APPENDIX E. Achieving overlap using BART

A serious concern in causal analysis regards the violation of the overlap (or common support) assumption. This occurs whenever it is not possible to find reliable counterfactuals for $Y_i(t_i)$, thereby forcing the model to make out-of-sample extrapolations. Extrapolation is highly risky when the model is misspecified, so inference over non-overlapping regions of $\boldsymbol{X}_i$ is generally inappropriate. The lack of overlap may represent a severe problem in the present application, as farmers in treatment group T_1 are likely to include some peculiar farms, at least for some of the covariates included in $\boldsymbol{X}_i$ (Esposti, 2017a; 2017b).

To satisfy the common support condition, researchers have come up with several methods to excise observations for neighbourhoods where only control or treated units exist. Some popular strategies include matching techniques based on the propensity score (Heckman et al., 1997; Nethery et al., 2019), or matching methods based on distance metrics (King and Nielsen, 2019). However, most of these approaches ignore the information in Y_i, which can provide useful insights about the relative importance of each potential exogenous confounder (Hill and Su, 2013). This can be particularly useful when the covariate set is large and some elements of $\boldsymbol{X}_i$ might be highly predictive of T_i, but less so for Y_i. In this case, the resulting common support could place more emphasis on regions of the confounders' space that weakly associate with the outcome of interest. In this respect, BART can exploit both the information in Y_i and T_i to identify areas of weak (or no) overlap through the posterior distribution of $f(\boldsymbol{x}_i, t_i = 1)$ and $f(\boldsymbol{x}_i, t_i = 0)$. Intuitively, when BART is used for CI (as in the case of BCFs), it produces individualised posterior distributions for each potential outcome, whose noisiness reflects how much information was used to identify either $Y_i(1)$ or $Y_i(0)$.

Suppose that for some individual i we only observe $Y_i(1)$, so the counterfactual, $Y_i(0)$, is unknown. Using BART, we can obtain the posterior distribution of the predicted factual outcome, $p_1 = p(f(\boldsymbol{x}_i, t_i = 1))$, as well as the posterior distribution of the corresponding counterfactual, $p_0 =$

$p(f(\boldsymbol{x}_i, t_i = 0))$, and their estimated posterior standard deviation (SD), $s_i^{p1} = \text{sd}(f(\boldsymbol{x}_i, t_i = 1))$ and $s_i^{p0} = \text{sd}(f(\boldsymbol{x}_i, t_i = 0))$, respectively. Consequently, the larger the dispersion of the posterior distribution p_0 (therefore, the larger s_i^{p0}), the more severe the lack of sufficient counterfactual evidence and, consequently, the higher the risk of extrapolation (Hill and Su, 2013; Hill et al., 2020; Hahn et al., 2020; Li et al., 2022). Following Hill and Su (2013), we implement three competing discrimination rules to guide data exclusion. First, when the squared ratio between the posterior SD of the estimated counterfactual and the SD of the corresponding imputed factual exceeds the 95% quantile of a χ^2 distribution with one degree of freedom, we drop that observation (Rule I). The second criterion is like Rule I, except we choose the 10% quantile, a slightly more conservative choice that results in dropping additional observations (Rule II). Third, if the posterior SD of an estimated counterfactual exceeds the maximum SD among all imputed factuals, we discard that observation (Rule III). The results reported in Section 6 are obtained using Rule I which, as illustrated in Appendix H, suggest keeping all the observations in our sample. We leave the remaining criteria as a robustness check.

References

Esposti, R. 2017a. "The Empirics of Decoupling: Alternative Estimation Approaches of the Farm-Level Production Response." *European Review of Agricultural Economics* 44 (3): 499–537.

———. 2017b. "The Heterogeneous Farm-Level Impact of the 2005 CAP-First Pillar Reform: A Multivalued Treatment Effect Estimation." *Agricultural Economics* 48 (3): 373–86.

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.

Hill, J., A. Linero, and J. Murray. 2020. "Bayesian Additive Regression Trees: A Review and Look Forward." *Annual Review of Statistics and Its Application* 7: 251–78.

Hill, J., and Y. S. Su. 2013. "Assessing Lack of Common Support in Causal Inference Using Bayesian Nonparametrics: Implications for Evaluating the Effect of Breastfeeding on Children's Cognitive Outcomes." *Annals of Applied Statistics* 7 (3): 1386–1420.

Li, F., P. Ding, and F. Mealli. 2022. "Bayesian Causal Inference: A Critical Review." arXiv preprint arXiv:2206.15460. https://doi.org/10.48550/arXiv.2206.15460.

Nethery, R. C., F. Mealli, and F. Dominici. 2019. "Estimating Population Average Causal Effects in the Presence of Non-overlap: The Effect of Natural Gas Compressor Station Exposure on Cancer Mortality." *Annals of Applied Statistics* 13 (2): 1242.