A Appendix

In this section, we include related proofs, additional simulation plots that demonstrate the necessity of our proposed debiased causal tree, and a brief introduction to our Supplementary materials.

A.1 Related Proofs

Proof of Theorem 1. By the triangular inequality, we obtain

$$|\eta(x)-\eta^*(x)| \le \sum_{j=1}^{q} \left|\delta_{t_m}(Q_j^{**})-\eta^*(Q_j^{**})\right|\mathrm{I}\{x\in Q_j^{**}\} + \sum_{j=1}^{q}\left|\eta^*(Q_j^{**})-\eta^*(x)\mathrm{I}\{x\in Q_j^{**}\}\right|$$

$$\le \sum_{j=1}^{q} \left|\boldsymbol{b}_{t_m}(Q_j^{**})\right|\mathrm{I}\{x\in Q_j^{**}\} + \max_{1\le j\le q}\sup_{x\in Q_j^{**}}\left|\eta^*(Q_j^{**})-\eta^*(x)\right|. \tag{15}$$

By the Cauchy–Schwarz inequality, we further obtain

$$\sum_{j=1}^{q} \left|\boldsymbol{b}_{t_m}(Q_j^{**})\right|\mathrm{I}\{x\in Q_j^{**}\} \le \left\{\sum_{j=1}^{q}\left|\boldsymbol{b}_{t_m}(Q_j^{**})\right|^2\right\}^{1/2}. \tag{16}$$

Next, we bound $\sum_{j=1}^{q}\left|\boldsymbol{b}_{t_m}(Q_j^{**})\right|^2 = \exp\{B(t_m)\}$. By Assumption 2, it holds that

$$B(t_m)\le B(t_k)+L|t_m-t_k|, \quad \text{i.e.,} \quad \sum_{j=1}^{q}\left|\boldsymbol{b}_{t_m}(Q_j^{**})\right|^2 \le e^{L|t_m-t_k|}\sum_{j=1}^{q}\left|\boldsymbol{b}_{t_k}(Q_j^{**})\right|^2,$$

which implies that

$$\sum_{j=1}^{q}\left|\boldsymbol{b}_{t_m}(Q_j^{**})\right|^2 \le e^{L(t_m-t_1)}\frac{1}{m-1}\sum_{k=1}^{m-1}\sum_{j=1}^{q}\left|\boldsymbol{b}_{t_k}(Q_j^{**})\right|^2. \tag{17}$$

Choosing $C_L \equiv e^{L(t_m-t_1)}$, and combining (15)–(17) give

$$\sup_{x\in[0,1]^p}\left|\eta(x)-\eta^*(x)\right| \le \left\{\frac{C_L}{m-1}\sum_{k=1}^{m-1}\sum_{j=1}^{q}\boldsymbol{b}_{t_k}^2(Q_j^{**})\right\}^{1/2} + \max_{1\le j\le q}\sup_{x\in Q_j^{**}}\left|\eta^*(x)-\eta^*(Q_j^{**})\right|.$$

□

Proof of Proposition 1. By (10), under the Gaussian specification of working densities,

$$\widehat{\mathrm{H}}(\boldsymbol{Q}) = \sum_{j=1}^{q}\left\{2\big(\widehat{\theta}_{t,j}^{(0)}-\widehat{\theta}_{t,j}^{(1)}\big)^2 + \sum_{i:X_i\in Q_j}\left\{\frac{D_i(Y_{i,t}-\widehat{\theta}_{t,j}^{(1)})^2}{|\{i:X_i\in Q_j, D_i=1\}|}+\frac{(1-D_i)(Y_{i,t}-\widehat{\theta}_{t,j}^{(0)})^2}{|\{i:X_i\in Q_j, D_i=0\}|}\right\}\right\}.$$

Under the unconfoundedness condition, if $Y_{i,t}\sim Y_t$ with mean μ_t and variance σ_t^2 independently, for each given Q_j,

$$\widehat{\theta}_{t,j}^{(d)} = \sum_{i=1}^{n}\frac{Y_{i,t}\mathrm{I}\{D_i=d, X_i\in Q_j\}}{n_d(Q_j)}, \quad \text{where} \quad n_d(Q_j)=|\{i:X_i\in Q_j, D_i=d\}|,$$

is the mean of i.i.d random variables with finite second moment and $\widehat{\theta}_{t,j}^{(0)}$ is independent of $\widehat{\theta}_{t,j}^{(1)}$. Thus,

$$E\big[\big(\widehat{\theta}_{t,j}^{(0)}-\widehat{\theta}_{t,j}^{(1)}\big)^2\big] = \sigma_t^2\left(\frac{1}{n_1(Q_j)}+\frac{1}{n_0(Q_j)}\right),$$

and

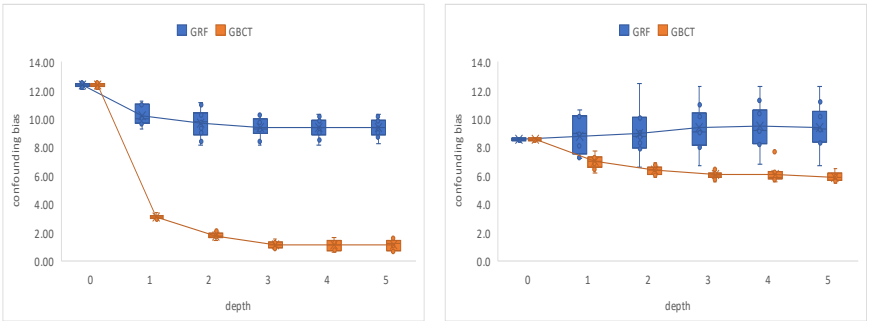
$$E\left[\sum_{i:X_i\in Q_j}\left\{\frac{D_i(Y_{i,t}-\widehat{\theta}_{t,j}^{(1)})^2}{|\{i:X_i\in Q_j, D_i=1\}|}+\frac{(1-D_i)(Y_{i,t}-\widehat{\theta}_{t,j}^{(0)})^2}{|\{i:X_i\in Q_j, D_i=0\}|}\right\}\right] = \sigma_t^2\left(\frac{n_1(Q_j)-1}{n_1(Q_j)}+\frac{n_0(Q_j)-1}{n_0(Q_j)}\right).$$

We then conclude that

$$E\big[\widehat{\mathrm{H}}(\boldsymbol{Q})\big] = 2\sigma_t^2+\sigma_t^2\sum_{j=1}^{q}\frac{n_0(Q_j)+n_1(Q_j)}{n_0(Q_j)n_1(Q_j)}.$$

□

A.2 Additional Simulation Plots



(a) ϕ=0.9 (b) ϕ=1

Figure 1: The confounding bias on the validation data under different depths of a single causal tree in the unmeasured confounding scenario. The y-axis represents the empirical confounding bias $\sum_{j=1}^{q_1} \left|\widehat{\boldsymbol{b}}_{t_m}(\widehat{Q}_j)\right|$, where $\{\widehat{Q}_j\}_j$ denotes the partition of a single tree in GRF or GBCT and $\widehat{\boldsymbol{b}}_{t_m}$ is defined in Eq. 7. The x-axis represents the depth of the causal tree. The data are generated in the same way as described in Section 4. The results show that `GBCT` performs better than `GRF` at the single-tree scale in reducing confounding bias.

A.3 Introduction to Supplementary Materials

Supplementary materials contain a more formal version of our response to official reviews from three anonymous referees and implementation details (including codes) of the data generating process in this paper.