Inference with Mondrian Random Forests

Supplementary material

Matias D. Cattaneo[1] Jason M. Klusowski[1] William G. Underwood[2]

November 6, 2025

Contents

[1]Department of Operations Research and Financial Engineering, Princeton University, USA
[2]Statistical Laboratory, University of Cambridge, UK

SA1 Proofs and technical results

In this section we present the full proofs of all our results, and also state some useful technical preliminary and intermediate lemmas. We use the following simplified notation for convenience, whenever it is appropriate: write $\mathbb{I}_{ib}(x) = \mathbb{I}\{X_i \in T_b(x)\}$ and $N_b(x) = \sum_{i=1}^n \mathbb{I}_{ib}(x)$, as well as $\mathbb{I}_b(x) = \mathbb{I}\{N_b(x) \geq 1\}$. We use C to denote a positive constant whose value may change from line to line, and write $a_n = O(b_n)$ for $a_n \lesssim b_n$. We begin with an overview of the main proof strategies and a discussion of the challenges involved in Section SA1.1. We then give some preliminary lemmas in Section SA1.2, and present the proofs for Section 3 (including Lemma 1, Lemma 2, Theorem 1, Theorem 2, Lemma 3, and Theorem 3) in Section SA1.3; the proofs for Section 4 (including Lemma 4, Lemma 5, Theorem 4, Theorem 5, Lemma 6, and Theorem 6) in Section SA1.4; and the proofs for Section 6 (including Lemma 7 and Lemma 8) in Section SA1.5.

SA1.1 Overview of proof strategies

This section provides some insight into the general approach we use to establish our main results. We focus on the technical innovations forming the core of our arguments, and refer the reader to the upcoming sections for full proofs.

SA1.1.1 Preliminary technical results

The starting point for our proofs is a result characterizing the distribution of the shape of a Mondrian cell $T(x)$. This property is a consequence of the fact that the restriction of a Mondrian process to a subcell remains a Mondrian process [Mourtada, Gaïffas, and Scornet, 2020]. We have

$$|T(x)_j| = \left(\frac{E_{j1}}{\lambda} \wedge x_j\right) + \left(\frac{E_{j2}}{\lambda} \wedge (1 - x_j)\right)$$

for all $1 \leq j \leq d$, recalling that $T(x)_j$ is the side of the cell $T(x)$ aligned with axis j, and where E_{j1} and E_{j2} are mutually independent $\mathrm{Exp}(1)$ random variables. Our assumptions that $x \in (0,1)$ and $\lambda \to \infty$ mean that the "boundary terms" x_j and $1 - x_j$ are eventually ignorable and so $|T(x)_j| = (E_{j1} + E_{j2})/\lambda$ with high probability. Controlling the size of the largest cell in the forest containing x is now straightforward with a union bound, giving

$$\max_{1 \leq b \leq B} \max_{1 \leq j \leq d} |T_b(x)_j| \lesssim_{\mathbb{P}} \frac{\log B}{\lambda}.$$

This shows that, up to logarithmic terms, none of the cells in the forest at x are significantly larger than average, ensuring that the Mondrian random forest estimator is "localized" around x on the scale of $1/\lambda$, an important property for our bias characterization.

Having provided upper bounds for the sizes of Mondrian cells, we also must establish some lower bounds in order to ensure a sufficient effective sample size and to quantify the "small cells" phenomenon mentioned previously. The first step towards this is to bound the first two moments of the truncated inverse Mondrian cell volume; we show that

$$\mathbb{E}\left[1 \wedge \frac{1}{n|T(x)|}\right] \asymp \frac{\lambda^d}{n} \qquad \text{and} \qquad \mathbb{E}\left[1 \wedge \frac{1}{n^2|T(x)|^2}\right] \asymp \frac{\lambda^{2d}(\log n)^d}{n^2}.$$

These bounds are computed using the exact distribution of $|T(x)|$. Note that $\mathbb{E}\left[1/|T(x)|^2\right] = \infty$ because $1/(E_{j1} + E_{j2})$ has only $2 - \delta$ finite moments, so the truncation is crucial here. Since we have "almost two" moments, this truncation is at the expense of only a logarithmic term. Nonetheless, third and higher truncated moments will not enjoy such tight bounds, demonstrating both the fragility of this result and the inadequacy of tools such as the Lyapunov central limit theorem which require $2 + \delta$ marginal moments.

To conclude this investigation into the "small cell" phenomenon, we apply the previous bounds to ensure that the empirical effective sample sizes $N_b(x) = \sum_{i=1}^n \mathbb{I}\left\{X_i \in T_b(x)\right\}$ are approximately of the order n/λ^d in an appropriate sense; we demonstrate that

$$\mathbb{E}\left[\frac{\mathbb{I}\{N_b(x) \geq 1\}}{N_b(x)}\right] \lesssim \frac{\lambda^d}{n} \qquad \text{and} \qquad \mathbb{E}\left[\frac{\mathbb{I}\{N_b(x) \geq 1\}}{N_b(x)^2}\right] \lesssim \frac{\lambda^{2d}(\log n)^d}{n^2},$$

as well as "mixed" bounds $\mathbb{E}\left[\mathbb{I}\{N_b(x) \geq 1\}\mathbb{I}\{N_{b'}(x) \geq 1\}/(N_b(x)N_{b'}(x))\right] \lesssim \lambda^{2d}/n^2$ when $b \neq b'$, which arise from covariance terms across multiple trees. The proof of this result is involved and technical, and proceeds by induction. The idea is to construct a class of subcells by taking all possible intersections of the cells in T_b and $T_{b'}$ (we show two trees here for clarity; there may be more) and noting that each $N_b(x)$ is the sum of the number of points in each such "refined cell" intersected with $T_b(x)$. We then swap out each refined cell one at a time and replace the number of data points it contains with its volume multiplied by $nf(x)$, showing that the expectation on the left hand side does not increase too much using a moment bound for inverse binomial random variables based on Bernstein's inequality. By induction and independence of the trees, eventually the problem is reduced to computing moments of truncated inverse Mondrian cell volumes, as above.

SA1.1.2 Bias characterization

Our first substantial result is the bias characterization given as Lemma 1, in which we precisely characterize the probability limit of the conditional bias

$$\mathbb{E}\left[\hat{\mu}(x) \mid \mathbf{X}, \mathbf{T}\right] - \mu(x) = \frac{1}{B}\sum_{b=1}^{B}\sum_{i=1}^{n}\left(\mu(X_i) - \mu(x)\right)\frac{\mathbb{I}\{X_i \in T_b(x)\}}{N_b(x)}.$$

The first step in this proof is to pass to the "infinite forest" limit by taking an expectation conditional on $\mathbf{X}$, or equivalently marginalizing over $\mathbf{T}$, applying the conditional Markov inequality to see

$$\left|\mathbb{E}\left[\hat{\mu}(x) \mid \mathbf{X}, \mathbf{T}\right] - \mathbb{E}\left[\hat{\mu}(x) \mid \mathbf{X}\right]\right| \lesssim_{\mathbb{P}} \frac{1}{\lambda^{1 \wedge \beta_\mu} \sqrt{B}}.$$

While this may seem a crude approximation, it is already known that fixed-size Mondrian forests have suboptimal bias properties when compared to forests with a diverging number of trees. In fact, when $\beta \geq 1$, the error $1/\left(\lambda^{1 \wedge \beta} \sqrt{B}\right)$ exactly accounts for the first-order bias of individual Mondrian trees [Mourtada et al., 2020].

Next we show that $\mathbb{E}\left[\hat{\mu}(x) \mid \mathbf{X}\right]$ converges in probability to its expectation, using the Efron–Stein theorem to handle this non-linear function of the i.i.d. variables X_i. The important insight here is that replacing a sample X_i with an independent copy $\tilde{X}_i$ can change the value of $N_b(x)$ by at most one. Further, this can happen only on the event $\{X_i \in T_b(x)\} \cup \{\tilde{X}_i \in T_b(x)\}$, which occurs with probability on the order $1/\lambda^d$ (the expected cell volume) for each tree $1 \leq b \leq B$. The Hölder property of μ and the upper bound on the maximum cell size then give $|\mu(X_i) - \mu(x)| \lesssim \max_{1 \leq j \leq d} |T_b(x)_j|^{1 \wedge \beta_\mu} \lesssim_{\mathbb{P}} 1/\lambda^{1 \wedge \beta_\mu}$ whenever $X_i \in T_b(x)$, so we combine this with moment bounds for the denominator $N_b(x)$ to see

$$\left|\mathbb{E}\left[\hat{\mu}(x) \mid \mathbf{X}\right] - \mathbb{E}\left[\hat{\mu}(x)\right]\right| \lesssim_{\mathbb{P}} \frac{1}{\lambda^{1 \wedge \beta_\mu}} \sqrt{\frac{\lambda^d}{n}}.$$

The next step is to approximate the resulting non-random bias $\mathbb{E}\left[\hat{\mu}(x)\right] - \mu(x)$ as a polynomial in $1/\lambda$. To this end, we firstly apply a concentration-type result for the binomial distribution to deduce that

$$\mathbb{E}\left[\frac{\mathbb{I}\{N_b(x) \geq 1\}}{N_b(x)} \;\middle|\; \mathbf{T}\right] \approx \frac{1}{n \int_{T_b(x)} f(s) \, \mathrm{d}s}$$

in an appropriate sense, and hence, by conditioning on $\mathbf{T}$ and $\mathbf{X}$ without X_i,

$$\mathbb{E}\left[\hat{\mu}(x)\right] - \mu(x) \approx \mathbb{E}\left[\frac{\int_{T_b(x)} (\mu(s) - \mu(x)) f(s) \, \mathrm{d}s}{\int_{T_b(x)} f(s) \, \mathrm{d}s}\right]. \tag{SA1}$$

Next we apply the multivariate version of Taylor's theorem to the integrands in both the numerator and the denominator in (SA1), and then apply the Maclaurin series of $1/(1+x)$ and the multinomial theorem to recover a single polynomial in $1/\lambda$. The error term is on the order of $1/\lambda^\beta$ and depends on the smoothness of μ and f, and the polynomial coefficients are given by various expectations involving exponential random variables. The final step is to verify using symmetry of Mondrian cells that all the odd monomial coefficients are zero, and to calculate some explicit examples of the form of the limiting bias.

SA1.1.3 Central limit theorem

To prove our second main result (Theorem 2), we apply a version of the Berry–Esseen theorem for i.n.i.d. random variables, conditional on $(\mathbf{X}, \mathbf{T})$, which only requires $2+\delta$ moments. Define the variables

$$S_i(x) = \sqrt{\frac{n}{\lambda^d}} \frac{1}{B} \sum_{b=1}^{B} \frac{\mathbb{I}\{X_i \in T_b(x)\} \varepsilon_i}{N_b(x)},$$

which are independent and zero-mean given $(\mathbf{X}, \mathbf{T})$, and further satisfy

$$\sqrt{\frac{n}{\lambda^d}} \left(\hat{\mu}(x) - \mathbb{E}\left[\hat{\mu}(x) \mid \mathbf{X}, \mathbf{T} \right] \right) = \sum_{i=1}^{n} S_i(x).$$

Thus by Petrov [1995, Theorem 5.7], conditional on $(\mathbf{X}, \mathbf{T})$, taking a marginal expectation,

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P}\left(\tilde{\Sigma}(x)^{-\frac{1}{2}} \sum_{i=1}^{n} S_i \le t \right) - \Phi(t) \right| \lesssim \mathbb{E}\left[1 \wedge \left(\tilde{\Sigma}(x)^{-\frac{2+(\delta \wedge 1)}{2}} \sum_{i=1}^{n} \mathbb{E}\left[|S_i|^{2+(\delta \wedge 1)} \mid \mathbf{X}, \mathbf{T} \right] \right) \right].$$

Bounding the right-hand side now reduces to establishing properties of $\tilde{\Sigma}(x)$ and its large-sample limit $\Sigma(x)$. To this end, we again use the Efron–Stein theorem to bound $\mathrm{Var}\left[\tilde{\Sigma}(x)\right]$ and then apply a careful sequence of approximations to control $\mathbb{E}\left[\tilde{\Sigma}(x)\right] - \Sigma(x)$. The final task is to calculate the limiting variance $\Sigma(x)$. It is a straightforward but tedious exercise to verify that each denominator $N_b(x)$ can be replaced by $nf(x)|T_b(x)|$, yielding

$$\Sigma(x) = \frac{\sigma^2(x)}{f(x)} \lim_{\lambda \to \infty} \frac{1}{\lambda^d} \mathbb{E}\left[\frac{|T_b(x) \cap T_{b'}(x)|}{|T_b(x)|\,|T_{b'}(x)|} \right] = \frac{\sigma^2(x)}{f(x)} \left(\mathbb{E}\left[\frac{(E_1 \wedge E_1') + (E_2 \wedge E_2')}{(E_1 + E_2)(E_1' + E_2')} \right] \right)^d,$$

where E_1, E_2, E_1', and E_2' are independent $\mathrm{Exp}(1)$, by the cell shape distribution and independence of the trees. This final expectation is calculated by integration, using various incomplete gamma function identities.

SA1.1.4 Confidence intervals

While Theorem 2 gives a distributional approximation for the infeasible t-statistic, in order to construct confidence intervals we must instead approximate the corresponding feasible t-statistic. To do this, first observe that if τ and $\hat{\tau}$ are real-valued random variables and $\varepsilon > 0$, then the following anti-concentration inequality holds:

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P}(\hat{\tau} \le t) - \Phi(t) \right| \le \sup_{t \in \mathbb{R}} \left| \mathbb{P}(\tau \le t) - \Phi(t) \right| + \varepsilon \sqrt{2/\pi} + \mathbb{P}(|\hat{\tau} - \tau| > \varepsilon).$$

We apply this result to

$$\hat{\tau} = \sqrt{\frac{n}{\lambda^d}} \left(\frac{\hat{\mu}(x) - \mu(x)}{\sqrt{\hat{\Sigma}(x)}} - \frac{\mathbb{E}\left[\hat{\mu}(x)\right] - \mu(x)}{\sqrt{\Sigma(x)}} \right) \qquad \text{and} \qquad \tau = \sqrt{\frac{n}{\lambda^d}} \frac{\hat{\mu} - \mathbb{E}\left[\hat{\mu}(x) \mid \mathbf{X}, \mathbf{T}\right]}{\sqrt{\tilde{\Sigma}(x)}},$$

bounding $\mathbb{P}(|\hat{\tau} - \tau| > \varepsilon)$ using our established results on $\hat{\mu}(x)$, $\tilde{\Sigma}(x)$ and $\hat{\Sigma}(x)$. Exploiting symmetry of the resulting confidence interval permits a quadratic dependence of the coverage error on the bias.

SA1.2 Preliminary lemmas

We begin by bounding the maximum size of any cell in a Mondrian forest containing x. This result is used regularly throughout many of our other proofs, and captures the "localizing" behavior of the Mondrian random forest estimator, showing that Mondrian cells have side lengths at most on the order of $1/\lambda$. For distributions P_1 and P_2 on $\mathbb{R}$, we write $P_1 \leq P_2$ if P_2 stochastically dominates P_1; that is, if there exist random variables $X_1 \sim P_1$ and $X_2 \sim P_2$ such that $X_1 \leq X_2$ almost surely. Likewise, if X_1 is a real-valued random variable and P_2 is a law on $\mathbb{R}$, we write $X_1 \leq P_2$ if there exists $X_2 \sim P_2$ such that $X_1 \leq X_2$ almost surely, on a possibly enlarged probability space. Observe that if $n_1, n_2 \in \mathbb{N}$ with $n_1 \leq n_2$ and $p_1, p_2 \in [0,1]$ with $p_1 \leq p_2$, then $\mathrm{Bin}(n_1, p_1) \leq \mathrm{Bin}(n_2, p_2)$.

Lemma SA1 (Upper bound on the largest cell in a Mondrian forest)**.** *Let $T_1, \ldots, T_B \sim \mathcal{M}([0,1]^d, \lambda)$ and take $x \in (0,1)^d$. Then for all $t > 0$*

$$\mathbb{P}\left(\max_{1 \leq b \leq B} \max_{1 \leq j \leq d} |T_b(x)_j| \geq \frac{t}{\lambda} \right) \leq 2dBe^{-t/2}.$$

Proof (Lemma SA1)**.** We use the explicit distribution of the shape of Mondrian cells given by Mourtada et al. [2020, Proposition 1]. In particular, we have $|T_b(x)_j| = \left(\frac{E_{bj1}}{\lambda} \wedge x_j \right) + \left(\frac{E_{bj2}}{\lambda} \wedge (1 - x_j) \right)$ where E_{bj1} and E_{bj2} are independent $\mathrm{Exp}(1)$ random variables for $1 \leq b \leq B$ and $1 \leq j \leq d$. Thus $|T_b(x)_j| \leq \frac{E_{bj1} + E_{bj2}}{\lambda}$ and so by a union bound

$$\begin{aligned}
\mathbb{P}\left(\max_{1 \leq b \leq B} \max_{1 \leq j \leq d} |T_b(x)_j| \geq \frac{t}{\lambda} \right) &\leq \mathbb{P}\left(\max_{1 \leq b \leq B} \max_{1 \leq j \leq d} (E_{bj1} \vee E_{bj2}) \geq \frac{t}{2} \right) \\
&\leq 2dB \, \mathbb{P}\left(E_{bj1} \geq \frac{t}{2} \right) \leq 2dBe^{-t/2}.
\end{aligned}$$

$\square$

The next result is another "localization" result, this time showing that the union over the forest of the cells $T_b(x)$ containing x do not contain "too many" samples X_i. In other words, the Mondrian random forest estimator fitted at x should only depend on n/λ^d (the effective sample size) data points up to logarithmic terms.

Lemma SA2 (Upper bound on the number of active data points)**.** *Suppose that Assumptions 1 and 2 hold and define $N_{\cup}(x) = \sum_{i=1}^n \mathbb{I}\left\{ X_i \in \bigcup_{b=1}^B T_b(x) \right\}$. Then for $t > 0$ and $n \geq \lambda^d$, with*

$\|f\|_\infty = \sup_{x \in [0,1]^d} f(x)$,

$$\mathbb{P}\left(N_\cup(x) > t^{d+1} \frac{n}{\lambda^d} \|f\|_\infty\right) \leq 4dBe^{-t/4}.$$

Proof (Lemma SA2)**.** Note that

$$N_\cup(x) \sim \mathrm{Bin}\left(n, \int_{\bigcup_{b=1}^B T_b(x)} f(s)\,\mathrm{d}s\right) \leq \mathrm{Bin}\left(n, 2^d \max_{1 \leq b \leq B} \max_{1 \leq j \leq d} |T_b(x)_j|^d \|f\|_\infty\right)$$

conditionally on $\mathbf{T}$. If $N \sim \mathrm{Bin}(n,p)$ then, by Bernstein's inequality, $\mathbb{P}\left(N \geq (1+t)np\right) \leq \exp\left(-\frac{t^2 n^2 p^2/2}{np(1-p)+tnp/3}\right) \leq \exp\left(-\frac{3t^2 np}{6+2t}\right)$. Thus for $t \geq 2$,

$$\mathbb{P}\left(N_\cup(x) > (1+t)n \frac{2^d t^d}{\lambda^d} \|f\|_\infty \;\Big|\; \max_{1 \leq b \leq B} \max_{1 \leq j \leq d} |T_j(x)| \leq \frac{t}{\lambda}\right) \leq \exp\left(-\frac{2^d t^d n}{\lambda^d}\right).$$

By Lemma SA1, $\mathbb{P}\left(\max_{1 \leq b \leq B} \max_{1 \leq j \leq d} |T_j(x)| > \frac{t}{\lambda}\right) \leq 2dBe^{-t/2}$. Hence

$$\begin{aligned}
&\mathbb{P}\left(N_\cup(x) > 2^{d+1} t^{d+1} \frac{n}{\lambda^d} \|f\|_\infty\right) \\
&\quad\leq \mathbb{P}\left(N_\cup(x) > 2tn \frac{2^d t^d}{\lambda^d} \|f\|_\infty \;\Big|\; \max_{1 \leq b \leq B} \max_{1 \leq j \leq d} |T_j(x)| \leq \frac{t}{\lambda}\right) + \mathbb{P}\left(\max_{1 \leq b \leq B} \max_{1 \leq j \leq d} |T_j(x)| > \frac{t}{\lambda}\right) \\
&\quad\leq \exp\left(-\frac{2^d t^d n}{\lambda^d}\right) + 2dBe^{-t/2}.
\end{aligned}$$

Noting the result is trivial for $t < 4$ and replacing t by $t/2$ gives that for $n \geq \lambda^d$,

$$\mathbb{P}\left(N_\cup(x) > t^{d+1} \frac{n}{\lambda^d} \|f\|_\infty\right) \leq 4dBe^{-t/4}.$$

□

Next we give a series of results culminating in a generalized moment bound for the denominator appearing in the Mondrian random forest estimator. We begin by providing a moment bound for the truncated inverse binomial distribution, which will be useful for controlling $\frac{\mathbb{I}_b(x)}{N_b(x)} \leq 1 \wedge \frac{1}{N_b(x)}$ because conditional on T_b we have $N_b(x) \sim \mathrm{Bin}\left(n, \int_{T_b(x)} f(s)\,\mathrm{d}s\right)$. Our constants could be suboptimal but they are sufficient for our applications.

Lemma SA3 (An inverse moment bound for the binomial distribution)**.** *For $n \geq 1$ and $p \in [0,1]$, let $N \sim \mathrm{Bin}(n,p)$. Take $a_1, \ldots, a_k \geq 0$ and $l_1, \ldots, l_k \geq 1$. Then with $L = \sum_{j=1}^k l_j$,*

$$\mathbb{E}\left[\prod_{j=1}^k \left(1 \wedge \frac{1}{N + a_j}\right)^{l_j}\right] \leq (9L)^{2L} \prod_{j=1}^k \left(1 \wedge \frac{1}{np + a_j}\right)^{l_j}.$$

Proof (Lemma SA3)**.** By Bernstein's inequality, $\mathbb{P}\left(N \leq np - t\right) \leq \exp\left(-\frac{3t^2}{6np+2t}\right)$. Therefore we have $\mathbb{P}(N \leq np/4) \leq \exp\left(-\frac{27n^2p^2/16}{6np+3np/2}\right) = e^{-9np/40}$. Partitioning by this event gives

$$
\begin{aligned}
\mathbb{E}\left[\prod_{j=1}^{k}\left(1 \wedge \frac{1}{N+a_j}\right)^{l_j}\right] &\leq e^{-9np/40}\prod_{j=1}^{k}\frac{1}{1 \vee a_j^{l_j}} + \prod_{j=1}^{k}\frac{1}{1 \vee (\frac{np}{4}+a_j)^{l_j}} \\
&\leq \prod_{j=1}^{k}\frac{1}{1 \vee \left(\frac{9np}{40kl_j}+a_j\right)^{l_j}} + \prod_{j=1}^{k}\frac{1}{1 \vee (\frac{np}{4}+a_j)^{l_j}} \\
&\leq 2\prod_{j=1}^{k}\frac{1}{1 \vee \left(\frac{9np}{40kl_j}+a_j\right)^{l_j}} \leq 2\prod_{j=1}^{k}\frac{(40kl_j/9)^{l_j}}{1 \vee (np+a_j)^{l_j}} \\
&\leq (9L)^{2L}\prod_{j=1}^{k}\left(1 \wedge \frac{1}{np+a_j}\right)^{l_j}.
\end{aligned}
$$

□

Our next result is probably the most technically involved in the paper, allowing one to bound moments of (products of) $\frac{\mathbb{I}_b(x)}{N_b(x)}$ by the corresponding moments of (products of) $\frac{1}{n|T_b(x)|}$, again based on the heuristic that $N_b(x)$ is conditionally binomial so concentrates around its conditional expectation $n\int_{T_b(x)} f(x)\,\mathrm{d}s \asymp n|T_b(x)|$. By independence of the trees, the latter expected products then factorize since the dependence on the data X_i has been eliminated. The proof is complicated, and relies on the following induction procedure. First we consider the common refinement consisting of the subcells $\mathcal{R}$ generated by all possible intersections of $T_b(x)$ over the selected trees (say $T_b(x), T_{b'}(x), T_{b''}(x)$ though there could be arbitrarily many). Note that $N_b(x)$ is the sum of the number of samples X_i in each such subcell in $\mathcal{R}$. We then apply Lemma SA3 repeatedly to each subcell in $\mathcal{R}$ in turn, replacing the number of samples X_i in that subcell with its volume multiplied by the sample size n, and controlling the error incurred at each step. We record the subcells which have been "checked" in this manner using the class $\mathcal{D} \subseteq \mathcal{R}$ and proceed by finite induction, beginning with $\mathcal{D} = \emptyset$ and ending at $\mathcal{D} = \mathcal{R}$.

Lemma SA4 (Generalized moment bound for Mondrian random forest denominators)**.** *Suppose Assumptions 1 and 2 hold. Let $T_b \sim \mathcal{M}([0,1]^d, \lambda)$ be independent and $k_b \geq 1$ for $1 \leq b \leq B_0$. Then with $k = \sum_{b=1}^{B_0} k_b$, for sufficiently large n,*

$$
\mathbb{E}\left[\prod_{b=1}^{B_0}\frac{\mathbb{I}_b(x)}{N_b(x)^{k_b}}\right] \leq \left(\frac{36k}{\inf_{x\in[0,1]^d} f(x)}\right)^{2^{2k}}\prod_{b=1}^{B_0}\mathbb{E}\left[1 \wedge \frac{1}{(n|T_b(x)|)^{k_b}}\right].
$$

Proof (Lemma SA4)**.** Define the common refinement of $\{T_b(x) : 1 \leq b \leq B_0\}$ as the class of sets

$$
\mathcal{R} = \left\{\bigcap_{b=1}^{B_0} D_b : D_b \in \{T_b(x), T_b(x)^{\mathrm{c}}\}\right\} \setminus \left\{\emptyset, \bigcap_{b=1}^{B_0} T_b(x)^{\mathrm{c}}\right\}
$$

where $T_b(x)^{\mathrm{c}} = [0,1]^d \setminus T_b(x)$, and let $\mathcal{D} \subset \mathcal{R}$. We will proceed by induction on the elements of $\mathcal{D}$, which represents the subcells we have checked, starting from $\mathcal{D} = \emptyset$ and finishing at $\mathcal{D} = \mathcal{R}$.

For $D \in \mathcal{R}$ let $\mathcal{A}(D) = \{1 \leq b \leq B_0 : D \subseteq T_b(x)\}$ be the indices of the trees which are active on subcell D, and for $1 \leq b \leq B_0$ let $\mathcal{A}(b) = \{D \in \mathcal{R} : D \subseteq T_b(x)\}$ be the subcells which are contained in $T_b(x)$, so that $b \in \mathcal{A}(D) \iff D \in \mathcal{A}(b)$. For a subcell $D \in \mathcal{R}$, write $N_b(D) = \sum_{i=1}^n \mathbb{I}\{X_i \in D\}$ so that $N_b(x) = \sum_{D \in \mathcal{A}(b)} N_b(D)$. Note that for any $D \in \mathcal{R} \setminus \mathcal{D}$,

$$\mathbb{E}\left[\prod_{b=1}^{B_0} \frac{1}{1 \vee \left(\sum_{D' \in \mathcal{A}(b) \setminus \mathcal{D}} N_b(D') + n \sum_{D' \in \mathcal{A}(b) \cap \mathcal{D}} |D'|\right)^{k_b}}\right]$$

$$= \mathbb{E}\left[\prod_{b \notin \mathcal{A}(D)} \frac{1}{1 \vee \left(\sum_{D' \in \mathcal{A}(b) \setminus \mathcal{D}} N_b(D') + n \sum_{D' \in \mathcal{A}(b) \cap \mathcal{D}} |D'|\right)^{k_b}}\right.$$

$$\left. \times \mathbb{E}\left[\prod_{b \in \mathcal{A}(D)} \frac{1}{1 \vee \left(\sum_{D' \in \mathcal{A}(b) \setminus \mathcal{D}} N_b(D') + n \sum_{D' \in \mathcal{A}(b) \cap \mathcal{D}} |D'|\right)^{k_b}} \;\middle|\; \mathbf{T}, N_b(D') : D' \in \mathcal{R} \backslash (\mathcal{D} \cup \{D\})\right]\right].$$

Now the inner conditional expectation is over $N_b(D)$ only. Since f is bounded away from zero,

$$\begin{aligned} N_b(D) &\sim \mathrm{Bin}\left(n - \sum_{D' \in \mathcal{R} \backslash (\mathcal{D} \cup \{D\})} N_b(D'),\ \frac{\int_D f(s)\,\mathrm{d}s}{1 - \int_{\bigcup (\mathcal{R} \backslash \mathcal{D}) \backslash D} f(s)\,\mathrm{d}s}\right) \\ &\geq \mathrm{Bin}\left(n - \sum_{D' \in \mathcal{R} \backslash (\mathcal{D} \cup \{D\})} N_b(D'),\ |D| \inf_{x \in [0,1]^d} f(x)\right) \end{aligned}$$

conditional on $\mathbf{T}$ and $N_b(D') : D' \in \mathcal{R} \setminus (\mathcal{D} \cup \{D\})$. Further, by Lemma SA2 with $n \geq \lambda^d$,

$$\mathbb{P}\left(\sum_{D' \in \mathcal{R} \backslash (\mathcal{D} \cup \{D\})} N_b(D') > t^{d+1} \frac{n}{\lambda^d} \|f\|_\infty\right) \leq \mathbb{P}\left(N_\cup(x) > t^{d+1} \frac{n}{\lambda^d} \|f\|_\infty\right) \leq 4 d B_0 e^{-t/4}.$$

Thus $N_b(D) \geq \mathrm{Bin}(n/2, |D| \inf_x f(x))$ conditional on $\{\mathbf{T}, N_b(D') : D' \in \mathcal{R} \setminus (\mathcal{D} \cup \{D\})\}$ with probability at least $1 - 4 d B_0 e^{\frac{-\sqrt{\lambda}}{8\|f\|_\infty}}$. So by Lemma SA3,

$$\mathbb{E}\left[\prod_{b \in \mathcal{A}(D)} \frac{1}{1 \vee \left(\sum_{D' \in \mathcal{A}(b) \setminus \mathcal{D}} N_b(D') + n \sum_{D' \in \mathcal{A}(b) \cap \mathcal{D}} |D'|\right)^{k_b}} \;\middle|\; \mathbf{T}, N_b(D') : D' \in \mathcal{R} \setminus (\mathcal{D} \cup \{D\})\right]$$

$$\begin{aligned} &\leq (9k)^{2k}\, \mathbb{E}\left[\prod_{b \in \mathcal{A}(D)} \frac{1}{1 \vee \left(\sum_{D' \in \mathcal{A}(b) \setminus (\mathcal{D} \cup \{D\})} N_b(D') + n|D| \inf_x f(x)/2 + n \sum_{D' \in \mathcal{A}(b) \cap \mathcal{D}} |D'|\right)^{k_b}}\right] \\ &\quad + 4 d B_0 e^{\frac{-\sqrt{\lambda}}{8\|f\|_\infty}} \\ &\leq \left(\frac{18k}{\inf_x f(x)}\right)^{2k} \mathbb{E}\left[\prod_{b \in \mathcal{A}(D)} \frac{1}{1 \vee \left(\sum_{D' \in \mathcal{A}(b) \setminus (\mathcal{D} \cup \{D\})} N_b(D') + n \sum_{D' \in \mathcal{A}(b) \cap (\mathcal{D} \cup \{D\})} |D'|\right)^{k_b}}\right] \\ &\quad + 4 d B_0 e^{\frac{-\sqrt{\lambda}}{8\|f\|_\infty}}. \end{aligned}$$

Therefore plugging this back into the marginal expectation yields

$$\begin{aligned}
&\mathbb{E}\left[\prod_{b=1}^{B_0} \frac{1}{1 \vee \left(\sum_{D' \in \mathcal{A}(b) \setminus \mathcal{D}} N_b(D') + n \sum_{D' \in \mathcal{A}(b) \cap \mathcal{D}} |D'|\right)^{k_b}}\right] \\
&\quad \leq \left(\frac{18k}{\inf_x f(x)}\right)^{2k} \mathbb{E}\left[\prod_{b=1}^{B_0} \frac{1}{1 \vee \left(\sum_{D' \in \mathcal{A}(b) \setminus (\mathcal{D} \cup \{D\})} N_b(D') + n \sum_{D' \in \mathcal{A}(b) \cap (\mathcal{D} \cup \{D\})} |D'|\right)^{k_b}}\right] \\
&\qquad + 4dB_0 e^{\frac{-\sqrt{\lambda}}{8\|f\|_\infty}}.
\end{aligned}$$

Now we apply induction, starting with $\mathcal{D} = \emptyset$ and adding $D \in \mathcal{R} \setminus \mathcal{D}$ to $\mathcal{D}$ until $\mathcal{D} = \mathcal{R}$. This takes at most $|\mathcal{R}| \leq 2^k$ steps and yields

$$\begin{aligned}
\mathbb{E}\left[\prod_{b=1}^{B_0} \frac{\mathbb{I}_b(x)}{N_b(x)^{k_b}}\right] &\leq \mathbb{E}\left[\prod_{b=1}^{B_0} \frac{1}{1 \vee N_b(x)^{k_b}}\right] = \mathbb{E}\left[\prod_{b=1}^{B_0} \frac{1}{1 \vee \left(\sum_{D \in \mathcal{A}(b)} N_b(D)\right)^{k_b}}\right] \leq \cdots \\
&\leq \left(\frac{18k}{\inf_x f(x)}\right)^{2^{2k}} \left(\prod_{b=1}^{B_0} \mathbb{E}\left[\frac{1}{1 \vee (n|T_b(x)|)^{k_b}}\right] + 4dB_0 2^k e^{\frac{-\sqrt{\lambda}}{8\|f\|_\infty}}\right),
\end{aligned}$$

where the expectation factorizes due to independence of $T_b(x)$. The last step is to remove the trailing exponential term. To do this, note that by Jensen's inequality,

$$\prod_{b=1}^{B_0} \mathbb{E}\left[\frac{1}{1 \vee (n|T_b(x)|)^{k_b}}\right] \geq \prod_{b=1}^{B_0} \frac{1}{\mathbb{E}\left[1 \vee (n|T_b(x)|)^{k_b}\right]} \geq \prod_{b=1}^{B_0} \frac{1}{n^{k_b}} = n^{-k}$$

while the assumption of $\lambda \gtrsim (\log n)^3$ implies $\lambda \geq (\log n)^3 / C^2$ eventually for some $C > 0$, giving

$$4dB_0 2^k e^{\frac{-\sqrt{\lambda}}{8\|f\|_\infty}} \leq 4dB_0 2^k e^{\frac{-(\log n)^{3/2}}{8C\|f\|_\infty}} \leq 4dB_0 2^k e^{-k \log n - \log(4dB_0 2^k)} \leq n^{-k}$$

for sufficiently large n because B_0, d, and k are fixed. $\square$

Now that moments of (products of) $\frac{\mathbb{I}_b(x)}{N_b(x)}$ have been bounded by moments of (products of) $\frac{1}{n|T_b(x)|}$, we establish further explicit bounds for these in the next result. Note that the problem has been reduced to determining properties of Mondrian cells, so once again we return to the exact cell shape distribution given by Mourtada et al. [2020], and evaluate the appropriate expectations by integration. Note that the truncation by taking the minimum with one inside the expectation is essential here, as otherwise second moment of the inverse Mondrian cell volume is not even finite. As such, there is a "penalty" of $(\log n)^d$ when bounding truncated second moments, and the upper bound for the kth moment is significantly larger than the naive assumption of $(\lambda^d/n)^k$ whenever $k > 2$. This "small cell" phenomenon in which the inverse volumes of Mondrian cells have heavy tails is a recurring challenge in our analysis.

Lemma SA5 (Inverse moments of the volume of a Mondrian cell)**.** *Suppose Assumption 2 holds* *and let $T \sim \mathcal{M}([0,1]^d, \lambda)$. Then with $k \geq 1$, for sufficiently large n,*

$$\mathbb{E}\left[1 \wedge \frac{1}{(n|T(x)|)^k}\right] \leq \left(\frac{2}{2-k}\frac{\lambda^{dk}}{n^k}\right)^{\mathbb{I}\{k<2\}} \left(\frac{3\lambda^{2d}(\log n)^d}{n^2}\right)^{\mathbb{I}\{k\geq 2\}} \prod_{j=1}^d \frac{1}{x_j(1-x_j)}.$$

Proof (Lemma SA5)**.** By Mourtada et al. [2020, Proposition 1], we have

$$|T(x)| = \prod_{j=1}^d \left\{\left(\frac{1}{\lambda}E_{j1}\right) \wedge x_j + \left(\frac{1}{\lambda}E_{j2}\right) \wedge (1-x_j)\right\},$$

where E_{j1} and E_{j2} are mutually independent $\mathrm{Exp}(1)$ random variables. Thus for any $0 < t < 1$, using the fact that $E_{j1} + E_{j2} \sim \mathrm{Gamma}(2,1)$,

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{1 \vee (n|T(x)|)^k}\right] &\leq \frac{1}{n^k}\mathbb{E}\left[\frac{\mathbb{I}\{\min_j(E_{j1}+E_{j2}) \geq t\}}{|T(x)|^k}\right] + \mathbb{P}\left(\min_{1\leq j\leq d}(E_{j1}+E_{j2}) < t\right) \\
&\leq \frac{1}{n^k}\prod_{j=1}^d \mathbb{E}\left[\frac{\mathbb{I}\{E_{j1}+E_{j2} \geq t\}}{\left(\frac{1}{\lambda}E_{j1} \wedge x_j + \frac{1}{\lambda}E_{j2} \wedge (1-x_j)\right)^k}\right] + d\,\mathbb{P}(E_{j1} < t) \\
&\leq \frac{\lambda^{dk}}{n^k}\prod_{j=1}^d \frac{1}{x_j(1-x_j)}\mathbb{E}\left[\frac{\mathbb{I}\{E_{j1}+E_{j2} \geq t\}}{(E_{j1}+E_{j2})^k \wedge 1}\right] + d(1-e^{-t}) \\
&\leq \frac{\lambda^{dk}}{n^k}\prod_{j=1}^d \frac{1}{x_j(1-x_j)}\left(\int_t^1 \frac{e^{-s}}{s^{k-1}}\,\mathrm{d}s + \int_1^\infty s e^{-s}\,\mathrm{d}s\right) + dt \\
&\leq \frac{\lambda^{dk}}{n^k}\prod_{j=1}^d \frac{1}{x_j(1-x_j)}\left(\int_t^1 s^{1-k}\,\mathrm{d}s + 1\right) + dt \\
&= dt + \frac{\lambda^{dk}}{n^k}\prod_{j=1}^d \frac{1}{x_j(1-x_j)} \times \begin{cases} 1 + \frac{1}{2-k} - \frac{t^{2-k}}{2-k} & \text{if } 1 \leq k < 2, \\ 1 - \log t & \text{if } k = 2. \end{cases}
\end{aligned}$$

If $k > 2$ we simply use $\frac{1}{1\vee(n|T(x)|)^k} \leq \frac{1}{1\vee(n|T(x)|)^2}$. Now if $1 \leq k < 2$ we let $t \to 0$, giving

$$\mathbb{E}\left[\frac{1}{1 \vee (n|T(x)|)^k}\right] \leq \frac{2}{2-k}\frac{\lambda^{dk}}{n^k}\prod_{j=1}^d \frac{1}{x_j(1-x_j)},$$

and if $k = 2$ then we set $t = 1/n^2$ so that for sufficiently large n,

$$\mathbb{E}\left[\frac{1}{1 \vee (n|T(x)|)^2}\right] \leq \frac{d}{n^2} + \frac{\lambda^{2d}(1+2(\log n)^d)}{n^2}\prod_{j=1}^d \frac{1}{x_j(1-x_j)} \leq \frac{3\lambda^{2d}(\log n)^d}{n^2}\prod_{j=1}^d \frac{1}{x_j(1-x_j)}.$$

Lower bounds which match up to constants for $1 \leq k < 2$ are easily obtained by noting $\mathbb{E}\left[1 \wedge \frac{1}{(n|T(x)|)^k}\right] \geq \mathbb{E}\left[1 \wedge \frac{1}{n|T(x)|}\right]^k$ by Jensen's inequality and

$$\mathbb{E}\left[1 \wedge \frac{1}{n|T(x)|}\right] \geq \frac{1}{1 + n\mathbb{E}\left[|T(x)|\right]} \geq \frac{1}{1 + 2^d n/\lambda^d} \gtrsim \frac{\lambda^d}{n}.$$

To obtain a lower bound when $k = 2$, note that

$$\mathbb{E}\left[\frac{1}{1 \vee (n|T(x)|)^2}\right] \geq \mathbb{E}\left[1 \wedge \frac{\lambda^d}{n^2} \prod_{j=1}^{d} \frac{1}{(E_{j1}+E_{j2})^2}\right] \geq \frac{\lambda^d}{n^2} \mathbb{E}\left[\frac{1}{(E_1+E_2)^2 \vee n^{-1/d}}\right]^d$$
$$\geq \frac{\lambda^d}{n^2}\left(\int_{n^{-\frac{1}{2d}}}^{1} \frac{e^{-s}}{s} \,\mathrm{d}s\right)^d \geq \frac{\lambda^d}{n^2}\frac{1}{e}\left(\int_{n^{-\frac{1}{2d}}}^{1} \frac{1}{s} \,\mathrm{d}s\right)^d \geq \frac{\lambda^d}{n^2}\frac{1}{e}\left(\frac{1}{2d}\log n\right)^d.$$

□

The ongoing endeavor to bound moments of (products of) $\frac{\mathbb{I}_b(x)}{N_b(x)}$ is concluded with the next result, chaining together the previous two lemmas to provide an explicit bound with no expectations on the right-hand side.

Lemma SA6 (Simplified generalized moment bound for Mondrian random forest denominators)**.** *Grant Assumptions 1 and 2. Let $T_b \sim \mathcal{M}([0,1]^d, \lambda)$ and $k_b \geq 1$ for $1 \leq b \leq B_0$. Then with $k = \sum_{b=1}^{B_0} k_b$, for sufficiently large n,*

$$\mathbb{E}\left[\prod_{b=1}^{B_0} \frac{\mathbb{I}_b(x)}{N_b(x)^{k_b}}\right]$$
$$\leq \left(\frac{36k}{\inf_{x\in[0,1]^d} f(x)}\right)^{2^{2k}} \left(\prod_{j=1}^{d} \frac{1}{x_j(1-x_j)}\right)^{B_0} \prod_{b=1}^{B_0} \left(\frac{2}{2-k_b}\frac{\lambda^{dk_b}}{n^{k_b}}\right)^{\mathbb{I}\{k_b<2\}} \left(\frac{3\lambda^{2d}(\log n)^d}{n^2}\right)^{\mathbb{I}\{k_b\geq 2\}}.$$

Proof (Lemma SA6)**.** This follows directly from Lemmas SA4 and SA5. □

Our final preliminary lemma is concerned with further properties of the inverse truncated binomial distribution, again with the aim of analyzing $\frac{\mathbb{I}_b(x)}{N_b(x)}$. This time, instead of merely upper bounding the moments, we aim to give convergence results for those moments, again in terms of moments of $\frac{1}{n|T_b(x)|}$. This time we only need to handle the first and second moment, so this result does not strictly generalize Lemma SA3 except in simple cases. The proof is by Taylor's theorem and the Cauchy–Schwarz inequality, using explicit expressions for moments of the binomial distribution and bounds from Lemma SA3.

Lemma SA7 (Expectation inequalities for the binomial distribution)**.** *Let $N \sim \mathrm{Bin}(n,p)$ and take $a, b \geq 1$. Then*

$$0 \leq \mathbb{E}\left[\frac{1}{N+a}\right] - \frac{1}{np+a} \leq \frac{2^{19}}{(np+a)^2},$$
$$0 \leq \mathbb{E}\left[\frac{1}{(N+a)(N+b)}\right] - \frac{1}{(np+a)(np+b)} \leq \frac{2^{27}}{(np+a)(np+b)}\left(\frac{1}{np+a}+\frac{1}{np+b}\right).$$

Proof (Lemma SA7)**.** For the first result, Taylor's theorem with Lagrange remainder applied to

$N \mapsto \frac{1}{N+a}$ around np gives

$$\mathbb{E}\left[\frac{1}{N+a}\right] = \mathbb{E}\left[\frac{1}{np+a} - \frac{N-np}{(np+a)^2} + \frac{(N-np)^2}{(\xi+a)^3}\right]$$

for some ξ between np and N. The second term on the right-hand side is zero-mean, clearly showing the non-negativity part of the result, and applying the Cauchy–Schwarz inequality to the remaining term gives

$$\begin{aligned}\mathbb{E}\left[\frac{1}{N+a}\right] - \frac{1}{np+a} &\leq \mathbb{E}\left[\frac{(N-np)^2}{(np+a)^3} + \frac{(N-np)^2}{(N+a)^3}\right] \\ &\leq \frac{\mathbb{E}[(N-np)^2]}{(np+a)^3} + \sqrt{\mathbb{E}[(N-np)^4]\mathbb{E}\left[\frac{1}{(N+a)^6}\right]}.\end{aligned}$$

Now we use $\mathbb{E}[(N-np)^4] \leq np(1+3np)$ and apply Lemma SA3 to see that

$$\mathbb{E}\left[\frac{1}{N+a}\right] - \frac{1}{np+a} \leq \frac{np}{(np+a)^3} + \sqrt{\frac{54^6 np(1+3np)}{(np+a)^6}} \leq \frac{2^{19}}{(np+a)^2}.$$

For the second result, Taylor's theorem applied to $N \mapsto \frac{1}{(N+a)(N+b)}$ around np gives

$$\begin{aligned}\mathbb{E}\left[\frac{1}{(N+a)(N+b)}\right] &= \mathbb{E}\left[\frac{1}{(np+a)(np+b)} - \frac{(N-np)(2np+a+b)}{(np+a)^2(np+b)^2}\right] \\ &\quad + \mathbb{E}\left[\frac{(N-np)^2}{(\xi+a)(\xi+b)}\left(\frac{1}{(\xi+a)^2} + \frac{1}{(\xi+a)(\xi+b)} + \frac{1}{(\xi+b)^2}\right)\right]\end{aligned}$$

for some ξ between np and N. The second term on the right-hand side is zero-mean, clearly showing the non-negativity part of the result, and applying the Cauchy–Schwarz inequality to the remaining term gives

$$\begin{aligned}\mathbb{E}\left[\frac{1}{(N+a)(N+b)}\right] - \frac{1}{np+a} &\leq \mathbb{E}\left[\frac{2(N-np)^2}{(N+a)(N+b)}\left(\frac{1}{(N+a)^2} + \frac{1}{(N+b)^2}\right)\right] \\ &\quad + \mathbb{E}\left[\frac{2(N-np)^2}{(np+a)(np+b)}\left(\frac{1}{(np+a)^2} + \frac{1}{(np+b)^2}\right)\right] \\ &\leq \sqrt{4\mathbb{E}\left[(N-np)^4\right]\mathbb{E}\left[\frac{1}{(N+a)^6(N+b)^2} + \frac{1}{(N+b)^6(N+a)^2}\right]} \\ &\quad + \frac{2\mathbb{E}[(N-np)^2]}{(np+a)(np+b)}\left(\frac{1}{(np+a)^2} + \frac{1}{(np+b)^2}\right).\end{aligned}$$

Now we use $\mathbb{E}[(N-np)^4] \leq np(1+3np)$ and apply Lemma SA3 to see that

$$
\begin{aligned}
\mathbb{E}\left[\frac{1}{(N+a)(N+b)}\right] - \frac{1}{np+a} &\leq \sqrt{\frac{4np(1+3np)\cdot 72^8}{(np+a)^2(np+b)^2}\left(\frac{1}{(np+a)^4}+\frac{1}{(np+b)^4}\right)} \\
&\quad + \frac{2np}{(np+a)(np+b)}\left(\frac{1}{(np+a)^2}+\frac{1}{(np+b)^2}\right) \\
&\leq \frac{2^{27}}{(np+a)(np+b)}\left(\frac{1}{np+a}+\frac{1}{np+b}\right).
\end{aligned}
$$

□

SA1.3 Proofs for Section 3

We give rigorous proofs of the bias and variance characterizations, rate of convergence, central limit theorem, variance estimation, and confidence interval validity results for the Mondrian random forest estimator.

Proof (Lemma 1)**.** We begin by showing that $\mathbb{E}[\hat{\mu}(x) \mid \mathbf{X}, \mathbf{T}]$ converges to $\mathbb{E}[\hat{\mu}(x) \mid \mathbf{X}]$.

Part 1: Removing the dependence on the trees

By measurability and with $\mu(X_i) = \mathbb{E}[Y_i \mid X_i]$ almost surely,

$$
\mathbb{E}[\hat{\mu}(x) \mid \mathbf{X}, \mathbf{T}] - \mu(x) = \frac{1}{B}\sum_{b=1}^{B}\sum_{i=1}^{n}(\mu(X_i)-\mu(x))\frac{\mathbb{I}_{ib}(x)}{N_b(x)}.
$$

Now conditional on $\mathbf{X}$, the terms in the outer sum depend only on T_b so are i.i.d. Since $\mu \in \mathcal{H}^{\beta_\mu}$,

$$
\begin{aligned}
\operatorname{Var}[\mathbb{E}[\hat{\mu}(x) \mid \mathbf{X}, \mathbf{T}] - \mu(x) \mid \mathbf{X}] &\leq \frac{1}{B}\mathbb{E}\left[\left(\sum_{i=1}^{n}(\mu(X_i)-\mu(x))\frac{\mathbb{I}_{ib}(x)}{N_b(x)}\right)^2 \;\middle|\; \mathbf{X}\right] \\
&\lesssim \frac{1}{B}\mathbb{E}\left[\max_{1\leq i\leq n}\left\{\mathbb{I}_{ib}(x)\|X_i-x\|_2^{2(1\wedge\beta_\mu)}\right\}\left(\sum_{i=1}^{n}\frac{\mathbb{I}_{ib}(x)}{N_b(x)}\right)^2 \;\middle|\; \mathbf{X}\right] \\
&\lesssim \frac{1}{B}\sum_{j=1}^{d}\mathbb{E}\left[|T(x)_j|^{2(1\wedge\beta_\mu)}\right] \lesssim \frac{1}{\lambda^{2(1\wedge\beta_\mu)}B},
\end{aligned}
$$

where we used the law of $T(x)_j$ from Mourtada et al. [2020, Proposition 1]. Hence

$$
\mathbb{E}\left[\left(\mathbb{E}[\hat{\mu}(x) \mid \mathbf{X}, \mathbf{T}] - \mathbb{E}[\hat{\mu}(x) \mid \mathbf{X}]\right)^2\right] \lesssim \frac{1}{\lambda^{2(1\wedge\beta_\mu)}B}.
$$

Part 2: Showing the conditional bias converges

Now $\mathbb{E}[\hat{\mu}(x) \mid \mathbf{X}]$ is a non-linear function of the i.i.d. random variables X_i, so we use the Efron–Stein inequality [Efron and Stein, 1981] to bound its variance. Let $\tilde{X}_{ij} = X_i$ if $i \neq j$ and be an independent copy of X_j, denoted $\tilde{X}_j$, if $i = j$. Write $\tilde{\mathbf{X}}_j = (\tilde{X}_{1j}, \ldots, \tilde{X}_{nj})$ and similarly

$\tilde{\mathbb{I}}_{ijb}(x) = \mathbb{I}\{\tilde{X}_{ij} \in T_b(x)\}$ and $N_{jb}(x) = \sum_{i=1}^n \tilde{\mathbb{I}}_{ijb}(x)$.

$$
\begin{aligned}
&\operatorname{Var}\left[\sum_{i=1}^n (\mu(X_i) - \mu(x))\mathbb{E}\left[\frac{\mathbb{I}_{ib}(x)}{N_b(x)} \mid \mathbf{X}\right]\right] \\
&\quad\leq \frac{1}{2}\sum_{j=1}^n \mathbb{E}\left[\left(\sum_{i=1}^n (\mu(X_i) - \mu(x))\mathbb{E}\left[\frac{\mathbb{I}_{ib}(x)}{N_b(x)} \mid \mathbf{X}\right] - \sum_{i=1}^n \left(\mu(\tilde{X}_{ij}) - \mu(x)\right)\mathbb{E}\left[\frac{\tilde{\mathbb{I}}_{ijb}(x)}{\tilde{N}_{jb}(x)} \mid \tilde{\mathbf{X}}_j\right]\right)^2\right] \\
&\quad\leq \frac{1}{2}\sum_{j=1}^n \mathbb{E}\left[\left(\sum_{i=1}^n \left(\left(\mu(X_i) - \mu(x)\right)\frac{\mathbb{I}_{ib}(x)}{N_b(x)} - \left(\mu(\tilde{X}_{ij}) - \mu(x)\right)\frac{\tilde{\mathbb{I}}_{ijb}(x)}{\tilde{N}_{jb}(x)}\right)\right)^2\right] \\
&\quad\leq \sum_{j=1}^n \mathbb{E}\left[\left(\sum_{i\neq j} (\mu(X_i) - \mu(x))\left(\frac{\mathbb{I}_{ib}(x)}{N_b(x)} - \frac{\mathbb{I}_{ib}(x)}{\tilde{N}_{jb}(x)}\right)\right)^2\right] \\
&\qquad + 2\sum_{j=1}^n \mathbb{E}\left[(\mu(X_j) - \mu(x))^2 \frac{\mathbb{I}_{jb}(x)}{N_b(x)^2}\right]. \qquad \text{(SA2)}
\end{aligned}
$$

For the first term in (SA2) to be non-zero, we must have $|N_b(x) - \tilde{N}_{jb}(x)| = 1$. Writing $N_{-jb}(x) = \sum_{i\neq j} \mathbb{I}_{ib}(x)$, we may assume by symmetry that $\tilde{N}_{jb}(x) = N_{-jb}(x)$ and $N_b(x) = N_{-jb}(x) + 1$, and also that $\mathbb{I}_{jb}(x) = 1$. Hence since f is bounded and $\mu \in \mathcal{H}^{\beta_\mu}$, writing $\mathbb{I}_{-jb}(x) = \mathbb{I}\{N_{-jb}(x) \geq 1\}$, by the Hölder inequality, Lemma SA1 and Lemma SA6,

$$
\begin{aligned}
&\sum_{j=1}^n \mathbb{E}\left[\left(\sum_{i\neq j} (\mu(X_i) - \mu(x))\left(\frac{\mathbb{I}_{ib}(x)}{N_b(x)} - \frac{\mathbb{I}_{ib}(x)}{\tilde{N}_{jb}(x)}\right)\right)^2\right] \\
&\quad\lesssim \sum_{j=1}^n \mathbb{E}\left[\max_{1\leq l\leq d} |T_b(x)_l|^{2(1\wedge\beta_\mu)}\left(\frac{\sum_{i\neq j}\mathbb{I}_{ib}(x)\mathbb{I}_{jb}(x)}{N_{-jb}(x)(N_{-jb}(x)+1)}\right)^2\right] \lesssim \mathbb{E}\left[\max_{1\leq l\leq d}|T_b(x)_l|^{2(1\wedge\beta_\mu)}\frac{\mathbb{I}_b(x)}{N_b(x)}\right] \\
&\lesssim \mathbb{E}\left[\max_{1\leq l\leq d}|T_b(x)_l|^{6(1\wedge\beta_\mu)}\right]^{1/3}\mathbb{E}\left[\frac{\mathbb{I}_b(x)}{N_b(x)^{3/2}}\right]^{2/3} \lesssim \frac{1}{\lambda^{2(1\wedge\beta_\mu)}}\frac{\lambda^d}{n}.
\end{aligned}
$$

For the second term in (SA2) we again use $\mu \in \mathcal{H}^{\beta_\mu}$ to see

$$
\sum_{j=1}^n \mathbb{E}\left[(\mu(X_j) - \mu(x))^2\frac{\mathbb{I}_{jb}(x)}{N_b(x)^2}\right] \lesssim \mathbb{E}\left[\max_{1\leq l\leq d}|T_b(x)_l|^{2(1\wedge\beta_\mu)}\frac{\mathbb{I}_b(x)}{N_b(x)}\right] \lesssim \frac{1}{\lambda^{2(1\wedge\beta_\mu)}}\frac{\lambda^d}{n}
$$

in the same manner. Hence

$$
\operatorname{Var}\left[\sum_{i=1}^n (\mu(X_i) - \mu(x))\,\mathbb{E}\left[\frac{\mathbb{I}_{ib}(x)}{N_b(x)} \mid \mathbf{X}\right]\right] \lesssim \frac{1}{\lambda^{2(1\wedge\beta_\mu)}}\frac{\lambda^d}{n},
$$

and so by the above and the previous part,

$$
\mathbb{E}\left[\left(\mathbb{E}\left[\hat{\mu}(x) \mid \mathbf{X}, \mathbf{T}\right] - \mathbb{E}\left[\hat{\mu}(x)\right]\right)^2\right] \lesssim \frac{1}{\lambda^{2(1\wedge\beta_\mu)}B} + \frac{1}{\lambda^{2(1\wedge\beta_\mu)}}\frac{\lambda^d}{n}.
$$

Part 3: Computing the limiting bias

It remains to compute the limiting value of $\mathbb{E}[\hat{\mu}(x)] - \mu(x)$. Let $\mathbf{X}_{-i} = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$ and $N_{-ib}(x) = \sum_{j=1}^n \mathbb{I}\{j \neq i\} \mathbb{I}\{X_j \in T_b(x)\}$. Then

$$\begin{aligned}\mathbb{E}[\hat{\mu}(x)] - \mu(x) = \mathbb{E}\left[\sum_{i=1}^n (\mu(X_i) - \mu(x)) \frac{\mathbb{I}_{ib}(x)}{N_b(x)}\right] &= \sum_{i=1}^n \mathbb{E}\left[\mathbb{E}\left[\frac{(\mu(X_i) - \mu(x)) \mathbb{I}_{ib}(x)}{N_{-ib}(x) + 1} \mid \mathbf{T}, \mathbf{X}_{-i}\right]\right] \\ &= n\, \mathbb{E}\left[\frac{\int_{T_b(x)} (\mu(s) - \mu(x)) f(s) \,\mathrm{d}s}{N_{-ib}(x) + 1}\right].\end{aligned}$$

By Lemma SA7, as $N_{-ib}(x) \sim \operatorname{Bin}\left(n-1, \int_{T_b(x)} f(s) \,\mathrm{d}s\right)$ given $\mathbf{T}$ and f is bounded away from zero,

$$\left|\mathbb{E}\left[\frac{1}{N_{-ib}(x) + 1} \mid \mathbf{T}\right] - \frac{1}{(n-1) \int_{T_b(x)} f(s) \,\mathrm{d}s + 1}\right| \lesssim \frac{1}{n^2 \left(\int_{T_b(x)} f(s) \,\mathrm{d}s\right)^2} \wedge 1 \lesssim \frac{1}{n^2 |T_b(x)|^2} \wedge 1$$

and also

$$\left|\frac{1}{(n-1) \int_{T_b(x)} f(s) \,\mathrm{d}s + 1} - \frac{1}{n \int_{T_b(x)} f(s) \,\mathrm{d}s}\right| \lesssim \frac{1}{n^2 \left(\int_{T_b(x)} f(s) \,\mathrm{d}s\right)^2} \wedge 1 \lesssim \frac{1}{n^2 |T_b(x)|^2} \wedge 1.$$

So by Lemma SA1 and Lemma SA5, since $\mu \in \mathcal{H}^{\beta_\mu}$ and f is bounded below, using the Hölder inequality,

$$\begin{aligned}&\left|\mathbb{E}[\hat{\mu}(x)] - \mu(x) - \mathbb{E}\left[\frac{\int_{T_b(x)} (\mu(s) - \mu(x)) f(s) \,\mathrm{d}s}{\int_{T_b(x)} f(s) \,\mathrm{d}s}\right]\right| \\ &\quad\lesssim \mathbb{E}\left[\frac{n \int_{T_b(x)} |\mu(s) - \mu(x)|\, f(s) \,\mathrm{d}s}{n^2 |T_b(x)|^2 \vee 1}\right] \lesssim \mathbb{E}\left[\frac{\max_{1 \leq l \leq d} |T_b(x)_l|^{1 \wedge \beta_\mu}}{n |T_b(x)| \vee 1}\right] \\ &\quad\lesssim \mathbb{E}\left[\max_{1 \leq l \leq d} |T_b(x)_l|^{3(1 \wedge \beta_\mu)}\right]^{1/3} \mathbb{E}\left[\frac{1}{n^{3/2} |T_b(x)|^{3/2} \vee 1}\right]^{2/3} \lesssim \frac{1}{\lambda^{1 \wedge \beta_\mu}} \frac{\lambda^d}{n}.\end{aligned}$$

Next set $A = \frac{1}{f(x) |T_b(x)|} \int_{T_b(x)} (f(s) - f(x)) \,\mathrm{d}s \geq \inf_{s \in [0,1]^d} \frac{f(s)}{f(x)} - 1 > -1$. Use the Maclaurin series of $\frac{1}{1+x}$ up to order $\underline{\beta} - 1$ to see $\frac{1}{1+A} = \sum_{k=0}^{\underline{\beta}-1} (-1)^k A^k + O\left(A^{\underline{\beta}}\right)$. Hence

$$\begin{aligned}\mathbb{E}\left[\frac{\int_{T_b(x)} (\mu(s) - \mu(x)) f(s) \,\mathrm{d}s}{\int_{T_b(x)} f(s) \,\mathrm{d}s}\right] &= \mathbb{E}\left[\frac{\int_{T_b(x)} (\mu(s) - \mu(x)) f(s) \,\mathrm{d}s}{f(x) |T_b(x)|} \frac{1}{1 + A}\right] \\ &= \mathbb{E}\left[\frac{\int_{T_b(x)} (\mu(s) - \mu(x)) f(s) \,\mathrm{d}s}{f(x) |T_b(x)|} \left(\sum_{k=0}^{\underline{\beta}-1} (-1)^k A^k + O\left(|A|^{\underline{\beta}}\right)\right)\right].\end{aligned}$$

Note that since $\mu \in \mathcal{H}^{\beta_\mu}$ and $f \in \mathcal{H}^{\beta_f}$, by integrating the tail probability given in Lemma SA1, the Maclaurin remainder term is bounded by

$$
\begin{aligned}
&\mathbb{E}\left[\frac{\int_{T_b(x)} |\mu(s) - \mu(x)|\, f(s)\,\mathrm{d}s}{f(x)|T_b(x)|}|A|^{\underline{\beta}}\right] \\
&\quad = \mathbb{E}\left[\frac{\int_{T_b(x)} |\mu(s) - \mu(x)|\, f(s)\,\mathrm{d}s}{f(x)|T_b(x)|}\left(\frac{1}{f(x)|T_b(x)|}\int_{T_b(x)} (f(s) - f(x))\,\mathrm{d}s\right)^{\underline{\beta}}\right] \\
&\quad \lesssim \mathbb{E}\left[\max_{1 \le l \le d} |T_b(x)_l|^{1 \wedge \beta_\mu}\left(\max_{1 \le l \le d} |T_b(x)_l|^{1 \wedge \beta_f}\right)^{\underline{\beta}}\right] \\
&\quad \lesssim \mathbb{E}\left[\max_{1 \le l \le d} |T_b(x)_l|^{1 \wedge \beta_\mu + \underline{\beta}(1 \wedge \beta_f)}\right] \lesssim \mathbb{E}\left[\max_{1 \le l \le d} |T_b(x)_l|^{\beta}\right] \lesssim \frac{d}{\lambda^\beta}\int_0^\infty e^{-t} t^\beta\,\mathrm{d}t \lesssim \frac{1}{\lambda^\beta},
\end{aligned}
$$

where we used that $1 \wedge \beta_\mu + \underline{\beta}(1 \wedge \beta_f) \ge \beta$. Hence to summarize the progress so far, we have

$$
\begin{aligned}
&\left|\mathbb{E}[\hat{\mu}(x)] - \mu(x) - \sum_{k=0}^{\underline{\beta}-1} (-1)^k\, \mathbb{E}\left[\frac{\int_{T_b(x)} (\mu(s) - \mu(x))\, f(s)\,\mathrm{d}s}{f(x)^{k+1}|T_b(x)|^{k+1}}\left(\int_{T_b(x)} (f(s) - f(x))\,\mathrm{d}s\right)^k\right]\right| \\
&\quad \lesssim \frac{1}{\lambda^{1 \wedge \beta_\mu}}\frac{\lambda^d}{n} + \frac{1}{\lambda^\beta}.
\end{aligned}
$$

We continue to evaluate this expectation. First, by Taylor's theorem and with ν a multi-index, since $f \in \mathcal{H}^{\beta_f}$,

$$
\left(\int_{T_b(x)} (f(s) - f(x))\,\mathrm{d}s\right)^k = \left(\sum_{|\nu|=1}^{\underline{\beta}_f} \frac{\partial^\nu f(x)}{\nu!}\int_{T_b(x)} (s - x)^\nu\,\mathrm{d}s\right)^k + O\left(|T_b(x)|^k \max_{1 \le l \le d} |T_b(x)_l|^{\beta_f}\right).
$$

Next, by the multinomial theorem with a multi-index u indexed by ν with $|\nu| \ge 1$,

$$
\left(\sum_{|\nu|=1}^{\underline{\beta}_f} \frac{\partial^\nu f(x)}{\nu!}\int_{T_b(x)} (s - x)^\nu\,\mathrm{d}s\right)^k = \sum_{|u|=k}\binom{k}{u}\left(\frac{\partial^\nu f(x)}{\nu!}\int_{T_b(x)} (s - x)^\nu\,\mathrm{d}s\right)^u
$$

where $\binom{k}{u}$ is a multinomial coefficient. By Taylor's theorem with $\mu \in \mathcal{H}^{\beta_\mu}$ and $f \in \mathcal{H}^{\beta_f}$, and using $\beta_\mu \wedge (1 \wedge \beta_\mu + \beta_f) \ge \beta$,

$$
\begin{aligned}
&\int_{T_b(x)} (\mu(s) - \mu(x))\, f(s)\,\mathrm{d}s \\
&\quad = \sum_{|\nu'|=1}^{\underline{\beta}_\mu}\sum_{|\nu''|=0}^{\underline{\beta}_f} \frac{\partial^{\nu'}\mu(x)}{\nu'!}\frac{\partial^{\nu''} f(x)}{\nu''!}\int_{T_b(x)} (s - x)^{\nu'+\nu''}\,\mathrm{d}s + O\left(|T_b(x)| \max_{1 \le l \le d} |T_b(x)_l|^{\beta}\right).
\end{aligned}
$$

Now by integrating the tail probabilities in Lemma SA1, $\mathbb{E}\left[\max_{1 \leq l \leq d} |T_b(x)_l|^\beta\right] \lesssim \frac{1}{\lambda^\beta}$. Therefore by Lemma SA5, writing $T_b(x)^\nu$ for $\int_{T_b(x)} (s-x)^\nu \, \mathrm{d}s$,

$$
\begin{aligned}
&\sum_{k=0}^{\beta-1} (-1)^k \, \mathbb{E}\left[\frac{\int_{T_b(x)} (\mu(s) - \mu(x)) \, f(s) \, \mathrm{d}s}{f(x)^{k+1} |T_b(x)|^{k+1}} \left(\int_{T_b(x)} (f(s) - f(x)) \, \mathrm{d}s\right)^k\right] \\
&= \sum_{k=0}^{\beta-1} (-1)^k \, \mathbb{E}\left[\frac{\sum_{|\nu'|=1}^{\beta_\mu} \sum_{|\nu''|=0}^{\beta_f} \frac{\partial^{\nu'} \mu(x)}{\nu'!} \frac{\partial^{\nu''} f(x)}{\nu''!} T_b(x)^{\nu'+\nu''}}{f(x)^{k+1} |T_b(x)|^{k+1}} \sum_{|u|=k} \binom{k}{u} \left(\frac{\partial^\nu f(x)}{\nu!} T_b(x)^\nu\right)^u\right] + O\left(\frac{1}{\lambda^\beta}\right) \\
&= \sum_{|\nu'|=1}^{\beta_\mu} \sum_{|\nu''|=0}^{\beta_f} \sum_{|u|=0}^{\beta-1} \frac{\partial^{\nu'} \mu(x)}{\nu'!} \frac{\partial^{\nu''} f(x)}{\nu''!} \left(\frac{\partial^\nu f(x)}{\nu!}\right)^u \binom{|u|}{u} \frac{(-1)^{|u|}}{f(x)^{|u|+1}} \mathbb{E}\left[\frac{T_b(x)^{\nu'+\nu''} (T_b(x)^\nu)^u}{|T_b(x)|^{|u|+1}}\right] + O\left(\frac{1}{\lambda^\beta}\right).
\end{aligned}
$$

Now we show this is a polynomial in λ. For $1 \leq j \leq d$, define the independent variables $E_{1j*} \sim \mathrm{Exp}(1) \wedge (\lambda x_j)$ and $E_{2j*} \sim \mathrm{Exp}(1) \wedge (\lambda(1-x_j))$ so $T_b(x) = \prod_{j=1}^d [x_j - E_{1j*}/\lambda, x_j + E_{2j*}/\lambda]$. Then

$$
\begin{aligned}
T_b(x)^\nu &= \int_{T_b(x)} (s-x)^\nu \, \mathrm{d}s = \prod_{j=1}^d \int_{x_j - E_{1j*}/\lambda}^{x_j + E_{2j*}/\lambda} (s - x_j)^{\nu_j} \, \mathrm{d}s = \prod_{j=1}^d \int_{-E_{1j*}}^{E_{2j*}} (s/\lambda)^{\nu_j} 1/\lambda \, \mathrm{d}s \\
&= \lambda^{-d-|\nu|} \prod_{j=1}^d \int_{-E_{1j*}}^{E_{2j*}} s^{\nu_j} \, \mathrm{d}s = \lambda^{-d-|\nu|} \prod_{j=1}^d \frac{E_{2j*}^{\nu_j+1} + (-1)^{\nu_j} E_{1j*}^{\nu_j+1}}{\nu_j + 1}.
\end{aligned}
$$

So by independence over j,

$$
\begin{aligned}
&\mathbb{E}\left[\frac{T_b(x)^{\nu'+\nu''} (T_b(x)^\nu)^u}{|T_b(x)|^{|u|+1}}\right] \\
&= \lambda^{-|\nu'|-|\nu''|-|\nu| \cdot u} \prod_{j=1}^d \mathbb{E}\left[\frac{E_{2j*}^{\nu'_j+\nu''_j+1} + (-1)^{\nu'_j+\nu''_j} E_{1j*}^{\nu'_j+\nu''_j+1}}{(\nu'_j + \nu''_j + 1)(E_{2j*} + E_{1j*})} \frac{\left(E_{2j*}^{\nu_j+1} + (-1)^{\nu_j} E_{1j*}^{\nu_j+1}\right)^u}{(\nu_j+1)^u (E_{2j*} + E_{1j*})^{|u|}}\right]. \qquad \text{(SA3)}
\end{aligned}
$$

The final step is to replace E_{1j*} by $E_{1j} \sim \mathrm{Exp}(1)$ and similarly for E_{2j*}. Note that for a positive constant C,

$$
\mathbb{P}\left(\bigcup_{j=1}^d \left(\{E_{1j*} \neq E_{1j}\} \cup \{E_{2j*} \neq E_{2j}\}\right)\right) \leq 2d \, \mathbb{P}\left(\mathrm{Exp}(1) \geq \lambda \min_{1 \leq j \leq d} (x_j \wedge (1-x_j))\right) \leq 2d e^{-C\lambda}.
$$

Further, the quantity inside the expectation in (SA3) is bounded almost surely by one and so the error incurred by replacing E_{1j*} and E_{2j*} by E_{1j} and E_{2j} in (SA3) is at most $2de^{-C\lambda} \lesssim \lambda^{-\beta}$. Thus the limiting bias is

$$\mathbb{E}\left[\hat{\mu}(x)\right] - \mu(x) = \sum_{|\nu'|=1}^{\beta_\mu} \sum_{|\nu''|=0}^{\beta_f} \sum_{|u|=0}^{\beta-1} \frac{\partial^{\nu'} \mu(x)}{\nu'!} \frac{\partial^{\nu''} f(x)}{\nu''!} \left(\frac{\partial^{\nu} f(x)}{\nu!} \right)^u \binom{|u|}{u} \frac{(-1)^{|u|}}{f(x)^{|u|+1}} \lambda^{-|\nu'|-|\nu''|-|\nu|\cdot u} \tag{SA4}$$

$$\times \prod_{j=1}^{d} \mathbb{E}\left[\frac{E_{2j}^{\nu_j'+\nu_j''+1} + (-1)^{\nu_j'+\nu_j''} E_{1j}^{\nu_j'+\nu_j''+1}}{(\nu_j'+\nu_j''+1)(E_{2j}+E_{1j})} \frac{\left(E_{2j}^{\nu_j+1} + (-1)^{\nu_j} E_{1j}^{\nu_j+1}\right)^u}{(\nu_j+1)^u (E_{2j}+E_{1j})^{|u|}} \right] + O\left(\frac{1}{\lambda^{1\wedge\beta_\mu}} \frac{\lambda^d}{n} + \frac{1}{\lambda^\beta} \right),$$

recalling that u is a multi-index which is indexed by the multi-index ν. This is a polynomial in λ of degree at most $\underline{\beta}$, since higher-order terms can be absorbed into $O(1/\lambda^\beta)$, which has finite coefficients depending only on the derivatives up to order $\underline{\beta} \leq \beta_\mu$ and $\underline{\beta} - 1 \leq \beta_f$ of μ and f respectively at x. Now we show that the odd-degree terms in this polynomial are all zero. Note that a term is of odd degree if and only if $|\nu'| + |\nu''| + |\nu| \cdot u$ is odd. This implies that there exists $1 \leq j \leq d$ such that exactly one of either $\nu_j' + \nu_j''$ is odd or $\sum_{|\nu|=1}^{\beta-1} \nu_j u_\nu$ is odd.

If $\nu_j' + \nu_j''$ is odd, then $\sum_{|\nu|=1}^{\beta-1} \nu_j u_\nu$ is even, so $|\{\nu : \nu_j u_\nu \text{ is odd}\}|$ is even. Consider the effect of swapping E_{1j} and E_{2j}, an operation which by independence preserves their joint law, in each of

$$\frac{E_{2j}^{\nu_j'+\nu_j''+1} + (-1)^{\nu_j'+\nu_j''} E_{1j}^{\nu_j'+\nu_j''+1}}{E_{2j}+E_{1j}} \tag{SA5}$$

and

$$\frac{\left(E_{2j}^{\nu_j+1} + (-1)^{\nu_j} E_{1j}^{\nu_j+1}\right)^u}{(E_{2j}+E_{1j})^{|u|}} = \prod_{\substack{|\nu|=1 \\ \nu_j u_\nu \text{ even}}}^{\beta-1} \frac{\left(E_{2j}^{\nu_j+1} + (-1)^{\nu_j} E_{1j}^{\nu_j+1}\right)^{u_\nu}}{(E_{2j}+E_{1j})^{u_\nu}} \prod_{\substack{|\nu|=1 \\ \nu_j u_\nu \text{ odd}}}^{\beta-1} \frac{\left(E_{2j}^{\nu_j+1} + (-1)^{\nu_j} E_{1j}^{\nu_j+1}\right)^{u_\nu}}{(E_{2j}+E_{1j})^{u_\nu}}. \tag{SA6}$$

Clearly $\nu_j' + \nu_j''$ being odd inverts the sign of (SA5). For (SA6), each term in the first product has either ν_j even or u_ν even, so its sign is preserved. Every term in the second product of (SA6) has its sign inverted due to both ν_j and u_ν being odd, but there are an even number of terms, preserving the overall sign. Therefore the expected product of (SA5) and (SA6) is zero by symmetry.

If however $\nu_j' + \nu_j''$ is even, then $\sum_{|\nu|=1}^{\beta-1} \nu_j u_\nu$ is odd so $|\{\nu : \nu_j u_\nu \text{ is odd}\}|$ is odd. Clearly the sign of (SA5) is preserved. Again the sign of the first product in (SA6) is preserved, and the sign of every term in (SA6) is inverted. However there are now an odd number of terms in the second product, so its overall sign is inverted. Therefore the expected product of (SA5) and (SA6) is again zero.

Part 4: Calculating the second-order bias

Next we calculate some special cases, beginning with the form of the leading second-order bias, where the exponent in λ is $|\nu'| + |\nu''| + u \cdot |\nu| = 2$, proceeding by cases on the values of $|\nu'|$, $|\nu''|$, and $|u|$. Firstly, if $|\nu'| = 2$ then $|\nu''| = |u| = 0$. Note that if any $\nu_j' = 1$ then the expectation in

(SA4) is zero. Hence we can assume $\nu'_j \in \{0, 2\}$, yielding

$$\frac{1}{2\lambda^2}\sum_{j=1}^{d}\frac{\partial^2\mu(x)}{\partial x_j^2}\frac{1}{3}\mathbb{E}\left[\frac{E_{2j}^3+E_{1j}^3}{E_{2j}+E_{1j}}\right] = \frac{1}{2\lambda^2}\sum_{j=1}^{d}\frac{\partial^2\mu(x)}{\partial x_j^2}\frac{1}{3}\mathbb{E}\left[E_{1j}^2+E_{2j}^2-E_{1j}E_{2j}\right] = \frac{1}{2\lambda^2}\sum_{j=1}^{d}\frac{\partial^2\mu(x)}{\partial x_j^2},$$

where we used that E_{1j} and E_{2j} are independent Exp(1). Next we consider $|\nu'| = 1$ and $|\nu''| = 1$, so $|u| = 0$. Note that if $\nu'_j = \nu''_{j'} = 1$ with $j \neq j'$ then the expectation in (SA4) is zero. So we need only consider $\nu'_j = \nu''_j = 1$, giving

$$\frac{1}{\lambda^2}\frac{1}{f(x)}\sum_{j=1}^{d}\frac{\partial\mu(x)}{\partial x_j}\frac{\partial f(x)}{\partial x_j}\frac{1}{3}\mathbb{E}\left[\frac{E_{2j}^3+E_{1j}^3}{E_{2j}+E_{1j}}\right] = \frac{1}{\lambda^2}\frac{1}{f(x)}\sum_{j=1}^{d}\frac{\partial\mu(x)}{\partial x_j}\frac{\partial f(x)}{\partial x_j}.$$

Finally we have the case where $|\nu'| = 1$, $|\nu''| = 0$ and $|u| = 1$. Then $u_\nu = 1$ for some $|\nu| = 1$ and zero otherwise. Note that if $\nu'_j = \nu_{j'} = 1$ with $j \neq j'$ then the expectation is zero. So we need only consider $\nu'_j = \nu_j = 1$, giving

$$\begin{aligned}
&-\frac{1}{\lambda^2}\frac{1}{f(x)}\sum_{j=1}^{d}\frac{\partial\mu(x)}{\partial x_j}\frac{\partial f(x)}{\partial x_j}\frac{1}{4}\mathbb{E}\left[\frac{(E_{2j}^2-E_{1j}^2)^2}{(E_{2j}+E_{1j})^2}\right] \\
&\quad = -\frac{1}{4\lambda^2}\frac{1}{f(x)}\sum_{j=1}^{d}\frac{\partial\mu(x)}{\partial x_j}\frac{\partial f(x)}{\partial x_j}\mathbb{E}\left[E_{1j}^2+E_{2j}^2-2E_{1j}E_{2j}\right] = -\frac{1}{2\lambda^2}\frac{1}{f(x)}\sum_{j=1}^{d}\frac{\partial\mu(x)}{\partial x_j}\frac{\partial f(x)}{\partial x_j}.
\end{aligned}$$

Hence the second-order bias term is

$$\frac{1}{2\lambda^2}\sum_{j=1}^{d}\frac{\partial^2\mu(x)}{\partial x_j^2} + \frac{1}{2\lambda^2}\frac{1}{f(x)}\sum_{j=1}^{d}\frac{\partial\mu(x)}{\partial x_j}\frac{\partial f(x)}{\partial x_j}.$$

Part 5: Calculating the bias if the data is uniformly distributed

If $X_i \sim \mathrm{Unif}\left([0,1]^d\right)$ then $f(x) = 1$ and the bias expansion from (SA4) becomes

$$\sum_{|\nu'|=1}^{\beta}\lambda^{-|\nu'|}\frac{\partial^{\nu'}\mu(x)}{\nu'!}\prod_{j=1}^{d}\mathbb{E}\left[\frac{E_{2j}^{\nu'_j+1}+(-1)^{\nu'_j}E_{1j}^{\nu'_j+1}}{(\nu'_j+1)(E_{2j}+E_{1j})}\right].$$

Note that this is zero if any ν'_j is odd. Therefore we can group these terms based on the exponent of λ to see

$$\frac{B_r(x)}{\lambda^{2r}} = \frac{1}{\lambda^{2r}}\sum_{|\nu|=r}\frac{\partial^{2\nu}\mu(x)}{(2\nu)!}\prod_{j=1}^{d}\frac{1}{2\nu_j+1}\mathbb{E}\left[\frac{E_{2j}^{2\nu_j+1}+E_{1j}^{2\nu_j+1}}{E_{2j}+E_{1j}}\right].$$

Since $\int_0^\infty \frac{e^{-t}}{a+t}\,\mathrm{d}t = e^a\Gamma(0,a)$ and $\int_0^\infty s^a\Gamma(0,a)\,\mathrm{d}s = \frac{a!}{a+1}$, with $\Gamma(0,a) = \int_a^\infty \frac{e^{-t}}{t}\,\mathrm{d}t$ the upper incomplete gamma function, the expectation is easily calculated as

$$\mathbb{E}\left[\frac{E_{2j}^{2\nu_j+1}+E_{1j}^{2\nu_j+1}}{E_{2j}+E_{1j}}\right] = 2\int_0^\infty s^{2\nu_j+1}e^{-s}\int_0^\infty \frac{e^{-t}}{s+t}\,\mathrm{d}t\,\mathrm{d}s = 2\int_0^\infty s^{2\nu_j+1}\Gamma(0,s)\,\mathrm{d}s = \frac{(2\nu_j+1)!}{\nu_j+1},$$

so

$$\frac{B_r(x)}{\lambda^{2r}} = \frac{1}{\lambda^{2r}}\sum_{|\nu|=r}\frac{\partial^{2\nu}\mu(x)}{(2\nu)!}\prod_{j=1}^d \frac{1}{2\nu_j+1}\frac{(2\nu_j+1)!}{\nu_j+1} = \frac{1}{\lambda^{2r}}\sum_{|\nu|=r}\partial^{2\nu}\mu(x)\prod_{j=1}^d\frac{1}{\nu_j+1}.$$

□

Proof (Lemma 2)**.** By Lemma 5 with $J=0$, $a_0=1$, and $\omega_0=1$. □

Proof (Theorem 1)**.** By Theorem 4 with $J=0$, $a_0=1$, and $\omega_0=1$. □

Proof (Theorem 2)**.** By Theorem 5 with $J=0$, $a_0=1$, and $\omega_0=1$. □

Proof (Lemma 3)**.** By Lemma 6 with $J=0$, $a_0=1$, and $\omega_0=1$. □

Proof (Theorem 3)**.** By Theorem 6 with $J=0$, $a_0=1$, and $\omega_0=1$, replacing β by $2\wedge\beta$. □

SA1.4 Proofs for Section 4

We give rigorous proofs of the bias and variance characterizations, minimax optimality, central limit theorem, variance estimation, and confidence interval validity results for the debiased Mondrian random forest estimator.

The bias characterization of Lemma 4 with debiasing is a purely algebraic consequence of the original bias characterization and the construction of the debiased Mondrian random forest estimator.

Proof (Lemma 4)**.** By the definition of the debiased estimator and Lemma 1, as J and a_r are fixed,

$$\begin{aligned}
&\mathbb{E}\left[\left(\mathbb{E}\left[\hat{\mu}_{\mathrm{d}}(x)\mid \mathbf{X},\mathbf{T}\right] - \sum_{l=0}^J \omega_l\left(\mu(x)+\sum_{r=1}^{\lfloor\beta/2\rfloor}\frac{B_r(x)}{a_l^{2r}\lambda^{2r}}\right)\right)^2\right] \\
&\quad= \mathbb{E}\left[\left(\sum_{l=0}^J\omega_l\left(\mathbb{E}\left[\hat{\mu}_l(x)\mid\mathbf{X},\mathbf{T}\right]-\mu(x)-\sum_{r=1}^{\lfloor\beta/2\rfloor}\frac{B_r(x)}{a_l^{2r}\lambda^{2r}}\right)\right)^2\right] \\
&\quad\lesssim \sum_{l=0}^J\mathbb{E}\left[\left(\mathbb{E}\left[\hat{\mu}_l(x)\mid\mathbf{X},\mathbf{T}\right]-\mu(x)-\sum_{r=1}^{\lfloor\beta/2\rfloor}\frac{B_r(x)}{a_l^{2r}\lambda^{2r}}\right)^2\right] \lesssim \frac{1}{\lambda^{2\beta}}+\frac{1}{\lambda^{2(1\wedge\beta)}B}+\frac{1}{\lambda^{2(1\wedge\beta)}}\frac{\lambda^d}{n}.
\end{aligned}$$

It remains to evaluate the resulting bias. Recalling that $A_{rs} = a_{r-1}^{2-2s}$ and $A\omega = e_0$, we have

$$\begin{aligned}
\sum_{l=0}^{J} \omega_l \left(\mu(x) + \sum_{r=1}^{\lfloor \beta/2 \rfloor} \frac{B_r(x)}{a_l^{2r} \lambda^{2r}} \right) &= \mu(x) \sum_{l=0}^{J} \omega_l + \sum_{r=1}^{\lfloor \beta/2 \rfloor} \frac{B_r(x)}{\lambda^{2r}} \sum_{l=0}^{J} \frac{\omega_l}{a_l^{2r}} \\
&= \mu(x) (A\omega)_1 + \sum_{r=1}^{\lfloor \beta/2 \rfloor \wedge J} \frac{B_r(x)}{\lambda^{2r}} (A\omega)_{r+1} + \sum_{r=(\lfloor \beta/2 \rfloor \wedge J)+1}^{\lfloor \beta/2 \rfloor} \frac{B_r(x)}{\lambda^{2r}} \sum_{l=0}^{J} \frac{\omega_l}{a_l^{2r}} \\
&= \mu(x) + \mathbb{I}\{\lfloor \beta/2 \rfloor \geq J+1\} \frac{B_{J+1}(x)}{\lambda^{2J+2}} \sum_{l=0}^{J} \frac{\omega_l}{a_l^{2J+2}} + O\left(\frac{1}{\lambda^{2J+4}} \right) \\
&= \mu(x) + \mathbb{I}\{2J+2 < \beta\} \frac{\bar{\omega} B_{J+1}(x)}{\lambda^{2J+2}} + O\left(\frac{1}{\lambda^{2J+4}} \right).
\end{aligned}$$

□

Proof (Lemma 5)**.** Firstly, note that with $\sigma_i^2 = \sigma^2(X_i)$ for brevity,

$$\begin{aligned}
\tilde{\Sigma}_{\mathrm{d}}(x) &= \frac{n}{\lambda^d} \operatorname{Var}\left[\sum_{r=0}^{J} \omega_r \frac{1}{B} \sum_{b=1}^{B} \sum_{i=1}^{n} \frac{Y_i \mathbb{I}\{X_i \in T_{br}(x)\}}{N_{br}(x)} \;\Big|\; \mathbf{X}, \mathbf{T} \right] \\
&= \frac{n}{\lambda^d} \sum_{i=1}^{n} \sum_{r=0}^{J} \sum_{r'=0}^{J} \omega_r \omega_{r'} \frac{1}{B^2} \sum_{b=1}^{B} \sum_{b'=1}^{B} \frac{\mathbb{I}_{ibr}(x) \mathbb{I}_{ib'r'}(x) \sigma_i^2}{N_{br}(x) N_{b'r'}(x)}.
\end{aligned}$$

Part 1: bounding the variance of $\tilde{\Sigma}_{\mathrm{d}}(x)$

$$\begin{aligned}
\operatorname{Var}\left[\tilde{\Sigma}_{\mathrm{d}}(x) \right] &= \operatorname{Var}\left[\frac{n}{\lambda^d} \sum_{i=1}^{n} \sum_{r=0}^{J} \sum_{r'=0}^{J} \omega_r \omega_{r'} \frac{1}{B^2} \sum_{b=1}^{B} \sum_{b'=1}^{B} \frac{\mathbb{I}_{ibr}(x) \mathbb{I}_{ib'r'}(x) \sigma_i^2}{N_{br}(x) N_{b'r'}(x)} \right] \\
&\lesssim \frac{n^2}{\lambda^{2d}} \frac{1}{B^4} \operatorname{Var}\left[\sum_{i=1}^{n} \sum_{b=1}^{B} \sum_{b'=1}^{B} \frac{\mathbb{I}_{ibr}(x) \mathbb{I}_{ib'r'}(x) \sigma_i^2}{N_{br}(x) N_{b'r'}(x)} \right] \\
&\lesssim \frac{n^2}{\lambda^{2d}} \frac{1}{B^4} \mathbb{E}\left[\operatorname{Var}\left[\sum_{i=1}^{n} \sum_{b=1}^{B} \sum_{b'=1}^{B} \frac{\mathbb{I}_{ibr}(x) \mathbb{I}_{ib'r'}(x) \sigma_i^2}{N_{br}(x) N_{b'r'}(x)} \;\Big|\; \mathbf{X} \right] \right] \\
&\quad + \frac{n^2}{\lambda^{2d}} \frac{1}{B^4} \operatorname{Var}\left[\mathbb{E}\left[\sum_{i=1}^{n} \sum_{b=1}^{B} \sum_{b'=1}^{B} \frac{\mathbb{I}_{ibr}(x) \mathbb{I}_{ib'r'}(x) \sigma_i^2}{N_{br}(x) N_{b'r'}(x)} \;\Big|\; \mathbf{X} \right] \right].
\end{aligned} \tag{SA7}$$

For the first term in (SA7),

$$\begin{aligned}
&\mathbb{E}\left[\operatorname{Var}\left[\sum_{i=1}^{n} \sum_{b=1}^{B} \sum_{b'=1}^{B} \frac{\mathbb{I}_{ibr}(x) \mathbb{I}_{ib'r'}(x) \sigma_i^2}{N_{br}(x) N_{b'r'}(x)} \;\Big|\; \mathbf{X} \right] \right] = \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{b=1}^{B} \sum_{b'=1}^{B} \sum_{\tilde{b}=1}^{B} \sum_{\tilde{b}'=1}^{B} \\
&\mathbb{E}\left[\sigma_i^2 \sigma_j^2 \left(\frac{\mathbb{I}_{ibr}(x)}{N_{br}(x)} \frac{\mathbb{I}_{ib'r'}(x)}{N_{b'r'}(x)} - \mathbb{E}\left[\frac{\mathbb{I}_{ibr}(x)}{N_{br}(x)} \frac{\mathbb{I}_{ib'r'}(x)}{N_{b'r'}(x)} \Big| \mathbf{X} \right] \right) \left(\frac{\mathbb{I}_{j\tilde{b}r}(x)}{N_{\tilde{b}r}(x)} \frac{\mathbb{I}_{j\tilde{b}'r'}(x)}{N_{\tilde{b}'r'}(x)} - \mathbb{E}\left[\frac{\mathbb{I}_{j\tilde{b}r}(x)}{N_{\tilde{b}r}(x)} \frac{\mathbb{I}_{j\tilde{b}'r'}(x)}{N_{\tilde{b}'r'}(x)} \Big| \mathbf{X} \right] \right) \right].
\end{aligned}$$

Since T_{br} is independent of $T_{b'r'}$ given $\mathbf{X}$, the summands are zero whenever $|\{b, b', \tilde{b}, \tilde{b}'\}| = 4$.

Further, by the Cauchy–Schwarz inequality and Lemma SA6,

$$\begin{aligned}
&\frac{n^2}{\lambda^{2d}} \frac{1}{B^4} \mathbb{E}\left[\operatorname{Var}\left[\sum_{i=1}^n \sum_{b=1}^B \sum_{b'=1}^B \frac{\mathbb{I}_{ibr}(x)}{N_{br}(x)} \frac{\mathbb{I}_{ib'r'}(x)}{N_{b'r'}(x)} \sigma^2(X_i) \;\Big|\; \mathbf{X}\right]\right] \\
&\quad \lesssim \frac{n^2}{\lambda^{2d}} \frac{1}{B^3} \sum_{b=1}^B \sum_{b'=1}^B \mathbb{E}\left[\left(\sum_{i=1}^n \frac{\mathbb{I}_{ibr}(x)}{N_{br}(x)} \frac{\mathbb{I}_{ib'r'}(x)}{N_{b'r'}(x)}\right)^2\right] \lesssim \frac{n^2}{\lambda^{2d}} \frac{1}{B^3} \sum_{b=1}^B \sum_{b'=1}^B \mathbb{E}\left[\frac{\mathbb{I}_{br}(x)}{N_{br}(x)} \frac{\mathbb{I}_{b'r'}(x)}{N_{b'r'}(x)}\right] \\
&\quad \lesssim \frac{n^2}{\lambda^{2d}} \frac{1}{B^3} \left(B^2 \frac{\lambda^{2d}}{n^2} + B \frac{\lambda^{2d} (\log n)^d}{n^2}\right) \lesssim \frac{1}{B} + \frac{(\log n)^d}{B^2} \lesssim \frac{1}{B}.
\end{aligned}$$

For the second term in (SA7), the random variable inside the variance is a nonlinear function of the i.i.d. variables X_i, so we apply the Efron–Stein inequality [Efron and Stein, 1981]. Let $\hat{X}_{ij} = X_i$ if $i \neq j$ and be an independent copy of X_j, denoted $\hat{X}_j$, if $i = j$, and define $\sigma_{ij}^2 = \sigma^2(\hat{X}_{ij})$. Write $\hat{\mathbb{I}}_{ijbr}(x) = \mathbb{I}\{\hat{X}_{ij} \in T_{br}(x)\}$ and $\hat{\mathbb{I}}_{jbr}(x) = \mathbb{I}\{\hat{X}_j \in T_{br}(x)\}$, and also $\hat{N}_{jbr}(x) = \sum_{i=1}^n \hat{\mathbb{I}}_{ijbr}(x)$, We use the leave-one-out notation $N_{-jbr}(x) = \sum_{i \neq j} \mathbb{I}_{ibr}(x)$ and also write $N_{-jbr \cap b'r'} = \sum_{i \neq j} \mathbb{I}_{ibr}(x) \mathbb{I}_{ib'r'}(x)$.

$$\begin{aligned}
&\frac{n^2}{\lambda^{2d}} \frac{1}{B^4} \operatorname{Var}\left[\mathbb{E}\left[\sum_{i=1}^n \sum_{b=1}^B \sum_{b'=1}^B \frac{\mathbb{I}_{ibr}(x)}{N_{br}(x)} \frac{\mathbb{I}_{ib'r'}(x)}{N_{b'r'}(x)} \sigma_i^2 \;\Big|\; \mathbf{X}\right]\right] \lesssim \frac{n^2}{\lambda^{2d}} \operatorname{Var}\left[\mathbb{E}\left[\sum_{i=1}^n \frac{\mathbb{I}_{ibr}(x)}{N_{br}(x)} \frac{\mathbb{I}_{ib'r'}(x)}{N_{b'r'}(x)} \sigma_i^2 \;\Big|\; \mathbf{X}\right]\right] \\
&\quad \lesssim \frac{n^2}{\lambda^{2d}} \sum_{j=1}^n \mathbb{E}\left[\left(\sum_{i=1}^n \left(\frac{\mathbb{I}_{ibr}(x) \mathbb{I}_{ib'r'}(x) \sigma_i^2}{N_{br}(x) N_{b'r'}(x)} - \frac{\hat{\mathbb{I}}_{ijbr}(x) \hat{\mathbb{I}}_{ijb'r'}(x) \hat{\sigma}_{ij}^2}{\hat{N}_{jbr}(x) \hat{N}_{jb'r'}(x)}\right)\right)^2\right] \\
&\quad \lesssim \frac{n^2}{\lambda^{2d}} \sum_{j=1}^n \mathbb{E}\left[\left(\left|\frac{1}{N_b(x) N_{b'r'}(x)} - \frac{1}{\hat{N}_{jbr}(x) \hat{N}_{jb'r'}(x)}\right| \sum_{i \neq j} \mathbb{I}_{ibr}(x) \mathbb{I}_{ib'r'}(x) \sigma_i^2\right)^2\right] \\
&\qquad + \frac{n^2}{\lambda^{2d}} \sum_{j=1}^n \mathbb{E}\left[\left(\frac{\mathbb{I}_{jbr}(x) \mathbb{I}_{jb'r'}(x) \sigma_j^2}{N_{br}(x) N_{b'r'}(x)} - \frac{\hat{\mathbb{I}}_{jbr}(x) \hat{\mathbb{I}}_{jb'r'}(x) \hat{\sigma}_{jj}^2}{\hat{N}_{jbr}(x) \hat{N}_{jb'r'}(x)}\right)^2\right] \\
&\quad \lesssim \frac{n^2}{\lambda^{2d}} \sum_{j=1}^n \mathbb{E}\left[N_{-jbr \cap b'r'}(x)^2 \left|\frac{1}{N_{br}(x) N_{b'r'}(x)} - \frac{1}{\hat{N}_{jbr}(x) \hat{N}_{jb'r'}(x)}\right|^2 + \frac{\mathbb{I}_{jbr}(x) \mathbb{I}_{jb'r'}(x)}{N_{br}(x)^2 N_{b'r'}(x)^2}\right]. \qquad \text{(SA8)}
\end{aligned}$$

For the first term in (SA8), note that since $|N_{br}(x) - \hat{N}_{jbr}(x)| \leq \mathbb{I}_{jbr}(x) + \hat{\mathbb{I}}_{jbr}(x)$ and similarly $|N_{b'r'}(x) - \hat{N}_{jb'r'}(x)| \leq \mathbb{I}_{jb'r'}(x) + \hat{\mathbb{I}}_{jb'r'}(x)$,

$$\begin{aligned}
&\left|\frac{1}{N_{br}(x) N_{b'r'}(x)} - \frac{1}{\hat{N}_{jbr}(x) \hat{N}_{jb'r'}(x)}\right| \\
&\quad \leq \frac{1}{N_{br}(x)} \left|\frac{1}{N_{b'r'}(x)} - \frac{1}{\hat{N}_{jb'r'}(x)}\right| + \frac{1}{\hat{N}_{jb'r'}(x)} \left|\frac{1}{N_{br}(x)} - \frac{1}{\hat{N}_{jbr}(x)}\right| \\
&\quad \leq \frac{\mathbb{I}_{jbr}(x) + \hat{\mathbb{I}}_{jbr}(x)}{N_{-jbr}(x) N_{-jb'r'}(x)^2} + \frac{\mathbb{I}_{jb'r'}(x) + \hat{\mathbb{I}}_{jb'r'}(x)}{N_{-jb'r'}(x) N_{-jbr}(x)^2}.
\end{aligned}$$

Therefore by Lemma SA6,

$$\begin{aligned}
&\frac{n^2}{\lambda^{2d}} \frac{1}{B^4} \operatorname{Var}\left[\mathbb{E}\left[\sum_{i=1}^{n} \sum_{b=1}^{B} \sum_{b'=1}^{B} \frac{\mathbb{I}_{ibr}(x)\mathbb{I}_{ib'r'}(x)\sigma^2(X_i)}{N_{br}(x)N_{b'r'}(x)} \;\Big|\; \mathbf{X}\right]\right] \\
&\quad\lesssim \frac{n^2}{\lambda^{2d}} \sum_{j=1}^{n} \mathbb{E}\left[\frac{\mathbb{I}_{jbr}(x)\mathbb{I}_{br\cap b'r'}(x)}{N_{br}(x)^2 N_{b'r'}(x)^2}\right] \lesssim \frac{n^2}{\lambda^{2d}} \mathbb{E}\left[\frac{\mathbb{I}_{br}(x)\mathbb{I}_{b'r'}(x)}{N_{br}(x)^{3/2} N_{b'r'}(x)^{3/2}}\right] \lesssim \frac{n^2}{\lambda^{2d}} \frac{\lambda^{3d}}{n^3} \lesssim \frac{\lambda^d}{n}.
\end{aligned}$$

We deduce that

$$\operatorname{Var}\left[\tilde{\Sigma}_{\mathrm{d}}(x)\right] \lesssim \frac{1}{B} + \frac{\lambda^d}{n}.$$

Part 2: controlling the expectation of $\tilde{\Sigma}_{\mathrm{d}}(x)$

$$\mathbb{E}\left[\tilde{\Sigma}_{\mathrm{d}}(x)\right] = \frac{n}{\lambda^d} \sum_{i=1}^{n} \sum_{r=0}^{J} \sum_{r'=0}^{J} \omega_r \omega_{r'} \frac{1}{B^2} \sum_{b=1}^{B} \sum_{b'=1}^{B} \mathbb{E}\left[\frac{\mathbb{I}_{ibr}(x)\mathbb{I}_{ib'r'}(x)\sigma^2(X_i)}{N_{br}(x)N_{b'r'}(x)}\right].$$

Firstly, by Lemma SA6, the diagonal terms in the forest are

$$\left|\frac{n}{\lambda^d} \sum_{i=1}^{n} \sum_{r=0}^{J} \sum_{r'=0}^{J} \omega_r \omega_{r'} \frac{1}{B^2} \sum_{b=1}^{B} \mathbb{E}\left[\frac{\mathbb{I}_{ibr}(x)\mathbb{I}_{ibr'}(x)\sigma^2(X_i)}{N_{br}(x)N_{br'}(x)}\right]\right| \lesssim \frac{n}{\lambda^d} \frac{1}{B} \mathbb{E}\left[\frac{\mathbb{I}_{br}(x)}{N_{br}(x)}\right] \lesssim \frac{1}{B},$$

so it suffices to take $b \neq b'$ since

$$\mathbb{E}\left[\tilde{\Sigma}_{\mathrm{d}}(x)\right] = \frac{n^2}{\lambda^d} \sum_{r=0}^{J} \sum_{r'=0}^{J} \omega_r \omega_{r'} \mathbb{E}\left[\frac{\mathbb{I}_{ibr}(x)\mathbb{I}_{ib'r'}(x)\sigma^2(X_i)}{N_{br}(x)N_{b'r'}(x)}\right] + O\left(\frac{1}{B}\right).$$

Next, note that

$$\mathbb{E}\left[\frac{\mathbb{I}_{ibr}(x)\mathbb{I}_{ib'r'}(x)\sigma^2(X_i)}{N_{br}(x)N_{b'r'}(x)}\right] = \sigma^2(x)\mathbb{E}\left[\frac{\mathbb{I}_{ibr}(x)\mathbb{I}_{ib'r'}(x)}{N_{br}(x)N_{b'r'}(x)}\right] + \mathbb{E}\left[\frac{\mathbb{I}_{ibr}(x)\mathbb{I}_{ib'r'}(x)\left(\sigma^2(X_i) - \sigma^2(x)\right)}{N_{br}(x)N_{b'r'}(x)}\right].$$

Since $\sigma^2 \in \mathcal{H}^{\beta_\sigma}$, we have by Lemma SA1 and Lemma SA6 that

$$\frac{n^2}{\lambda^d}\mathbb{E}\left[\frac{\mathbb{I}_{ibr}(x)\mathbb{I}_{ib'r'}(x)\left|\sigma^2(X_i) - \sigma^2(x)\right|}{N_{br}(x)N_{b'r'}(x)}\right] \lesssim \frac{n^2}{\lambda^d}\frac{1}{n}\mathbb{E}\left[\frac{\mathbb{I}_{b'r'}(x)\max_j |T_{br}(x)_j|}{N_{b'r'}(x)}\right] \lesssim \frac{1}{\lambda^{1\wedge\beta_\sigma}}.$$

Therefore

$$\frac{n^2}{\lambda^d}\mathbb{E}\left[\frac{\mathbb{I}_{ibr}(x)\mathbb{I}_{ib'r'}(x)\sigma^2(X_i)}{N_{br}(x)N_{b'r'}(x)}\right] = \sigma^2(x)\frac{n^2}{\lambda^d}\mathbb{E}\left[\frac{\mathbb{I}_{ibr}(x)\mathbb{I}_{ib'r'}(x)}{N_{br}(x)N_{b'r'}(x)}\right] + O\left(\frac{1}{\lambda^{1\wedge\beta_\sigma}}\right).$$

Next, by conditioning on T_{br}, $T_{b'r'}$, $N_{-ibr}(x)$, and $N_{-ib'r'}(x)$,

$$
\begin{aligned}
\mathbb{E}\left[\frac{\mathbb{I}_{ibr}(x)\mathbb{I}_{ib'r'}(x)}{N_{br}(x)N_{b'r'}(x)}\right] &= \mathbb{E}\left[\frac{\int_{T_{br}(x)\cap T_{b'r'}(x)} f(\xi)\,\mathrm{d}\xi}{(N_{-ibr}(x)+1)(N_{-ib'r'}(x)+1)}\right] \\
&= \mathbb{E}\left[\frac{f(x)|T_{br}(x)\cap T_{b'r'}(x)|}{(N_{-ibr}(x)+1)(N_{-ib'r'}(x)+1)}\right] + \mathbb{E}\left[\frac{\int_{T_{br}(x)\cap T_{b'r'}(x)}(f(\xi)-f(x))\,\mathrm{d}\xi}{(N_{-ibr}(x)+1)(N_{-ib'r'}(x)+1)}\right] \\
&= f(x)\,\mathbb{E}\left[\frac{|T_{br}(x)\cap T_{b'r'}(x)|}{(N_{-ibr}(x)+1)(N_{-ib'r'}(x)+1)}\right] + O\left(\frac{\lambda^d}{n^2}\frac{1}{\lambda^{1\wedge\beta_f}}\right)
\end{aligned}
$$

by an argument based on Lemma SA1, the Hölder property of $f(x)$, and the proof of Lemma SA6. Hence

$$
\begin{aligned}
\frac{n^2}{\lambda^d}\mathbb{E}\left[\frac{\mathbb{I}_{ibr}(x)\mathbb{I}_{ib'r'}(x)\sigma^2(X_i)}{N_{br}(x)N_{b'r'}(x)}\right] &= \sigma^2(x)f(x)\frac{n^2}{\lambda^d}\mathbb{E}\left[\frac{|T_{br}(x)\cap T_{b'r'}(x)|}{(N_{-ibr}(x)+1)(N_{-ib'r'}(x)+1)}\right] \\
&\quad + O\left(\frac{1}{\lambda^{1\wedge\beta_f\wedge\beta_\sigma}}\right).
\end{aligned}
$$

Apply Lemma SA7 to approximate the expectation with $N_{-ib'r'\setminus br}(x) = \sum_{j\neq i}\mathbb{I}\{X_j \in T_{b'r'}(x)\setminus T_{br}(x)\}$:

$$
\begin{aligned}
&\mathbb{E}\left[\frac{|T_{br}(x)\cap T_{b'r'}(x)|}{(N_{-ibr}(x)+1)(N_{-ib'r'}(x)+1)}\right] \\
&\quad = \mathbb{E}\left[\frac{|T_{br}(x)\cap T_{b'r'}(x)|}{N_{-ibr}(x)+1}\,\mathbb{E}\left[\frac{1}{N_{-ib'r'\cap br}(x)+N_{-ib'r'\setminus br}(x)+1}\,\Big|\,\mathbf{T}, N_{-ib'r'\cap br}(x), N_{-ibr\setminus b'r'}(x)\right]\right].
\end{aligned}
$$

Now conditional on $\mathbf{T}$, $N_{-ib'r'\cap br}(x)$, and $N_{-ibr\setminus b'r'}(x)$,

$$
N_{-ib'r'\setminus br}(x) \sim \mathrm{Bin}\left(n-1-N_{-ibr}(x),\ \frac{\int_{T_{b'r'}(x)\setminus T_{br}(x)} f(\xi)\,\mathrm{d}\xi}{1-\int_{T_{br}(x)} f(\xi)\,\mathrm{d}\xi}\right).
$$

We bound these parameters above and below. Firstly, by applying Lemma SA2 with $B=1$, we have

$$
\mathbb{P}\left(N_{-ibr}(x) > t^{d+1}\frac{n}{\lambda^d}\right) \leq 4de^{-t/(4\|f\|_\infty(1+1/a_r))} \leq e^{-t/C}
$$

for some $C > 0$ and sufficiently large t. Next, note if f is β_f-Hölder with constant L, by Lemma SA1,

$$
\begin{aligned}
&\mathbb{P}\left(\left|\frac{\int_{T_{b'r'}(x)\setminus T_{br}(x)} f(\xi)\,\mathrm{d}\xi}{1-\int_{T_{br}(x)} f(\xi)\,\mathrm{d}\xi} - f(x)|T_{b'r'}(x)\setminus T_{br}(x)|\right| > 2L\,|T_{b'r'}(x)\setminus T_{br}(x)|\sum_{j=1}^{d}|T_{b'r'}(x)_j|^{1\wedge\beta_f}\right) \\
&\quad\leq \mathbb{P}\left(\frac{\int_{T_{b'r'}(x)\setminus T_{br}(x)} |f(\xi)-f(x)|\,\mathrm{d}\xi}{1-\int_{T_{br}(x)} f(\xi)\,\mathrm{d}\xi} > 2L\,|T_{b'r'}(x)\setminus T_{br}(x)|\sum_{j=1}^{d}|T_{b'r'}(x)_j|^{1\wedge\beta_f}\right) \\
&\quad\leq \mathbb{P}\left(\frac{1}{1-\int_{T_{br}(x)} f(\xi)\,\mathrm{d}\xi} > 2\right) \leq \mathbb{P}\left(\int_{T_{br}(x)} f(\xi)\,\mathrm{d}\xi > \frac{1}{2}\right) \\
&\quad\leq \mathbb{P}\left(|T_{br}(x)| > \frac{1}{2\|f\|_\infty}\right) \lesssim \mathbb{P}\left(\max_{1\leq j\leq d}|T_{br}(x)_j|^{1\wedge\beta_f} > \frac{1}{2\|f\|_\infty}\right) \lesssim 2de^{-\lambda/(4\|f\|_\infty)} \lesssim e^{-\lambda/C},
\end{aligned}
$$

increasing C as necessary. Thus with probability at least $1 - e^{-t/C} - e^{-\lambda/C}$,

$$
\begin{aligned}
N_{-ib'r'\setminus br}(x) &\leq \mathrm{Bin}\left(n,\, |T_{b'r'}(x)\setminus T_{br}(x)|\left(f(x) + 2L\sum_{j=1}^{d}|T_{b'r'}(x)_j|^{1\wedge\beta_f}\right)\right) \\
N_{-ib'r'\setminus br}(x) &\geq \mathrm{Bin}\left(n\left(1-\frac{t^{d+1}}{\lambda^d}-\frac{1}{n}\right),\, |T_{b'r'}(x)\setminus T_{br}(x)|\left(f(x) - 2L\sum_{j=1}^{d}|T_{b'r'}(x)_j|^{1\wedge\beta_f}\right)\right).
\end{aligned}
$$

So by Lemma SA7 conditionally on $\mathbf{T}$, $N_{-ib'r'\cap br}(x)$, and $N_{-ibr\setminus b'r'}(x)$, taking $t = 4C\log n$ and recalling $\lambda \gtrsim (\log n)^3$, with probability at least $1 - n^{-3}$,

$$
\begin{aligned}
&\Bigg|\mathbb{E}\left[\frac{1}{N_{-ib'r'\cap br}(x) + N_{-ib'r'\setminus br}(x) + 1}\;\Big|\; \mathbf{T}, N_{-ib'r'\cap br}(x), N_{-ibr\setminus b'r'}(x)\right] \\
&\qquad - \frac{1}{N_{-ib'r'\cap br}(x) + nf(x)|T_{b'r'}(x)\setminus T_{br}(x)| + 1}\Bigg| \lesssim \frac{1 + n|T_{b'r'}(x)_j|^{1\wedge\beta_f}|T_{b'r'}(x)\setminus T_{br}(x)|}{\left(N_{-ib'r'\cap br}(x) + n|T_{b'r'}(x)\setminus T_{br}(x)| + 1\right)^2}.
\end{aligned}
$$

Therefore by the same approach as the proof of Lemma SA4,

$$
\begin{aligned}
&\left|\mathbb{E}\left[\frac{|T_{br}(x)\cap T_{b'r'}(x)|}{(N_{-ibr}(x)+1)(N_{-ib'r'}(x)+1)} - \frac{|T_{br}(x)\cap T_{b'r'}(x)|}{(N_{-ibr}(x)+1)(N_{-ib'r'\cap br}(x) + nf(x)|T_{b'r'}(x)\setminus T_{br}(x)| + 1)}\right]\right| \\
&\quad\lesssim \mathbb{E}\left[\frac{|T_{br}(x)\cap T_{b'r'}(x)|}{N_{-ibr}(x)+1}\frac{1 + n|T_{b'r'}(x)_j|^{1\wedge\beta_f}|T_{b'r'}(x)\setminus T_{br}(x)|}{\left(N_{-ib'r'\cap br}(x) + n|T_{b'r'}(x)\setminus T_{br}(x)| + 1\right)^2}\right] + n^{-3} \\
&\quad\lesssim \mathbb{E}\left[\frac{|T_{br}(x)\cap T_{b'r'}(x)|}{n|T_{br}(x)|+1}\frac{1 + n|T_{b'r'}(x)_j|^{1\wedge\beta_f}|T_{b'r'}(x)\setminus T_{br}(x)|}{(n|T_{b'r'}(x)|+1)^2}\right] + n^{-3} \\
&\quad\lesssim \mathbb{E}\left[\frac{1}{n}\frac{1}{(n|T_{br}(x)|+1)(n|T_{b'r'}(x)|+1)} + \frac{1}{n}\frac{|T_{b'r'}(x)_j|^{1\wedge\beta_f}}{n|T_{b'r'}(x)|+1}\right] + n^{-3} \\
&\quad\lesssim \frac{\lambda^{2d}}{n^3} + \frac{1}{n\lambda^{1\wedge\beta_f}}\frac{\lambda^d}{n} \lesssim \frac{\lambda^d}{n^2}\left(\frac{\lambda^d}{n} + \frac{1}{\lambda^{1\wedge\beta_f}}\right).
\end{aligned}
$$

Now apply the same argument to the other term in the expectation, to see that

$$\left|\mathbb{E}\left[\frac{1}{N_{-ibr\cap b'r'}(x)+N_{-ibr\setminus b'r'}(x)+1}\;\Big|\;\mathbf{T},N_{-ibr\cap b'r'}(x),N_{-ib'r'\setminus br}(x)\right]\right.$$
$$\left.-\frac{1}{N_{-ibr\cap b'r'}(x)+nf(x)|T_{br}(x)\setminus T_{b'r'}(x)|+1}\right|\lesssim\frac{1+n|T_{br}(x)_j|^{1\wedge\beta_f}|T_{br}(x)\setminus T_{b'r'}(x)|}{(N_{-ibr\cap b'r'}(x)+n|T_{br}(x)\setminus T_{b'r'}(x)|+1)^2}.$$

with probability at least $1-n^{-3}-e^{-\lambda/C}$, and so likewise again with $t=4C\log n$,

$$\frac{n^2}{\lambda^d}\left|\mathbb{E}\left[\frac{|T_{br}(x)\cap T_{b'r'}(x)|}{N_{-ibr}(x)+1}\frac{1}{N_{-ib'r'\cap br}(x)+nf(x)|T_{b'r'}(x)\setminus T_{br}(x)|+1}\right]\right.$$
$$\left.-\mathbb{E}\left[\frac{|T_{br}(x)\cap T_{b'r'}(x)|}{N_{-ibr\cap b'r'}(x)+nf(x)|T_{br}(x)\setminus T_{b'r'}(x)|+1}\frac{1}{N_{-ib'r'\cap br}(x)+nf(x)|T_{b'r'}(x)\setminus T_{br}(x)|+1}\right]\right|$$
$$\lesssim\frac{n^2}{\lambda^d}\mathbb{E}\left[\frac{1+n|T_{br}(x)_j|^{1\wedge\beta_f}|T_{br}(x)\setminus T_{b'r'}(x)|}{(N_{-ibr\cap b'r'}(x)+n|T_{br}(x)\setminus T_{b'r'}(x)|+1)^2}\frac{|T_{br}(x)\cap T_{b'r'}(x)|}{N_{-ib'r'\cap br}(x)+nf(x)|T_{b'r'}(x)\setminus T_{br}(x)|+1}\right]$$
$$+\frac{n^2}{\lambda^d}n^{-3}\lesssim\frac{\lambda^d}{n}+\frac{1}{\lambda^{1\wedge\beta_f}}.$$

Thus far we have proven that

$$\frac{n^2}{\lambda^d}\mathbb{E}\left[\frac{\mathbb{I}_{ibr}(x)\mathbb{I}_{ib'r'}(x)\sigma^2(X_i)}{N_{br}(x)N_{b'r'}(x)}\right]=\sigma^2(x)f(x)\frac{n^2}{\lambda^d}$$
$$\times\mathbb{E}\left[\frac{|T_{br}(x)\cap T_{b'r'}(x)|}{N_{-ibr\cap b'r'}(x)+nf(x)|T_{br}(x)\setminus T_{b'r'}(x)|+1}\frac{1}{N_{-ib'r'\cap br}(x)+nf(x)|T_{b'r'}(x)\setminus T_{br}(x)|+1}\right]$$
$$+O\left(\frac{1}{\lambda^{1\wedge\beta_f\wedge\beta_\sigma}}+\frac{\lambda^d}{n}\right).$$

Next we remove the $N_{-ibr\cap b'r'}(x)$ terms. As before, with probability at least $1-e^{-t/C}-e^{-\lambda/C}$, conditional on $\mathbf{T}$,

$$N_{-ibr\cap b'r'}(x)\le\operatorname{Bin}\left(n,\,|T_{br}(x)\cap T_{b'r'}(x)|\left(f(x)+2L\sum_{j=1}^d|T_{br}(x)_j|^{1\wedge\beta_f}\right)\right),$$
$$N_{-ibr\cap b'r'}(x)\ge\operatorname{Bin}\left(n\left(1-\frac{t^{d+1}}{\lambda^d}-\frac{1}{n}\right),\,|T_{br}(x)\cap T_{b'r'}(x)|\left(f(x)-2L\sum_{j=1}^d|T_{br}(x)_j|^{1\wedge\beta_f}\right)\right).$$

So by Lemma SA7 conditionally on $\mathbf{T}$, with $t=4C\log n$ and with probability at least $1-n^{-3}$,

$$\left|\mathbb{E}\left[\frac{1}{N_{-ibr\cap b'r'}(x)+nf(x)|T_{br}(x)\setminus T_{b'r'}(x)|+1}\frac{1}{N_{-ib'r'\cap br}(x)+nf(x)|T_{b'r'}(x)\setminus T_{br}(x)|+1}\;\Big|\;\mathbf{T}\right]\right.$$
$$\left.-\frac{1}{nf(x)|T_{br}(x)|+1}\frac{1}{nf(x)|T_{b'r'}(x)|+1}\right|$$
$$\lesssim\frac{1+n|T_{br}(x)_j|^{1\wedge\beta_f}|T_{br}(x)\cap T_{b'r'}(x)|}{(n|T_{br}(x)|+1)(n|T_{b'r'}(x)|+1)}\left(\frac{1}{n|T_{br}(x)|+1}+\frac{1}{n|T_{b'r'}(x)|+1}\right).$$

Now by Lemma SA5,

$$
\begin{aligned}
&\frac{n^2}{\lambda^d}\bigg|\mathbb{E}\left[\frac{|T_{br}(x)\cap T_{b'r'}(x)|}{N_{-ibr\cap b'r'}(x)+nf(x)|T_{br}(x)\setminus T_{b'r'}(x)|+1}\frac{1}{N_{-ib'r'\cap br}(x)+nf(x)|T_{b'r'}(x)\setminus T_{br}(x)|+1}\right]\\
&\quad-\mathbb{E}\left[\frac{|T_{br}(x)\cap T_{b'r'}(x)|}{nf(x)|T_{br}(x)|+1}\frac{1}{nf(x)|T_{b'r'}(x)|+1}\right]\bigg|\\
&\quad\lesssim \frac{n^2}{\lambda^d}\mathbb{E}\left[\frac{1+n|T_{br}(x)_j|^{1\wedge\beta_f}|T_{br}(x)\cap T_{b'r'}(x)|}{(n|T_{br}(x)|+1)(n|T_{b'r'}(x)|+1)}\left(\frac{|T_{br}(x)\cap T_{b'r'}(x)|}{n|T_{br}(x)|+1}+\frac{|T_{br}(x)\cap T_{b'r'}(x)|}{n|T_{b'r'}(x)|+1}\right)\right]+\frac{1}{n\lambda^d}\\
&\quad\lesssim \frac{n^2}{\lambda^d}\frac{1}{n^3}\mathbb{E}\left[\frac{1+n|T_{br}(x)_j|^{1\wedge\beta_f}|T_{br}(x)\cap T_{b'r'}(x)|}{|T_{br}(x)||T_{b'r'}(x)|}\right]+\frac{1}{n\lambda^d}\\
&\quad\lesssim \frac{1}{n\lambda^d}\mathbb{E}\left[\frac{1}{|T_{br}(x)||T_{b'r'}(x)|}\right]+\frac{1}{\lambda^d}\mathbb{E}\left[\frac{|T_{br}(x)_j|^{1\wedge\beta_f}}{|T_{b'r'}(x)|}\right]+\frac{1}{n\lambda^d}\lesssim\frac{\lambda^d}{n}+\frac{1}{\lambda^{1\wedge\beta_f}}.
\end{aligned}
$$

This allows us to deduce that

$$
\begin{aligned}
\frac{n^2}{\lambda^d}\mathbb{E}\left[\frac{\mathbb{I}_{ibr}(x)\mathbb{I}_{ib'r'}(x)\sigma^2(X_i)}{N_{br}(x)N_{b'r'}(x)}\right]&=\sigma^2(x)f(x)\frac{n^2}{\lambda^d}\mathbb{E}\left[\frac{|T_{br}(x)\cap T_{b'r'}(x)|}{(nf(x)|T_{br}(x)|+1)(nf(x)|T_{b'r'}(x)|+1)}\right]\\
&\quad+O\left(\frac{1}{\lambda^{1\wedge\beta_f\wedge\beta_\sigma}}+\frac{\lambda^d}{n}\right),
\end{aligned}
$$

and so

$$
\begin{aligned}
\mathbb{E}\left[\tilde{\Sigma}_{\mathrm{d}}(x)\right]&=\sigma^2(x)f(x)\frac{n^2}{\lambda^d}\sum_{r=0}^{J}\sum_{r'=0}^{J}\omega_r\omega_{r'}\mathbb{E}\left[\frac{|T_{br}(x)\cap T_{b'r'}(x)|}{(nf(x)|T_{br}(x)|+1)(nf(x)|T_{b'r'}(x)|+1)}\right]\\
&\quad+O\left(\frac{1}{\lambda^{1\wedge\beta_f\wedge\beta_\sigma}}+\frac{\lambda^d}{n}+\frac{1}{B}\right).
\end{aligned}
$$

Part 3: calculating the limiting variance $\Sigma_{\mathrm{d}}(x)$

Now that we have reduced the variance to an expression only involving the sizes of Mondrian cells, we can exploit their exact distribution to compute this expectation. Recall from Mourtada et al. [2020, Proposition 1] that we can write

$$
\begin{aligned}
|T_{br}(x)|&=\prod_{j=1}^{d}\left(\frac{E_{1j}}{a_r\lambda}\wedge x_j+\frac{E_{2j}}{a_r\lambda}\wedge(1-x_j)\right), \qquad |T_{b'r'}(x)|=\prod_{j=1}^{d}\left(\frac{E_{3j}}{a_{r'}\lambda}\wedge x_j+\frac{E_{4j}}{a_{r'}\lambda}\wedge(1-x_j)\right),\\
|T_{br}(x)\cap T_{b'r'}(x)|&=\prod_{j=1}^{d}\left(\frac{E_{1j}}{a_r\lambda}\wedge\frac{E_{3j}}{a_{r'}\lambda}\wedge x_j+\frac{E_{2j}}{a_r\lambda}\wedge\frac{E_{4j}}{a_{r'}\lambda}\wedge(1-x_j)\right)
\end{aligned}
$$

where E_{1j}, E_{2j}, E_{3j}, and E_{4j} are independent and Exp(1). Define their non-truncated versions as

$$
\begin{aligned}
|\tilde{T}_{br}(x)|&=a_r^{-d}\lambda^{-d}\prod_{j=1}^{d}\left(E_{1j}+E_{2j}\right), \qquad |\tilde{T}_{b'r'}(x)|=a_{r'}^{-d}\lambda^{-d}\prod_{j=1}^{d}\left(E_{3j}+E_{4j}\right),\\
|\tilde{T}_{br}(x)\cap\tilde{T}_{b'r'}(x)|&=\lambda^{-d}\prod_{j=1}^{d}\left(\frac{E_{1j}}{a_r}\wedge\frac{E_{3j}}{a_{r'}}+\frac{E_{2j}}{a_r}\wedge\frac{E_{4j}}{a_{r'}}\right),
\end{aligned}
$$

and note that

$$
\begin{aligned}
&\mathbb{P}\left((\tilde{T}_{br}(x), \tilde{T}_{b'r'}(x), \tilde{T}_{br}(x) \cap T_{b'r'}(x)) \neq (T_{br}(x), T_{b'r'}(x), T_{br}(x) \cap T_{b'r'}(x))\right) \\
&\quad \leq \sum_{j=1}^{d} \left(\mathbb{P}(E_{1j} \geq a_r \lambda x_j) + \mathbb{P}(E_{3j} \geq a_{r'} \lambda x_j) + \mathbb{P}(E_{2j} \geq a_r \lambda (1 - x_j)) + \mathbb{P}(E_{4j} \geq a_{r'} \lambda (1 - x_j))\right) \\
&\quad \leq e^{-C\lambda}
\end{aligned}
$$

for some $C > 0$ and sufficiently large λ. Hence by the Cauchy–Schwarz inequality and Lemma SA5,

$$
\begin{aligned}
&\frac{n^2}{\lambda^d} \left| \mathbb{E}\left[\frac{|T_{br}(x) \cap T_{b'r'}(x)|}{nf(x)|T_{br}(x)| + 1} \frac{1}{nf(x)|T_{b'r'}(x)| + 1}\right] - \mathbb{E}\left[\frac{|\tilde{T}_{br}(x) \cap T_{b'r'}(x)|}{nf(x)|\tilde{T}_{br}(x)| + 1} \frac{1}{nf(x)|\tilde{T}_{b'r'}(x)| + 1}\right]\right| \\
&\quad \lesssim \frac{n^2}{\lambda^d} e^{-C\lambda} \lesssim \frac{1}{n\lambda^d}
\end{aligned}
$$

as $\lambda \gtrsim (\log n)^3$. Therefore

$$
\begin{aligned}
\frac{n^2}{\lambda^d} \mathbb{E}\left[\frac{\mathbb{I}_{ibr}(x)\mathbb{I}_{ib'r'}(x)\sigma^2(X_i)}{N_{br}(x)N_{b'r'}(x)}\right] &= \sigma^2(x) f(x) \frac{n^2}{\lambda^d} \mathbb{E}\left[\frac{|\tilde{T}_{br}(x) \cap \tilde{T}_{b'r'}(x)|}{(nf(x)|\tilde{T}_{br}(x)| + 1)(nf(x)|\tilde{T}_{b'r'}(x)| + 1)}\right] \\
&\quad + O\left(\frac{1}{\lambda^{1 \wedge \beta_f \wedge \beta_\sigma}} + \frac{\lambda^d}{n}\right).
\end{aligned}
$$

Now we remove the superfluous units in the denominators. Firstly, by independence of the trees,

$$
\begin{aligned}
&\frac{n^2}{\lambda^d} \left| \mathbb{E}\left[\frac{|\tilde{T}_{br}(x) \cap \tilde{T}_{b'r'}(x)|}{(nf(x)|\tilde{T}_{br}(x)| + 1)(nf(x)|\tilde{T}_{b'r'}(x)| + 1)}\right] - \mathbb{E}\left[\frac{|\tilde{T}_{br}(x) \cap \tilde{T}_{b'r'}(x)|}{(nf(x)|\tilde{T}_{br}(x)| + 1)(nf(x)|\tilde{T}_{b'r'}(x)|)}\right]\right| \\
&\quad \lesssim \frac{n^2}{\lambda^d} \mathbb{E}\left[\frac{|\tilde{T}_{br}(x) \cap \tilde{T}_{b'r'}(x)|}{n|\tilde{T}_{br}(x)|} \frac{1}{n^2|\tilde{T}_{b'r'}(x)|^2}\right] \lesssim \frac{1}{n\lambda^d} \mathbb{E}\left[\frac{1}{|T_{br}(x)|}\right] \mathbb{E}\left[\frac{1}{|T_{b'r'}(x)|}\right] \lesssim \frac{\lambda^d}{n}.
\end{aligned}
$$

Secondly, we have in exactly the same manner that

$$
\frac{n^2}{\lambda^d} \left| \mathbb{E}\left[\frac{|\tilde{T}_{br}(x) \cap T_{b'r'}(x)|}{(nf(x)|\tilde{T}_{br}(x)| + 1)(nf(x)|\tilde{T}_{b'r'}(x)|)}\right] - \mathbb{E}\left[\frac{|\tilde{T}_{br}(x) \cap T_{b'r'}(x)|}{n^2 f(x)^2 |\tilde{T}_{br}(x)||\tilde{T}_{b'r'}(x)|}\right]\right| \lesssim \frac{\lambda^d}{n}.
$$

Therefore

$$
\frac{n^2}{\lambda^d} \mathbb{E}\left[\frac{\mathbb{I}_{ibr}(x)\mathbb{I}_{ib'r'}(x)\sigma^2(X_i)}{N_{br}(x)N_{b'r'}(x)}\right] = \frac{\sigma^2(x)}{f(x)} \frac{1}{\lambda^d} \mathbb{E}\left[\frac{|\tilde{T}_{br}(x) \cap \tilde{T}_{b'r'}(x)|}{|\tilde{T}_{br}(x)||\tilde{T}_{b'r'}(x)|}\right] + O\left(\frac{1}{\lambda^{1 \wedge \beta_f \wedge \beta_\sigma}} + \frac{\lambda^d}{n}\right).
$$

It remains to compute this integral. By independence over $1 \le j \le d$,

$$
\begin{aligned}
&\mathbb{E}\left[\frac{|\tilde{T}_{br}(x) \cap \tilde{T}_{b'r'}(x)|}{|\tilde{T}_{br}(x)||\tilde{T}_{b'r'}(x)|}\right] \\
&\quad = a_r^d a_{r'}^d \lambda^d \prod_{j=1}^d \mathbb{E}\left[\frac{(E_{1j}/a_r) \wedge (E_{3j}/a_{r'}) + (E_{2j} a_r) \wedge (E_{4j}/a_{r'})}{(E_{1j}+E_{2j})\,(E_{3j}+E_{4j})}\right] \\
&\quad = 2^d a_r^d a_{r'}^d \lambda^d \prod_{j=1}^d \mathbb{E}\left[\frac{(E_{1j}/a_r) \wedge (E_{3j}/a_{r'})}{(E_{1j}+E_{2j})\,(E_{3j}+E_{4j})}\right] \\
&\quad = 2^d a_r^d a_{r'}^d \lambda^d \prod_{j=1}^d \int_0^\infty \int_0^\infty \int_0^\infty \int_0^\infty \frac{(t_1/a_r) \wedge (t_3/a_{r'})}{(t_1+t_2)\,(t_3+t_4)} e^{-t_1-t_2-t_3-t_4} \,\mathrm{d}t_1 \,\mathrm{d}t_2 \,\mathrm{d}t_3 \,\mathrm{d}t_4 \\
&\quad = 2^d a_r^d a_{r'}^d \lambda^d \prod_{j=1}^d \int_0^\infty \int_0^\infty ((t_1/a_r) \wedge (t_3/a_{r'})) e^{-t_1-t_3} \left(\int_0^\infty \frac{e^{-t_2}}{t_1+t_2} \,\mathrm{d}t_2\right) \left(\int_0^\infty \frac{e^{-t_4}}{t_3+t_4} \,\mathrm{d}t_4\right) \mathrm{d}t_1 \,\mathrm{d}t_3 \\
&\quad = 2^d a_r^d a_{r'}^d \lambda^d \prod_{j=1}^d \int_0^\infty \int_0^\infty ((t/a_r) \wedge (s/a_{r'})) \Gamma(0,t)\Gamma(0,s) \,\mathrm{d}t \,\mathrm{d}s,
\end{aligned}
$$

where we used $\int_0^\infty \frac{e^{-t}}{a+t}\,\mathrm{d}t = e^a \Gamma(0,a)$ with $\Gamma(0,a) = \int_a^\infty \frac{e^{-t}}{t}\,\mathrm{d}t$ the upper incomplete gamma function. Now

$$
\begin{aligned}
&2\int_0^\infty \int_0^\infty ((t/a_r) \wedge (s/a_{r'})) \Gamma(0,t)\Gamma(0,s) \,\mathrm{d}t \,\mathrm{d}s \\
&\quad = \int_0^\infty \Gamma(0,t) \left(\frac{1}{a_{r'}} \int_0^{a_{r'}t/a_r} 2s\Gamma(0,s)\,\mathrm{d}s + \frac{t}{a_r} \int_{a_{r'}t/a_r}^\infty 2\Gamma(0,s)\,\mathrm{d}s\right) \mathrm{d}t \\
&\quad = \int_0^\infty \Gamma(0,t) \left(\frac{t}{a_r} e^{-\frac{a_{r'}}{a_r}t} - \frac{1}{a_{r'}} e^{-\frac{a_{r'}}{a_r}t} + \frac{1}{a_{r'}} - \frac{a_{r'}}{a_r^2} t^2 \Gamma\left(0, \frac{a_{r'}}{a_r}t\right)\right) \mathrm{d}t \\
&\quad = \frac{1}{a_r}\int_0^\infty t e^{-\frac{a_{r'}}{a_r}t} \Gamma(0,t)\,\mathrm{d}t - \frac{1}{a_{r'}} \int_0^\infty e^{-\frac{a_{r'}}{a_r}t}\Gamma(0,t)\,\mathrm{d}t \\
&\qquad + \frac{1}{a_{r'}} \int_0^\infty \Gamma(0,t)\,\mathrm{d}t - \frac{a_{r'}}{a_r^2} \int_0^\infty t^2 \Gamma\left(0, \frac{a_{r'}}{a_r}t\right)\Gamma(0,t)\,\mathrm{d}t,
\end{aligned}
$$

since $\int_0^a 2t\Gamma(0,t)\,\mathrm{d}t = a^2\Gamma(0,a) - ae^{-a} - e^{-a} + 1$ and $\int_a^\infty \Gamma(0,t)\,\mathrm{d}t = e^{-a} - a\Gamma(0,a)$. Next, we use $\int_0^\infty \Gamma(0,t)\,\mathrm{d}t = 1$, $\int_0^\infty e^{-at}\Gamma(0,t)\,\mathrm{d}t = \frac{\log(1+a)}{a}$, $\int_0^\infty te^{-at}\Gamma(0,t)\,\mathrm{d}t = \frac{\log(1+a)}{a^2} - \frac{1}{a(a+1)}$ and $\int_0^\infty t^2\Gamma(0,t)\Gamma(0,at)\,\mathrm{d}t = -\frac{2a^2+a+2}{3a^2(a+1)} + \frac{2(a^3+1)\log(a+1)}{3a^3} - \frac{2\log a}{3}$ to see

$$
\begin{aligned}
&2\int_0^\infty \int_0^\infty ((t/a_r) \wedge (s/a_{r'})) \Gamma(0,t)\Gamma(0,s) \,\mathrm{d}t \,\mathrm{d}s \\
&\quad = \frac{a_r \log(1 + a_{r'}/a_r)}{a_{r'}^2} - \frac{a_r/a_{r'}}{a_r + a_{r'}} - \frac{a_r \log(1 + a_{r'}/a_r)}{a_{r'}^2} + \frac{1}{a_{r'}} \\
&\qquad + \frac{2a_{r'}^2 + a_r a_{r'} + 2a_r^2}{3a_r a_{r'}(a_r + a_{r'})} - \frac{2(a_{r'}^3 + a_r^3)\log(a_{r'}/a_r + 1)}{3a_r^2 a_{r'}^2} + \frac{2a_{r'}\log(a_{r'}/a_r)}{3a_r^2} \\
&\quad = \frac{2}{3a_r}\left(1 - \frac{a_{r'}}{a_r}\log\left(\frac{a_r}{a_{r'}} + 1\right)\right) + \frac{2}{3a_{r'}}\left(1 - \frac{a_r}{a_{r'}}\log\left(\frac{a_{r'}}{a_r} + 1\right)\right).
\end{aligned}
$$

Finally we conclude this part by giving the limiting variance.

$$
\begin{aligned}
&\mathbb{E}\left[\tilde{\Sigma}_{\mathrm{d}}(x)\right] \\
&= \frac{\sigma^2(x)}{f(x)} \sum_{r=0}^{J} \sum_{r'=0}^{J} \omega_r \omega_{r'} \left(\frac{2a_{r'}}{3} \left(1 - \frac{a_{r'}}{a_r} \log\left(\frac{a_r}{a_{r'}} + 1 \right) \right) + \frac{2a_r}{3} \left(1 - \frac{a_r}{a_{r'}} \log\left(\frac{a_{r'}}{a_r} + 1 \right) \right) \right)^d \\
&\quad + O\left(\frac{1}{\lambda^{1 \wedge \beta_f \wedge \beta_\sigma}} + \frac{\lambda^d}{n} + \frac{1}{B} \right) \\
&= \Sigma_{\mathrm{d}}(x) + O\left(\frac{1}{\lambda^{1 \wedge \beta_f \wedge \beta_\sigma}} + \frac{\lambda^d}{n} + \frac{1}{B} \right).
\end{aligned}
$$

It follows from this and the previous part that

$$
\mathbb{E}\left[\left(\tilde{\Sigma}_{\mathrm{d}}(x) - \Sigma_{\mathrm{d}}(x) \right)^2 \right] \lesssim \frac{\lambda^d}{n} + \frac{1}{B} + \frac{1}{\lambda^{2(1 \wedge \beta_f \wedge \beta_\sigma)}}.
$$

Part 4: a lower bound for the second moment with a single tree

We finally show here that if $B = 1$ (a forest with a single tree), then $\tilde{\Sigma}_{\mathrm{d}}(x)$ has a divergent second moment. We take $J = 0$ for brevity (no debiasing), and recall that $\sigma^2(x)$ and $f(x)$ are bounded below. Further, since $\lambda T(x)_j \leq \Gamma(2, 1)$, we have by Jensen's inequality that

$$
\begin{aligned}
\mathbb{E}\left[\tilde{\Sigma}_{\mathrm{d}}(x)^2 \right] &= \mathbb{E}\left[\left(\frac{n}{\lambda^d} \sum_{i=1}^n \frac{\mathbb{I}\{X_i \in T(x)\} \sigma^2(X_i)}{N(x)^2} \right)^2 \right] \gtrsim \frac{n^2}{\lambda^{2d}} \mathbb{E}\left[\frac{1}{N(x)^4} \left(\sum_{i=1}^n \mathbb{I}\{X_i \in T(x)\} \right)^2 \right] \\
&= \frac{n^2}{\lambda^{2d}} \sum_{i=1}^n \mathbb{E}\left[\frac{\mathbb{I}\{X_i \in T(x)\}}{N(x)^3} \right] = \frac{n^3}{\lambda^{2d}} \mathbb{E}\left[\frac{\mathbb{I}\{X_i \in T(x)\}}{(N_{-i}(x) + 1)^3} \right] \gtrsim \frac{n^3}{\lambda^{2d}} \mathbb{E}\left[\frac{|T(x)|}{\mathbb{E}\left[(N_{-i}(x) + 1)^3 \mid T \right]} \right] \\
&\gtrsim \frac{n^3}{\lambda^{2d}} \mathbb{E}\left[\frac{|T(x)|}{n^3 |T(x)|^3 + 1} \right] \gtrsim \frac{1}{\lambda^{2d}} \mathbb{E}\left[\frac{\mathbb{I}\{|T(x)| \geq 1/n\}}{|T(x)|^2} \right] \gtrsim \frac{1}{\lambda^{2d}} \mathbb{E}\left[\frac{\mathbb{I}\{|T(x)_j| \geq 1/n\}}{|T(x)_j|^2} \right]^d \\
&\gtrsim \left(\int_{1/n}^1 \frac{s e^{-s}}{s^2} \, \mathrm{d}s \right)^d \gtrsim \left(\int_{1/n}^1 \frac{1}{s} \, \mathrm{d}s \right)^d \gtrsim (\log n)^d.
\end{aligned}
$$

□

Proof (Theorem 4)**.** The bias–variance decomposition with Lemma 4 and (9), along with the proof of Theorem 5, setting $J = \lfloor \beta/2 \rfloor$, gives

$$
\begin{aligned}
\mathbb{E}\left[\left(\hat{\mu}_{\mathrm{d}}(x) - \mu(x) \right)^2 \right] &= \mathbb{E}\left[\left(\hat{\mu}_{\mathrm{d}}(x) - \mathbb{E}\left[\hat{\mu}_{\mathrm{d}}(x) \mid \mathbf{X}, \mathbf{T} \right] \right)^2 \right] + \mathbb{E}\left[\left(\mathbb{E}\left[\hat{\mu}_{\mathrm{d}}(x) \mid \mathbf{X}, \mathbf{T} \right] - \mu(x) \right)^2 \right] \\
&\lesssim \frac{\lambda^d}{n} + \frac{1}{\lambda^{2\beta}} + \frac{1}{\lambda^{2(1 \wedge \beta)} B}.
\end{aligned}
$$

As $\lambda \asymp n^{\frac{1}{d+2\beta}}$ and $B \gtrsim n^{\frac{2\beta - 2(1 \wedge \beta)}{d+2\beta}}$, we have

$$
\mathbb{E}\left[\left(\hat{\mu}_{\mathrm{d}}(x) - \mu(x) \right)^2 \right] \lesssim n^{-\frac{2\beta}{d+2\beta}}.
$$

□

Proof (Theorem 5)**.** Define $S_i(x) = \sqrt{n/\lambda^d} \sum_{r=0}^{J} \omega_r \frac{1}{B} \sum_{b=1}^{B} \frac{\mathbb{I}_{ibr}(x)\varepsilon_i}{N_{br}(x)}$, which are independent and zero mean conditional on $(\mathbf{X}, \mathbf{T})$, and satisfy

$$\sqrt{\frac{n}{\lambda^d}} \big(\hat{\mu}_{\mathrm{d}}(x) - \mathbb{E}\left[\hat{\mu}_{\mathrm{d}}(x) \mid \mathbf{X}, \mathbf{T} \right] \big) = \sum_{i=1}^{n} S_i(x).$$

Therefore by Petrov [1995, Theorem 5.7] conditional on $(\mathbf{X}, \mathbf{T})$, with $\zeta = \delta \wedge 1$,

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P}\left(\tilde{\Sigma}_{\mathrm{d}}(x)^{-1/2} \sum_{i=1}^{n} S_i \leq t \;\Big|\; \mathbf{X}, \mathbf{T} \right) - \Phi(t) \right| \lesssim 1 \wedge \left(\tilde{\Sigma}_{\mathrm{d}}(x)^{-1-\zeta/2} \sum_{i=1}^{n} \mathbb{E}\left[|S_i|^{2+\zeta} \mid \mathbf{X}, \mathbf{T} \right] \right).$$

It immediately follows that

$$\begin{aligned}
\sup_{t \in \mathbb{R}} \left| \mathbb{P}\left(\tilde{\Sigma}_{\mathrm{d}}(x)^{-1/2} \sum_{i=1}^{n} S_i \leq t \right) - \Phi(t) \right| &= \sup_{t \in \mathbb{R}} \left| \mathbb{E}\left[\mathbb{P}\left(\tilde{\Sigma}_{\mathrm{d}}(x)^{-1/2} \sum_{i=1}^{n} S_i \leq t \;\Big|\; \mathbf{X}, \mathbf{T} \right) \right] - \Phi(t) \right| \\
&\leq \mathbb{E}\left[\sup_{t \in \mathbb{R}} \left| \mathbb{P}\left(\tilde{\Sigma}_{\mathrm{d}}(x)^{-1/2} \sum_{i=1}^{n} S_i \leq t \;\Big|\; \mathbf{X}, \mathbf{T} \right) - \Phi(t) \right| \right] \\
&\lesssim \mathbb{E}\left[1 \wedge \left(\tilde{\Sigma}_{\mathrm{d}}(x)^{-1-\zeta/2} \sum_{i=1}^{n} \mathbb{E}\left[|S_i|^{2+\zeta} \mid \mathbf{X}, \mathbf{T} \right] \right) \right].
\end{aligned}$$

To bound this quantity, we first partition by the event that $\tilde{\Sigma}_{\mathrm{d}}(x)$ is bounded away from zero:

$$\begin{aligned}
\mathbb{E}\left[1 \wedge \left(\tilde{\Sigma}_{\mathrm{d}}(x)^{-1-\zeta/2} \sum_{i=1}^{n} \mathbb{E}\left[|S_i|^{2+\zeta} \mid \mathbf{X}, \mathbf{T} \right] \right) \right] &\lesssim \mathbb{E}\big[\tilde{\Sigma}_{\mathrm{d}}(x)\big]^{-1-\zeta/2} \sum_{i=1}^{n} \mathbb{E}\left[|S_i|^{2+\zeta} \right] \\
&\quad + \mathbb{P}\left(\left| \tilde{\Sigma}_{\mathrm{d}}(x) - \mathbb{E}\big[\tilde{\Sigma}_{\mathrm{d}}(x)\big] \right| > \frac{\mathbb{E}\big[\tilde{\Sigma}_{\mathrm{d}}(x)\big]}{2} \right). \qquad \text{(SA9)}
\end{aligned}$$

The first term in (SA9) is bounded as follows. We already have from the proof of Lemma 5 that eventually $\mathbb{E}[\tilde{\Sigma}_{\mathrm{d}}(x)] \geq \Sigma_{\mathrm{d}}(x)/2 \gtrsim 1$. Since $\mathbb{E}[\varepsilon_i^{2+\delta} \mid \mathbf{X}]$ is bounded, by Jensen's inequality,

$$\begin{aligned}
\sum_{i=1}^{n} \mathbb{E}\left[|S_i|^{2+\zeta} \right] &= \sum_{i=1}^{n} \mathbb{E}\left[\left| \sqrt{\frac{n}{\lambda^d}} \sum_{r=0}^{J} \omega_r \frac{1}{B} \sum_{b=1}^{B} \frac{\mathbb{I}_{ibr}(x)\varepsilon_i}{N_{br}(x)} \right|^{2+\zeta} \right] \\
&\leq \left(\frac{n}{\lambda^d} \right)^{1+\zeta/2} (J+1)^{1+\zeta} \sum_{r=0}^{J} |\omega_r|^3 \, \mathbb{E}\left[\sum_{i=1}^{n} \left(\frac{1}{B} \sum_{b=1}^{B} \frac{\mathbb{I}_{ibr}(x)}{N_{br}(x)} \right)^{2+\zeta} \right] \\
&\lesssim \left(\frac{n}{\lambda^d} \right)^{1+\zeta/2} \mathbb{E}\left[\sum_{i=1}^{n} \left(\frac{1}{B} \sum_{b=1}^{B} \frac{\mathbb{I}_{ibr}(x)}{N_{br}(x)} \right)^{2+\zeta} \right].
\end{aligned}$$

We now proceed by cases. If $\zeta = 1$, note that by Lemma SA6 and with $B \gtrsim (\log n)^d$,

$$\mathbb{E}\left[\sum_{i=1}^n \left(\frac{1}{B}\sum_{b=1}^B \frac{\mathbb{I}_{ibr}(x)}{N_{br}(x)}\right)^3\right] = \mathbb{E}\left[\sum_{i=1}^n \frac{1}{B^3}\sum_{b=1}^B\sum_{b'=1}^B\sum_{b''=1}^B \frac{\mathbb{I}_{ibr}(x)}{N_{br}(x)}\frac{\mathbb{I}_{ib'r'}(x)}{N_{b'r'}(x)}\frac{\mathbb{I}_{ib''r''}(x)}{N_{b''r''}(x)}\right]$$
$$\le \frac{1}{B^2}\sum_{b=1}^B\sum_{b'=1}^B \mathbb{E}\left[\frac{\mathbb{I}_{br}(x)}{N_{br}(x)}\frac{\mathbb{I}_{b'r'}(x)}{N_{b'r'}(x)}\right] \lesssim \frac{\lambda^{2d}}{n^2} + \frac{1}{B}\frac{\lambda^{2d}(\log n)^d}{n^2} \lesssim \frac{\lambda^{2d}}{n^2}.$$

Alternatively, if $\zeta \in (0,1)$, by Jensen's inequality and Lemma SA6,

$$\mathbb{E}\left[\sum_{i=1}^n \left(\frac{1}{B}\sum_{b=1}^B \frac{\mathbb{I}_{ibr}(x)}{N_{br}(x)}\right)^{2+\zeta}\right] \le \mathbb{E}\left[\sum_{i=1}^n \left(\frac{\mathbb{I}_{ibr}(x)}{N_{br}(x)}\right)^{2+\zeta}\right] \le \mathbb{E}\left[\sum_{i=1}^n \frac{\mathbb{I}_{ibr}(x)}{N_{br}(x)}\left(\frac{\mathbb{I}_{br}(x)}{N_{br}(x)}\right)^{1+\zeta}\right]$$
$$\le \mathbb{E}\left[\left(\frac{\mathbb{I}_{br}(x)}{N_{br}(x)}\right)^{1+\zeta}\right] \lesssim \left(\frac{\lambda^d}{n}\right)^{1+\zeta}.$$

Both cases lead to the conclusion that

$$\mathbb{E}[\tilde{\Sigma}_{\mathrm{d}}(x)]^{-1-\zeta/2}\sum_{i=1}^n \mathbb{E}\left[|S_i|^{2+\zeta}\right] \lesssim \left(\frac{n}{\lambda^d}\right)^{1+\zeta/2}\left(\frac{\lambda^d}{n}\right)^{1+\zeta} = \left(\frac{\lambda^d}{n}\right)^{\zeta/2}.$$

For the second term in (SA9), Chebyshev's inequality along with the proof of Lemma 5 give

$$\mathbb{P}\left(\left|\tilde{\Sigma}_{\mathrm{d}}(x) - \mathbb{E}[\tilde{\Sigma}_{\mathrm{d}}(x)]\right| > \frac{\mathbb{E}[\tilde{\Sigma}_{\mathrm{d}}(x)]}{2}\right) \lesssim \mathrm{Var}\left[\tilde{\Sigma}_{\mathrm{d}}(x)\right] \lesssim \frac{1}{B} + \frac{\lambda^d}{n}.$$

Therefore

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(\tilde{\Sigma}_{\mathrm{d}}(x)^{-1/2}\sum_{i=1}^n S_i \le t\right) - \Phi(t)\right| \lesssim \left(\frac{\lambda^d}{n}\right)^{\frac{1\wedge\delta}{2}} + \frac{1}{B}.$$

□

Proof (Lemma 6)**.** We begin by showing that $\hat{\sigma}^2(x)$ is consistent for $\sigma^2(x)$.

Part 1: consistency of $\hat{\sigma}^2(x)$

Recall that

$$\hat{\sigma}^2(x) = \frac{1}{B}\sum_{b=1}^B\sum_{i=1}^n \frac{Y_i^2\,\mathbb{I}_{ib}(x)}{N_b(x)} - \hat{\mu}(x)^2. \tag{SA10}$$

The first term in (SA10) is simply a Mondrian forest estimator of $\mathbb{E}[Y_i^2 \mid X_i = x] = \sigma^2(x) + \mu(x)^2$, which is bounded and in $\mathcal{H}^{\beta_\mu \wedge \beta_\sigma}$. Therefore, by Lemma 1, its conditional bias is

$$\mathbb{E}\left[\left(\frac{1}{B}\sum_{b=1}^{B}\sum_{i=1}^{n}\frac{\left(\sigma^2(X_i)+\mu(X_i)^2-\sigma^2(x)-\mu(x)^2\right)\mathbb{I}_{ib}(x)}{N_b(x)}\right)^2\right]$$
$$\lesssim \frac{1}{\lambda^{2(2\wedge\beta\wedge\beta_\sigma)}} + \frac{1}{\lambda^{2(1\wedge\beta\wedge\beta_\sigma)}B} + \frac{1}{\lambda^{2(1\wedge\beta\wedge\beta_\sigma)}}\frac{\lambda^d}{n}.$$

We handle the stochastic part with a truncation argument. Let $S_i = Y_i^2 - \sigma^2(X_i) - \mu(X_i)^2$ and $\tilde{S}_i = S_i\,\mathbb{I}\{|S_i| \le M\} - \mathbb{E}\left[S_i\,\mathbb{I}\{|S_i| \le M\} \mid X_i\right]$ where $M > 0$ is to be determined. We bound

$$\frac{1}{B}\sum_{b=1}^{B}\sum_{i=1}^{n}\frac{\tilde{S}_i\,\mathbb{I}_{ib}(x)}{N_b(x)} + \frac{1}{B}\sum_{b=1}^{B}\sum_{i=1}^{n}\frac{(S_i-\tilde{S}_i)\,\mathbb{I}_{ib}(x)}{N_b(x)}. \tag{SA11}$$

The first term in (SA11) is controlled with a variance bound, noting $\tilde{S}_i \le 2M$ almost surely.

$$\mathrm{Var}\left[\frac{1}{B}\sum_{b=1}^{B}\sum_{i=1}^{n}\frac{\tilde{S}_i\,\mathbb{I}_{ib}(x)}{N_b(x)}\right] \le \sum_{i=1}^{n}\mathbb{E}\left[\frac{\tilde{S}_i^2\,\mathbb{I}_{ib}(x)}{N_b(x)^2}\right] \le 4M^2\sum_{i=1}^{n}\mathbb{E}\left[\frac{\mathbb{I}_b(x)}{N_b(x)}\right] \lesssim M^2\frac{\lambda^d}{n}.$$

For the second term in (SA11), note that $S_i - \tilde{S}_i = S_i\,\mathbb{I}\{|S_i| > M\} - \mathbb{E}\left[S_i\,\mathbb{I}\{|S_i| > M\} \mid X_i\right]$ because $\mathbb{E}[S_i \mid X_i] = 0$. Since $\mathbb{E}\left[|Y_i|^{2+\delta} \mid X_i\right]$ is bounded, so is $\mathbb{E}\left[|S_i|^{1+\delta/2} \mid X_i\right]$. Thus

$$\begin{aligned}
\mathbb{E}\left[\left|\frac{1}{B}\sum_{b=1}^{B}\sum_{i=1}^{n}\frac{(S_i-\tilde{S}_i)\,\mathbb{I}_{ib}(x)}{N_b(x)}\right|\right] &\le \sum_{i=1}^{n}\mathbb{E}\left[\frac{\mathbb{I}_{ib}(x)}{N_b(x)}\mathbb{E}\left[|S_i-\tilde{S}_i| \mid X_i\right]\right] \\
&\le 2\sum_{i=1}^{n}\mathbb{E}\left[\frac{\mathbb{I}_{ib}(x)}{N_b(x)}\mathbb{E}\left[|S_i|\,\mathbb{I}\{|S_i| > M\} \mid X_i\right]\right] \\
&\le \frac{2}{M^{\delta/2}}\sum_{i=1}^{n}\mathbb{E}\left[\frac{\mathbb{I}_{ib}(x)}{N_b(x)}\mathbb{E}\left[|S_i|^{1+\delta/2} \mid X_i\right]\right] \lesssim \frac{1}{M^{\delta/2}}.
\end{aligned}$$

Consistency of the second term in (SA10) follows directly from Lemma 1 and Theorem 5 with the same bias and variance bounds. Therefore

$$\begin{aligned}
\mathbb{E}\left[\left|\hat{\sigma}^2(x)-\sigma^2(x)\right|\right] &\lesssim \frac{1}{\lambda^{2\wedge\beta\wedge\beta_\sigma}} + \frac{1}{\lambda^{1\wedge\beta\wedge\beta_\sigma}\sqrt{B}} + \frac{1}{\lambda^{1\wedge\beta\wedge\beta_\sigma}}\sqrt{\frac{\lambda^d}{n}} + M\sqrt{\frac{\lambda^d}{n}} + \frac{1}{M^{\delta/2}} \\
&\lesssim \frac{1}{\lambda^{2\wedge\beta\wedge\beta_\sigma}} + \frac{1}{\lambda^{1\wedge\beta\wedge\beta_\sigma}\sqrt{B}} + \left(\frac{\lambda^d}{n}\right)^{\frac{\delta}{4+2\delta}},
\end{aligned}$$

where we set $M = \left(\frac{\lambda^d}{n}\right)^{-\frac{1}{2+\delta}}$. Note that if $\delta \ge 2$ then the variance argument applies directly, without the need for truncation, yielding

$$\mathbb{E}\left[\left|\hat{\sigma}^2(x)-\sigma^2(x)\right|^2\right] \lesssim \frac{1}{\lambda^{2(2\wedge\beta\wedge\beta_\sigma)}} + \frac{1}{\lambda^{2(1\wedge\beta\wedge\beta_\sigma)}B} + \frac{\lambda^d}{n}.$$

Part 2: consistency of the sum

Note that

$$\frac{n}{\lambda^d}\sum_{i=1}^{n}\left(\sum_{r=0}^{J}\omega_r\frac{1}{B}\sum_{b=1}^{B}\frac{\mathbb{I}\{X_i\in T_{br}(x)\}}{\sum_{i=1}^{n}\mathbb{I}\{X_i\in T_{br}(x)\}}\right)^2=\frac{n}{\lambda^d}\frac{1}{B^2}\sum_{i=1}^{n}\sum_{r=0}^{J}\sum_{r'=0}^{J}\omega_r\omega_{r'}\sum_{b=1}^{B}\sum_{b'=1}^{B}\frac{\mathbb{I}_{ibr}(x)\mathbb{I}_{ib'r'}(x)}{N_{br}(x)N_{b'r'}(x)}.$$

This is exactly the same as $\tilde{\Sigma}_{\mathrm{d}}(x)$, if we were to take $\sigma^2(x)=1$. Thus by Lemma 5, we obtain

$$\mathbb{E}\left[\left(\frac{n}{\lambda^d}\frac{1}{B^2}\sum_{i=1}^{n}\sum_{r=0}^{J}\sum_{r'=0}^{J}\omega_r\omega_{r'}\sum_{b=1}^{B}\sum_{b'=1}^{B}\frac{\mathbb{I}_{ibr}(x)\mathbb{I}_{ib'r'}(x)}{N_{br}(x)N_{b'r'}(x)}-\frac{\Sigma_{\mathrm{d}}(x)}{\sigma^2(x)}\right)^2\right]\lesssim\frac{\lambda^d}{n}+\frac{1}{B}+\frac{1}{\lambda^{2(1\wedge\beta_f\wedge\beta_\sigma)}}.$$

Part 3: conclusion

By the previous parts and the Cauchy–Schwarz inequality,

$$\mathbb{E}\left[\left|\hat{\Sigma}_{\mathrm{d}}(x)-\Sigma_{\mathrm{d}}(x)\right|^{1/2}\right]^2\lesssim\left(\frac{\lambda^d}{n}\right)^{\frac{\delta}{4+2\delta}}+\frac{1}{\sqrt{B}}+\frac{1}{\lambda^{1\wedge\beta_\mu\wedge\beta_f\wedge\beta_\sigma}}.$$

If $\delta\geq 2$ then we obtain

$$\mathbb{E}\left[\left|\hat{\Sigma}_{\mathrm{d}}(x)-\Sigma_{\mathrm{d}}(x)\right|\right]\lesssim\sqrt{\frac{\lambda^d}{n}}+\frac{1}{\sqrt{B}}+\frac{1}{\lambda^{1\wedge\beta_\mu\wedge\beta_f\wedge\beta_\sigma}}.$$

Combining these yields

$$\mathbb{E}\left[\left|\hat{\Sigma}_{\mathrm{d}}(x)-\Sigma_{\mathrm{d}}(x)\right|^{\frac{2-\mathbb{I}\{\delta<2\}}{2}}\right]^{\frac{2}{2-\mathbb{I}\{\delta<2\}}}\lesssim\left(\frac{\lambda^d}{n}\right)^{\frac{1}{2}-\frac{\mathbb{I}\{\delta<2\}}{2+\delta}}+\frac{1}{\sqrt{B}}+\frac{1}{\lambda^{1\wedge\beta_\mu\wedge\beta_f\wedge\beta_\sigma}}.$$

□

Proof (Theorem 6)**.** Let τ and $\hat{\tau}$ be real-valued random variables. Then for any $\varepsilon>0$,

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}(\hat{\tau}\leq t)-\Phi(t)\right|\leq\sup_{t\in\mathbb{R}}\left|\mathbb{P}(\tau\leq t)-\Phi(t)\right|+\varepsilon\sqrt{2/\pi}+\mathbb{P}(|\hat{\tau}-\tau|>\varepsilon).$$

Defining $a/0=0$ for all $a\in\mathbb{R}$ to accommodate the event $\hat{\Sigma}_{\mathrm{d}}(x)=0$, we apply this result to

$$\hat{\tau}=\sqrt{\frac{n}{\lambda^d}}\left(\frac{\hat{\mu}_{\mathrm{d}}(x)-\mu(x)}{\sqrt{\hat{\Sigma}_{\mathrm{d}}(x)}}-\frac{\mathbb{E}\left[\hat{\mu}_{\mathrm{d}}(x)\right]-\mu(x)}{\sqrt{\Sigma_{\mathrm{d}}(x)}}\right)\qquad\text{and}\qquad\tau=\sqrt{\frac{n}{\lambda^d}}\frac{\hat{\mu}_{\mathrm{d}}(x)-\mathbb{E}\left[\hat{\mu}_{\mathrm{d}}(x)\mid\mathbf{X},\mathbf{T}\right]}{\sqrt{\Sigma_{\mathrm{d}}(x)}}$$

respectively, noting that $\sup_{t\in\mathbb{R}}\left|\mathbb{P}(\tau\leq t)-\Phi(t)\right|\lesssim\left(\frac{\lambda^d}{n}\right)^{\frac{1\wedge\delta}{2}}+\frac{1}{B}$ by Theorem 5. With

$$v=\sqrt{\frac{n}{\lambda^d}}\frac{\mathbb{E}\left[\hat{\mu}_{\mathrm{d}}(x)\right]-\mu(x)}{\sqrt{\Sigma_{\mathrm{d}}(x)}}\lesssim\sqrt{\frac{\lambda^d}{n}}\frac{1}{\lambda^{1\wedge\beta_\mu}}+\sqrt{\frac{n}{\lambda^d}}\frac{1}{\lambda^{\beta}}$$

by the proof of Lemma 1, and by Taylor's theorem, for some $s, s' \in \mathbb{R}$, we have

$$
\begin{aligned}
&\left|\mathbb{P}(\mu(x) \in \mathrm{CI}_{\mathrm{d}}(x)) - (1-\alpha)\right| \\
&\quad= \left|\mathbb{P}\left(q_{\alpha/2} \leq \sqrt{\frac{n}{\lambda^d}} \frac{\hat{\mu}_{\mathrm{d}}(x) - \mu(x)}{\sqrt{\hat{\Sigma}_{\mathrm{d}}(x)}} \leq q_{1-\alpha/2}\right) - (1-\alpha)\right| \\
&\quad= \left|\mathbb{P}\left(q_{\alpha/2} - v \leq \hat{\tau} \leq q_{1-\alpha/2} - v\right) - (1-\alpha)\right| \\
&\quad\leq \left|\mathbb{P}\left(\hat{\tau} \leq q_{1-\alpha/2} - v\right) - \Phi(q_{1-\alpha/2} - v)\right| + \left|\mathbb{P}\left(\hat{\tau} < q_{\alpha/2} - v\right) - \Phi(q_{\alpha/2} - v)\right| \\
&\qquad+ \left|\Phi(q_{1-\alpha/2} - v) - (1-\alpha/2) - \Phi(q_{\alpha/2} - v) + \alpha/2\right| \\
&\quad\lesssim \left(\frac{\lambda^d}{n}\right)^{\frac{1 \wedge \delta}{2}} + \frac{1}{B} + \varepsilon + \mathbb{P}\left(|\hat{\tau} - \tau| > \varepsilon\right) + \left|-v\phi(1-\alpha/2) + v^2\phi'(s)/2 + v\phi(\alpha/2) - v^2\phi'(s')/2\right| \\
&\quad\lesssim \left(\frac{\lambda^d}{n}\right)^{\frac{1 \wedge \delta}{2}} + \frac{1}{B} + \varepsilon + \frac{n}{\lambda^d}\frac{1}{\lambda^{2\beta}} + \mathbb{P}\left(|\hat{\tau} - \tau| > \varepsilon\right).
\end{aligned}
$$

It remains to bound $\mathbb{P}\left(|\hat{\tau} - \tau| > \varepsilon\right)$. Observe that

$$
\begin{aligned}
|\hat{\tau} - \tau| &\leq R_1 + R_2 + R_3, \\
R_1 &= \sqrt{\frac{n}{\lambda^d}} |\hat{\mu}_{\mathrm{d}}(x) - \mu(x)| \left|\frac{1}{\sqrt{\hat{\Sigma}_{\mathrm{d}}(x)}} - \frac{1}{\sqrt{\tilde{\Sigma}_{\mathrm{d}}(x)}}\right|, \\
R_2 &= \sqrt{\frac{n}{\lambda^d}} \left|\mathbb{E}\left[\hat{\mu}_{\mathrm{d}}(x) \mid \mathbf{X}, \mathbf{T}\right] - \mu(x)\right| \left|\frac{1}{\sqrt{\tilde{\Sigma}_{\mathrm{d}}(x)}} - \frac{1}{\sqrt{\Sigma_{\mathrm{d}}(x)}}\right|, \\
R_3 &= \sqrt{\frac{n}{\lambda^d}} \frac{\left|\mathbb{E}\left[\hat{\mu}_{\mathrm{d}}(x) \mid \mathbf{X}, \mathbf{T}\right] - \mathbb{E}\left[\hat{\mu}_{\mathrm{d}}(x)\right]\right|}{\sqrt{\Sigma_{\mathrm{d}}(x)}}.
\end{aligned}
$$

We begin with R_1. Take $a > 1$ and $b^{2/3} = \Sigma_{\mathrm{d}}(x)/2$, so by the proof of Lemma 5,

$$
\begin{aligned}
\mathbb{P}\left(R_1 > \varepsilon\right) &\leq \mathbb{P}\left(\sqrt{\frac{n}{\lambda^d}} |\hat{\mu}_{\mathrm{d}}(x) - \mu(x)| > a\varepsilon\right) + \mathbb{P}\left(\left|\frac{1}{\sqrt{\hat{\Sigma}_{\mathrm{d}}(x)}} - \frac{1}{\sqrt{\tilde{\Sigma}_{\mathrm{d}}(x)}}\right| > \frac{1}{a}\right) \\
&\leq \frac{n}{a^2\varepsilon^2\lambda^d} \mathbb{E}\left[(\hat{\mu}_{\mathrm{d}}(x) - \mu(x))^2\right] + \mathbb{P}\left(\tilde{\Sigma}_{\mathrm{d}}(x) < b^{2/3}\right) + \mathbb{P}\left(\hat{\Sigma}_{\mathrm{d}}(x) < b^{2/3}\right) \\
&\quad+ \mathbb{P}\left(\left|\hat{\Sigma}_{\mathrm{d}}(x) - \tilde{\Sigma}_{\mathrm{d}}(x)\right| > \frac{b}{a}\right) \\
&\lesssim \frac{n}{a^2\varepsilon^2\lambda^d}\left(\frac{\lambda^d}{n} + \frac{1}{\lambda^{2\beta}} + \frac{1}{\lambda^{2(1\wedge\beta)}B}\right) + \mathrm{Var}\left[\tilde{\Sigma}_{\mathrm{d}}(x)\right] + \mathbb{P}\left(|\hat{\Sigma}_{\mathrm{d}}(x) - \Sigma_{\mathrm{d}}(x)| > \Sigma_{\mathrm{d}}(x)/2\right) \\
&\quad+ \mathbb{P}\left(\left|\tilde{\Sigma}_{\mathrm{d}}(x) - \Sigma_{\mathrm{d}}(x)\right| > \frac{b}{2a}\right) + \mathbb{P}\left(\left|\hat{\Sigma}_{\mathrm{d}}(x) - \Sigma_{\mathrm{d}}(x)\right| > \frac{b}{2a}\right) \\
&\lesssim \frac{1}{a^2\varepsilon^2} + \frac{1}{B} + \frac{\lambda^d}{n} + a^2\left(\frac{\lambda^d}{n} + \frac{1}{B} + \frac{1}{\lambda^{2(1\wedge\beta_f\wedge\beta_\sigma)}}\right) + \mathbb{P}\left(\left|\hat{\Sigma}_{\mathrm{d}}(x) - \Sigma_{\mathrm{d}}(x)\right| > \frac{b}{2a}\right).
\end{aligned}
$$

If $\delta < 2$ then the proof of Lemma 6 along with Markov's inequality gives

$$\mathbb{P}\left(\left|\hat{\Sigma}_{\mathrm{d}}(x) - \Sigma_{\mathrm{d}}(x)\right| > \frac{b}{2a}\right) \lesssim \sqrt{a}\,\mathbb{E}\left[\left|\hat{\Sigma}_{\mathrm{d}}(x) - \Sigma_{\mathrm{d}}(x)\right|^{1/2}\right]$$
$$\lesssim \sqrt{a}\left(\left(\frac{\lambda^d}{n}\right)^{\frac{\delta}{8+4\delta}} + \frac{1}{B^{1/4}} + \frac{1}{\lambda^{(1\wedge\beta_\mu\wedge\beta_f\wedge\beta_\sigma)/2}}\right),$$

and whenever this converges to zero, minimizing over a yields

$$\mathbb{P}(R_1 > \varepsilon) \lesssim \frac{1}{a^2\varepsilon^2} + \frac{1}{B} + \frac{\lambda^d}{n} + \sqrt{a}\left(\left(\frac{\lambda^d}{n}\right)^{\frac{\delta}{8+4\delta}} + \frac{1}{B^{1/4}} + \frac{1}{\lambda^{(1\wedge\beta_\mu\wedge\beta_f\wedge\beta_\sigma)/2}}\right)$$
$$\lesssim \frac{1}{\varepsilon^{2/5}}\left(\left(\frac{\lambda^d}{n}\right)^{\frac{\delta}{10+5\delta}} + \frac{1}{B^{1/5}} + \frac{1}{\lambda^{2(1\wedge\beta_\mu\wedge\beta_f\wedge\beta_\sigma)/5}}\right).$$

If $\delta \geq 2$ however then we instead obtain

$$\mathbb{P}\left(\left|\hat{\Sigma}_{\mathrm{d}}(x) - \Sigma_{\mathrm{d}}(x)\right| > \frac{b}{2a}\right) \lesssim a\,\mathbb{E}\left[\left|\hat{\Sigma}_{\mathrm{d}}(x) - \Sigma_{\mathrm{d}}(x)\right|\right] \lesssim a\left(\sqrt{\frac{\lambda^d}{n}} + \frac{1}{\sqrt{B}} + \frac{1}{\lambda^{1\wedge\beta_\mu\wedge\beta_f\wedge\beta_\sigma}}\right),$$

and again if this converges to zero then minimizing over a gives

$$\mathbb{P}(R_1 > \varepsilon) \lesssim \frac{1}{a^2\varepsilon^2} + \frac{1}{B} + \frac{\lambda^d}{n} + a\left(\sqrt{\frac{\lambda^d}{n}} + \frac{1}{\sqrt{B}} + \frac{1}{\lambda^{1\wedge\beta_\mu\wedge\beta_f\wedge\beta_\sigma}}\right)$$
$$\lesssim \frac{1}{\varepsilon^{2/3}}\left(\left(\frac{\lambda^d}{n}\right)^{1/3} + \frac{1}{B^{1/3}} + \frac{1}{\lambda^{2(1\wedge\beta_\mu\wedge\beta_f\wedge\beta_\sigma)/3}}\right).$$

For R_2, note that the same arguments used for R_1 apply again, yielding a bound no worse than that for R_1. Finally, for R_3, we have by Lemma 4 that

$$\mathbb{P}(R_3 > \varepsilon) \leq \frac{1}{\varepsilon^2}\mathbb{E}\left[R_3^2\right] \lesssim \frac{1}{\varepsilon^2}\frac{n}{\lambda^d}\mathbb{E}\left[\left(\mathbb{E}\left[\hat{\mu}_{\mathrm{d}}(x) \mid \mathbf{X}, \mathbf{T}\right] - \mathbb{E}\left[\hat{\mu}_{\mathrm{d}}(x)\right]\right)^2\right]$$
$$\lesssim \frac{1}{\varepsilon^2}\frac{n}{\lambda^d}\left(\frac{1}{\lambda^{2(1\wedge\beta_\mu)}B} + \frac{1}{\lambda^{2(1\wedge\beta_\mu)}}\frac{\lambda^d}{n}\right) \lesssim \frac{1}{\varepsilon^2}\left(\frac{n}{\lambda^d}\frac{1}{\lambda^{2(1\wedge\beta_\mu)}B} + \frac{1}{\lambda^{2(1\wedge\beta_\mu)}}\right).$$

So far we have shown that if $\varepsilon \to 0$, then for $\delta < 2$,

$$\left|\mathbb{P}(\mu(x) \in \mathrm{CI}_{\mathrm{d}}(x)) - (1-\alpha)\right| \lesssim \left(\frac{\lambda^d}{n}\right)^{\frac{1\wedge\delta}{2}} + \frac{1}{B} + \varepsilon + \frac{n}{\lambda^d}\frac{1}{\lambda^{2\beta}} + \mathbb{P}(|\hat{\tau} - \tau| > \varepsilon)$$
$$\lesssim \varepsilon + \frac{n}{\lambda^d}\frac{1}{\lambda^{2\beta}} + \frac{1}{\varepsilon^{2/5}}\left(\left(\frac{\lambda^d}{n}\right)^{\frac{\delta}{10+5\delta}} + \frac{1}{B^{1/5}} + \frac{1}{\lambda^{2(1\wedge\beta_\mu\wedge\beta_f\wedge\beta_\sigma)/5}} + \left(\frac{n}{\lambda^d}\right)^{1/5}\frac{1}{\lambda^{2(1\wedge\beta_\mu)/5}B^{1/5}}\right)$$
$$\lesssim \frac{n}{\lambda^d}\frac{1}{\lambda^{2\beta}} + \left(\left(\frac{\lambda^d}{n}\right)^{\frac{\delta}{2+\delta}} + \frac{1}{B} + \frac{1}{\lambda^{2(1\wedge\beta_\mu\wedge\beta_f\wedge\beta_\sigma)}} + \frac{n}{\lambda^d}\frac{1}{\lambda^{2(1\wedge\beta)}B}\right)^{1/7},$$

while for $\delta \geq 2$,

$$
\begin{aligned}
|\mathbb{P}(\mu(x) \in \mathrm{CI}_{\mathrm{d}}(x)) - (1-\alpha)| &\lesssim \left(\frac{\lambda^d}{n}\right)^{\frac{1 \wedge \delta}{2}} + \frac{1}{B} + \varepsilon + \frac{n}{\lambda^d} \frac{1}{\lambda^{2\beta}} + \mathbb{P}(|\hat{\tau} - \tau| > \varepsilon) \\
&\lesssim \varepsilon + \frac{n}{\lambda^d} \frac{1}{\lambda^{2\beta}} + \frac{1}{\varepsilon^{2/3}} \left(\left(\frac{\lambda^d}{n}\right)^{1/3} + \frac{1}{B^{1/3}} + \frac{1}{\lambda^{2(1 \wedge \beta_\mu \wedge \beta_f \wedge \beta_\sigma)/3}} + \left(\frac{n}{\lambda^d}\right)^{1/3} \frac{1}{\lambda^{2(1 \wedge \beta_\mu)/3} B^{1/3}} \right) \\
&\lesssim \frac{n}{\lambda^d} \frac{1}{\lambda^{2\beta}} + \left(\frac{\lambda^d}{n} + \frac{1}{B} + \frac{1}{\lambda^{2(1 \wedge \beta_\mu \wedge \beta_f \wedge \beta_\sigma)}} + \frac{n}{\lambda^d} \frac{1}{\lambda^{2(1 \wedge \beta_\mu)} B} \right)^{1/5},
\end{aligned}
$$

where in both displays we minimized over $\varepsilon > 0$. □

SA1.5 Proofs for Section 6

Proof (Lemma 7)**.** All computational complexities in this proof are understood to be upper bounds up to constants. The first step is to select λ using polynomial fitting as in Section 5.2. Constructing the design matrix $\mathbf{P}$ requires raising a number to a power of at most $J+1$ a total of $nd(J+1)$ times, giving a complexity of $nd(J+1)^2$. Multiplying the design matrix to obtain $\mathbf{P}^\top \mathbf{P}$ is $nd^2(J+1)^2$, and inverting this is $d^3(J+1)^3 \lesssim nd^2(J+1)^2$, giving an overall complexity of $nd^2(J+1)^2$ for selecting the lifetime.

Calculating the debiasing coefficients ω_r as in Section 4 involves inverting a $(J+1) \times (J+1)$ matrix, so is $(J+1)^3 \lesssim n(J+1)^2$. Next, constructing $U(x)$ as in (12) requires $Bd(J+1)$ comparisons, and forming $I(x)$ then needs nd comparisons.

Once $I(x)$ is available, Calculating $N_{br}(x)$, $S_{br}(x)$ and $V_{br}(x)$ as in (13) each take $Bd(J+1)|I(x)|$ operations, and from these we compute $\hat{\mu}(x)$ and $\hat{\sigma}^2(x)$ in $(J+1)B$ using (14). Constructing $\hat{\Sigma}_{\mathrm{d}}(x)$ as in (15) is $Bd(J+1)|I(x)|$, and calculating $\mathrm{CI}_{\mathrm{d}}(x)$ with (11) has complexity 1.

Thus the overall complexity of Algorithm 1 is $nd^2(J+1)^2 + Bd(J+1) + Bd(J+1)|I(x)|$. To obtain the average case behavior we present a bound for $\mathbb{E}[|I(x)|]$. Firstly, since $f(x)$ is bounded and by the distribution of Mondrian cells,

$$
\mathbb{E}[|I(x)|] = \mathbb{E}\left[\sum_{i=1}^n \mathbb{I}\{X_i \in U(x)\}\right] = \sum_{i=1}^n \mathbb{E}[\mathbb{P}(X_i \in U(x) \mid U(x))] \lesssim n\,\mathbb{E}[|U(x)|] \lesssim n\,\mathbb{E}[|U(x)_j|]^d .
$$

Next, by Lemma SA1, we have that

$$
\mathbb{P}\left(|U(x)_j| \geq \frac{4t + 4\log(2B(J+1))}{\lambda}\right) \leq \mathbb{P}\left(\max_{0 \leq r \leq J} \max_{1 \leq b \leq B} |T_b(x)_j| \geq \frac{2t + 2\log(2B(J+1))}{\lambda}\right) \leq e^{-t},
$$

and integrating the tail probability yields

$$
\mathbb{E}[|U(x)_j|] \lesssim \frac{\log(2B(J+1))}{\lambda}, \qquad \text{so that} \qquad \mathbb{E}[|I(x)|] \lesssim \frac{n \log(2B(J+1))^d}{\lambda^d}.
$$

□

Proof (Lemma 8)**.** As in the proof of Lemma 7, the complexity of selecting the lifetime is $(n+k)d^2(J+1)^2$. Since this occurs with probability at most k/K, and $k \leq n$, the average case time complexity is $\frac{knd^2(J+1)^2}{K}$.

To update the trees, we sample and perform comparisons with at most $B^* d(J+1) \lesssim Bd(J+1)$ exponential random variables. We verify here that the resulting trees have the correct distribution, since by Mourtada et al. [2020, Proposition 1] and the memoryless property of the exponential distribution, with E'_{brj1} and E'_{brj1} i.i.d. copies of E_{brj1},

$$\begin{aligned}
T^*_{br}(x)^-_j &= T_{br}(x)^-_j \vee \left(x_j - \frac{E_{brj1}}{\lambda^* - \lambda}\right) = 0 \vee \left(x_j - \frac{E'_{brj1}}{\lambda}\right) \vee \left(x_j - \frac{E_{brj1}}{\lambda^* - \lambda}\right) \\
&= x_j - \left(x_j \wedge \frac{E'_{brj1}}{\lambda} \wedge \frac{E_{brj1}}{\lambda^* - \lambda}\right) = x_j - \left(x_j \wedge \frac{E''_{brj1}}{\lambda^*}\right),
\end{aligned}$$

as required. The same argument applies to $T^*_{br}(x)^+_j$. We also bound the expected number of trees which have changed. By a union bound and with E_1 and E_2 i.i.d. $\mathrm{Exp}(1)$,

$$\begin{aligned}
&\mathbb{E}\left[\sum_{r=0}^{J} \sum_{b=1}^{B} \mathbb{I}\{T^*_{br}(x) \neq T_{br}(x)\}\right] = B(J+1)\,\mathbb{P}(T^*_{br}(x) \neq T_{br}(x)) \\
&\quad \leq 2dB(J+1)\,\mathbb{P}\left(\frac{E_1}{\lambda^* - \lambda} < \frac{E_2}{\lambda}\right) \leq 2dB(J+1)\frac{\lambda^* - \lambda}{\lambda^*} \lesssim \frac{B(J+1)k}{n},
\end{aligned}$$

since $\frac{\lambda^* - \lambda}{\lambda} = \left(\frac{n+k}{n}\right)^{\zeta} - 1 \leq \frac{k\zeta}{n} \leq \frac{k}{nd}$. Constructing $U(x)$ requires $Bd(J+1)$ comparisons, and since for $b \leq B$ we have $T^*_{br}(x) \subseteq T_{br}(x)$,

$$\begin{aligned}
\mathbb{P}\left(U^*(x) \nsubseteq U(x)\right) &\leq d\,\mathbb{P}\left(U^*(x)^-_j < U(x)^-_j\right) + d\,\mathbb{P}\left(U^*(x)^+_j > U(x)^+_j\right) \\
&\leq 2d\,\mathbb{P}\left(\max_{B+1 \leq b \leq B^*} \max_{0 \leq r \leq J} T^*_{br}(x)^+_j > \max_{1 \leq b \leq B} \max_{0 \leq r \leq J} T^*_{br}(x)^+_j\right) \\
&\leq 2(B^* - B)d(J+1)\,\mathbb{P}\left(T^*_{br}(x)^+_j > \max_{1 \leq b \leq B} \max_{0 \leq r \leq J} T^*_{br}(x)^+_j\right) \\
&\lesssim \frac{(B^* - B)d(J+1)}{B} \lesssim \frac{kd(J+1)}{n},
\end{aligned}$$

since $\frac{B^* - B}{B} \leq \left(\frac{n+k}{n}\right)^{\xi} - 1 \leq \frac{k}{n}$. The average case complexity of calculating $I^*(x)$ is therefore

$$\begin{aligned}
&\mathbb{E}\left[d|I(x)| + dk + dn\mathbb{I}\{U^*(x) \nsubseteq U(x)\}\right] \\
&\quad \leq d\,\mathbb{E}\left[|I(x)|\right] + dk + dn\,\mathbb{P}\left(U^*(x) \nsubseteq U(x)\right) \lesssim \frac{nd\log(2B(J+1))^d}{\lambda^d} + kd^2(J+1).
\end{aligned}$$

A similar calculation shows the cost of calculating all of $N^*_{br}(x)$, $S^*_{br}(x)$, and $V^*_{br}(x)$ is at most

$$\begin{aligned}
&\sum_{b=1}^{B}\sum_{r=0}^{J}\mathbb{E}\left[\frac{dk}{n}|I^*(x)| + d|I^*(x)|\,\mathbb{I}\{T_{br}(x) \neq T^*_{br}(x)\}\right] + \sum_{b=B+1}^{B^*}\sum_{r=0}^{J}\mathbb{E}\left[d|I^*(x)|\right] \\
&\quad\lesssim \mathbb{E}\left[\frac{B(J+1)dk}{n}|I^*(x)| + d|I^*(x)|\sum_{b=1}^{B}\sum_{r=0}^{J}\mathbb{I}\{T_{br}(x) \neq T^*_{br}(x)\} + d(B^*-B)(J+1)|I^*(x)|\right] \\
&\quad\lesssim \frac{Bkd(J+1)\log(2B(J+1))^d}{\lambda^d}.
\end{aligned}$$

where we used that the bounds for $|I(x)|$ and $\sum_{r=0}^{J}\sum_{b=1}^{B}\mathbb{I}\{T^*_{br}(x) \neq T_{br}(x)\}$ hold also in L^2 and applied the Cauchy–Schwarz inequality.

Finally, updating $\hat{\Sigma}_{\mathrm{d}}(x)$ takes $Bd(J+1)|I(x)|$ computations, which is done with probability at most k/K, yielding a time complexity of $\frac{nBd(J+1)\log(2B(J+1))^d}{K\lambda^d}$ on average. The overall average case time complexity is therefore bounded by

$$d(J+1)\left(\frac{knd(J+1)}{K} + kd + B\right) + \frac{d(J+1)\log(2B(J+1))^d}{\lambda^d}\left(n + Bk + \frac{nB}{K}\right).$$

□

SA2 Additional empirical results

Tables SA1, SA2, SA3 and SA4 present some additional empirical results not given in the main paper. The data generating process is identical to that in Section 5, and we demonstrate here the effect of a smaller forest size B, taking $B = 1$ in Tables SA1 and SA2, and $B = 10$ in Tables SA3 and SA4. Note that the bias in Table SA1 is not significantly larger than that in Table 1, even though Lemma 1 suggests that the bias should be much greater. This is because Lemma 1 is stated for the *conditional* bias. In fact, the repeated experiments (we use 3000 independent trials) have the same effect as using a large forest in reducing the apparent bias of the estimator. As such, the error incurred appears in the standard deviation column instead; indeed the standard deviations in Table SA1 are substantially higher than those in Table 1.

References

B. Efron and C. Stein. The jackknife estimate of variance. *The Annals of Statistics*, 9(3):586–596, 1981.

J. Mourtada, S. Gaïffas, and E. Scornet. Minimax optimal rates for Mondrian trees and forests. *The Annals of Statistics*, 48(4):2253–2276, 2020.

V. V. Petrov. *Limit Theorems of Probability Theory: Sequences of Independent Random Variables*, volume 4 of *Oxford Studies in Probability*. Oxford University Press, Oxford, 1995.

	J	LS	LM	λ	RMSE	Bias	SD	Bias/SD	$\widehat{\mathrm{SD}}$	$\hat{\sigma}^2$	ARMSE	ABias	ASD	CR	CIW
No debiasing	0	$\hat{\lambda}_0$	1.0	14.72	0.0606	-0.0235	0.0558	0.4211	0.0322	0.0907	0.0358	-0.0259	0.0232	77.1%	0.126
		λ_0	1.2	23.10	0.0526	-0.0093	0.0518	0.1792	0.0397	0.0883	0.0306	-0.0092	0.0292	88.9%	0.156
			1.1	21.18	0.0565	-0.0102	0.0555	0.1828	0.0381	0.0886	0.0300	-0.0110	0.0279	87.2%	0.149
			1.0	19.25	0.0551	-0.0150	0.0530	0.2834	0.0361	0.0890	0.0298	-0.0133	0.0266	84.9%	0.142
			0.9	17.33	0.0504	-0.0157	0.0479	0.3268	0.0345	0.0895	0.0301	-0.0164	0.0253	83.3%	0.135
			0.8	15.40	0.0568	-0.0202	0.0530	0.3814	0.0323	0.0902	0.0316	-0.0208	0.0238	79.2%	0.127
Debiasing	1	$\hat{\lambda}_1$	1.0	11.20	0.1150	-0.0058	0.1149	0.0508	0.0633	0.1184	0.0302	-0.0031	0.0287	83.0%	0.248
		λ_1	1.2	7.86	0.1274	-0.0046	0.1274	0.0362	0.0568	0.1389	0.0245	-0.0038	0.0242	71.7%	0.223
			1.1	7.21	0.1360	-0.0097	0.1357	0.0719	0.0546	0.1471	0.0238	-0.0053	0.0232	66.1%	0.214
			1.0	6.55	0.1465	-0.0147	0.1457	0.1008	0.0531	0.1551	0.0235	-0.0078	0.0221	63.5%	0.208
			0.9	5.90	0.1595	-0.0150	0.1587	0.0946	0.0543	0.1692	0.0241	-0.0119	0.0210	60.1%	0.213
			0.8	5.24	0.1799	-0.0256	0.1781	0.1439	0.0531	0.1862	0.0275	-0.0191	0.0198	54.0%	0.208
Robust BC	1	$\hat{\lambda}_0$	1.0	14.72	0.1156	-0.0004	0.1156	0.0036	0.0702	0.1111	0.0332	-0.0006	0.0330	90.1%	0.275
		λ_0	1.2	23.10	0.1187	-0.0042	0.1186	0.0353	0.0855	0.1044	0.0415	-0.0001	0.0415	96.2%	0.335
			1.1	21.18	0.1093	-0.0011	0.1093	0.0101	0.0842	0.1025	0.0398	-0.0001	0.0398	95.5%	0.330
			1.0	19.25	0.1006	-0.0026	0.1005	0.0259	0.0808	0.1020	0.0379	-0.0001	0.0379	94.4%	0.317
			0.9	17.33	0.0873	0.0007	0.0873	0.0077	0.0751	0.1012	0.0360	-0.0002	0.0360	93.4%	0.294
			0.8	15.40	0.1032	-0.0018	0.1032	0.0174	0.0710	0.1050	0.0339	-0.0003	0.0339	92.3%	0.278

Table SA1: Simulation results with $d = 1$, $n = 1000$, and $B = 1$, over 3000 repeats

	J	LS	LM	λ	RMSE	Bias	SD	Bias/SD	$\widehat{\mathrm{SD}}$	$\hat{\sigma}^2$	ARMSE	ABias	ASD	CR	CIW
No debiasing	0	$\hat{\lambda}_0$	1.0	12.32	0.3192	-0.1082	0.3003	0.3603	0.0709	0.0846	0.0831	-0.0670	0.0478	64.2%	0.278
		λ_0	1.2	18.39	0.5522	-0.1706	0.5252	0.3248	0.0788	0.0658	0.0771	-0.0292	0.0714	67.1%	0.309
			1.1	16.85	0.4737	-0.1344	0.4542	0.2960	0.0780	0.0707	0.0741	-0.0347	0.0654	70.3%	0.306
			1.0	15.32	0.4220	-0.1178	0.4052	0.2908	0.0764	0.0743	0.0728	-0.0420	0.0595	69.6%	0.299
			0.9	13.79	0.3473	-0.1027	0.3318	0.3096	0.0746	0.0796	0.0746	-0.0519	0.0535	70.4%	0.293
			0.8	12.26	0.3137	-0.1021	0.2966	0.3442	0.0702	0.0835	0.0811	-0.0657	0.0476	65.5%	0.275
Debiasing	1	$\hat{\lambda}_1$	1.0	9.22	0.7451	-0.1351	0.7328	0.1843	0.2742	0.6519	0.0731	-0.0093	0.0692	85.5%	1.075
		λ_1	1.2	7.18	0.5078	-0.0781	0.5018	0.1556	0.2272	0.4124	0.0550	-0.0108	0.0540	81.7%	0.891
			1.1	6.58	0.4920	-0.0686	0.4872	0.1408	0.2227	0.4197	0.0518	-0.0154	0.0495	79.7%	0.873
			1.0	5.99	0.4391	-0.0700	0.4335	0.1616	0.2083	0.3817	0.0503	-0.0225	0.0450	74.9%	0.816
			0.9	5.39	0.4185	-0.0725	0.4121	0.1759	0.1954	0.3862	0.0530	-0.0343	0.0405	72.5%	0.766
			0.8	4.79	0.3645	-0.0789	0.3559	0.2217	0.1766	0.3631	0.0656	-0.0549	0.0360	68.5%	0.692
Robust BC	1	$\hat{\lambda}_0$	1.0	12.32	0.9612	-0.1957	0.9411	0.2079	0.3495	0.9536	0.0926	-0.0014	0.0926	88.3%	1.370
		λ_0	1.2	18.39	1.3984	-0.4142	1.3357	0.3101	0.5041	1.8378	0.1381	-0.0003	0.1381	80.1%	1.976
			1.1	16.85	1.2900	-0.3571	1.2396	0.2881	0.4627	1.5834	0.1266	-0.0004	0.1266	83.3%	1.814
			1.0	15.32	1.1782	-0.3052	1.1379	0.2682	0.4299	1.3422	0.1151	-0.0005	0.1151	85.4%	1.685
			0.9	13.79	1.0995	-0.2733	1.0650	0.2567	0.3901	1.2125	0.1036	-0.0008	0.1036	87.7%	1.529
			0.8	12.26	0.9288	-0.1874	0.9097	0.2060	0.3361	0.8958	0.0921	-0.0013	0.0921	89.3%	1.318

Table SA2: Simulation results with $d = 2$, $n = 1000$, and $B = 1$, over 3000 repeats

	J	LS	LM	λ	RMSE	Bias	SD	Bias/SD	$\widehat{\text{SD}}$	$\hat{\sigma}^2$	ARMSE	ABias	ASD	CR	CIW
No debiasing	0	$\hat{\lambda}_0$	1.0	14.72	0.0376	-0.0241	0.0289	0.8324	0.0250	0.0931	0.0363	-0.0263	0.0232	79.9%	0.098
		λ_0	1.2	23.10	0.0325	-0.0089	0.0313	0.2837	0.0310	0.0895	0.0306	-0.0092	0.0292	93.8%	0.122
			1.1	21.18	0.0317	-0.0104	0.0300	0.3462	0.0296	0.0898	0.0300	-0.0110	0.0279	92.6%	0.116
			1.0	19.25	0.0320	-0.0126	0.0293	0.4306	0.0285	0.0903	0.0298	-0.0133	0.0266	91.5%	0.112
			0.9	17.33	0.0324	-0.0153	0.0286	0.5368	0.0270	0.0907	0.0301	-0.0164	0.0253	89.0%	0.106
			0.8	15.40	0.0334	-0.0200	0.0268	0.7464	0.0255	0.0917	0.0316	-0.0208	0.0238	85.6%	0.100
Debiasing	1	$\hat{\lambda}_1$	1.0	11.05	0.0434	-0.0029	0.0433	0.0678	0.0350	0.1023	0.0297	-0.0027	0.0285	89.0%	0.137
		λ_1	1.2	7.86	0.0458	-0.0062	0.0454	0.1376	0.0316	0.1138	0.0245	-0.0038	0.0242	82.0%	0.124
			1.1	7.21	0.0484	-0.0088	0.0476	0.1843	0.0307	0.1181	0.0238	-0.0053	0.0232	78.2%	0.120
			1.0	6.55	0.0511	-0.0119	0.0497	0.2404	0.0301	0.1239	0.0235	-0.0078	0.0221	75.3%	0.118
			0.9	5.90	0.0572	-0.0187	0.0541	0.3466	0.0291	0.1299	0.0241	-0.0119	0.0210	68.5%	0.114
			0.8	5.24	0.0641	-0.0249	0.0591	0.4209	0.0284	0.1389	0.0275	-0.0191	0.0198	63.0%	0.111
Robust BC	1	$\hat{\lambda}_0$	1.0	14.72	0.0419	-0.0010	0.0419	0.0229	0.0393	0.0951	0.0334	-0.0009	0.0330	93.7%	0.154
		λ_0	1.2	23.10	0.0500	-0.0001	0.0500	0.0014	0.0483	0.0910	0.0415	-0.0001	0.0415	95.4%	0.189
			1.1	21.18	0.0483	0.0009	0.0483	0.0193	0.0468	0.0912	0.0398	-0.0001	0.0398	95.4%	0.183
			1.0	19.25	0.0460	0.0008	0.0460	0.0164	0.0445	0.0917	0.0379	-0.0001	0.0379	95.2%	0.174
			0.9	17.33	0.0439	0.0006	0.0439	0.0134	0.0424	0.0923	0.0360	-0.0002	0.0360	94.7%	0.166
			0.8	15.40	0.0424	0.0004	0.0424	0.0084	0.0400	0.0932	0.0339	-0.0003	0.0339	93.9%	0.157

Table SA3: Simulation results with $d = 1$, $n = 1000$, and $B = 10$, over 3000 repeats

	J	LS	LM	λ	RMSE	Bias	SD	Bias/SD	$\widehat{\text{SD}}$	$\hat{\sigma}^2$	ARMSE	ABias	ASD	CR	CIW
No debiasing	0	$\hat{\lambda}_0$	1.0	12.27	0.0893	-0.0648	0.0614	1.0566	0.0551	0.0979	0.0836	-0.0678	0.0476	72.7%	0.216
		λ_0	1.2	18.39	0.0857	-0.0310	0.0799	0.3884	0.0707	0.0866	0.0771	-0.0292	0.0714	87.7%	0.277
			1.1	16.85	0.0826	-0.0346	0.0750	0.4619	0.0675	0.0881	0.0741	-0.0347	0.0654	87.9%	0.265
			1.0	15.32	0.0819	-0.0435	0.0694	0.6267	0.0635	0.0906	0.0728	-0.0420	0.0595	85.2%	0.249
			0.9	13.79	0.0815	-0.0505	0.0640	0.7891	0.0595	0.0934	0.0746	-0.0519	0.0535	81.4%	0.233
			0.8	12.26	0.0863	-0.0626	0.0593	1.0557	0.0553	0.0972	0.0811	-0.0657	0.0476	75.0%	0.217
Debiasing	1	$\hat{\lambda}_1$	1.0	9.24	0.1034	-0.0138	0.1024	0.1348	0.1017	0.1356	0.0723	-0.0082	0.0694	92.4%	0.399
		λ_1	1.2	7.18	0.0959	-0.0221	0.0933	0.2368	0.0925	0.1577	0.0550	-0.0108	0.0540	88.8%	0.363
			1.1	6.58	0.0964	-0.0284	0.0921	0.3082	0.0901	0.1698	0.0518	-0.0154	0.0495	87.1%	0.353
			1.0	5.99	0.1007	-0.0378	0.0934	0.4046	0.0855	0.1831	0.0503	-0.0225	0.0450	83.3%	0.335
			0.9	5.39	0.1079	-0.0503	0.0955	0.5269	0.0814	0.2002	0.0530	-0.0343	0.0405	77.4%	0.319
			0.8	4.79	0.1208	-0.0748	0.0949	0.7879	0.0740	0.2138	0.0656	-0.0549	0.0360	68.3%	0.290
Robust BC	1	$\hat{\lambda}_0$	1.0	12.27	0.1231	-0.0049	0.1230	0.0399	0.1182	0.1146	0.0923	-0.0015	0.0922	95.1%	0.463
		λ_0	1.2	18.39	0.1577	-0.0036	0.1576	0.0226	0.1361	0.1058	0.1381	-0.0003	0.1381	92.7%	0.534
			1.1	16.85	0.1491	-0.0021	0.1491	0.0139	0.1338	0.1065	0.1266	-0.0004	0.1266	93.4%	0.525
			1.0	15.32	0.1404	-0.0012	0.1404	0.0088	0.1287	0.1071	0.1151	-0.0005	0.1151	94.3%	0.504
			0.9	13.79	0.1294	-0.0018	0.1294	0.0138	0.1231	0.1093	0.1036	-0.0008	0.1036	95.3%	0.482
			0.8	12.26	0.1179	-0.0028	0.1178	0.0240	0.1172	0.1130	0.0921	-0.0013	0.0921	95.6%	0.459

Table SA4: Simulation results with $d = 2$, $n = 1000$, and $B = 10$, over 3000 repeats