\} \mid D_n\right) = o_p(n^{-1})$$

and that, using Lemma A.3 (below), $\Pi\left(S : \{n_{I_x^S} \leq s_n(\delta)\} \mid D_n\right) = o_p(n^{-1})$. Combining these two results with Lemma A.1, we then have on Ω_n

$$\begin{aligned}
&\Pi\left(|f_0(x) - f_\beta^S(x)| > 2M_1\varepsilon_n | D_n\right) \leq \sum_{S\in B_n} \Pi\left(|f_0(x) - f_\beta^S(x)| > 2M_1\varepsilon_n | D_n, S\right)\Pi(S|D_n) \\
&\quad + \Pi\left(S : \{|\bar{y}_{I_x^S} - f_0(x)| > M_1\varepsilon_n\} \cap \{n_{I_x^S} > s_n(\delta)\} \mid D_n\right) + \Pi\left(S : \{n_{I_x^S} \leq s_n(\delta)\} \mid D_n\right) \\
&\quad = o_p(n^{-1}).
\end{aligned}$$

Lemma A.2. *Assume the prior (4.11) with (4.9) and (4.12). Let $x \in (0,1)$, then for all $a, u_0, u_1, u_2 > 0$ and for all $\delta > 0$ small enough, there exists a constant $C(a, u_1, u_2) > 0$ such that if $M_1 > \max(2u_0/\sqrt{\delta}, C(a, u_1, u_2))$,*

$$E\left[\mathbb{I}_{\Omega_n} \Pi\left(S : \{|\bar{y}_{I_x^S} - f_0(x)| > M_1\varepsilon_n\} \cap \{n_{I_x^S} > s_n(\delta)\} \mid D_n\right)\right] = O(n^{-a}).$$

Lemma A.3. *Assume the prior (4.11) with (4.9) and (4.12). With $\delta > 0$ as in Lemma A.2 we have if $B > 9$,*

$$E\left[\mathbb{I}_{\Omega_{n,x}(u_1)} \Pi\Big(\Big\{n_{I_x^S} \le s_n(\delta)\Big\} \mid D_n\Big)\right] = o(1/n).$$

Lemma A.2 is proved by showing that if $|\bar{y}_{I_x^S} - f_0(x)|$ and $n_{I_x^S} > s_n(\delta)$ then the partition has much smaller posterior probability than the one obtained by splitting I_x into smaller intervals. The proof of Lemma A.2 is given below while the proof of Lemma A.3 is given in Section E of the Supplementary Material [**?**]. The idea of the proof of Lemma A.3 is that partitions verifying $\{n_{I_x^S} > s_n(\delta)\}$ have either much smaller probability than the one resulting from merging I_x^S with a neighboring interval, say $I_{x,1}$, or much smaller probability than the one resulting from splitting $I_{x,1}^S$ into smaller intervals. The latter result comes from the fact that if $I_{x,1}^S$ is too large then there is a point x_1 in $I_{x,1}^S$, such that $|\bar{y}_{I_{x,1}^S} - f_0(x_1)| > M_0\varepsilon_n(x_1)$ and $n_{I_{x,1}^S} > s_n(\delta_1)$ for some appropriate values M_0, δ_1 and Lemma A.2 can then be used.

Proof of Lemma A.2. Throughout the rest of the proof, we suppress the index S when referring to intervals I_x^S or I_j^S. On the event $\Omega_{n,y}(u_0)$, and if $n_{I_x} > s_n(\delta)$ for a given δ, we have for $\bar{\beta}_{0,I_x} = \sum_{x_i \in I_x} f_0(x_i)/n_{I_x}$

$$|\bar{y}_{I_x} - \bar{\beta}_{0,I_x}| = |\bar{\epsilon}_{I_x}| \le u_0 \frac{\sqrt{\log n}}{\sqrt{n_{I_x}}} \le \frac{u_0 \varepsilon_n}{\sqrt{\delta}} \le \frac{M_1 \varepsilon_n}{2}$$

as soon as $M_1 > 2u_0/\sqrt{\delta}$. In particular if $|\bar{y}_{I_x} - f_0(x)| > M_1\varepsilon_n$ then we have from Assumption 1 that as soon as $|I_x| \le \underline{\eta}$,

$$M_1\varepsilon_n/2 \le |\bar{\beta}_{0,I_x} - f_0(x)| \le M|I_x|^{t(x)}$$

so that in all cases $|I_x| \ge (M_1\varepsilon_n/2M)^{1/t(x)}$.

Since the cell I_x has a large bias, we compare the partition S with a partition obtained by splitting I_x into 2 or 3 intervals, say I_1, I_2, and possibly I_3 if x is too far from the boundary of I_x. We do the splitting[3] so that $x \in I_1$ and $|I_1| = (\tau M_1 \varepsilon_n / 2M)^{1/t(x)}$ for some $\tau < 1$. We choose also $\tau > 0$ small so that both $|I_2|, |I_3| \geq (\tau M_1 \varepsilon_n / 2M)^{1/t(x)}$. Then on $\Omega_{n,y}(u_0)$,

$$|\bar{\beta}_{0,I_1} - f_0(x)| \leq \tau M_1 \varepsilon_n / 2 \quad \text{and} \quad |\bar{y}_{I_1} - f_0(x)| \leq \tau M_1 \varepsilon_n / 2 + \frac{u_0 \sqrt{\log n}}{\sqrt{n_{I_1}}}.$$

In the following, we write the computations in the case where we have split I_x into 3 intervals. Computations for the case of 2 intervals can be derived similarly. Note that, by construction, $|I_2| \geq |I_1|$ and $|I_3| \geq |I_1|$. In addition, on the event $\Omega_{n,x}(u_1)$ defined in (4.15) we have $n_{I_j} \geq n p_0 |I_j|/2$ for $j = 1, 2, 3$. Hence, there exists a constant $C_0 > 0$ such that

$$\frac{u_0 \sqrt{\log n}}{\sqrt{n_{I_1}}} \leq \frac{u_0 C_0}{(\tau M_1)^{\frac{1}{2t(x)}}} \varepsilon_n \leq M_1 \varepsilon_n / 2 \tag{A.11}$$

by choosing M_1 large enough so that $|\bar{y}_{I_1} - f_0(x)| \leq M_1 \varepsilon_n (1+\tau)/2$. On the event $\Omega_{n,y,2}(u_2)$, for all $u_2 > 0$, we have

$$|\bar{y}_{I_1}| \leq |f_0(x)| + \epsilon \quad \text{and} \quad |\bar{y}_{I_2}| \leq \|f_0\|_\infty + \epsilon \leq B_0$$

for any $\epsilon > 0$ small when n is large enough since $\|f_0\|_\infty < B_0$. Hence using (A.10),

$$\begin{aligned}
\frac{m(I_x)}{m(I_1)m(I_2)m(I_3)} &\leq \frac{2c_1 \sqrt{n_{I_1} n_{I_2} n_{I_3}}}{2\pi c_0^3 \sqrt{n_{I_x}}} \exp\left(-\frac{\sum_{i \in I_x} (y_i - \bar{y}_{I_x})^2}{2} + \frac{\sum_{j=1}^3 \sum_{i \in I_j} (y_i - \bar{y}_{I_j})^2}{2} \right) \\
&= \frac{2c_1 \sqrt{n_{I_1} n_{I_2} n_{I_3}}}{2\pi c_0^3 \sqrt{n_{I_x}}} \exp\left(-\sum_{j=1}^3 \frac{n_{I_j} (\bar{y}_{I_x} - \bar{y}_{I_j})^2}{2} \right).
\end{aligned}$$

Moreover, we have

$$|\bar{y}_{I_x} - \bar{y}_{I_1}| > |\bar{y}_{I_x} - f_0(x)| - |f_0(x) - \bar{y}_{I_1}| \geq M_1 (1 - \tau) \varepsilon_n / 2 \geq M_1 \varepsilon_n / 4$$

by choosing $\tau \leq 1/2$. Finally, by noting that $n_I \asymp n|I|$ on the event $\Omega_{n,x}(u_1)$ we obtain

$$\frac{m(I_x)}{m(I_1)m(I_2)m(I_3)} \lesssim n \sqrt{|I_1||I_2|} \exp\left(-n_{I_1} M_1^2 \varepsilon_n^2 / 8 \right)$$

Noting that

$$n_{I_1} M_1^2 \varepsilon_n^2 \geq \frac{n p_0 |I_1| M_1^2 \varepsilon_n^2}{2} \geq p_0 (\tau/(2M))^{1/t(x)} M_1^{(2t(x)+1)/t(x)} \log n / 2$$

[3]Without loss of generality, we can assume that cutting an interval of such a size is possible otherwise we would replace it with $|I_1| = (\tau M_1 \varepsilon_n / 2M)^{1/t(x)} (1 + o(1))$ which makes no difference.

Then we have

$$Z_x(S) \equiv \frac{m(I_x)|I_x|^B}{m(I_1)m(I_2)m(I_3)|I_1|^B|I_2|^B|I_3|^B} \lesssim n(|I_1||I_2|)^{-(B-1/2)} \exp\left(-n_{I_1} M_1^2 \varepsilon_n^2/8\right)$$
$$\lesssim (|I_1||I_2|)^{-B} n^{1-M_1^2(M_1\tau)^{1/t(x)}C(p_0,M)}$$
$$\lesssim n^{-[M_1^2(M_1\tau/(2M))^{1/t(x)}2p_0-2B]} = O(n^{-a})$$

as soon as $M_1^2 \geq \max(2B/p_0 + a, 2M/\tau)$ since $|I_1||I_2| \gtrsim \epsilon_n^{1/t(x)}$. This implies that on Ω_n we have for

$$\Pi_1 \equiv \Pi(S : \{|\bar{y}_{I_x} - f_0(x)| > M_1\varepsilon_n\} \cap \{n_{I_x} > s_n(\delta)\} \mid D_n)$$

and $\mathbb{I}_1(S) \equiv \mathbb{I}\{S : |\bar{y}_{I_x} - f_0(x)| > M_1\varepsilon_n\}$ and $\mathbb{I}_2(S) \equiv \mathbb{I}\{S : n_{I_x} > s_n\}$ the following bound

$$\Pi_1 = \frac{\sum\limits_{S=S'\cup I_x} \mathbb{I}_1(S) \times \mathbb{I}_2(S) \times m(S') \times m(I_x) \times \pi_S(S' \cup I_x)}{\sum\limits_{S=S'\cup I_x} m(S') \times m(I_x) \times \pi_S(S' \cup I_x)}$$
$$\leq \frac{\sum\limits_{S=S'\cup I_x} \mathbb{I}_1(S) \times \mathbb{I}_2(S) \times m(S')m(I_1)m(I_2)m(I_3) \times \pi_S(S' \cup I_1 \cup I_2 \cup I_3) \times Z_x(S)}{\sum\limits_{S=S'\cup I_1\cup I_2\cup I_3} \mathbb{I}_2(S) \times m(S') \times m(I_1)m(I_2)m(I_3) \times \pi_S(S' \cup I_1 \cup I_2 \cup I_3)}$$
$$\leq Cn^{-a/2}.$$

□

B Proofs for the White Noise Model

B.1 Proof of Theorem 1

The proof is similar to the proof in the regression case but is simpler. For the sake of self-sufficiency, we recall some the definitions used in the proof of Theorem 5, see also Section A.1. We write $L_{max} = \lfloor \log_2 n \rfloor$ and denote with $\mathbb{T}$ the set of binary trees whose deepest internal node depth is smaller than L_{max}. Recall the notation from Section 3.1 of the manuscript where we denoted the set of internal tree nodes with $\mathcal{T}_{int}$ and the set of external tree nodes with $\mathcal{T}_{ext}$. Using again the definition of M_{lk}, η_{lk} and $k_l(x)$ in Lemma 1 we first define, for some $\bar{\gamma} > 0$,

$$d_l(x) = \left\lfloor \log_2 \left[C_l(x) \left(\frac{n}{\log n} \right)^{\frac{1}{2t(x)+1}} \right] \right\rfloor \quad \text{where} \quad C_l(x) = (2M_{lk_l(x)}/\bar{\gamma})^{\frac{1}{t(x)+1/2}}. \tag{B.1}$$

Using the fact that when

$$l \geq \widetilde{d}_l(x) \equiv \max\{\log_2(1/2\eta_{lk_l(x)}), d_l(x)\}, \tag{B.2}$$

the multiscale coefficient satisfies (from Lemma 1)

$$|\beta^0_{lk_l(x)}| \leq \bar{\gamma}\sqrt{\frac{\log n}{n}}. \tag{B.3}$$

Moreover, (B.2) implies that $|\beta^0_{l'k_{l'}(x)}| \leq \bar{\gamma}\sqrt{\frac{\log n}{n}}$ for *all* $(l', k_{l'}(x))$ where $l' > l$ as explained in Section A.1. For a tree $\mathcal{T}$, we denote with $\widetilde{\mathcal{T}}_{int}$ a set of *pre-terminal nodes* defined in (A.4). Note that for all $x \in [0,1]$ we have

$$\widetilde{d}_l(x) \geq \widetilde{d}_{l+1}(x).$$

In the sequel, T denotes a set of trees $\mathcal{T} \in \mathbb{T}$ that (a) *capture signal* and (b) *that are suitably small locally.* Formally, we define the set T as

$$T = \left\{\mathcal{T} \in \mathbb{T} : l \leq \min_{x \in \mathrm{I}_{lk}} \widetilde{d}_l(x) \quad \forall (l,k) \in \widetilde{\mathcal{T}}_{int} \quad \text{and} \quad S(f_0, A) \subseteq \mathcal{T}_{int}\right\} \tag{B.4}$$

for some $A > 0$ where

$$S(f_0; A) \equiv \{(l,k) : |\beta^0_{lk}| > A \log n/\sqrt{n}\}. \tag{B.5}$$

Going further, with $\mathcal{E}(\mathcal{T})$ we denote the set of functions $f = \sum_{(l,k)\in\mathcal{T}_{int}} \psi_{lk}\beta_{lk}$ that live on the tree skeleton $\mathcal{T}$ and

$$\mathcal{E} = \bigcup_{\mathcal{T} \in T} \mathcal{E}(\mathcal{T}) = \{f : \mathcal{T} \in T\}. \tag{B.6}$$

First, we show that $E_{f_0}\Pi(\mathcal{E}^c \mid Y) \to 0$. To begin, in Section B.1.1 below we show that the posterior concentrates on locally small trees.

B.1.1 Posterior Concentrates on $\mathcal{E}$

Our considerations will be conditional on the event

$$\mathcal{A}_n = \left\{\max_{-1 \leq l \leq L_{max}, 0 \leq k < 2^l} \epsilon^2_{lk} \leq 2\log(2^{L_{max}+1})\right\} \tag{B.7}$$

which has a large probability in the sense that $P(\mathcal{A}^c_n) \lesssim (\log n)^{-1}$.

Lemma B.1. *Let $\widetilde{d}_l(x)$ be as in (B.2). For the Bayesian CART prior from Section 3.1 with a split probability $p_l = (1/\Gamma)^l$ we have, on the event $\mathcal{A}_n$ in (B.7), for $\Gamma > 0$ large enough*

$$\Pi\left[\mathcal{T} : \exists (l,k) \in \widetilde{\mathcal{T}}_{int} \;\; s.t. \;\; l > \min_{x \in \mathrm{I}_{lk}} \widetilde{d}_l(x) \mid Y\right] \to 0. \tag{B.8}$$

Proof. We can write

$$\Pi\left[\mathcal{T} : \exists(l,k) \in \widetilde{\mathcal{T}}_{int} \;\; s.t. \;\; l > \min_{x\in \mathrm{I}_{lk}} \widetilde{d}_l(x) \,|\, Y\right] \le \sum_{l\le L_{max}} \sum_{k=0}^{2^l-1} \mathbb{I}[l > \min_{x\in \mathrm{I}_{lk}} \widetilde{d}_l(x)]\Pi[(l,k) \in \widetilde{\mathcal{T}}_{int} \,|\, Y].$$

We denote with $\mathbb{T}_{lk}$ the set of all trees $\mathcal{T}$ such that $(l,k) \in \widetilde{\mathcal{T}}_{int}$. Then

$$\Pi\left[(l,k) \in \widetilde{\mathcal{T}}_{int} \,|\, Y\right] = \frac{\sum_{\mathcal{T}\in\mathbb{T}_{lk}} W_Y(\mathcal{T})}{\sum_{\mathcal{T}} W_Y(\mathcal{T})} \tag{B.9}$$

where, for $\boldsymbol{\beta}_{\mathcal{T}} = (\beta_{lk} : (l,k) \in \mathcal{T}_{int})'$ and $\boldsymbol{Y}_{\mathcal{T}} = (Y_{lk} : (l,k) \in \mathcal{T}_{int})'$,

$$W_Y(\mathcal{T}) = \Pi(\mathcal{T})N_Y(\mathcal{T}) \quad \text{with} \quad N_Y(\mathcal{T}) = \int \mathrm{e}^{-\frac{n}{2}\|\boldsymbol{\beta}_{\mathcal{T}}\|_2^2 + n\boldsymbol{Y}_{\mathcal{T}}'\boldsymbol{\beta}_{\mathcal{T}}}\pi(\boldsymbol{\beta}_{\mathcal{T}})d\boldsymbol{\beta}_{\mathcal{T}}.$$

For a tree $\mathcal{T} \in \mathbb{T}_{lk}$, denote with $\mathcal{T}^-$ the smallest subtree of $\mathcal{T}$ that does *not* contain (l,k) as a pre-terminal node, i.e. $\mathcal{T}^-$ is obtained from $\mathcal{T}$ by turning (l,k) into a terminal node. We can then rewrite (B.9) as

$$\Pi\left[(l,k) \in \widetilde{\mathcal{T}}_{int} \,|\, Y\right] = \frac{\sum_{\mathcal{T}\in\mathbb{T}_{lk}} \frac{W_Y(\mathcal{T})}{W_Y(\mathcal{T}^-)} W_Y(\mathcal{T}^-)}{\sum_{\mathcal{T}} W_Y(\mathcal{T})}. \tag{B.10}$$

Assuming an independent product prior $\beta_{lk} \overset{iid}{\sim} \mathcal{N}(0,1)$, we have

$$\frac{W_Y(\mathcal{T})}{W_Y(\mathcal{T}^-)} = \frac{\Pi(\mathcal{T})}{\Pi(\mathcal{T}^-)} \frac{\mathrm{e}^{\frac{n^2}{2(n+1)}Y_{lk}^2}}{\sqrt{n+1}}. \tag{B.11}$$

See Section 3.1 in [18] for details on this derivation. Since (l,k) is such that $l \ge \widetilde{d}_l(x)$ for some $x \in \mathrm{I}_{lk}$, we have $|\beta^0_{l\,k_l(x)}| \le \bar{\gamma}\sqrt{\log n/n}$ from (B.3) and thereby $Y_{lk}^2 = (\beta^0_{lk} + \frac{1}{\sqrt{n}}\epsilon_{lk})^2 \le C_y \log n/n$ on the event $\mathcal{A}_n$ for some $C_y > 0$. Next, the prior ratio (under the Galton-Watson process prior) equals

$$\frac{\Pi(\mathcal{T})}{\Pi(\mathcal{T}^-)} = \frac{p_l(1-p_{l+1})^2}{1-p_l}.$$

For $p_l = \Gamma^{-l} \le 1/2$, we can bound this from above with $2\Gamma^{-l}$. Since for each (ℓ,k), the mapping $\mathcal{T} \to \mathcal{T}^-$ is injective, we can bound (B.10) with

$$2\,\Gamma^{-l}\mathrm{e}^{C_y/2\log n}\frac{\sum_{\mathcal{T}\in\mathbb{T}_{lk}^-} W_Y(\mathcal{T})}{\sum_{\mathcal{T}} W_Y(\mathcal{T})} \le 2\,\Gamma^{-l}\mathrm{e}^{C_y/2\log n}, \tag{B.12}$$

where $\mathbb{T}_{lk}^-$ corresponds to trimmed trees inside $\mathbb{T}_{lk}$ whose pre-terminal node (l,k) has been turned into a terminal node. Writing $\bar{d}_{lk} = \min_{x\in\mathrm{I}_{lk}} \widetilde{d}_l(x)$ and $\bar{d} = \min_{0\le l\le L_{max}} \min_{0\le k<2^l} \bar{d}_{lk}$, we can then bound the probability in (B.8) with, for $\Gamma > 2$

$$2\,\mathrm{e}^{C_y/2\log n}\sum_{l=\bar{d}}^{L_{max}} \Gamma^{-l}\sum_{k=0}^{2^l-1} \mathbb{I}[l > \min_{x\in\mathrm{I}_{lk}} \widetilde{d}_l(x)] = \mathrm{e}^{C_y/2\log n}\sum_{l=\bar{d}}^{L_{max}} (\Gamma/2)^{-l} \lesssim \mathrm{e}^{C_y/2\log n - \bar{d}\log(\Gamma/2)}$$

Since $M(\cdot)$ and $\eta(\cdot)$ are bounded away from zero and $t(x) \geq t_1$ (see Assumption 1), for a sufficiently large n we have $\widetilde{d}_l(x) = d_l(x)$ for all $x \in [0,1]$ and

$$\bar{d} \geq \underline{\mathrm{C}} + \frac{1}{3}\log n - \frac{1}{3}\log\log n,$$

where $d^* < \underline{\mathrm{C}} = \min_{0\leq l\leq L_{max}} \min_{0\leq k<2^l} \min_{x\in \mathrm{I}_{lk}} \log C_l(x)$ for some $d^* \in \mathbb{R}$. For a sufficiently large Γ, the right side goes to zero. $\square$

Next, with our Bayesian CART prior we can deploy Lemma 2 of [18] to find that, on the event $\mathcal{A}_n$,

$$\Pi\left[\mathcal{T} : S(f_0; A) \not\subseteq \mathcal{T}_{int} \,|\, Y\right] \to 0 \quad \text{as } n \to \infty. \tag{B.13}$$

where $S(f_0; A)$ was defined in (B.5). We can thus conclude, together with Lemma B.1 above, that $E_{f_0}\Pi(\mathcal{E}^c \,|\, Y) \to 0$.

B.1.2 Controlling the Bias Term

The next step in the proof is to show that the class of trees T in (A.5) are good approximators of locally Hölder functions.

Lemma B.2. *Let f_0 satisfy Assumption 1 and let $\widetilde{d}_l(x)$ be as in (B.2). We define the* local bias *as*

$$f_0^{\backslash d}(x) = \sum_{l\leq L_{max}} \sum_{k=0}^{2^l-1} \mathbb{I}[l > \widetilde{d}_l(x)]\, \psi_{lk}(x)\beta^0_{lk}. \tag{B.14}$$

With $\zeta_n(x) = (n/\log n)^{t(x)/[2t(x)+1]}$, the local bias is uniformly small in the sense that

$$B \equiv \sup_{x\in[0,1]} \left[\zeta_n(x)|f_0^{\backslash d}|\right] \leq \bar{C} \quad \textit{for some } \bar{C} > 0. \tag{B.15}$$

Proof. Using Lemma 1 and assuming $M(x) \leq \bar{M}$ we have for some $C_1 > 0$

$$\begin{aligned} B \equiv \sup_{x\in[0,1]} \left[\zeta_n(x)|f_0^{\backslash d}|\right] &\leq \sup_{x\in[0,1]} \left[\zeta_n(x) \sum_{l\leq L_{max}} \sum_{k=0}^{2^l-1} 2^{l/2}\mathbb{I}[l > \widetilde{d}_l(x)]|\beta^0_{lk}|\right] \\ &\leq 2\bar{M} \sup_{x\in[0,1]} \left[\zeta_n(x) \sum_{l>\widetilde{d}_l(x)} 2^{-lt(x)}\right] \leq 2\bar{M}\, C_1 \sup_{x\in[0,1]} \left[\zeta_n(x) 2^{-\widetilde{d}_l(x)t(x)}\right]. \end{aligned}$$

From the definition of $\widetilde{d}_l(x)$ and $C_l(x)$ in (B.2) and (B.1), we have under Assumption 1 for some $\bar{C} > 0$

$$2^{-\widetilde{d}_l(x)t(x)} \leq (C_l(x))^{-t(x)} \left(\frac{\log n}{n}\right)^{\frac{t(x)}{2t(x)+1}} \leq \frac{\bar{C}}{2\bar{M}\,C_1} \left(\frac{n}{\log n}\right)^{\frac{t(x)}{2t(x)+1}}.$$

B.1.3 The Main Proof

With $\mathcal{E}$ introduced in (B.6), we have shown in Section B.1.1 that $E_{f_0}\Pi(\mathcal{E}^c \mid Y) \to 0$. We can then write, for $\mathcal{A}_n$ introduced in (B.7),

$$E_{f_0}\Pi\left[f : \sup_{x\in[0,1]} \zeta_n(x)|f(x)-f_0(x)| > M_n \,\Big|\, Y\right] \leq P_{f_0}[\mathcal{A}_n^c] + E_{f_0}\Pi[\mathcal{E}^c \mid Y] + E_{f_0}\Pi_{\mathcal{E}}\mathbb{I}_{\mathcal{A}_n}$$

where

$$\Pi_{\mathcal{E}} \equiv \Pi\left[f \in \mathcal{E} : \sup_{x\in[0,1]} \zeta_n(x)|f(x)-f_0(x)| > M_n \,\Big|\, Y\right]. \tag{B.16}$$

Using the Markov's inequality, one can bound the display above with

$$\begin{aligned}\Pi_{\mathcal{E}} &\leq M_n^{-1}\int_{\mathcal{E}} \sup_{x\in[0,1]} \zeta_n(x)|f(x)-f_0(x)| d\Pi(f \mid Y)\\ &\leq M_n^{-1}\int_{\mathcal{E}} \sup_{x\in[0,1]} \zeta_n(x)|f(x)-f_0^d(x)| d\Pi(f \mid Y) + M_n^{-1}B,\end{aligned}$$

where $f_0^d = f_0 - f_0^{\backslash d}$ with $f_0^{\backslash d}$ introduced in (B.14) and where B was defined in (B.15) and was shown to be $\mathcal{O}(1)$ in Lemma B.2. We now focus on the integrand in the last display above. For a function $f \in \mathcal{E}(\mathcal{T})$ supported on $\mathcal{T} \in T$ we have

$$\left|f(x) - f_0^d(x)\right| \leq \sum_{(l,k)\in\mathcal{T}_{int}} \mathbb{I}_{x\in \mathrm{I}_{lk}} 2^{l/2}|\beta_{lk} - \beta_{lk}^0| + \sum_{(l,k)\notin\mathcal{T}_{int}; l\leq \widetilde{d}_l(x)} \mathbb{I}_{x\in \mathrm{I}_{lk}} 2^{l/2}|\beta_{lk}^0|. \tag{B.17}$$

We now focus on the second term above. Since trees $\mathcal{T} \in T$ catch large signals (definition of T in (A.5)), we have $|\beta_{lk}^0| < A\log n/\sqrt{n}$ for $(l,k) \notin \mathcal{T}_{int}$ and thereby

$$\sup_{x\in[0,1]}\left[\zeta_n(x)\sum_{(l,k)\notin\mathcal{T}_{int}; l\leq\widetilde{d}_l(x)} 2^{l/2}\mathbb{I}_{x\in\mathrm{I}_{lk}}|\beta_{lk}^0|\right] \lesssim \frac{\log n}{\sqrt{n}}\sup_{x\in[0,1]}\zeta_n(x)2^{\frac{\widetilde{d}_l(x)}{2}} \lesssim \sqrt{\log n}. \tag{B.18}$$

Above, we have used the fact that (for $M(x) \leq \bar{M}$ and $t_1 \leq t(x) \leq 1$)

$$\zeta_n(x)2^{\widetilde{d}_l(x)/2} \leq (2\bar{M}/\bar{\gamma})^{1/[2t(x)+1]}\left(\frac{n}{\log n}\right)^{1/2} \lesssim \left(\frac{n}{\log n}\right)^{1/2}.$$

Regarding the first term in (B.17), we can write for a given tree $\mathcal{T} \in T$

$$\begin{aligned}A(\mathcal{T}) &\equiv \int \sup_{x\in[0,1]}\left[\zeta_n(x)\sum_{(l,k)\in\mathcal{T}_{int}}\mathbb{I}_{x\in\mathrm{I}_{lk}}2^{l/2}|\beta_{lk}-\beta_{lk}^0|\right] d\,\Pi(\boldsymbol{\beta} \mid \mathcal{T}, Y) \\ &\lesssim \int \max_{(l,k)\in\mathcal{T}_{int}} |\beta_{lk}-\beta_{lk}^0| \sup_{x\in[0,1]}\left[\zeta_n(x)2^{\widetilde{d}_l(x)/2}\right] d\,\Pi(\boldsymbol{\beta} \mid \mathcal{T}, Y)\\ &\lesssim \sqrt{\frac{n}{\log n}}\int \max_{(l,k)\in\mathcal{T}_{int}} |\beta_{lk}-\beta_{lk}^0| d\,\Pi(\boldsymbol{\beta} \mid \mathcal{T}, Y).\end{aligned} \tag{B.19}$$

According to Lemma 3 of [18], the integral above is bounded by $C'\sqrt{\log n/n}$, which implies $A(\mathcal{T}) \leq B_A$ uniformly for all $\mathcal{T} \in T$ for some $B_A > 0$. We now put the pieces together. From the considerations above, we continue the calculations in (B.16) using (B.17) and (B.18) to obtain

$$\begin{aligned}\Pi_{\mathcal{E}} &\leq M_n^{-1} \sum_{\mathcal{T}\in T} \Pi[\mathcal{T} \mid Y] \int_{\mathcal{E}(\mathcal{T})} \sup_{x\in[0,1]} \zeta_n(x) \left| f(x) - f_0^{\setminus d}(x) \right| d\,\Pi(f \mid Y, \mathcal{T}) + o(1) \\ &\leq M_n^{-1} \left[\mathcal{O}(\sqrt{\log n}) + B_A \right] + o(1).\end{aligned}$$

The upper bound goes to zero as long as M_n is strictly faster than $\sqrt{\log n}$.

B.2 Proof of Theorem 2

We follow the strategy of the proof of Theorem 2 in [18]. We first show the set $\mathcal{C}_n$ has an optimal diameter, uniformly over the domain $[0, 1]$.

B.2.1 Optimal Diameter

We will use the following Lemma (a simple modification of Lemma S-11 in [18]).

Lemma B.3. *(Median Tree Estimator) Consider the prior distribution as in Theorem 1 and let $\mathcal{T}_Y^*$ be as in (3.5). Then there exists an event $\mathcal{A}_n^*$ such that $P_{f_0}[\mathcal{A}_n^*] = 1 + o(1)$ as $n \to \infty$ on which the tree $\mathcal{T}_Y^*$ has the following two properties*

(1) With $S(f_0; A)$ defined in (B.5), we have

$$\mathcal{T}_Y^* \supseteq S(f_0; A).$$

(2) With $\widetilde{d}_l(x)$ as in (B.2) and with $\widetilde{\mathcal{T}}_{Y\,int}^*$ denoting the pre-terminal nodes of $\mathcal{T}_{Y\,int}^*$ as defined in (A.4)

$$l \leq \min_{x\in \mathrm{I}_{lk}} \widetilde{d}_l(x) \quad \forall (l, k) \in \widetilde{\mathcal{T}}_{Y\,int}^*.$$

Proof. Recall the notation $L_{max} = \lfloor \log_2 n \rfloor$. We denote with

$$T_1 = \{\mathcal{T} : \mathcal{T}_{int} \supseteq S(f_0; A)\} \quad \text{and} \quad T_2 = \{\mathcal{T} : l \leq \min_{x\in \mathrm{I}_{lk}} \widetilde{d}_l(x) \ \forall (l, k) \in \widetilde{\mathcal{T}}_{int}\}.$$

We define an event $\mathcal{A}_n^* = \{Y : \Pi(T_1 \cap T_2 \mid Y) \geq 3/4\}$. Using (B.13) and (B.8) we know that $P_{f_0}(\mathcal{A}_n^{*c}) = o(1)$ as $n \to \infty$. Then, on the event $\mathcal{A}_n^*$, for any node $(l_1, k_1) \in S(f_0; A)$ we have

$$\Pi\left((l_1, k_1) \in \mathcal{T}_{int} \mid Y\right) \geq \Pi(T_1 \mid Y) \geq 3/4 > 1/2$$

which implies that $(l_1, k_1) \in \mathcal{T}^*_{Y\,int}$. Thereby, on the event $\mathcal{A}^*_n$, we have $\mathcal{T}^*_Y \in T_1$. Similarly, for any (l_1, k_1) such that $l_1 > \min_{x \in \mathrm{I}_{l_1 k_1}} \widetilde{d}_{l_1}(x)$ we have $\Pi((l_1, k_1) \in \mathcal{T}_{int} \,|\, Y) < 1/4 < 1/2$ and thereby $(l_1, k_1) \notin \mathcal{T}^*_Y$ on the event $\mathcal{A}^*_n$. This yields that $\mathcal{T}^*_Y \in T_2$ on the event $\mathcal{A}^*_n$. Since $P_{f_0}(\mathcal{A}^*_n) = 1 + o(1)$, one obtains $P_{f_0}[\{\mathcal{T}^*_Y \notin T_1\} \cup \{\mathcal{T}^*_Y \notin T_2\}] = o(1)$. $\square$

From Lemma B.3 it follows that, for some suitable sequence v_n, that increases at least as fast as $\log n$ (as shown below),

$$\sup_{f,g \in \mathcal{C}_n} \left[\sup_{x \in [0,1]} \frac{\zeta_n(x)}{v_n} |f(x) - g(x)| \right] = \mathcal{O}_{P_{f_0}}(1). \tag{B.20}$$

Indeed, for any $f, g \in \mathcal{C}_n$ we have

$$\begin{aligned} \sup_{x \in [0,1]} \left[\frac{\zeta_n(x)}{v_n} |f(x) - g(x)| \right] &\leq \sup_{x \in [0,1]} \left[\frac{\zeta_n(x)}{v_n} \Big(|f(x) - \widehat{f}_T(x)| + |\widehat{f}_T(x) - g(x)| \Big) \right] \\ &\leq 2 \sup_{x \in [0,1]} \left[\frac{\zeta_n(x) \sigma_n(x)}{v_n} \right]. \end{aligned}$$

where $\sigma_n(x)$ was defined in (3.6). From the properties of the median tree in Lemma B.3, we know that there exists an event $\mathcal{A}^*_n$ such that $P_{f_0}(\mathcal{A}^*_n) = 1 + o(1)$ where the median tree satisfies $2^l \leq 2^{\widetilde{d}_l(x)} \lesssim (n/\log n)^{1/[2t(x)+1]}$ for all $(l, k_l(x)) \in \widetilde{\mathcal{T}}^*_{Y\,int}$. For any $x \in [0,1]$ and we then have

$$\frac{\zeta_n(x)\sigma_n(x)}{v_n} \leq \left(\frac{n}{\log n} \right)^{\frac{t(x)}{2t(x)+1} - \frac{1}{2}} \sum_{l=0}^{L_{max}} 2^{l/2} \mathbb{I}[(l, k_l(x)) \in \mathcal{T}^*_{Y\,int}] \lesssim 2^{\widetilde{d}_l(x)/2} \left(\frac{n}{\log n} \right)^{\frac{t(x)}{2t(x)+1} - \frac{1}{2}}.$$

From the definition of $\widetilde{d}_l(x)$ we conclude that the right-hand-side is $\mathcal{O}(1)$ on the event $\mathcal{A}^*_n$. This concludes the statement (B.20).

B.2.2 Confidence of the set $\mathcal{C}_n$

We first show that the median tree is a (nearly) rate-optimal estimator. Denote with $\widehat{f}^T_{lk} = \langle \widehat{f}_T, \psi_{lk} \rangle$ and recall $\mathcal{S}(f_0; A) = \{(l,k) : \ |\beta^0_{lk}| \geq A \log n / \sqrt{n}\}$. Recall the definition of trees T in (B.4). Let us consider the event

$$B_n = \{\mathcal{T}^*_Y \in T\} \cap \mathcal{A}_n, \tag{B.21}$$

where the noise-event $\mathcal{A}_n$ is defined in (B.7). According to Lemma B.3, we have $P_{f_0}(B_n) = 1 + o(1)$. Using similar arguments as around the inequality (B.17), on the event B_n, we have for some $M > 0$

$$\sup_{x \in [0,1]} \zeta_n(x) |\widehat{f}_T(x) - f_0(x)| \leq M \sqrt{\log n}. \tag{B.22}$$

Next, one needs to show that $\sigma_n(x)$ is appropriately large for each $x \in [0,1]$.

Let $\Lambda_n(x)$ be defined by, for $\mu(x) > 0$ to be chosen below,

$$\frac{\mu(x)}{\log n}\left(\frac{n}{\log n}\right)^{\frac{1}{2t(x)+1}} \leq 2^{\Lambda_n(x)} \leq \frac{2\,\mu(x)}{\log n}\left(\frac{n}{\log n}\right)^{\frac{1}{2t(x)+1}}. \tag{B.23}$$

We will use the following lemma which follows from the proof of Proposition 3 in [**?**]

Lemma B.4. *Assume $f_0 \in \mathcal{C}_{SS}(t(x), x, M(x), \eta(x))$. Then for the sequence $\Lambda_n(x)$ in (B.23) there exists $C > 0$ and $l \geq \Lambda_n(x)$ such that*

$$|\beta^0_{lk_l(x)}| \geq C\,2^{-\Lambda_n(x)[t(x)+1/2]}. \tag{B.24}$$

Proof. From the definition of local self-similarity, we have for some $c_1 > 0$

$$2^{-jt(x)}c_1 \leq |K_j(f_0)(x) - f_0(x)| \leq \sum_{l \geq j} 2^{l/2}|\beta^0_{lk_l(x)}|. \tag{B.25}$$

Now, for all $N \geq 1$ there exists $j \geq \Lambda_n(x)$ such that, using (B.25)

$$\begin{aligned}
|\beta^0_{jk_j(x)}| &\geq \frac{1}{N}\sum_{l=\Lambda_n(x)}^{\Lambda_n(x)+N-1} |\beta^0_{lk_l(x)}| \\
&\geq \frac{2^{-(\Lambda_n(x)+N)/2}}{N}\left(\sum_{l=\Lambda_n(x)}^{\infty} 2^{l/2}|\beta^0_{lk_l(x)}| - \sum_{l=\Lambda_n(x)+N}^{\infty} 2^{l/2}|\beta^0_{lk_l(x)}|\right) \\
&\geq \frac{2^{-(\Lambda_n(x)+N)/2}}{N}\left(2^{-\Lambda_n(x)t(x)}c_1 - c(t(x),N)2^{-(\Lambda_n(x)+N)t(x)}\right) \\
&\geq \frac{2^{-(\Lambda_n(x)+N)/2}}{2N}2^{-\Lambda_n(x)t(x)}c_1 > \underline{c_1}2^{-\Lambda_n(x)[t(x)+1/2]}.
\end{aligned}$$

□

where $\underline{c_1} = 2^{-N/2}c_1/(2N)$ and N is large enough.

Combining (B.23) with (B.24), one can choose $\mu(x)$ such that for each $x \in [0,1]$ there exists $l \geq \Lambda_n(x)$ such that

$$|\beta^0_{lk_l(x)}| > C[2\mu(x)]^{-t(x)-1/2}\sqrt{\frac{\log n}{n}}(\log n)^{t(x)+1/2} \geq A\log n/\sqrt{n}.$$

Since this is a signal node (i.e. $\beta^0_{lk_l(x)} \in S(f_0; A)$), it will be captured by the median tree. One deduces that the term $(l, k_l(x))$ in the sum defining $\sigma_n(x)$ is nonzero on the event B_n, so that

$$\sigma_n(x) \geq v_n\sqrt{\frac{\log n}{n}}|\psi_{lk_l(x)}| \geq v_n\sqrt{\frac{\log n}{n}}2^{\Lambda_n(x)/2} \geq \frac{\sqrt{\mu(x)}v_n}{\sqrt{\log n}}\left(\frac{\log n}{n}\right)^{\frac{t(x)}{2t(x)+1}}. \tag{B.26}$$

For v_n faster than $\log n$ for all $x \in [0,1]$ one has $\sigma_n(x) \geq \sqrt{\log n}/\zeta_n(x)$ and from (B.22) one obtains

$$B_n \subset \left\{ \sup_{x\in[0,1]} \left[\frac{1}{\sigma_n(x)} |\widehat{f}_T(x) - f_0(x)| \right] \leq 1/2 \right\} \tag{B.27}$$

and the desired coverage property, since

$$P_{f_0}(f_0 \in C_n) = P_{f_0}\left(\sup_{x\in[0,1]} \left[\frac{1}{\sigma_n(x)} |\widehat{f}_T(x) - f_0(x)| \right] \leq 1 \right) \geq P_{f_0}(B_n) = 1 + o(1).$$

B.2.3 Credibility of the set $\mathcal{C}_n$

We want to show that

$$\Pi[\mathcal{C}_n \mid Y] = 1 + o_{P_{f_0}}(1).$$

We note that the posterior distribution and the median estimator $\widehat{f}_T$ converge at a rate at $x \in [0,1]$ strictly faster than $\sigma_n(x)$ on the event B_n, using again the lower bound on $\sigma_n(x)$ in (B.26). In particular, because of (B.27) we can write

$$E_{f_0}\left(\Pi\left[\sup_{x\in[0,1]} \frac{1}{\sigma_n(x)} |f(x) - \widehat{f}_T(x)| \leq 1 \,\Big|\, Y \right] \right) \geq \tag{B.28}$$

$$E_{f_0}\left(\Pi\left[\sup_{x\in[0,1]} \frac{1}{\sigma_n(x)} |f(x) - f_0(x)| \leq 1/2 \,\Big|\, Y \right] \mathbb{I}_{B_n} \right) + o(1). \tag{B.29}$$

The right side converges to 1 in P_{f_0}-probability, which concludes the proof of the theorem.

B.3 Proof of Theorem 3

The proof follows the lines of [41], with several refinements to allow for weaker constraints on the inclusion probabilities ω_l's. For some suitable $B > 0$, we define an event (for $L_{max} = \lfloor \log_2 n \rfloor$)

$$\mathcal{A}_{n,B} = \{ |\varepsilon_{lk}| \leq \sqrt{2[\log 2^l + B \log n]} \quad \forall (l,k) \quad \text{such that} \quad l \leq L_{max} \} \tag{B.30}$$

which satisfies $P_{f_0}(\mathcal{A}_{n,B}^c) \leq \frac{2 \log n}{n^B}$. First, we show an auxiliary Lemma which is reminiscent of Lemma 1 in [41]. We define $S(f_0; A) = \{(l,k) : l \leq L_{max} \text{ and } |\beta_{lk}^0| > A\sqrt{\log n/n}\}$.

Lemma B.5. *Under the assumptions of Theorem 3 there exists $a > 0$ determined by $\delta > 0$ defined in (3.9) in Assumption 2 and $A > a$ such that, uniformly over $\mathcal{C}(t, M, \eta)$, we have*

$$E_{f_0} \Pi \left[\mathcal{T} \cap S(f_0; a)^c \neq \emptyset \mid Y \right] = o(1) \quad \textit{as } n \to \infty \tag{B.31}$$

and

$$E_{f_0}\Pi\left[S(f_0;A)\not\subseteq \mathcal{T} \mid Y\right] = o(1) \quad \textit{as } n\to\infty. \tag{B.32}$$

Proof. We first prove (B.31) folowing [41], except that we have a weaker condition on the prior on $\mathcal{T}$. Note that

$$\begin{aligned}\Pi\left[\mathcal{T}\cap S(f_0;a)^c\neq\emptyset \mid Y\right] &= \Pi\left[\exists (l,k)\in\mathcal{T}\cap S(f_0;a)^c \mid Y\right]\\ &\leq \sum_{l\leq L_{max}}\sum_{k=0}^{2^l-1}\mathbb{I}\left[(l,k)\notin S(f_0;a)\right]\Pi\left[(l,k)\in\mathcal{T}\mid Y\right].\end{aligned}$$

We denote by $\mathbb{T}$ the set of all subsets of wavelet coefficients β_{lk} up to the maximal depth $L_{max}=\lfloor\log_2 n\rfloor$. Then

$$\Pi\left[(l,k)\in\mathcal{T}\mid Y\right] = \frac{\sum_{\mathcal{T}\in\mathbb{T}}\mathbb{I}\left[(l,k)\in\mathcal{T}\right]\Pi(\mathcal{T})m_n(\mathcal{T})}{\sum_{\mathcal{T}\in\mathbb{T}}\Pi(\mathcal{T})m_n(\mathcal{T})}$$

where

$$m_n(\mathcal{T}) = \prod_{(l,k)\notin\mathcal{T}} \mathrm{e}^{-\frac{n}{2}Y_{lk}^2}\times\prod_{(l,k)\in\mathcal{T}}\int \mathrm{e}^{-\frac{n}{2}|Y_{lk}-\beta_{lk}|^2}\pi_{lk}(\beta_{lk})d\beta_{lk}.$$

For a set $\mathcal{T}$ such that $(l,k)\in\mathcal{T}$ we denote with $\mathcal{T}^-=\mathcal{T}\backslash\{(l,k)\}$. Due to the fact that the marginal likelihood factorizes, we obtain

$$\Pi\left[(l,k)\in\mathcal{T}\mid Y\right] = \frac{\sum_{\mathcal{T}\in\mathbb{T}}\mathbb{I}\left[(l,k)\in\mathcal{T}\right]R(\mathcal{T},\mathcal{T}^-)\Pi(\mathcal{T}^-)m_n(\mathcal{T}^-)}{\sum_{\mathcal{T}\in\mathbb{T}}\Pi(\mathcal{T})m_n(\mathcal{T})},$$

where, invoking the prior assumption (3.8), we obtain

$$R(\mathcal{T},\mathcal{T}^-) \coloneqq \frac{\Pi(\mathcal{T})m_n(\mathcal{T})}{\Pi(\mathcal{T}^-)m_n(\mathcal{T}^-)} \leq \sqrt{2\pi}n^{-1/2}\,C\times C_T\times w_l\times \mathrm{e}^{\frac{n}{2}Y_{lk}^2}.$$

This yields

$$\Pi\left[(l,k)\in\mathcal{T}\mid Y\right] \leq \sqrt{2\pi}n^{-1/2}\,C\times C_T\times w_l\times \mathrm{e}^{\frac{n}{2}Y_{lk}^2}.$$

On the event $\mathcal{A}_{n,B}$ in (B.30) we have $|\varepsilon_{lk}|\leq\sqrt{2(1+B)\log n}$ and for $(l,k)\notin S(f_0;a)$ we have $|\beta_{lk}|<a\sqrt{\log n/n}$. We use the fact that for any $b\in(0,1)$

$$\frac{n}{2}Y_{lk}^2\leq\frac{1-b}{2}\varepsilon_{lk}^2+\frac{b}{2}\,\varepsilon_{lk}^2+|\varepsilon_{lk}|a\sqrt{\log n}+\frac{a^2}{2}\log n$$

to find that for $\widetilde{C}\equiv\sqrt{2\pi}\times C\times C_T$

$$\Pi\left[\mathcal{T}\cap S(f_0;a)^c\neq\emptyset\mid Y\right]\leq\widetilde{C}\,n^{\frac{a^2-1}{2}+b(1+B)}\sum_{l\leq L_{max}}w_l\sum_{k:(l,k)\notin S(f_0;a)}\mathrm{e}^{\frac{1-b}{2}\varepsilon_{lk}^2+\frac{a}{2}|\varepsilon_{lk}|\sqrt{\log n}}.$$

Since, using the prior assumption (3.9),

$$E_{f_0} \sum_{l \le L_{max}} w_l \sum_{k:(l,k) \notin S(f_0;a)} \mathrm{e}^{\frac{1-b}{2}\varepsilon_{lk}^2 + \frac{a}{2}|\varepsilon_{lk}|\sqrt{\log n}} \le 2 \sum_{l \le L_{max}} 2^l \omega_l \int_0^\infty \mathrm{e}^{-\frac{b}{2}x^2 + \frac{a}{2}x\sqrt{\log n}} \mathrm{d}\, x$$
$$\le \frac{2\sqrt{2\pi} n^{a^2/(2b)}}{\sqrt{b}} \sum_{l \le L_{max}} 2^l \omega_l.$$

This yields

$$E_{f_0} \Pi\left[\mathcal{T} \cap S(f_0;a)^c \neq \emptyset \,|\, Y\right] \le \frac{2\sqrt{2\pi}\widetilde{C}\, n^{\frac{a^2-1}{2} + b(1+B) + a^2/(2b)}}{\sqrt{b}} \sum_{l \le L_{max}} 2^l \omega_l.$$

We can find b and a such that $a^2 + 2b(1+B) + a^2/b < \delta$ and thereby, using the assumption (3.9),

$$E_{f_0} \Pi\left[\mathcal{T} \cap S(f_0;a)^c \neq \emptyset \,|\, Y\right] \lesssim L_{max} n^{-c/2}$$

for $c = \delta - [a^2 + 2b(1+B) + a^2/b] > 0$. This proves the first statement (B.31).

We now prove that there exists $A > 0$ such that on the event $\mathcal{A}_{n,B}$ we have (B.32). We have

$$\Pi\left[S(f_0;A) \not\subseteq \mathcal{T} \,|\, Y\right] \le \sum_{(l,k) \in S(f_0;A)} \Pi[(l,k) \notin \mathcal{T} \,|\, Y]$$

For $\mathcal{T}$ such that $(l,k) \notin \mathcal{T}$, denote with $\mathcal{T}^+ = \mathcal{T} \cup \{(l,k)\}$. Then

$$\Pi\left[(l,k) \notin \mathcal{T} \,|\, Y\right] = \frac{\sum_{\mathcal{T} \in \mathbb{T}} \mathbb{I}\left[(l,k) \notin \mathcal{T}\right] R(\mathcal{T},\mathcal{T}^+) \Pi(\mathcal{T}^+) m_n(\mathcal{T}^+)}{\sum_{\mathcal{T} \in \mathbb{T}} \Pi(\mathcal{T}) m_n(\mathcal{T})},$$

where (choosing $R > C_\beta + \sqrt{2(1+B)\log n/n}$ for a suitably large C_β)

$$R(\mathcal{T},\mathcal{T}^+) := \frac{\Pi(\mathcal{T}) m_n(\mathcal{T})}{\Pi(\mathcal{T}^+) m_n(\mathcal{T}^+)} \le \frac{n^{1/2}}{\sqrt{2\pi}\, c_T\, w_l\, c_R\, c} \mathrm{e}^{-\frac{n}{2} Y_{lk}^2}.$$

Above, we have used the fact that on the event $\mathcal{A}_{n,B}$ we have $|Y_{lk}| \le U \equiv C_\beta + \sqrt{2(1+B)\log n/n} \le R$ and for some $c > 0$

$$\int \mathrm{e}^{-\frac{n}{2}|Y_{lk} - \beta_{lk}|^2} \pi_{lk}(\beta_{lk}) d\beta_{lk} \ge c_R \int_{-R}^{R} \mathrm{e}^{-\frac{n}{2}|Y_{lk} - \beta_{lk}|^2} d\beta_{lk} = c_R \frac{\sqrt{2\pi}}{\sqrt{n}} [\Phi(R; Y_{lk}, 1/n) - \Phi(-R; Y_{lk}, 1/n)]$$
$$\ge c_R \frac{\sqrt{2\pi}}{\sqrt{n}} [\Phi(U; Y_{lk}, 1/n) - \Phi(-U; Y_{lk}, 1/n)] \ge c_R \frac{\sqrt{2\pi}}{\sqrt{n}} (\Phi(2U\sqrt{n}; 0, 1) - 1/2)$$
$$\ge c_R \frac{\sqrt{2\pi}}{\sqrt{n}} c$$

where $\Phi(x;\mu,\sigma)$ is a cdf of a normal distribution with mean μ and variance σ. On the event $\mathcal{A}_{n,B}$ (in (B.30)) we have $|\varepsilon_{lk}| \leq \sqrt{2(1+B)\log n}$ and

$$Y_{lk}^2 = [(\beta_{lk}^0)^2 + \varepsilon_{lk}/\sqrt{n}]^2 \geq (\beta_{lk}^0)^2/2 - 4(1+B)\log n/n > [A^2 - 4(1+B)]\log n/n.$$

This yields (using the prior assumption (3.9))

$$\Pi\left[\mathcal{T} : S(f_0;A) \not\subseteq \mathcal{T} \mid Y\right] \leq \frac{|S(f_0;A)|}{\sqrt{2\pi}\, c_T\, c_R} n^{B_\omega + 1/2 - [A^2 - 4(1+B)]/2} < n^{-A^2/4}$$

for $A^2/4 > 2(1+B) + B_\omega + 1/2$. □

Now, we complete the proof of Theorem 3. We deploy Lemma B.5 to find $A > a > 0$ such that for $S(f_0;b) = \{(l,k) : |\beta_{lk}^0| \geq b\sqrt{\log n/n}\}$ we obtain, on the event $\mathcal{A}_{n,B}$,

$$\Pi\left[\mathcal{T} : \mathcal{T} \subset S(f_0;a) \mid Y\right] \to 1 \quad \text{and} \quad \Pi\left[\mathcal{T} : S(f_0;A) \not\subseteq \mathcal{T} \mid Y\right] \to 0 \quad \text{as } n \to \infty.$$

Similarly as in the proof of Theorem 1, we then define

$$T = \{\mathcal{T} : S(f_0;A) \subset \mathcal{T} \subset S(f_0;a)\},$$

and we denote with $\mathcal{E}(\mathcal{T})$ the set of of functions $f(x) = \sum_{(l,k)\in\mathcal{T}} \psi_{lk}(x)\beta_{lk}$. We also assume that the bound $A > 0$ can be chosen large enough such that

$$\sup_{f_0 \in \mathcal{C}(t,M,\eta)} E_{f_0}\Pi\left[\max_{(l,k)\in S(f_0;A)} |\beta_{lk} - \beta_{lk}^0| > A\sqrt{\log n/n} \,\Big|\, Y\right] \lesssim \frac{\log n}{n^B}$$

This is indeed the case, as shown in the proof of Theorem 3.1 in [41]. From the definition of local Hölder balls we have

$$\{(l,k) : l \leq \min_{x\in \mathrm{I}_{lk}} \widetilde{d}_l(x,A)\} \subset S(f_0;a) \subset \{(l,k) : l \leq \min_{x\in \mathrm{I}_{lk}} \widetilde{d}_l(x,a)\},$$

where $\widetilde{d}_l(x,a)$ is defined as in (B.1) using a instead of $\bar{\gamma}$. We note that for each $f \in \mathcal{E}(\mathcal{T})$ with a coefficient sequence $\{\beta_{lk}\}$ that satisfies

$$\max_{(l,k)\in S(f_0;A)} |\beta_{lk} - \beta_{lk}^0| \leq A\sqrt{\log n/n} \tag{B.33}$$

and we thereby have for $\zeta_n(x) = \left(\frac{n}{\log n}\right)^{\frac{t(x)}{2t(x)+1}}$

$$\begin{aligned}\zeta_n(x)|f(x) - f_0(x)| \leq{}& \zeta_n(x)|f_0(x)^{\setminus d}| \\ &+ \zeta_n(x)\left[\sum_{(l,k)\in\mathcal{T}} \mathbb{I}(x\in \mathrm{I}_{lk})2^{l/2}|\beta_{lk} - \beta_{lk}^0| + \sum_{(l,k)\notin\mathcal{T}} \mathbb{I}(x\in \mathrm{I}_{lk})2^{l/2}|\beta_{lk}^0|\right]\end{aligned} \tag{B.34}$$

where $f_0^d(x) = f_0 - f_0^{\setminus d}$ and where $f_0^{\setminus d}$ is the bias term defined in (B.14) which satisfies $\sup_{x\in[0,1]} \zeta_n(x)|f_0(x)^{\setminus d}| = \mathcal{O}(1)$ according to Lemma B.2. Focusing on the last term in (B.34), we know that for each $\mathcal{T} \in T$, we have $|\beta_{lk}^0| < A\sqrt{\log n/n}$ for $(l,k) \notin \mathcal{T}$ and $l \leq \min_{x\in \mathrm{I}_{lk}} \widetilde{d}_l(x, A)$. Thereby, we obtain

$$\sup_{x\in[0,1]} \zeta_n(x) \sum_{(l,k)\notin\mathcal{T}} \mathbb{I}(x \in \mathrm{I}_{lk}) 2^{l/2} |\beta_{lk}^0| \lesssim \sqrt{\log n/n} \sup_{x\in[0,1]} \zeta_n(x) 2^{\widetilde{d}_l(x,A)/2} = \mathcal{O}(1).$$

Regarding the middle term in (B.34), we use the property (B.33) and the fact that $(l,k) \in \mathcal{T}$ and $x \in \mathrm{I}_{lk}$ implies $l \leq \widetilde{d}_l(x, a)$. Then

$$\sup_{x\in[0,1]} \zeta_n(x) \sum_{(l,k)\in\mathcal{T}} \mathbb{I}(x \in \mathrm{I}_{lk}) 2^{l/2} |\beta_{lk} - \beta_{lk}^0| \lesssim \sqrt{\log n/n} \sup_{x\in[0,1]} \zeta_n(x) 2^{\widetilde{d}_l(x,a)/2} = \mathcal{O}(1).$$

This completes the proof of Theorem 3.

C Proof of Theorem 4

The proof of Theorem 4 is based on Corollary 2.1 and Theorem 2.2 of [60] for the regression case with wavelet priors and the proof is very similar to the proof of Propositions 3.1 and 3.2 of [60] . We first determine $\epsilon_n(\lambda)$ defined by

$$\Pi\left[\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|_2 \leq K\epsilon_n(\lambda) \mid \lambda\right] = \mathrm{e}^{-n\epsilon_n(\lambda)^2} \tag{C.1}$$

for some $K > 0$ and where λ is the unknown hyper-parameter L, τ and α in cases T1, T2 and T3, respectively. We assume that f_0 satisfies (4.2) and determine $\epsilon_n(\lambda)$ and $\epsilon_{n,0} = \inf_\lambda \epsilon_n(\lambda)$ in all 3 cases (T1)-(T3).

Case T1. The main difference with Lemma 3.1 of [60] (further referred to as RS17) is that the parameter space is different. We denote with $\boldsymbol{\beta}_L = (\beta_{lk} : l \leq L, k \in I_l)'$ for $I_l = \{0, 1, \ldots, 2^l - 1\}$ where $2^l = |I_l|$. Since $\beta_{lk} = 0$ for $l > L$, we can write $\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|_2^2 = \|\boldsymbol{\beta}_L - \boldsymbol{\beta}_L^0\|_2^2 + \sum_{l>L}\sum_{k=0}^{2^l-1}(\beta_{lk}^0)^2$. For $s_n^2 = K^2\epsilon_n(L)^2 - \sum_{l>L}\sum_{k=0}^{2^l-1}(\beta_{lk}^0)^2$ we use the same arguments as in Lemma 3.1 of RS17 to conclude that

$$\Pi\left(\|\boldsymbol{\beta}_L - \boldsymbol{\beta}_L^0\|_2 \leq s_n \mid L\right) \asymp \mathrm{e}^{2^L \log(s_n\, 2^{-L/2})(1+o(1))}$$

as in the proof of Lemma 3.1 of [60]. We then obtain a variant of equation (A1) in RS17

$$s_n^2 + \sum_{l>L}\sum_{k=0}^{2^l-1}(\beta_{lk}^0)^2 = \frac{K^2 2^L}{n} \log\left(\frac{2^{L/2}}{s_n}\right)(1 + o(1)).$$

In addition,

$$\sum_{l>L}\sum_{k=0}^{2^l-1}(\beta_{lk}^0)^2 \leq \frac{M_1^2 2^{-2\alpha_1 L}}{2(1-2^{-2\alpha_1})}(1+o(1))$$

so that

$$\epsilon_{n,0} \lesssim (n/\log n)^{-\alpha_1/(2\alpha_1+1)} \quad \text{and} \quad \epsilon_n(L)^2 \asymp \sum_{l>L}\sum_k \beta_{lk}^{02} + \frac{2^L \log n}{n}. \tag{C.2}$$

For all β_0 satisfying also (4.5), we obtain that

$$\epsilon_{n,0} \asymp (n/\log n)^{-\alpha_1/(2\alpha_1+1)} \quad \text{and} \quad \epsilon_n(L)^2 \gtrsim M_1^2 2^{-2\alpha_1 L} + \frac{2^L \log n}{n}.$$

Case T2 and T3. Similarly as in the proof of Lemma 3.2 in [60], it can be seen that (see equation (3.4) in RS17 or Theorem 4 of [**?**])

$$-\log \Pi(\|\boldsymbol{\beta}\|_2 \leq K\epsilon \mid \alpha, \tau) \asymp (K\epsilon/\tau)^{-1/\alpha}.$$

Note that this equivalence is valid for the non-truncated prior and remains valid under the priors defined in T2 and T3 for any positive α.[4] Similarly as in the proof of Lemma 3.2 [60], we bound from above

$$\inf_{h\in\mathbb{H}^{\alpha,\tau};\|h-\boldsymbol{\beta}_0\|_2\leq\epsilon} \|h\|^2_{\mathbb{H}^{\alpha,\tau}}$$

by $\|\boldsymbol{\beta}_0\|^2_{H^{\alpha,\tau}}$ if $\alpha_1 > \alpha + 1/2$ where, under (4.2),

$$\|\boldsymbol{\beta}_0\|^2_{H^{\alpha,\tau}} = \tau^{-2}\sum_{l,k} 2^{(2\alpha+1)l}\beta_{lk}^{02} \leq \frac{M_1}{\tau^2}\sum_l 2^{(2\alpha+1-2\alpha_1)l} \lesssim \frac{M_1}{\tau^2[2\alpha_1-1-2\alpha]}$$

and by

$$\|\boldsymbol{\beta}_0\|^2_{H^{\alpha,\tau}} = \tau^{-2}\sum_l^{L_\epsilon}\sum_k 2^{(2\alpha+1)l}\beta_{lk}^{02} \lesssim \frac{M_1 2^{(2\alpha+1-2\alpha_1)L_\epsilon}}{\tau^2[2\alpha+1-2\alpha_1]}, \quad M_1 2^{-2\alpha_1 L_\epsilon} = \epsilon^2$$

if $\alpha_1 < \alpha + 1/2$ or by

$$\|\boldsymbol{\beta}_0\|^2_{H^{\alpha,\tau}} \lesssim \frac{M_1 L_\epsilon}{\tau^2}, \quad \text{if} \quad \alpha_1 = \alpha + 1/2.$$

We thus obtain that if $\alpha_1 \neq \alpha + 1/2$ then

$$\epsilon_n(\alpha,\tau) \lesssim n^{-\alpha/(2\alpha+1)}\tau^{1/(2\alpha+1)} + \left(\frac{1}{n\tau^2(\alpha_1-\alpha-1/2)}\right)^{\frac{\alpha_1}{2\alpha+1}\wedge\frac{1}{2}}$$

[4] The case α close to 0 is of no importance here since the associated $\epsilon_n(\lambda)$ is much bigger than $\epsilon_{n,0}$

while if $\alpha_1 = \alpha + 1/2$

$$\epsilon_n(\alpha,\tau) \lesssim n^{-\alpha/(2\alpha+1)}\tau^{1/(2\alpha+1)} + \left(\frac{\log(n\tau^2)}{n\tau^2}\right)^{\frac{1}{2}}.$$

Note that if f_0 also follows (4.5), we can bound from below (similarly to [60]) for all $h \in \mathbb{H}^{\alpha,\tau}$; $\|h - \boldsymbol{\beta}_0\|_2 \le \epsilon_n(\alpha,\tau)$ when $\alpha_1 < \alpha + 1/2$, $L_{max} > 1$,

$$\begin{aligned}
\|h\|^2_{\mathbb{H}^{\alpha,\tau}} &\ge \tau^{-2} \sum_{l \le L_{max}} \sum_{k \in I_{l1}} 2^{(2\alpha+1)l}[\beta^{02}_{lk} - 2|\beta^0_{lk}||\beta^0_{lk} - h_{lk}|] \\
&\ge \frac{m_1}{2\tau^2(2\alpha+1-2\alpha_1)} 2^{(2\alpha+1-2\alpha_1)L_{max}} - \frac{2M_1}{\tau^2} 2^{(2\alpha-\alpha_1+1/2)L_{max}} \sum_{l \le L_{max}} \sum_{k \in I_{l1}} |\beta^0_{lk} - h_{lk}| \\
&\ge \frac{m_1}{2\tau^2(2\alpha+1-2\alpha_1)} 2^{(2\alpha+1-2\alpha_1)L_{max}} - C 2^{(2\alpha-\alpha_1+1)L_{max}} \epsilon_n(\alpha,\tau) \\
&= \frac{m_1}{2\tau^2(2\alpha+1-2\alpha_1)} 2^{(2\alpha+1-2\alpha_1)L_{max}} \left(1 - \frac{C 2^{\alpha_1 L_{max}}}{m_1} \epsilon_n(\alpha,\tau)\right)
\end{aligned}$$

for some $C > 0$ and choosing L_{max} equal to $L_{max} = \left\lfloor \frac{\log\left(\frac{m_1}{2C\epsilon_n(\alpha,\tau)}\right)}{\alpha_1 \log 2} \right\rfloor$, we bound

$$\|h\|^2_{\mathbb{H}^{\alpha,\tau}} \gtrsim \epsilon_n(\alpha,\tau)^{-(2\alpha+1-2\alpha_1)/\alpha_1}$$

which leads to

$$\epsilon_n(\alpha,\tau) \gtrsim n^{-\alpha/(2\alpha+1)}\tau^{1/(2\alpha+1)} + \left(\frac{1}{n\tau^2 \mid \alpha_1 - \alpha - 1/2 \mid}\right)^{\frac{\alpha_1}{2\alpha+1}}$$

while if $\alpha_1 = \alpha + 1/2$

$$\epsilon_n(\alpha,\tau) \gtrsim n^{-\alpha/(2\alpha+1)}\tau^{1/(2\alpha+1)} + \left(\frac{\log(n\tau^2)}{n\tau^2}\right)^{\frac{\alpha_1}{2\alpha+1} \wedge \frac{1}{2}}$$

and if $\alpha_1 > \alpha + 1/2$, since $\|\boldsymbol{\beta}_0\|_2 \ge c > 0$ for some fixed c,

$$\epsilon_n(\alpha,\tau) \gtrsim n^{-\alpha/(2\alpha+1)}\tau^{1/(2\alpha+1)} + \left(\frac{\|\boldsymbol{\beta}_0\|_2}{n\tau^2(\alpha_1 - \alpha - 1/2)}\right)^{1/2}.$$

The lower bounds thus match the previous upper bounds. Minimizing in α in the case T3 these upper and lower bounds lead to choosing $\alpha = \alpha_1$ and $\epsilon_{n,0} \asymp n^{-\alpha_1/(2\alpha_1+1)}$ while minimizing in τ (Case T2) with $\alpha_1 < \alpha + 1/2$, the minimum is obtained by considering $\tau \asymp n^{-(\alpha_1-\alpha)/(2\alpha_1+1)}$ and $\epsilon_{n,0} \asymp n^{-\alpha_1/(2\alpha_1+1)}$ while if $\alpha_1 \ge \alpha + 1/2$ it is obtained with $\tau \asymp n^{-1/(4\alpha+4)}$ leading to $\epsilon_{n,0} \asymp n^{-(2\alpha+1)/(4\alpha+4)}$ up to a $\log n$ term.

Having quantified $\epsilon_{n,0}$ for the three cases, the upper bound (4.4) follows directly from Theorem 2.3 of RS17.

Regarding the lower bound, we let $\Lambda_0 = \{\lambda : \epsilon_n(\lambda) \le M_n \epsilon_{n,0}\}$ with M_n going to infinity and λ either L, τ or α in cases T1, T2 and T3, respectively. Then following from the proofs of Propositions 3.1 and 3.2 of [60] and since the priors on λ satisfy condition H1 (using Lemma 3.5 and 3.6 of [60]) one obtains

$$\Pi(\lambda \in \Lambda_0 \mid Y) = 1 + o_{P_{f_0}}(1).$$

From this and the remark that for all $\boldsymbol{\beta} = \{\beta_{lk}\}$ such that $\beta_{lk} = 0$ for $l \ge L_{max} = \lfloor \log_2 n \rfloor$ we have for some $C_1 > 0$

$$\begin{aligned}\|f_{\beta_0} - f_\beta\|_n &\le \|f_{\beta_0, L_{max}} - f_\beta\|_n + C_1\, M_1 n^{-\alpha_1} = \|f_{\beta_0, L_{max}} - f_\beta\|_2 + C_1\, M_1 n^{-\alpha_1} \\ &\ge \|f_{\beta_0, L_{max}} - f_\beta\|_n - C_1\, M_1 n^{-\alpha_1} = \|f_{\beta_0, L_{max}} - f_\beta\|_2 - C_1\, M_1 n^{-\alpha_1}.\end{aligned}$$

Together with the fact $\|f_{\beta_0, L_{max}} - f_\beta\|_2 = \|\boldsymbol{\beta}^0_{L_{max}} - \boldsymbol{\beta}\|_2$ we obtain for some $\delta > 0$

$$B_n = \left\{ f : \|f - f_0\|_{1/2,1} \le n^{-\delta} \epsilon_n(\alpha_1) \right\}$$

the following (with $l_n(\boldsymbol{\beta}) = \log f(Y \mid \boldsymbol{\beta})$ and $m_n(\lambda) = \int_{\boldsymbol{\beta}} \mathrm{e}^{l_n(\boldsymbol{\beta}) - l_n(\boldsymbol{\beta}_0)} d\,\Pi(\boldsymbol{\beta} \mid \lambda)$)

$$\begin{aligned}\Pi(B_n \mid Y) &= \Pi\left(B_n \cap \{\lambda \in \Lambda_0\} \mid Y\right) + o_{P_{f_0}}(1) \\ &\le \frac{\int_{\lambda \in \Lambda_0} \int_{B_n} \mathrm{e}^{l_n(\beta) - l_n(\beta_0)} d\,\Pi(\boldsymbol{\beta} \mid \lambda)\, d\Pi(\lambda)}{\int_{\lambda \in \Lambda_0} m_n(\lambda) d\,\Pi(\lambda)}.\end{aligned}$$

The Case T1. We have $\lambda = L$ and set $L_{n,1}$ such that $L_{n,1} = \lfloor \log(L_0 (n/\log n)^{1/(2\alpha_1+1)})/\log 2 \rfloor$ for some suitable $L_0 > 0$. Then $\Pi(L \in \Lambda_0 \mid Y) = 1 + o_{P_{f_0}}(1)$ and and for all $L \in \Lambda_0$ under (4.2) and (4.5) we have

$$2^{L_{n,1}} M_n^{-2/\alpha_1} \lesssim 2^L \lesssim 2^{L_{n,1}} M_n^2.$$

Moreover

$$\Pi(B_n \mid Y) = \sum_{L \in \Lambda_0} \Pi(B_n \mid Y, L) \Pi(L \mid Y) + o_{P_{f_0}}(1).$$

It will be useful to rewrite (4.1) in a vector notation. For any $L \ge 1$, we denote with $\boldsymbol{\beta}_L$ a vector with coordinates $\beta_j = \beta_{lk}$ for $j = 2^{l-1} + k - 1$. If $L \le L_{max}$ and $\boldsymbol{\beta}$ is according to the *model* L, i.e. $\beta_{lk} = 0$ for all $l > L$, the log-likelihood at $\boldsymbol{\beta}$ (conditionally on the model L) can be written as $\ell_n(\boldsymbol{\beta}) = \ell_n(\hat{\boldsymbol{\beta}}_L) - \frac{(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_L)^t \Psi_L^t \Psi_L (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_L)}{2\sigma^2} + C$, where

$$\Psi_L(i,j) = \psi_{\ell k}(x_i) \quad \text{for} \quad j = 2^{\ell-1} + k - 1, \quad i \le n, \quad \text{and for} \quad \hat{\boldsymbol{\beta}}_L = \frac{\Psi_L^t Y}{n},$$

since $\Psi_L^t \Psi_L = nI$ so that $\ell_n(\boldsymbol{\beta}) = -\frac{n\|\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}_L\|^2}{2\sigma^2} + C'$. This implies, together with the Gaussian prior on the β_{lk} when $l \leq L$, that for all $L \leq L_{max}$ the conditional posterior given L is Gaussian with a mean $\tilde{\boldsymbol{\beta}}_L = \frac{n\hat{\boldsymbol{\beta}}_L}{1+n}$ and a variance $\sigma^2 I/(n+1)$. In the following, we write $\tilde{\theta}_{L,2}$ as the subvector of $\tilde{\boldsymbol{\beta}}_L$ whose coordinates correspond to $j = 2^{l-1} + k - 1$ with $k \in I_{l2}$ and $\mathcal{N}(\tilde{\theta}_{L,2}, \sigma^2 I_2/(n+1))$ as a Gaussian vector with a mean $\tilde{\theta}_{L,2}$ and a covariance matrix $\sigma^2 I_2/(n+1)$, where I_2 is the identity matrix of dimension $|I_2| = \sum_{l\leq L} I_{l2} \asymp c2^L$ for some $c > 0$ and $L \in \Lambda_0$. We have

$$\begin{aligned}\Pi(B_n \mid Y, L) &\leq P(\|\mathcal{N}(\tilde{\theta}_{L,2}, \sigma^2 I_2/(n+1))\| \leq n^{-\delta}\epsilon_{n,0}) \\ &\leq P(\|\mathcal{N}(0, I_2)\|^2 \leq (n+1)\frac{n^{-2\delta}\epsilon_{n,0}^2}{\sigma^2}) \\ &= P\left(\mathcal{X}^2(|I_2|) \leq (n+1)\frac{n^{-2\delta}\epsilon_{n,0}^2}{\sigma^2}\right) \lesssim \mathrm{e}^{-c'|I_2|}\end{aligned}$$

for some $c' > 0$, since

$$(n+1)\frac{n^{-2\delta}\epsilon_{n,0}^2}{\sigma^2} \lesssim n^{-2\delta}(\log n)^q n^{1/(2\alpha_1+1)} \lesssim n^{-\delta}|I_2| \quad \text{for some } q > 0$$

and

$$\Pi(B_n \mid Y) = o_p(1).$$

Note that the same holds true if the prior is not Gaussian and if $\alpha_1 > 1/2$.

The cases T2 and T3 We write $(\alpha, \tau) \in \Lambda_0$ to denote $\alpha \in \Lambda_0$ in the case T2 and $\tau \in \Lambda_0$ in the case T3. We have $n^{-\alpha_1} = o(n^{-\delta}\epsilon_n(\alpha_1))$ as soon as $\delta < 2\alpha_1^2/(2\alpha_1+1)$. Moreover, given λ the conditional prior probability

$$\begin{aligned}\Pi(\|\boldsymbol{\beta} - \boldsymbol{\beta}_{0L}\|_2 \leq 2n^{-\delta}\epsilon_n(\alpha_1)|\alpha, \tau) &\leq \Pi(\|\boldsymbol{\beta}\|_2 \leq 2n^{-\delta}\epsilon_n(\alpha_1)|\alpha, \tau) \\ &\leq C' e^{-C(n^{-\delta}\epsilon_n(\alpha_1)/\tau)^{-1/\alpha}},\end{aligned}$$

for some $C, C' > 0$. We also have using the notations $\lambda_n = \alpha_1$ in case T3, $\lambda_n = n^{-(\alpha_1-\alpha)/(2\alpha_1+1)}$ in case T2 with $\alpha_1 < \alpha + 1/2$ and $\lambda_n = n^{-1/(4\alpha+4)}$ in case T2 with $\alpha_1 \geq \alpha + 1/2$. Set

$$D_n = \int_\Lambda m_n(\lambda)\pi(\lambda)d\lambda$$

then

$$D_n \geq \int_{\lambda_n(1-1/n)}^{\lambda_n(1+1/n)} m_n(\lambda)d\Pi(\lambda)$$

and using

$$\Pi([\lambda_n(1-1/n), \lambda_n(1+1/n)]) \geq e^{-n\epsilon_n(\alpha_1)^2/2},$$

we obtain that for some $\tilde{M}_1 > 0$,

$$
\begin{aligned}
&P_{f_0}\left(D_n \leq e^{-\tilde{M}_1 n\epsilon_n(\alpha_1)^2}\right) \\
&\quad \leq \frac{2\int_{\lambda_n(1-1/n)}^{\lambda_n(1+1/n)} \int_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|\leq\epsilon_n(\lambda)} P_0(l_n(\boldsymbol{\beta}) - l_n(\boldsymbol{\beta}_0) \leq -\tilde{M}_1 n\epsilon_n(\alpha_1)^2/4) d\Pi(\boldsymbol{\beta}\mid\lambda)d\Pi(\lambda)}{\int_{\lambda_n(1-1/n)}^{\lambda_n(1+1/n)} \Pi(\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\| \leq \epsilon_n(\lambda)\mid\lambda)d\Pi(\lambda)} \\
&\quad \lesssim \frac{1}{n\epsilon_n(\alpha_1)^2}.
\end{aligned}
$$

This leads to

$$
\begin{aligned}
P_{f_0}\left(\Pi(B_n|Y^n) > \epsilon\right) \leq\ & P_{f_0}\left(D_n < e^{-\tilde{M}_1 n\epsilon_n(\alpha_1)^2}\right) \\
&+ \frac{e^{\tilde{M}_1 n\epsilon_n(\alpha_1)^2}}{\epsilon}\int_{\Lambda_0} \Pi(\|\boldsymbol{\beta}-\boldsymbol{\beta}_{0L}\|_2 \leq 2n^{-\delta}\epsilon_n(\alpha_1)|\lambda)\pi(\lambda)d\lambda \\
\leq\ & o(1) + C'\frac{e^{\tilde{M}_1 n\epsilon_n(\alpha_1)^2}}{\epsilon}\sup_{\lambda\in\Lambda_0} e^{-C(n^{-\delta}\epsilon_n(\alpha_1)/\tau)^{-1/\alpha}}.
\end{aligned}
$$

Moreover for all $\lambda \in \Lambda_0$, $(\epsilon_n(\alpha_1)/\tau)^{-1/\alpha} \gtrsim n\epsilon_n^2(\alpha_1)(\log n)^q$ for some $q \in \mathbb{R}$, therefore

$$
P_{f_0}\left(\Pi(B_n|Y^n) > \epsilon\right) = o(1).
$$

This concludes the proof of Theorem 4.

D Intermediate Results for Theorem 5

We first describes some properties of the Gram matrix induced by irregular designs $\mathcal{X} = \{x_i \in [0,1] : 1 \leq i \leq n\}$. Note that Lemma F.1 implies that, under the balancing Assumption 4, we have for the j^{th} column X_j of X with $j = 2^l + k$

$$
\|X_j\|_2^2 = 2^l n_{lk} \leq 2n\,(C+l) \quad \text{and} \quad \|X_j\|_1 = 2^{l/2} n_{lk} \leq \frac{2n\,(C+l)}{2^{l/2}}. \tag{D.1}
$$

and for $i = 2^{l_2} + k_2$

$$
|X_j' X_i| \leq C_d\sqrt{n}\log^v n\, 2^{\frac{l}{2}}\mathbb{I}\{(l_2,k_2)\,\text{is a descendant of}\,(l,k)\}. \tag{D.2}
$$

Recall the notation of pre-terminal nodes $\widetilde{\mathcal{T}}_{int}$ in (A.4) and let $\mathcal{X} = \{x_i : 1 \leq i \leq n\}$. We will also be denoting with $\lambda_{min}(A)$ and $\lambda_{max}(A)$ the minimal and maximal eigenvalues of a matrix A. The idea behind the proof is similar to the one of Theorem 1. We will be using the same definition of $\widetilde{d}_l(x)$ in (B.2), T in (A.5), $\mathcal{E}$ in (A.6) and $S(f_0; A; \upsilon)$ in (D.14). First, we show that $E_{f_0}\Pi(\mathcal{E}^c \mid Y) \to 0$.

To this end, in Section D.1 we show that the posterior concentrates on locally small trees and in Section D.2 we show that the posterior trees catch signal nodes. These results will be conditional on the set $\mathcal{A}$ in (D.4). The complement of this set has a vanishing probability $P_{f_0}(\mathcal{A}^c) \leq 2/p \to 0$ where $p = 2^{L_{max}} = \lfloor C^* \sqrt{n/\log n} \rfloor$ for some suitable $C^* > 0$.

D.1 Posterior Concentrates on Locally Small Trees

We now show that

$$\Pi\left[\mathcal{T} : \exists (l,k) \in \widetilde{\mathcal{T}}_{int} \;\; s.t. \;\; l > \min_{x \in \mathrm{I}_{lk} \cap \mathcal{X}} \widetilde{d}_l(x) \mid Y\right] \to 0. \tag{D.3}$$

on the set

$$\mathcal{A} = \{\boldsymbol{\varepsilon} : \|X'\boldsymbol{\varepsilon}\|_\infty \leq 2\|X\|\sqrt{\log p}\}, \tag{D.4}$$

where $\|X\| = \max_{1 \leq j \leq p} \|X_j\|_2$.

To prove this statement, we follow the route of Lemma B.1 for the white noise model. The irregular design requires non-trivial modifications of the proof due to the induced correlation among predictors. Similarly as in the proof of Lemma B.1, we denote with $\mathcal{T}^-$ the sub-tree of $\mathcal{T}$ obtained by deleting a deep node (l_1, k_1) which corresponds to the column X_j where $j = 2^{l_1} + k_1$ and which satisfies $l_1 \geq \widetilde{d}_l(x)$ (as in (B.2)) for some $x \in \mathrm{I}_{l_1 k_1}$ and thereby $|\beta_{l_1 k_1}| \lesssim \sqrt{\log n/n}$. Then we have

$$\frac{N_Y(\mathcal{T})}{N_Y(\mathcal{T}^-)} = \frac{1}{\sqrt{1+g_n}} \exp\left\{\frac{1}{2} Y'[X_{\mathcal{T}}\Sigma_{\mathcal{T}}X'_{\mathcal{T}} - X_{\mathcal{T}^-}\Sigma_{\mathcal{T}^-}X'_{\mathcal{T}^-}]Y\right\}. \tag{D.5}$$

Using Lemma F.2, we simplify the exponent in (D.5) to find for $c_n = g_n/(g_n+1)$

$$\frac{N_Y(\mathcal{T})}{N_Y(\mathcal{T}^-)} = \frac{1}{\sqrt{1+g_n}} \exp\left\{\frac{c_n |X'_j(I - P_{\mathcal{T}^-})Y|^2}{2Z}\right\}.$$

First, we bound the term

$$\begin{aligned}
|X'_j(I - P_{\mathcal{T}^-})Y|^2 =& |X'_j(I - P_{\mathcal{T}^-})(X_j\beta^0_j + X_{\backslash\mathcal{T}}\boldsymbol{\beta}^0_{\backslash\mathcal{T}} + \boldsymbol{\nu})|^2 \\
\leq& 2|X'_j(I - P_{\mathcal{T}^-})X_j|^2|\beta^0_j|^2 && \text{(D.6)}\\
&+ 4|X'_j(X_{\backslash\mathcal{T}}\boldsymbol{\beta}^0_{\backslash\mathcal{T}} + \boldsymbol{\nu})|^2 && \text{(D.7)}\\
&+ 4|X'_jP_{\mathcal{T}^-}(X_{\backslash\mathcal{T}}\boldsymbol{\beta}^0_{\backslash\mathcal{T}} + \boldsymbol{\nu})|^2. && \text{(D.8)}
\end{aligned}$$

Using the design assumption (D.1), the first term satisfies (since $\lambda_{max}(I - P_{\mathcal{T}^-}) = 1$)

$$\frac{|X_j'(I - P_{\mathcal{T}^-})X_j|^2|\beta_j^0|^2}{Z} = Z|\beta_j^0|^2 \leq \|X_j\|_2^2 \log n/n \lesssim \log^2 n,$$

where we used the fact that (l_1, k_1) is deep and thereby $|\beta_j^0| \lesssim \sqrt{\log n/n}$. Next

Using (D.2), the Hölder condition and the assumption $t_1 > 1/2$, we obtain

$$\begin{aligned} |X_j' X_{\backslash\mathcal{T}} \boldsymbol{\beta}_{\backslash\mathcal{T}}^0| &\leq C_d \sqrt{n} \log^{\upsilon} n \sum_{(l_2,k_2)} \mathbb{I}[(l_2, k_2) \text{ is a descendant of } (l_1, k_1)] 2^{l_1/2} 2^{-l_2(t_1+1/2)} \\ &\leq C_d \sqrt{n} \log^{\upsilon} n \sum_{l_2=l_1+1}^{L_{max}} 2^{l_2-l_1}\, 2^{l_1/2}\, 2^{-l_2(t_1+1/2)} \\ &\lesssim \frac{C_d \sqrt{n} \log^{\upsilon} n}{2^{l_1/2}} \lesssim \sqrt{n} \log^{\upsilon} n. \end{aligned} \tag{D.9}$$

Regarding the second term in (D.7), on the event $\mathcal{A}$, we have from the Lemma F.3

$$|X_j' \boldsymbol{\nu}| \leq |X_j'(F_0 - X\boldsymbol{\beta}_0 + \boldsymbol{\varepsilon})| \lesssim \sqrt{n} \log^{1\vee\upsilon} n. \tag{D.10}$$

We again split the term (D.8) into two and upper-bound each summand separately. Using the fact (from Lemma F.4) that $(X_{\mathcal{T}}' X_{\mathcal{T}})^{-1}$ is positive definite for any $\mathcal{T}$ and thereby $|\, \boldsymbol{u}'(X_{\mathcal{T}}' X_{\mathcal{T}})^{-1} \boldsymbol{v}\,| \leq \lambda_{max}((X_{\mathcal{T}}' X_{\mathcal{T}})^{-1}) \times |\boldsymbol{u}'\boldsymbol{v}|$ for any $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^{|\mathcal{T}_{int}|}$, we have

$$|\, X_j' P_{\mathcal{T}^-} \boldsymbol{\nu}\,| \leq \frac{1}{\lambda_{min}(X_{\mathcal{T}^-}' X_{\mathcal{T}^-})} \,|\, (X_{\mathcal{T}^-}' X_j)'(X_{\mathcal{T}^-}' \boldsymbol{\nu})\,|\,.$$

Note that the matrix $X_{\mathcal{T}^-}$ has $|\mathcal{T}_{int}^-|$ columns, one for each active wavelet coefficient. Using Lemma F.1, we know that the $|\mathcal{T}_{int}^-| \times 1$ vector $(X_{\mathcal{T}^-}' X_j)$ has only l_1 nonzero entries due to orthogonality of (l_1, k_1) to non-ancestors. In other words, there is one ancestor for each layer in $\mathcal{T}^-$ that is not orthogonal to (l_1, k_1). Using (D.2), we thus find that

$$\|X_{\mathcal{T}^-}' X_j\|_1 \leq C_d \sqrt{n} \log^{\upsilon} n \sum_{l=0}^{l_1-1} 2^{l/2} \leq \frac{3C_d\sqrt{n}}{4} 2^{l_1/2} \log^{\upsilon} n.$$

Under our design Assumption 4 and using Lemma F.3, we then also find that for each column X_m of $X_{\mathcal{T}^-}$ we have $|X_m' \boldsymbol{\nu}| \lesssim \sqrt{n} \log^{1\vee\upsilon} n$. Altogether, using Lemma F.4, we conclude

$$|\, X_j' P_{\mathcal{T}^-} \boldsymbol{\nu}\,| \lesssim \frac{2^{l_1/2} \log^{\upsilon} n}{\sqrt{n}} \times \sqrt{n} \log^{1\vee\upsilon} n \lesssim \sqrt{n} \log^{\upsilon+1\vee\upsilon} n. \tag{D.11}$$

Similarly, using the fact that the only nonzero entries of the vector $X'_{\mathcal{T}^-}X_j$ correspond to the l_1 ancestors of (l_1, k_1) inside the tree $\mathcal{T}^-$ and using (D.9) we obtain

$$
\begin{aligned}
|\, X_j' P_{\mathcal{T}^-} X_{\backslash \mathcal{T}} \boldsymbol{\beta}^0_{\backslash \mathcal{T}} \,| \;&\leq \frac{1}{\lambda_{min}(X'_{\mathcal{T}^-}X_{\mathcal{T}^-})} \,|\, (X'_{\mathcal{T}^-}X_j)'(X'_{\mathcal{T}^-}X_{\backslash \mathcal{T}}\boldsymbol{\beta}^0_{\backslash \mathcal{T}}) \,| \\
&\leq \frac{C_d\sqrt{n}}{\underaccent{\sim}{\lambda}\, n} \log^{\upsilon} n \sum_{l=0}^{l_1-1} 2^{l/2} \frac{C_d \sqrt{n} \log^{\upsilon} n}{2^{l/2}} \lesssim \log^{2\upsilon+1} n.
\end{aligned}
$$

This completes the bound for the term in (D.8).

Now, we find a lower bound for $Z = X_j'(I - P_{\mathcal{T}^-})X_j$. From the proof of Lemma F.2, we can see that $1/Z$ is a 'submatrix' of $(X'_{\mathcal{T}}X_{\mathcal{T}})^{-1}$. The eigenvalue of this 'submatrix' will be smaller than the maximal eigenvalue of the entire matrix $(X'_{\mathcal{T}}X_{\mathcal{T}})^{-1}$ (from the interlacing eigenvalue theorem [?]) and thereby

$$1/Z \leq \lambda_{\max}(X'_{\mathcal{T}}X_{\mathcal{T}})^{-1} = 1/\lambda_{min}(X'_{\mathcal{T}}X_{\mathcal{T}}).$$

From Lemma F.4 we have

$$\lambda_{min}(X'_{\mathcal{T}}X_{\mathcal{T}}) \geq \underaccent{\sim}{\lambda}\, n \quad \text{for some} \quad \underaccent{\sim}{\lambda} > 0. \tag{D.12}$$

From $Z \geq \underaccent{\sim}{\lambda}\, n$ we then obtain for some suitable $C > 0$

$$\frac{N_Y(\mathcal{T})}{N_Y(\mathcal{T}^-)} \leq \exp\left(C \log^{2[\upsilon + (1\vee \upsilon)]} n \right).$$

We can now continue as in the proof of Theorem 1 by plugging-in the likelihood ratio above in the expression (B.12). Earlier in the proof of Lemma B.1, the likelihood ratio was of the order $\mathrm{e}^{C\log n}$. Here, we have a larger logarithmic factor which can be taken care off by choosing $p_l = (\Gamma)^{-l^{2[\upsilon+(1\vee\upsilon)]}}$ as the split probability. We then conclude (D.3) using the same strategy as in the proof of Lemma B.1 for the white noise.

D.2 Catching Signal

We now show that, on the event $\mathcal{A}$ in (D.4),

$$\Pi\left[\mathcal{T} : S(f_0; A; \upsilon) \not\subseteq \mathcal{T} \,|\, Y\right] \to 0 \quad \text{as } n \to \infty \tag{D.13}$$

for

$$S(f_0, A; \upsilon) \equiv \{(l,k) : |\beta^0_{lk}| > A \log^{1+\upsilon+1\vee\upsilon} n/\sqrt{n}\}, \tag{D.14}$$

where υ is the balancing constant in the design Assumption 4.

The proof of (D.13) follows the route of Lemma 3 in [18] with nontrivial alterations due to the fact that we now have the regression model where the regression matrix is not orthogonal. Suppose that $(l_1, k_1) \in S(f_0; A; \upsilon)$ is a signal node for some $A > 0$ and let $\mathcal{T}$ be such that $(l_1, k_1) \notin \mathcal{T}$. We grow a branch from $\mathcal{T}$ that extends towards (l_1, k_1) to obtain an enlarged tree $\mathcal{T}^+ \supset \mathcal{T}$. In other words $\mathcal{T}^+$ is the smallest tree that contains $\mathcal{T}$ and (l_1, k_1) as an internal node. For details, we refer to Lemma 3 in [18]. We define $K = |\mathcal{T}^+_{int} \backslash \mathcal{T}_{int}|$ and write

$$\frac{N_Y(\mathcal{T})}{N_Y(\mathcal{T}^+)} = (1+g_n)^{K/2} \exp\left\{\frac{1}{2} Y'[X_{\mathcal{T}} \Sigma_{\mathcal{T}} X'_{\mathcal{T}} - X_{\mathcal{T}^+} \Sigma_{\mathcal{T}^+} X'_{\mathcal{T}^+}] Y\right\}. \tag{D.15}$$

We denote with $\mathcal{T} = \mathcal{T}^- \to \mathcal{T}^1 \to \cdots \to \mathcal{T}^K = \mathcal{T}^+$ the sequence of nested trees obtained by adding one additional internal node towards (l_1, k_1). Then using Lemma F.2 we find

$$\frac{N_Y(\mathcal{T})}{N_Y(\mathcal{T}^+)} = (1+g_n)^{K/2} \prod_{j=1}^{K} \exp\left\{\frac{c_n Y'(P_{\mathcal{T}^{j-1}} - P_{\mathcal{T}^j}) Y}{Z_j}\right\} \tag{D.16}$$

$$= (1+g_n)^{K/2} \prod_{j=1}^{K} \exp\left\{-\frac{c_n |X'_{[j]}(I - P_{j-1}) Y|^2}{Z_j}\right\}, \tag{D.17}$$

where

$$P_j = X_{\mathcal{T}^j}(X'_{\mathcal{T}^j} X_{\mathcal{T}^j})^{-1} X'_{\mathcal{T}^j} \quad \text{and} \quad Z_j = X'_{[j]}(I - P_{j-1}) X_{[j]}$$

and where $X_{[j]}$ is the column added at the j^{th} step of branch growing. Let $X_{[K]}$ be the *last* column to be added to $X_{\mathcal{T}^+}$, i.e. the *signal* column associated with (l_1, k_1). We will be denoting simply $\beta^0_{[K]} \equiv \beta^0_{l_1 k_1}$ the coefficient associated with $X_{[K]}$. Then (using the fact that P_{K-1} is a projection matrix onto the columns of $X_{\mathcal{T}^{K-1}}$)

$$|X'_{[K]}(I - P_{K-1}) Y|^2 = |X'_{[K]}(I - P_{K-1}) X_{[K]} \beta^0_{[K]} + X'_{[K]}(I - P_{K-1}) X_{\backslash \mathcal{T}^K} \boldsymbol{\beta}^0_{\backslash \mathcal{T}^K} + X'_{[K]}(I - P_{K-1}) \boldsymbol{\nu}|^2$$

Using the inequality $(a+b)^2 \geq a^2/2 - b^2$, we find that

$$\frac{|X'_{[K]}(I - P_{K-1}) Y|^2}{Z_K} \geq \frac{Z_K |\beta^0_{[K]}|^2}{2} - \frac{1}{Z_K} |X'_{[K]}(I - P_{K-1}) X_{\backslash \mathcal{T}^K} \boldsymbol{\beta}^0_{\backslash \mathcal{T}^K} + X'_{[K]}(I - P_{K-1}) \boldsymbol{\nu}|^2.$$

Next, since all entries in $X_{\backslash \mathcal{T}^K}$ are either descendants of (l_1, k_1) or are orthogonal to $X_{[K]}$ we have (using similar arguments as before in Section D.1)

$$|X'_{[K]}(I - P_{K-1}) X_{\backslash \mathcal{T}^K} \boldsymbol{\beta}^0_{\backslash \mathcal{T}^K}| \leq |X'_{[K]} X_{\backslash \mathcal{T}^K} \boldsymbol{\beta}^0_{\backslash \mathcal{T}^K}| + |X'_{[K]} P_{K-1} X_{\backslash \mathcal{T}^K} \boldsymbol{\beta}^0_{\backslash \mathcal{T}^K}| \lesssim \sqrt{n} \log^{\upsilon} n.$$

Finally, using Lemma F.3 and similar arguments as in (D.11), we find that

$$|X'_{[K]}(I-P_{K-1})\boldsymbol{\nu}| \leq |X'_{[K]}\boldsymbol{\nu}| + |X'_{[K]}P_{K-1}\boldsymbol{\nu}| \lesssim \sqrt{n}\log^{\upsilon+1\vee\upsilon} n$$

which yields

$$\frac{|X'_{[K]}(I-P_{K-1})Y|^2}{Z_K} \geq \frac{Z_K|\beta^0_{[K]}|^2}{2} - \frac{C_1 n\log^{2[\upsilon+(1\vee\upsilon)]} n}{Z_K} \quad \text{for some } C_1>0.$$

The term $Z_K = X'_{[K]}(I-P_{K-1})X_{[K]}$ is a submatrix of the matrix $(X'_{\mathcal{T}^K}X_{\mathcal{T}^K})^{-1}$ and by our assumption (D.12) we have $Z_K \geq \underline{\lambda} n$ which yields (for $g_n = n$) and from the assumption $|\beta^0_{[K]}| > A\log^{1+\upsilon+(1\vee\upsilon)} n/\sqrt{n}$ for some sufficiently large $A>0$ and some $C_2, C_3 > 0$

$$\frac{N_Y(\mathcal{T})}{N_Y(\mathcal{T}^+)} \leq \mathrm{e}^{K\log(1+g_n) - C_2\log^{2+2[\upsilon+(1\vee\upsilon)]} n} = \exp\left\{-C_3\log^{2+2[\upsilon+(1\vee\upsilon)]} n\right\}.$$

Similarly as was shown in the proof of Lemma 3 in [18], we have for $p_l = (1/\Gamma)^{l^a}$ with $a = 2[\upsilon+(1\vee\upsilon)]$ the following bound for the prior ratio $\Pi(\mathcal{T})/\Pi(\mathcal{T}^+) \lesssim \Gamma^{2l_1^{a+1}}$, and thereby for some $C_4>0$

$$\Pi[(l_1,k_1)\notin\mathcal{T}_{int}\,|\,Y] \leq \exp\left\{C_4(\log\Gamma)\log^{1+2[\upsilon+(1\vee\upsilon)]} n - C_3\log^{2+2[\upsilon+(1\vee\upsilon)]} n\right\}.$$

Thereby, for some $C_5>0$,

$$\sum_{(l_1,k_1)\in S(f_0;A;\upsilon)} \Pi[(l_1,k_1)\notin\mathcal{T}_{int}\,|\,Y] \leq \mathrm{e}^{-C_5\log^2 n}2^{L_{max}+1} \lesssim \mathrm{e}^{-C_5/2\log^2 n} \to 0.$$

This concludes the proof of (D.13).

E Proof of Lemma A.3 in Section A.2

To prove Lemma A.3, we split $\{n_{I_x} \leq s_n(\delta)\}$ into $B_{n,1} = \{n_{I_x} \leq s_n(\delta)\} \cap \{|\bar{y}_{I_{x,1}} - f_0(x)| \leq M_0\varepsilon_n\}$, $B_{n,2} = \{n_{I_x} \leq s_n(\delta)\} \cap \{|\bar{y}_{I_{x,1}} - f_0(x)| > M_0\varepsilon_n\} \cap \{n_{I_{x,1}} \leq s_n(\delta_1)\}$ and $B_{n,3} = \{n_{I_x} \leq s_n(\delta)\} \cap \{|\bar{y}_{I_{x,1}} - f_0(x)| > M_0\varepsilon_n\} \cap \{n_{I_{x,1}} > s_n(\delta_1)\}$ where $\sqrt{\delta_1}M_0 > 2u_0$.

We first consider $B_{n,1}$. We have for $\bar{I} = I_x \cup I_{x,1}$ and writing $S = S' \cup I_x \cup I_{x,1}$

$$\Pi(B_{n,1}|D_n) = \frac{\sum_{S=S'\cup I_x\cup I_{x,1}} m(S')m(\bar{I})\pi_S(S'\cup\bar{I})\mathbb{I}_{B_{n,1}}\frac{m(I_x)m(I_{x,1})\pi_S(S'\cup I_x\cup I_{x,1})}{m(\bar{I})\pi_S(S'\cup\bar{I})}}{\sum_S m(S)\pi_S(S)}.$$

Moreover on $\Omega_{n,x}(u_1) \cap \Omega_{n,y}(u_0)$,

$$\begin{aligned}
\frac{m(I_x)m(I_{x,1})}{m(\bar{I})} &\leq \frac{c_1^2\sqrt{2\pi}\sqrt{n_{\bar{I}}}}{c_0\sqrt{n_{I_x}}\sqrt{n_{I_{x,1}}}} \exp\left\{\frac{n_{I_x}}{2}(\bar{y}_{I_x} - \bar{y}_{\bar{I}})^2 + \frac{n_{I_{x,1}}}{2}(\bar{y}_{I_{x,1}} - \bar{y}_{\bar{I}})^2\right\} \\
&= \frac{c_1^2\sqrt{2\pi}\sqrt{n_{\bar{I}}}}{c_0\sqrt{n_{I_x}}\sqrt{n_{I_{x,1}}}} \exp\left\{\frac{n_{I_{x,1}}n_x}{2n_{\bar{I}}}(\bar{y}_{I_x} - \bar{y}_{I_{x,1}})^2\right\}
\end{aligned}$$

Note that on $\Omega_{n,x}(u_1)$ we have

$$p_{I_x}\left(1 - u_1\sqrt{\frac{\log n}{np_{I_x}}}\right) \leq \frac{n_{I_x}}{n} \leq p_{I_x}\left(1 + u_1\sqrt{\frac{\log n}{np_{I_x}}}\right)$$

and since $p_{I_x} \geq p_0|I_x| \geq p_0 C_1 \log n/n$ with $u_1 \leq \sqrt{p_0 C_1}/2$ we obtain

$$|I_x| \leq \frac{2n_{I_x}}{np_0} \leq \frac{2s_n(\delta)}{n}.$$

Also $n_{I_{x,1}}n_{I_x}/n_{\bar{I}} \leq n_{I_x} \leq s_n(\delta)$. Moreover,

$$\bar{y}_{I_x} - \bar{y}_{I_{x,1}} = \bar{\epsilon}_{I_x} + \bar{\beta}_{0,I_x} - f_0(x) + f_0(x) - \bar{y}_{I_{x,1}} \tag{E.1}$$

and $|\bar{\beta}_{0,I_x} - f_0(x)| \leq M|I_x|^{t(x)} \leq \delta^{t(x)}C\varepsilon_n$ for some C independent on δ and n. Therefore when $|\bar{y}_{I_{x,1}} - f_0(x)| \leq M_0\varepsilon_n$

$$\begin{aligned}
\frac{n_{I_{x,1}}n_{I_x}}{2n_{\bar{I}}}(\bar{y}_{I_x} - \bar{y}_{I_{x,1}})^2 &\leq \frac{n_{I_x}\bar{\epsilon}_{I_x}^2}{2} + C^2\delta^{2t(x)+1}\log n + \delta M_0^2\log n + \sqrt{n_{I_x}}|\bar{\epsilon}_{I_x}|\sqrt{\delta}[M_0 + C\delta^{t(x)}]\sqrt{\log n} \\
&\leq \frac{n_{I_x}\bar{\epsilon}_{I_x}^2}{2} + \delta\log n[C^2\delta^{2t(x)} + M_0^2] + \sqrt{\delta}[M_0 + C\delta^{t(x)}]u_0\log n \\
&\leq \frac{n_{I_x}\bar{\epsilon}_{I_x}^2}{2} + 2\sqrt{\delta}M_0^2\log n
\end{aligned}$$

on Ω_n, as soon as M_0 is large enough (independently of δ) and δ is small enough.

Moreover, on Ω_n we can also bound $n_{I_x}\bar{\epsilon}_{I_x}^2$ by $u_0^2\log n$ so that for all $b \in (0,1)$, so that

$$\begin{aligned}
\frac{m(I_x)m(I_{x,1})}{m(\bar{I})} &\leq \frac{c_1^2\sqrt{2\pi}\sqrt{n_{\bar{I}}}}{c_0\sqrt{n_{I_x}}\sqrt{n_{I_{x,1}}}} n^{2\sqrt{\delta}M_0^2} e^{\frac{n_{I_x}\bar{\epsilon}_{I_x}^2}{2}} \\
&\leq \frac{c_1^2\sqrt{2\pi}\sqrt{n_{\bar{I}}}}{c_0\sqrt{n_{I_x}}\sqrt{n_{I_{x,1}}}} n^{[2\sqrt{\delta}M_0^2 + bu_0^2/2]} e^{\frac{(1-b)n_{I_x}\bar{\epsilon}_{I_x}^2}{2}}
\end{aligned}$$

and denoting $Z_b^{I_x} \equiv \exp\left\{\frac{(1-b)n_{I_x}\bar{\epsilon}_{I_x}^2}{2}\right\}$ and using the fact that

$$E(Z_b^{I_x}|X) = \int \frac{e^{(1-b)u^2/2 - u^2/2}}{\sqrt{2\pi}} du = 1/\sqrt{b} < \infty$$

we obtain on Ω_n,

$$\Pi(B_{n,1}|D_n) \leq \frac{\sqrt{2\pi} n^{bu_0^2/2+2\sqrt{\delta}M_0^2} c_1^2}{c_0} \times \frac{\sum\limits_{S=S'\cup\bar{I}} m(S')m(\bar{I})\pi_S(S'\cup\bar{I}) \sum\limits_{\bar{I}=I_x\cup I_{x,1}} \mathbb{I}_{B_{n,1}} \frac{|I_{x,1}|^B |I_x|^B \sqrt{n_{\bar{I}}}}{|\bar{I}|^B \sqrt{n_{I_x}}\sqrt{n_{I_{x,1}}}} Z_b^{I_x}}{\sum\limits_{S=S'\cup\bar{I}} m(S')m(\bar{I})\pi_S(S'\cup\bar{I})}.$$

Note that for any $\bar{I}$ containing x, there are many possible choices for $(I_x, I_{x,1})$ such that $\bar{I} = I_{x,1} \cup I_x$. Also $n_{\bar{I}}/[n_{I_{x,1}} n_{I_x}] \leq 2/(n_{I_{x,1}} \wedge n_{I_x})$ so that, choosing without loss of generality $n_{I_x} \leq n_{I_{x,1}}$,

$$\frac{|I_{x,1}|^B |I_x|^B \sqrt{n_{\bar{I}}}}{|\bar{I}|^B \sqrt{n_{I_x}}\sqrt{n_{I_{x,1}}}} \leq \frac{\sqrt{2}|I_x|^B}{\sqrt{n_{I_x}}} \leq \frac{2|I_x|^{B-1/2}}{\sqrt{p_0 n}}$$

Hence, there exists $\gamma < 0$ such that for any $u_n = o(1)$, writing $I_{x,1} = \bar{I} \setminus I_x$ and using Markov inequality,

$$\begin{aligned} P\left(\Pi(B_{n,1}|D_n) > u_n\right) &\lesssim o(1/n) + \sum_{\bar{I}: x\in\bar{I}} P\left[\sum_{I_x\subset\bar{I}} \mathbb{I}_{n_{I_x}\leq s_n(\delta)} |I_x|^{B-1/2} Z_b^{I_x} > \frac{u_n c_0 \sqrt{p_0} n^{1/2-bu_0^2-2\sqrt{\delta}M_0^2}}{4c_1^2\sqrt{2\pi}}\right] \\ &\lesssim o(1/n) + \frac{n^{bu_0^2+2\sqrt{\delta}M_0^2}}{\sqrt{b}u_n\sqrt{n}} \sum_{\bar{I}: x\in\bar{I}} \sum_{l_x, I_{x,1}} \mathbb{I}_{n_{I_x}\leq s_n(\delta)} |l_x|^{B-1/2} \end{aligned}$$

Given that each interval is made of a number of units of size $\asymp \log n/n$, the number of intervals $I_x, I_{x,1}$ where $|I_x|$ is composed of ℓ *units* (i.e. elementary intervals (z_l, z_{l+1})) is bounded by $O(\ell \times n/\log n)$ and since $n_{I_x} \leq s_n(\delta)$, $\ell \lesssim \delta\varepsilon_n^{-2}$ so that

$$\sum_{l_x, I_{x,1}} \mathbb{I}_{n_{I_x}\leq s_n(\delta)} |l_x|^{B-1/2} \lesssim \left(\frac{\log n}{n}\right)^{B-3/2} \sum_{\ell \leq O(\varepsilon_n^{-2})} \ell^{B+1/2} \lesssim \left(\frac{\log n}{n}\right)^{B-3/2} \varepsilon_n^{-2B-3}.$$

Hence we obtain

$$P\left(\Pi(B_{n,1}|D_n) > u_n\right) \lesssim o(1/n) + \frac{n^{bu_0^2+2\sqrt{\delta}M_0^2}}{\sqrt{b}u_n\sqrt{n}} \left(\frac{\log n}{n}\right)^{B-3/2} \varepsilon_n^{-2B-3} = o(1/n)$$

as soon as $B > 7t(x) + 2$ by choosing b, δ small enough.This is verified as soon as $B > 9$.

We now study $B_{n,2}$. When $n_{I_{x,1}} < s_n(\delta_1)$ with $\delta_1 \geq \delta$ we have $|I_x \cup I_{x,1}| \leq p_1 s_n(\delta+\delta_1)/n$ and by the Hölder condition on f_0 at x we obtain for some $M > 0$

$$|\bar{\beta}_{0,I_x} - \bar{\beta}_{0,I_{x,1}}| \leq 2M[p_1 s_n(\delta+\delta_1)/n]^{t(x)}$$

so that

$$\bar{y}_{I_x} - \bar{y}_{I_{x,1}} = \bar{\epsilon}_{I_x} - \bar{\epsilon}_{I_{x,1}} + \bar{\beta}_{0,I_x} - \bar{\beta}_{0,I_{x,1}} = \bar{\epsilon}_{I_x} - \bar{\epsilon}_{I_{x,1}} + O(\delta_1^{t(x)}\varepsilon_n).$$

Consider the event

$$\bar{\Omega}_{n,2} = \left\{\forall \bar{I} \ s.t.\ n_{\bar{I}} \leq s_n(\delta+\delta_1) \text{ and } x \in \bar{I} \ : \frac{\sqrt{n_{I_x}}\sqrt{n_I}|\bar{\epsilon}_{I_x} - \bar{\epsilon}_I|}{\sqrt{n_{I_x}+n_I}} \leq u_1'\sqrt{\log n}\right\}.$$

Then since for each $(\bar{I}, I_x)$, $\sqrt{n_{I_x}}\sqrt{n_I}(\bar{\epsilon}_{I_x} - \bar{\epsilon}_I)/\sqrt{n_{I_x}+n_I} \sim \mathcal{N}(0,1)$ and since the number of $(\bar{I}, I_x)$ satisfying $\bar{I} = I_x \cup I_{x,1}$ and $n_{I_x}, n_{I_{x,1}} \leq s_n(\delta)$ is bounded by a term of order

$$\sum_{\ell \lesssim \varepsilon_n^{-2}} \underbrace{\varepsilon_n^{-2}}_{\text{bound on number of } I_{x,1}} \ell \lesssim \varepsilon_n^{-6},$$

as soon as $(u_1')^2 > 6t(x)/(2t(x)+1)+2$ we have $P(\bar{\Omega}_{n,2}) = 1 + o(1/n)$. On $\bar{\Omega}_{n,2}$,

$$\frac{n_{I_{x,1}}n_{I_x}}{2n_{\bar{I}}}(\bar{y}_{I_x} - \bar{y}_{I_{x,1}})^2 \leq \frac{n_{I_{x,1}}n_x}{2n_{\bar{I}}}(\bar{\epsilon}_{I_x} - \bar{\epsilon}_{I_{x,1}})^2 + a(\delta_1)\log n$$

for some $a(\delta_1) > 0$ which goes to 0 when δ_1 goes to 0 and similarly to before, for all $1 > b > 0$

$$\begin{aligned}
&\mathbb{P}(\Pi(B_{n,2}|D_n) > u_n) \\
&\quad \leq \mathbb{P}\left(\frac{n^{b(u_1')^2+a(\delta_1)}}{\sqrt{n}} \max_{\bar{I}:x\in\bar{I}, n_{\bar{I}}\leq s_n(\delta_1+\delta)} \sum_{I_x\subset\bar{I}} \mathbb{I}_{n_{I_x}\leq s_n(\delta)}|I_x|^{B-1/2}\exp\left\{\frac{(1-b)n_{I_{x,1}}n_x}{2n_{\bar{I}}}(\bar{\epsilon}_{I_x} - \bar{\epsilon}_{I_{x,1}})^2\right\} > u_n\right) \\
&\quad \leq \frac{n^{b(u_1')^2+a(\delta_1)}}{\sqrt{b}u_n\sqrt{n}} \sum_{\bar{I}:x\in\bar{I}} \mathbb{I}_{n_{\bar{I}}\leq 2s_n(\delta_1)} \sum_{I_x\subset I} \mathbb{I}_{n_{I_x}\leq s_n(\delta_1)}|I_x|^{B-1/2} \\
&\quad \leq \frac{n^{b(u_1')^2+a(\delta_1)}}{\sqrt{b}u_n\sqrt{n}}\left(\frac{\log n}{n}\right)^{B-1/2} \varepsilon_n^{-2}\varepsilon_n^{-2(B+3/2)} = O(n^{\frac{-B+5t(x)}{2t(x)+1}+bu_1'+a(\delta_1)}) = o(1/n)
\end{aligned}$$

since $B > 8$ by choosing b, δ_1 small enough.

Finally, we study $B_{n,3}$. Since $|I_x| \leq p_1 s_n(\delta)/n$ we can choose a point in the grid $x_1 \in I_{x,1}$, such that $|x - x_1| \leq 2p_1 s_n(\delta)/n$, so that the Hölder condition of f_0 at x implies that

$$|f_0(x) - f_0(x_1)| \leq M(2p_1)^{t(x)}\delta^{t(x)}\varepsilon_n(x).$$

Moreover, since t is Hölder α for some $\alpha > 0$ on $[x, x_1]$ (note that for n large enough $|x - x_1|$ is arbitrarily small) we have

$$|t(x_1) - t(x)| \leq L_0\delta^\alpha(n/\log n)^{-\alpha/(2t(x)+1)}$$

and $\varepsilon_n(x_1)^2 = (n/\log n)^{-2t(x_1)/(2t(x_1)+1)} = (n/\log n)^{-2t(x)/(2t(x)+1)}(1+o(1))$. Hence, choosing $M_0 > 2(3p_1)^{t(x)}$, for n large enough,

$$\begin{aligned}
&\Pi\left(B_{n,3} \mid D_n\right)\\
&\quad\le \Pi\left(\{|\bar{y}_{I_{x,1}} - f_0(x_1)| > M_0\varepsilon_n(x_1)/2\} \cap \{n_{I_{x,1}} > s_n(\delta_1)\}|D_n\right) = o_P(1/n)
\end{aligned}$$

from Lemma A.2 and Theorem 6 is proved by choosing $M_0 > 4/\sqrt{\delta_1}$.

F Auxiliary Results

Lemma F.1. *Let X_i and X_j be columns of X that correspond to nodes (l_2, k_2) and (l_1, k_1), respectively. Then we have*

$$|X_j'X_i| = 0 \text{ when } (l_2,k_2) \text{ is not a descendant of } (l_1,k_1),$$
$$|X_j'X_i| \le 2^{\frac{l_1+l_2}{2}}|n^L_{l_2k_2} - n^R_{l_2k_2}| \text{ when } (l_2,k_2) \text{ is a descendant of } (l_1,k_1).$$

Proof. When (l_2, k_2) is *not* a descendant of (l_1, k_1), the domains of $\psi_{l_1k_1}$ and $\psi_{l_2k_2}$ do not overlap, yielding orthogonality. When (l_2, k_2) is a descendant of (l_1, k_1), the wavelet domains satisfy $\mathrm{I}_{l_2k_2} \subset \mathrm{I}_{l_1k_1}$ and $X_j'X_i$ will be (up to a sign) equal to the size of the amplitude product $2^{(l_1+l_2)/2}$ multiplied by the excess number of observations falling inside the longer wavelet piece ψ_{l_2,k_2}. □

Lemma F.2. *We denote with $P_{\mathcal{T}} = X_{\mathcal{T}}(X_{\mathcal{T}}'X_{\mathcal{T}})^{-1}X_{\mathcal{T}}'$ the projection matrix and with $Z = \|X_j\|_2^2 - X_j'P_{\mathcal{T}^-}X_j$. Then*

$$Y'[P_{\mathcal{T}} - P_{\mathcal{T}^-}]Y = \frac{Y'(I - P_{\mathcal{T}^-})X_jX_j'(I - P_{\mathcal{T}^-})Y}{Z}. \tag{F.1}$$

Proof. We can write, for $\widetilde{\Sigma}_{\mathcal{T}} = (X_{\mathcal{T}}'X_{\mathcal{T}})^{-1}$,

$$(X_{\mathcal{T}}'X_{\mathcal{T}})^{-1} = \begin{pmatrix} X_{\mathcal{T}^-}'X_{\mathcal{T}^-} & X_{\mathcal{T}^-}'X_j \\ X_j'X_{\mathcal{T}^-} & \|X_j\|_2^2 \end{pmatrix}^{-1} = \begin{pmatrix} \widetilde{\Sigma}_{\mathcal{T}^-} + \widetilde{\Sigma}_{\mathcal{T}^-}X_{\mathcal{T}^-}'X_jX_j'X_{\mathcal{T}^-}\widetilde{\Sigma}_{\mathcal{T}^-}/Z & -\widetilde{\Sigma}_{\mathcal{T}^-}X_{\mathcal{T}^-}'X_j/Z \\ -X_j'X_{\mathcal{T}^-}\widetilde{\Sigma}_{\mathcal{T}^-}/Z & 1/Z \end{pmatrix}$$

Next, noting that $X_{\mathcal{T}} = (X_{\mathcal{T}^-}, X_j)$

$$X_{\mathcal{T}}\widetilde{\Sigma}_{\mathcal{T}} = \left(X_{\mathcal{T}^-}\left[\widetilde{\Sigma}_{\mathcal{T}^-} + \tfrac{\widetilde{\Sigma}_{\mathcal{T}^-}X_{\mathcal{T}^-}'X_jX_j'X_{\mathcal{T}^-}\widetilde{\Sigma}_{\mathcal{T}^-}}{Z}\right] - \tfrac{X_jX_j'X_{\mathcal{T}^-}'\widetilde{\Sigma}_{\mathcal{T}^-}}{Z}, \qquad -\tfrac{P_{\mathcal{T}^-}X_j}{Z} + \tfrac{X_j}{Z}\right)$$

which yields

$$P_{\mathcal{T}} = P_{\mathcal{T}^-} + \frac{1}{Z}\left[P_{\mathcal{T}^-}X_jX_j'P_{\mathcal{T}^-} - X_jX_j'P_{\mathcal{T}^-} - P_{\mathcal{T}^-}'X_jX_j' + X_jX_j'\right].$$

We then obtain (F.1). □

Lemma F.3. *Let X_j be the j^{th} column in the matrix X and let $\boldsymbol{\beta}_0 = (\beta_1^0, \dots, \beta_p^0)'$ be the vector of multiscale coefficients $\langle \psi_{lk}, f_0 \rangle$ for $f_0 \in \mathcal{C}(t, M, \eta)$ where t, M, η satisfy Assumption 1 with $t_1 > 1/2$. Then, on the event $\mathcal{A}$, we have*

$$|X_j' \boldsymbol{\nu}| = |X_j'(F_0 - X\boldsymbol{\beta}_0 + \boldsymbol{\varepsilon})| \lesssim \sqrt{n} \log^{1 \vee \upsilon} n.$$

Proof. From the definition of the set $\mathcal{A}$ in (D.4) we know that $|X_j' \boldsymbol{\varepsilon}| \lesssim \sqrt{n} \log n$. Next, we decompose the bias term $|X_j'(F_0 - X\boldsymbol{\beta}_0)|$ into resolutions $L_{max} < l \leq \widetilde{L}_{max}$ that are within the spam of the matrix X and higher resolutions $l > \widetilde{L}_{max}$ for which the balancing Assumption 4 is no longer required. Then, using (D.2), we obtain

$$\begin{aligned}|X_j'(F_0 - X\boldsymbol{\beta}_0)| \leq C_d \sqrt{n} \log^{\upsilon} n &\sum_{l=L_{max}+1}^{\widetilde{L}_{max}} 2^{l-l_1} 2^{l_1/2} 2^{-l(t_1+1/2)} \\ &+ \|X_j\|_1 \left\| \sum_{l > \widetilde{L}_{max}} \sum_k \psi_{lk}(x) \beta_{lk}^0 \right\|_\infty .\end{aligned}$$

The first term above can be bounded by a constant multiple of $2^{-l_1/2} \sqrt{n} \log^{\upsilon} n$ when $t_1 > 1/2$. Regarding the second term, under the assumption $t_1 > 1/2$ and using the fact that $\widetilde{L}_{max} = \mathcal{O}[\log_2(n/\log n)]$, we obtain for each $x \in [0, 1]$

$$\left| \sum_{l > \widetilde{L}_{max}} \sum_k \psi_{lk}(x) \beta_{lk}^0 \right| \leq \sum_{l > \widetilde{L}_{max}} 2^{l/2} |\beta_{lk_l(x)}^0| \leq \sum_{l > \widetilde{L}_{max}} 2^{-lt(x)} \lesssim 2^{-\widetilde{L}_{max}/2} \lesssim \sqrt{\frac{\log n}{n}}.$$

Using (D.1) we find that $\|X_j\|_1 \lesssim n$ and conclude that

$$|X_j' \boldsymbol{\nu}| \lesssim \sqrt{n} \log^{\upsilon} n + \sqrt{n} \log n \lesssim \sqrt{n} \log^{1 \vee \upsilon} n. \qquad \square$$

Lemma F.4. *(Eigenvalue Bounds) Under the Assumption 4 with $0 \leq \upsilon < 1/2$ and with $\upsilon = 1/2$ for $c > 2C_d C^*$, the eigen-spectrum of $X_{\mathcal{T}}' X_{\mathcal{T}}$ for each $\mathcal{T} \in \mathbb{T}$ satisfies (for n large enough)*

$$\underline{\lambda}\, n \leq \lambda_{min}(X_{\mathcal{T}}' X_{\mathcal{T}}) \leq \lambda_{max}(X_{\mathcal{T}}' X_{\mathcal{T}}) \leq \bar{\lambda}\, n \log n \quad \textit{for some } 0 < \underline{\lambda} \leq \bar{\lambda}. \tag{F.2}$$

Proof. The diagonal elements of $X'X$, denoted with $a(i)$, satisfy

$$2\, c\, n \leq a(i) \equiv \|X_i\|_2^2 \leq 2n[C + \log_2 \lfloor C^* \sqrt{n/\log n} \rfloor].$$

For a given node $(l_1, k_1) \in \mathcal{T}$ with $i = 2^{l_1} + k_1$ we denote with $a(\backslash i)$ the sum of absolute off-diagonal terms in the i^{th} row of $X'X$. Under the Assumption 4, we show below (for $C_m = 2C_d\, C^*$)

$$a(\backslash i) = \sum_{(l,k)\neq(l_1,k_1)} |X_i' X_{2^l+k}| \leq C_m\, n \log^{\upsilon-1/2} n. \tag{F.3}$$

In order to show (F.3), we split the sum into nodes $(l, k) \in P(l, k)$ that are predecessors of (l_1, k_1) and nodes $(l, k) \in D(l, k)$ that are descendants of (l_1, k_1). Using (D.2) and the fact that there are 2^{l-l_1} descendants at each layer $l > l_1$ we have (using the fact that $2^{L_{max}} = \lfloor C^* \sqrt{n/\log n} \rfloor$)

$$\sum_{(l,k)\in D(l_1,k_1)} |X_i' X_{2^l+k}| \leq C_d \sqrt{n} \log^\upsilon n \sum_{l=l_1+1}^{L_{max}} 2^{l-l_1}\, 2^{\frac{l_1}{2}} \leq C_d \sqrt{n} \log^\upsilon n 2^{L_{max}} \leq C_d C^*\, n \log^{\upsilon-1/2} n$$

and (using the fact that $l_1 \leq L_{max}$)

$$\sum_{(l,k)\in P(l_1,k_1)} |X_i' X_{2^l+k}| \leq C_d \sqrt{n} \log^\upsilon n \sum_{l=0}^{l_1-1} 2^{\frac{l}{2}} \leq C_d \sqrt{n} \log^\upsilon n\, 2^{l_1/2} < C_d C^*\, n \log^{\upsilon-1/2} n.$$

From (F.3) one obtains $a(i) - a(\backslash i) > 2n[c - C_m\, \log^{\upsilon-1/2} n] > 0$ for n large enough when $0 \leq \upsilon < 1/2$ and for $c > C_m$ for $\upsilon = 1/2$. The Gershgorin circle theorem [?] then yields

$$\min[a(i) - a(\backslash i)] \leq \lambda_{min}(X'X) \leq \lambda_{max}(X'X) \leq \max[a(i) + a(\backslash i)]. \qquad \square$$

Lemma F.5. *Assume that* $x_i \stackrel{iid}{\sim} U[0, 1]$. *Then for* $t_l = \frac{C_d}{2} \frac{\log^\upsilon n}{\sqrt{n}\, 2^{l/2}}$ *we have*

$$P\left(|\bar{n}_{lk} - \underline{n}_{lk}| \leq 2nt_l \quad \forall (l, k)\ s.t.\ l \leq \widetilde{L}_{max}\right) = 1 + o(1)$$

for $\upsilon > 1/2$ *and for* $\upsilon = 1/2$ *when* $C_d^2/[4(1 + C_d/3)] \geq 1$.

Proof. Under the uniform random design, both n_{lk}^R and n_{lk}^L are distributed according to $Bin(n, 2^{-(l+1)})$. We can write

$$\begin{aligned}
&P(|\bar{n}_{lk} - \underline{n}_{lk}| \leq 2nt_l \quad \forall (l, k)\ s.t.\ l \leq \widetilde{L}_{max}) \geq \\
&P(|n_{lk}^R - n2^{-(l+1)}| \leq nt_l \quad \text{and} \quad |n_{lk}^L - n2^{-(l+1)}| \leq nt_l \quad \forall (l, k)\ s.t.\ l \leq \widetilde{L}_{max}) = \\
&1 - P\left(\cup_{l,k}\{|n_{lk}^R - n2^{-(l+1)}| > nt_l\} \cup \{|n_{lk}^L - n2^{-(l+1)}| > nt_l\}\right).
\end{aligned}$$

We show that the probability on the right-hand side above is $o(1)$. We note that $2^{\widetilde{L}_{max}} = \mathcal{O}(n/\log n)$. The Bernstein inequality tailored to iid Bernoulli random variables X_i with a mean μ and variance σ^2 (Theorem 2.8.4 in [**?**]) states that

$$P\left(|\bar{X}_n - \mu| > \epsilon\right) \leq 2\exp\left\{-\frac{n\epsilon^2}{2\sigma^2 + 2\epsilon/3}\right\} \ \forall\epsilon > 0$$

where $\bar{X}_n$ is the mean of Bernoulli random variables X_i's. Applied to our context, we obtain

$$\begin{aligned}
\sum_{l,k} P\left(|n_{lk}^R - n2^{-(l+1)}| > nt_l\right) &\leq \sum_{l=0}^{\widetilde{L}_{max}} 2^{l+1}\exp\left(-\frac{nt_l^2}{2(\sigma^2 + t_l/3)}\right)\\
&\leq \sum_{l=0}^{\widetilde{L}_{max}} 2^{l+1}\exp\left(-\frac{C_d^2\log^{2\upsilon} n}{4\times 2^l} \times \frac{1}{2^{-l}(1-2^{-(l+1)}) + C_d/3\log^{\upsilon} n/(\sqrt{n}2^{l/2})}\right)\\
&\leq \sum_{l=0}^{\widetilde{L}_{max}} 2^{l+1}\exp\left(-\frac{C_d^2\log^{2\upsilon} n}{4} \times \frac{1}{1 + C_d/3\log^{\upsilon} n \times 2^{l/2}/\sqrt{n}}\right)\\
&\leq \sum_{l=0}^{\widetilde{L}_{max}} 2^{l+1}\exp\left(-\frac{C_d^2\log^{2\upsilon} n}{4} \times \frac{1}{1 + C_d/3\log^{\upsilon-1/2} n}\right)
\end{aligned}$$

For $\upsilon > 1/2$, the sum can be bounded by (for large enough n)

$$\sum_{l=0}^{\widetilde{L}_{max}} 2^{l+1}\exp\left(-C_d^2/8\log^{2\upsilon} n\right) = \mathcal{O}\left(\frac{n}{\log n}\right) \times n^{-C_d^2/8\log^{2\upsilon-1} n} = o(1).$$

When $\upsilon = 1/2$, the sum can be bounded by (for $\widetilde{C}_d = C_d^2/[4(1+C_d/3)]$)

$$\sum_{l=0}^{\widetilde{L}_{max}} 2^{l+1}\exp\left(-C_d^2/[4(1+C_d/3)]\log^{2\upsilon} n\right) = \mathcal{O}\left(\frac{n}{\log n}\right) \times n^{-\widetilde{C}_d\log^{2\upsilon-1} n} = o(1) \text{ when } \widetilde{C}_d \geq 1.$$

Lemma F.6. *(Random Design) Assume that the design points x_i's are iid with density p bounded from above and below on $[0,1]$ by p_1 and p_0, respectively. Then*

$$\mathbb{P}\left(|n_I - n\times p(I)| \leq u_1\sqrt{np(I)\log n}\right) \leq \mathrm{e}^{-\gamma u_1\log n}, \quad \textit{where} \quad \gamma = 3/4\sqrt{p_0 C_1}, \tag{F.4}$$

when u_1 is chosen large enough.

Proof. The Bernstein inequality (Theorem 2.8.4 in [**?**]) implies that for all possible intervals I in the partition, we have

$$\mathbb{P}\left(|n_I - n\times p(I)| \leq u_1\sqrt{np(I)\log n}\right) \leq 2\exp\left\{-\frac{np(I)u_1^2\log n}{2np(I)(1-p(I)) + 2/3u_1\sqrt{np(I)\log n}}\right\}$$

For $p(I) \geq p_0 C_1 \sqrt{\log n/n}$ we obtain

$$np(I)(1-p(I)) + u_1\sqrt{np(I)\log n}/3 \leq np(I)\left[1 + \frac{u_1}{3\sqrt{p_0 C_1}}\right].$$

With $u_1 \geq 3\sqrt{p_0 C_1}$ we obtain the desired statement (F.4). □

G Details of the Simulation Study

Three of the test functions we used have been investigated before in [29]:

(1) $f_0(x) = 3\sin[4/(x+0.2)] + 1.5$ according to Example 1 followed from [29],

(2) $f_0(x)$ is a simulated Brownian motion on $[0, 1/2)$ (a cumulative sum of iid increments $5 \times \mathcal{N}(0, 1/\sqrt{n})$) and a constant on $[1/2, 1]$,

(3) the “Bumps" test function from [29].

(4) the “Blocks" test function from [29].

The cases (3) and (4) exhibit substantial spatial inhomogeneity and should best showcase the benefits of our locally-adaptive bands. The case (1) is relatively smooth and thereby smooth estimation methods (such as Symmlets [14] or local polynomials [19]) have clear advantages over Haar wavelets when approximating the true signal. As will be seen from the plots, however, these global adaptation methods adapt to the worse regularity, under-smoothing large portions of the signal. Due to adaptive placement of the splits (compared to the binscatter [19]), our method is very competitive and performs well in terms of coverage. We elaborate on the Doppler example below.

Example G.1. *(Doppler Curve) Similarly as in [30] and [71], we generate* $n = 2048$ *observations from a noisy Doppler curve* (2.1) *with* $f_0(x) = 3\sin[4/(x+0.2)] + 1.5$ *and* $\sigma = 1$ *with* $x_i = i/n$. *This function has heterogeneous smoothness which cannot be captured with prototypical global smoothing methods such as global kernel regression (Figure 1 on the left) which leads to over/undersmoothing depending on the choice of a fixed kernel width. Tree-based methods, such as Bayesian CART, are better suited for this task by placing the splits more often in areas where the function is less smooth (Figure 1 on the right).*

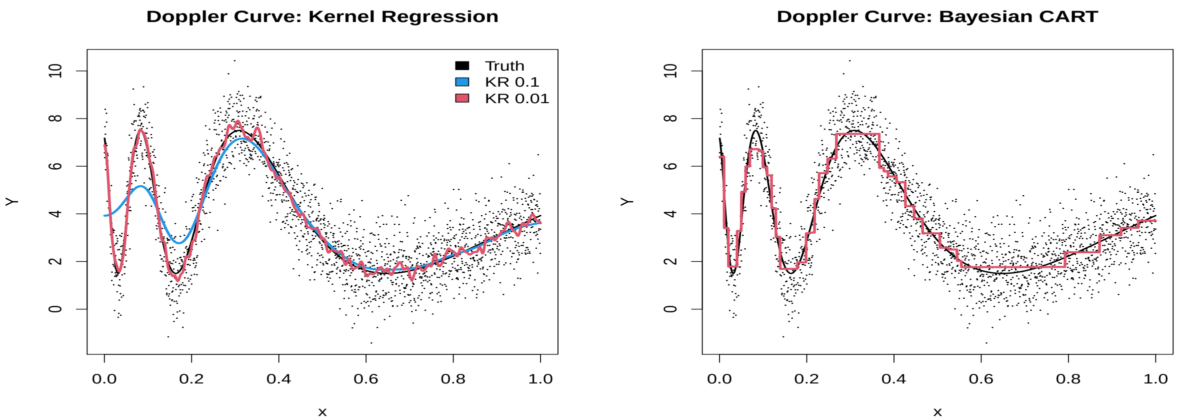

Figure 1: Doppler curve and (left) kernel regression estimates (`ksmooth` in R) with a bandwidth 0.1 (capturing well the flat part) and 0.01 (capturing well the wiggly part), (right) Bayesian CART posterior mean fit (`wbart` in the R package `BART`).

We implemented the Bayesian (dyadic) CART (Section 3.1) as well as the Spike-and-Slab wavelet reconstruction (Section 3.2) using the Metropolis-Hastings (MH) algorithm. Dyadic Bayesian CART is implemented according to [20] with a proposal distribution consisting of two steps: grow (splitting a randomly chosen bottom node) and prune (collapsing two children bottom nodes into one). The implementation is fairly straightforward due to immediate access to the posterior tree probabilities. For the spike-and-slab prior (the point-mass mixture version [41]), we use a one-site proposal for adding and removing one wavelet coefficient at a time. We run our Bayesian CART procedure with a split probability $p_l = a(1/\Gamma)^l$ with $a = 0.95$ and $\Gamma = 1.001$ which resulted in the MH acceptance rate in between $15\% - 25\%$. For the point-mass spike-and-slab prior, we used a split probability $p_l = a(1/\Gamma)^l$ with $a = 0.95$ and $\Gamma = 2$. This choice again resulted in the MH acceptance rate in between $15 - 20\%$. We found that it is important to penalize the inclusion of deep coefficients in order to prevent from erratic inclusion of spurious high-resolution signals. This is why the inclusion probability is smaller for deeper coefficients than in the Bayesian CART, where the tree has to grow into the deeper signal. We found that the tree-shaped regularization has smoothing benefits compared to the spike-and-slab prior which may decide to include deeper wavelet coefficients without including the ancestors. This results in less smooth reconstructions and wider confidence bands for the spike-and-slab prior. We simulated $M = 5\,000$ posterior samples for both Bayesian CART and spike-and-slab and summarized them after a 1000 burn-in period. All of the procedures we chose for comparisons estimate the residual variance σ^2. While in our theory we set it equal to one, in our implementations we treat is as unknown with the traditional inverse gamma (IG) prior (shape and rate equal to 1/2 as in [**?**]).

We construct our confidence bands according to (3.8) using an adaptive choice of v_n in (3.7) using posterior information. In particular, we choose v_n to be the smallest number that yields a band that contains $(1-\alpha)\%$ of posterior draws. We implement this optimization in practice by taking a fine grid of values $v_n = \{0.5 + k \times 0.005 : 1 \leq k \leq 100\}$ and computing the amount of simulated posterior probability contained in the set for each value on the grid. We have chosen $\alpha \in \{0.05, 0\}$ for the locally adaptive bands in our simulations. We denote these two bands with $\mathcal{C}_n^1$ (with $\alpha = 0.05$) and $\mathcal{C}_n^2$ (with $\alpha = 0$) in our tables. In addition, we compare our bands with the locally *non-adaptive* band [18] which uses $\sup_{x\in[0,1]} \sigma_n(x)$ as the locally non-adaptive diameter in (3.8). Again, we choose v_n adaptively so that the band contains $(1-\alpha)\%$ of the posterior probability. We denote this band by $\widetilde{\mathcal{C}}_n$ using $\alpha = 0.05$ in our tables. This band is a direct relative to the globally adaptive construction in [14] where the global level of truncation is estimated by performing tests on individual wavelet coefficients. We included this globally adaptive (non-locally adaptive) band in our comparisons. We considered this band to be one of the closest non-Bayesian counterparts to our approach in the literature. We used authors' Matlab code which implements a Symmlet 8 basis with default tuning parameter options ($\beta_0 = 3$ and $M_0 = 100$). We denote this method by CLM in our tables, using $\alpha = 0.05$. Next, we compare our bands to $(1-\alpha)\%$ credible L_∞ bands centered at the posterior mean estimator $\widehat{f}$ (i.e. $\{f : \sup_{x\in[0,1]} |\, f(x) - \widehat{f}(x)\,| \ \leq R_\alpha\}$, where R_α is the $(1-\alpha)\%$ sample quantile of $\max_{x\in\mathcal{X}} |\, f_i(x) - \widehat{f}(x)\,|$ where f_i for $1 \leq i \leq M$ are the posterior samples of f and $\mathcal{X} = \{x_i : 1 \leq i \leq n\}$. We denote this band by L_∞ in our tables with $\alpha = 0.05$. This construction is locally non-adaptive. The L_∞ band is somewhat similar to the multiscale credible band in [18] but its coverage properties are not theoretically understood. In our comparisons, we also included two point-wise bands. One natural candidate is the pointwise $(1-\alpha)\%$ credible bands obtained directly from our posterior output. The pointwise credible bands are denoted by $\mathcal{P}_n$ in our tables using $\alpha = 0.05$. Next, we also include the (pointwise) bands implemented in an R package `nprobust` [15]. This is a recent package which implements robust bias-corrected bands for inference in non-parametric regression using local polynomial regression. We have used their default settings. This method is denoted by CCF in our tables. Another natural method for comparisons is the regressogram (or binscatter [19]) which we implemented using piece-wise step functions with non-adaptive placement of splits (option $p = s = 0$ in the R package `binsreg` and BIN1 in our tables). This histogram implementation is the closest frequentist non-adaptive counterpart where the splits are on a regular grid and not data-driven. We also considered the recommended default option $p = s = 2$ (method BIN2 in our tables) based on smoother local polynomials with a global smoothing penalty across the bins. This package provides confidence bands based on bias correction and adaptive selection of the number of bins.

For each method, we evaluated the coverage of $f_0(x)$ at all design points $x_i \in \mathcal{X}$ (regular grid $x_i = i/n$ for $n = 2^{10}$). We report the average proportion of non-covered points (averaged after 100 repetitions). Next, we look into the average band size (both average size over all design points as well as minimal and maximal width over design points). In addition, we keep track of the estimation error of the point (centering) estimator. This is the median estimator for $\mathcal{C}_n, \widetilde{\mathcal{C}}_n$, the posterior mean estimator for L_∞ and $\mathcal{P}_n$, the centering point for CLM, the point estimator of the regression function based on local polynomial of order p estimated by their default method for CCF and the centering point of the binscatter bands for BIN1 and BIN2.

The results are summarized in Table 1 in the main manuscript. Our adaptive band constructions $\mathcal{C}_n^1$ (v_n chosen with $\alpha = 0.05$) and $\mathcal{C}_n^2$ (v_n chosen with $\alpha = 0$) perform very well in terms of the average percentage of non-covered points. The comparisons are particularly striking in the bumps and block examples where the competing methods (as well as point-wise credible intervals $\mathcal{P}_n$ and CCF) grossly under-cover. The performance of the L_∞ band is also very good but it is not locally adaptive and, again, there is no theoretical justification. The non-adaptive band $\widetilde{\mathcal{C}}_n$ from [18] with an adaptively chosen v_n ($\alpha = 0.05$) also performs well, but it may be unnecessarily wide. Comparisons of Bayesian CART with Spike-and-Slab priors are quite interesting. Tree-shaped regularization may be beneficial when the signal itself has a hierarchical tree structure. With hierarchically separated higher-resolution signals, spike-and-slab priors are more likely to mix better and capture these signals. With a tree prior, one may need to initiate MCMC at richer (deeper) trees so that the trees can grow into the signal throughout the computation. For smoother signals, on the other hand, spike-and-slab priors may include too many spurious high-resolution coefficients, causing the bands to widen.

It is interesting to compare the various band constructions visually. Figure 2 shows one realization for signals (1) and (2), Figure 3 shows one realization for signals (3) and (4). For example, the binscatter with a step function (BIN1 method) build on a regular partition does not achieve uniform coverage. This is in line with the conclusion in [18] (Theorem 5) showing that regular (equispaced) partitions fail to achieve minimax ℓ_∞ adaptation.

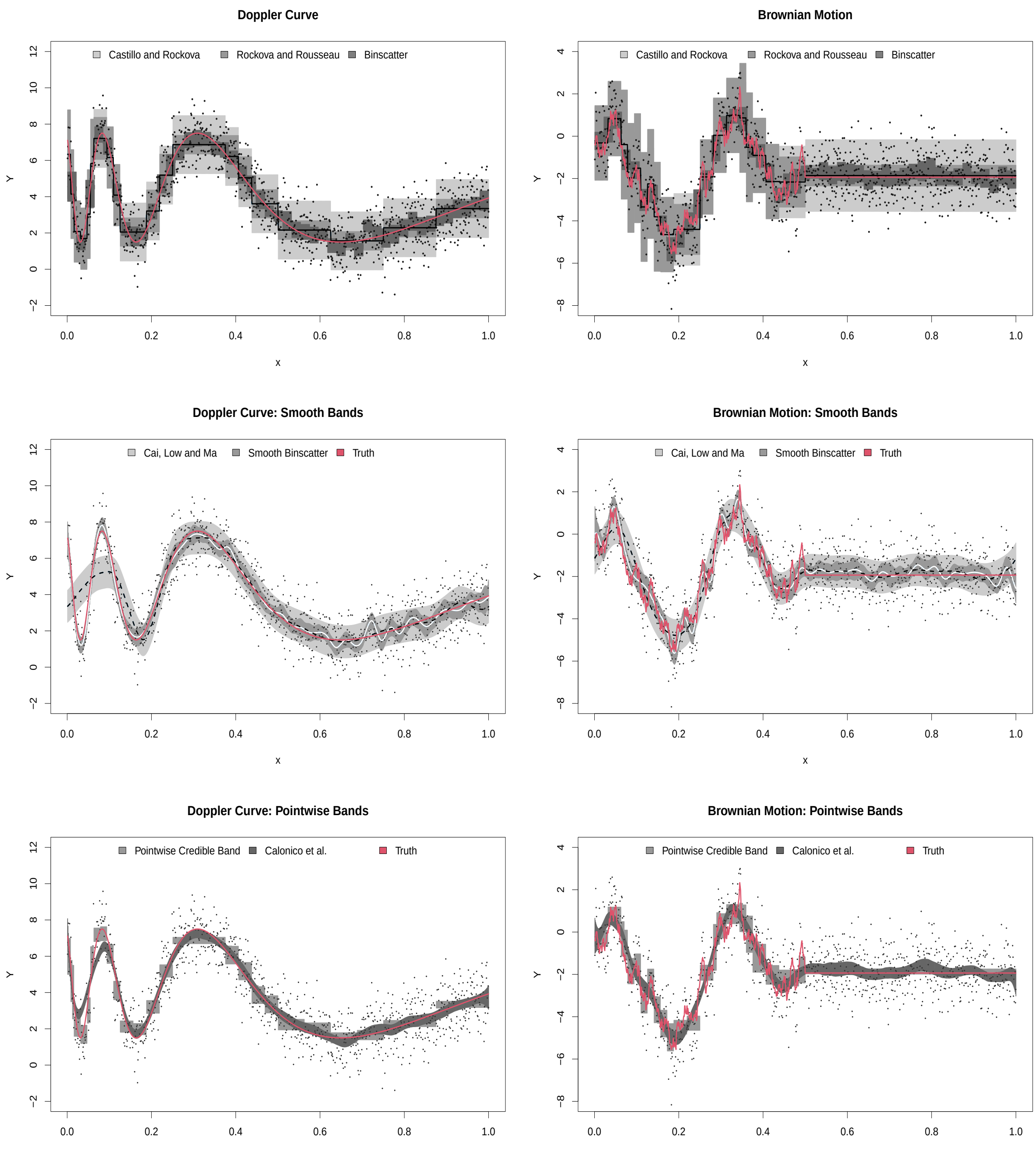

Figure 2: Plots of recovered bands with true curve marked in red color. The top panel displays (a) the non-adaptive band of [18] with an adaptively chosen v_n so that the band contains 95% posterior probability, (b) our adaptive band with an adaptively chosen v_n so that the band contains 95% posterior probability and the binscatter bands with $s = p = 0$. The black broken line is the posterior median estimator. The middle panel displays smooth bands together with their centerings obtained with (a) symmlet 8 basis ([14] with $\alpha = 0.05$) and (b) the smooth binscatter [19] with $s = p = 2$. The bottom panel displays point-wise bands: (a) pasted 95% posterior credible intervals and the bands in [15] with $\alpha = 0.05$.

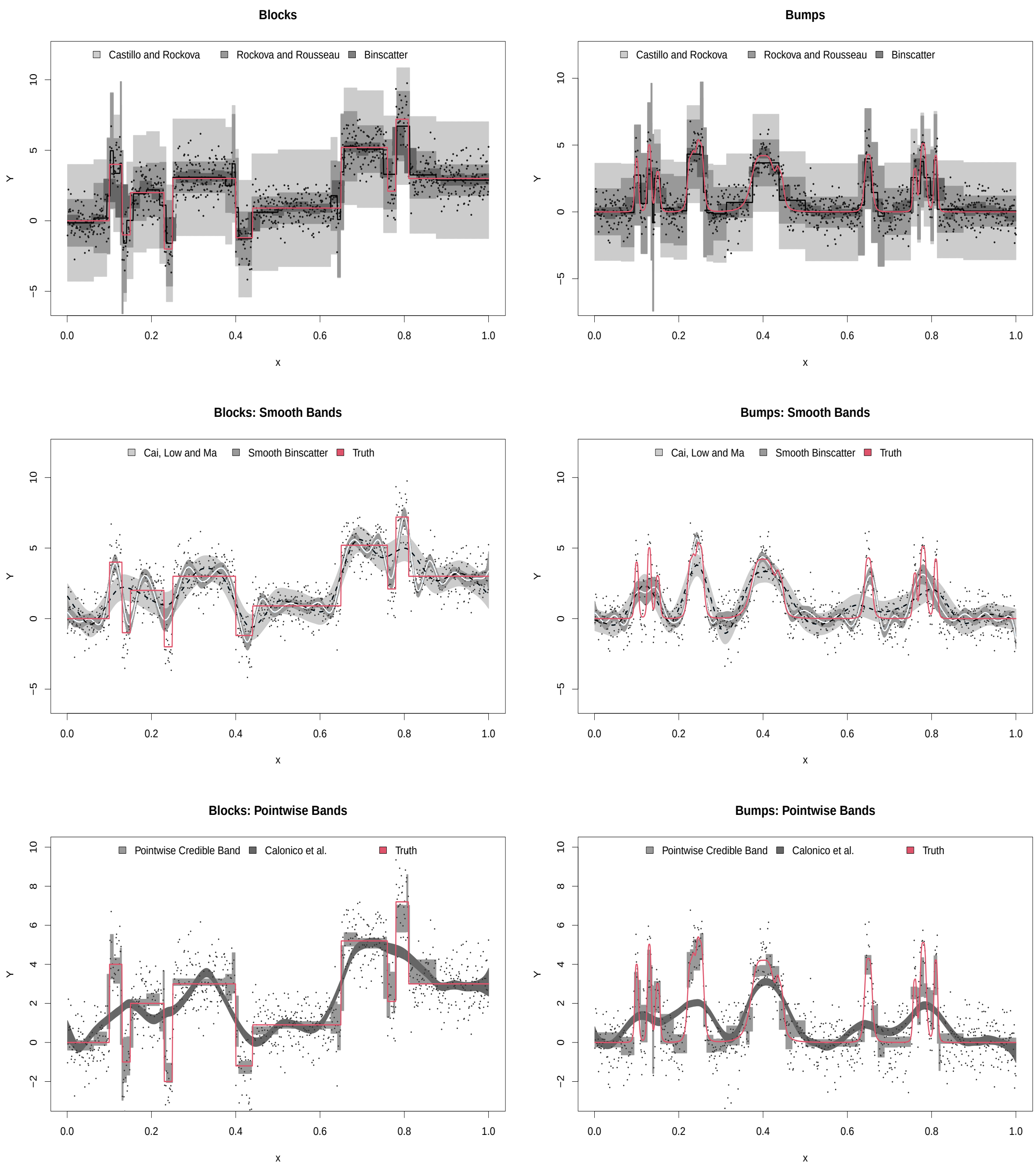

Figure 3: Plots of recovered bands with true curve marked in red color. The top panel displays (a) the non-adaptive band of [18] with an adaptively chosen v_n so that the band contains 95% posterior probability, (b) our adaptive band with an adaptively chosen v_n so that the band contains 95% posterior probability and the binscatter bands with $s = p = 0$. The black broken line is the posterior median estimator. The middle panel displays smooth bands together with their centerings obtained with (a) symmlet 8 basis ([14] with $\alpha = 0.05$) and (b) the smooth binscatter [19] with $s = p = 2$. The bottom panel displays point-wise bands: (a) pasted 95% posterior credible intervals and the bands in [15] with $\alpha = 0.05$.

References

[14] Cai, T., M. Low, and Z. Ma (2014). Adaptive confidence bands for non-parametric regression functions. *Journal of the American Statistical Association 109*, 1054–1070.

[15] Calonico, S., M. D. Cattaneo, and M. H. Farrell (2019). nprobust: Nonparametric kernel-based estimation and robust bias-corrected inference. *Journal of Statistical Software 91*(8), 1–33.

[18] Castillo, I. and V. Ročková (2020). Uncertainty quantification for Bayesian CART. *The Annals of Statistics (to appear).*

[19] Cattaneo, M., R. Crump, M. Farrell, and Y. Feng (2019). On binscatter. *ArXiv*.

[20] Chipman, H., E. George, and R. McCulloch (1997). Bayesian CART model search. *Journal of the American Statistical Association 93*, 935–960.

[29] Donoho, D. and I. Johnstone (1995a). Ideal spatial adaptation by wavelet shrinkage. *Biometrika 81*, 425–455.

[30] Donoho, D. and I. Johnstone (1995b). Wavelet shrinkage: Asymptopia? *Journal of the Royal Statistical Society. Series B 57*, 301–369.

[] George, E. and R. McCulloch (1997). Approaches for Bayesian variable selection. *Statistica Sinica 7*, 339–373.

[] Gine, E. and R. Nickl (2011). On adaptive inference and confidence bands. *The Annals of Statistics 39*, 2383–2409.

[41] Hoffmann, M., J. Rousseau, and J. Schmidt-Hieber (2015). On adaptive posterior concentration rates. *The Annals of Statistics 43*, 2259–2295.

[] Horn, R. and C. Johnson (2013). *Matrix Analysis, second edition.*

[] Kuelbs, J. and W. Li (1993). Metric entropy and the small ball problem for gaussian measures. *Journal of Functional Analysis 116*, 133–157.

[60] Rousseau, J. and B. Szabo (2017). Asymptotic behavior of the empirical Bayes posteriors associated to maximum marginal likelihood estimator. *The Annals of Statistics 45*, 833–865.

[] Ročková, V. and J. Rousseau (2021). Ideal bayesian spatial adaptation. *Supplemental Materials*, 1–22.

[71] Tibshirani, R. (2014). Adaptive piecewise polynomial estimation via trend filtering. *Annals of Statistics 42*, 285–323.

[] Vershynin, R. (2020). *High-Dimensional Probability An Introduction with Applications in Data Science*. Springer.