APPENDIX D. Regularisation and the role of the propensity score

In investigating HTEs, all ML estimation approaches entail some form of complexity penalty that discourages learning an overly complex or flexible model. This prevents overfitting by striking a balance between bias and variance (in a mean squared error sense). Regularisation is the term indicating this set of techniques. Regularisation, however, can also represent a major limitation in using off-the-shelf ML methods for CI. Under non-random treatment assignment (i.e., observational studies), individuals may self-select into a treatment group according to some observable exogenous characteristics. As already discussed, conditioning on such features is necessary to ensure unconfoundedness and avoid selection bias. Unfortunately, the naïve application of regularised estimators may still produce severely biased estimates of τ, even after conditioning on $\boldsymbol{X}_i$ (see Belloni et al., 2014; Chernozhukov et al., 2018; Hahn et al., 2018; Hahn et al., 2020, and references therein). Hahn et al. (2020) illustrate this issue from a BART perspective. Suppose that the conditional probability of being selected into one treatment group, $\mathrm{PS}(\boldsymbol{x}_i)$, is partially or strictly based on the expected potential outcome under no treatment, i.e., $\mathrm{PS}(\boldsymbol{x}_i) \approx \mathrm{PS}(m(\boldsymbol{x}_i))$ roughly monotonically, and that m is a function that is difficult to learn. Then, regression tree's splits will disproportionally target t_i when BART tries to learn the complex function f, because partitioning along t_i will require fewer cuts to achieve the same loss reduction. As a result, BART would almost completely ignore the confounding relationship, m, thereby introducing bias. This is the potential artefact of the procedure designated as RIC. Notice that RIC is likely to be a relevant issue in our analysis since farms may self-select into a certain AEP based on expectations of future outcomes in the absence of treatment, thereby leading to a form of RIC known as targeted selection.

The BCF proposed by Hahn et al. (2020), and adopted here, address RIC by separating f into a prognostic and causal component as in Equation (6) and by controlling for an estimated PS, $\widehat{\mathrm{PS}}(\boldsymbol{x}_i)$, in the prognostic function. In this context, the PS plays the role of an additional synthetic control that helps the trees in the ensemble perform partitions along both m and t_i. Technically, this strategy can also be thought of as imposing a covariate-dependent prior over m, where a classical BART prior is given an extra hyperparameter: the estimated PS. Unlike other causal models such as PS Matching, Inverse Probability Weighting and doubly robust estimators, here estimating the PS is purely a prediction problem and, as such, it can be directly addressed through standard ML methods such BART (our choice in this paper).

Finally, it is worth mentioning that the literature proposes alternative ways to address RIC when other ML methods are used to estimate CATEs. These include data splitting and orthogonalization to estimate nuisance functions in meta-learners (Chernozhukov et al., 2018; Nie and Wager, 2021), and honest splitting rules for random forests (Athey and Imbens, 2016; Athey et al., 2019). However, these methods only consider asymptotic bias in semi- and non-parametric CI steups, whereas BCFs address the problem in finite samples (Hahn et al., 2020). Also, as shown by Zhou et al. (2019), including a nonparametric PS as in Equation (7) can provide a form of double robustness to misspecification of the PS itself or of the prognostic function.

References

Bartolini, F., D. Vergamini, D. Longhitano, and A. Povellato. 2021. "Do Differential Payments for Agri-environment Schemes Affect the Environment Benefits? A Case Study in Northeastern Italy." *Land Use Policy* 107: 104862.

Belloni, A., V. Chernozhukov, and C. Hansen. 2014. "High-Dimensional Methods and Inference on Structural and Treatment Effects." *Journal of Economic Perspectives* 28 (2): 29–50.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins. 2018. "Double/Debiased Machine Learning for Treatment and Structural Parameters." Econometrics Journal 21 (1): C1–68. https://doi.org/10.1111/ectj.12097.

Hahn, P. R., C. M. Carvalho, D. Puelz, and J. He. 2018. "Regularisation and Confounding in Linear Regression for Treatment Effect Estimation." *Bayesian Analysis* 13 (1): 163–82.

Hahn, P. R., J. S. Murray, and C. M. Carvalho. 2020. "Bayesian Regression Tree Models for Causal Inference: Regularisation, Confounding, and Heterogeneous Effects (With Discussion)." *Bayesian Analysis* 15 (3): 965–1056.

Nie, X., and S. Wager. 2021. "Quasi-oracle Estimation of Heterogeneous Treatment Effects." *Biometrika* 108 (2): 299–319.

Zhou, T., M. R. Elliott, and R. J. Little. 2019. "Penalized Spline of Propensity Methods for Treatment Comparison." *Journal of the American Statistical Association* 114 (525): 1–19.