Online Appendix Material for "Using Causal Forests to Predict Treatment Heterogeneity: An Application to Summer Jobs"

How Splits are Made

Step 3 of the causal forest described in the main text explains that splits in the tree are made to maximize the variance of treatment effects across leaves subject to a within-leaf variance penalty. This works as follows:

1) Starting with the unpartitioned data, first calculate the value of the objective function: $O = (n_T + n_C)\hat{\tau}_l^2 - 2(\frac{\hat{Var}(Y_{Tl})}{n_T} + \frac{\hat{Var}(Y_{Cl})}{n_C})$ in the current leaf, where $\hat{Var}(Y_{Tl})$ and $\hat{Var}(Y_{Cl})$ are the variances of treatment and control outcomes in leaf l, and n_T and n_C are the number of treatment and control observations in leaf l, respectively.
2) Randomly select v baseline covariates to consider as candidates for a split.[1]
3) At every unique value of each covariate, $X_j = x$, form a candidate split by placing all observations with $X_j \leq x$ in a left child leaf and all observations with $X_j > x$ in a right child leaf.
4) If there are at least 10 treatment and 10 control observations in both child leaves after a candidate split, calculate $O' = O_{left} + O_{right}$ (the sum of the leaf-specific quantity in step 1 across the two new leaves).
5) If there is at least one candidate split where $O' > O$, implement the single split that maximizes O'.
6) If a split is made, repeat this process in each child leaf. If no split is made, this is a terminal leaf.

[1] Considering only a random subset of covariates at each split ensures that a few strong predictors are not used over and over again in the same way across trees. De-correlating trees in this way improves prediction. We use the square root of the total number of covariates, $v = \lceil \sqrt{|X|} \rceil$.