Assessing the Impact of Farm Ponds on Agricultural Productivity in Northern India

May 12, 2024

Supplementary document

Our work employs the difference-in-differences (DiD) methodology to study the impact of farm ponds on agricultural productivity. The DiD method relies on some identifying assumptions such as unconfoundedness, overlap, balance, consistency and no interference. Here, we include the tests performed by us to ensure that the identifying assumptions in causal inference are not violated in our framework.

- **Unconfoundedness:** We considered an extensive list of covariates in our analysis in order to capture the most significant features that would affect the agricultural productivity of an area, as well as treatment assignment. More specifically, we included covariates under the broad categories of soil (which includes soil type and its characteristics), water (availability of water, proximity to water bodies, drainage density, flow accumulation, precipitation), and natural terrain (slope; elevation; proximity to roads, lineament, forests), which cover important contributing factors that affect vegetation growth dynamics.
- **Overlap:** Figure 1 depicts the distribution of propensity scores of the treatment and control groups. It can be seen that there is a significant overlap or common support in the propensity scores distributions of the two groups, implying that, for each location in our analysis the probability of receiving either treatment is positive. Further, we also sample a large enough number of control points which allows for a more accurate matching of treated and control points to form the counterfactual pairs.

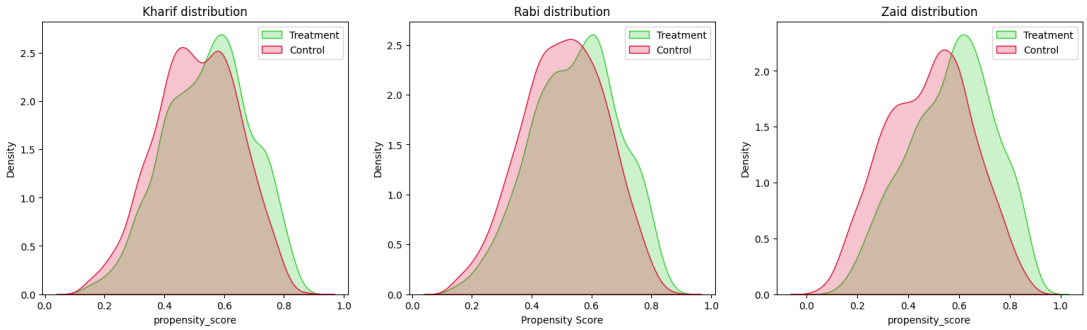

Figure 1: Distribution of propensity scores for the treatment and control groups

- **Balance:**

  Figure 2 depicts the distributions of covariates for the treatment and control groups in our analysis. For all covariates, we observe a significant overlap in the distributions of the two groups, thereby implying that the treatment and control groups are balanced in their covariate distributions.

Table 1: P-value statistics for t-test of various covariates

Covariate	**P-value**
dist_closest_lin	0.1582
dist_closest_upstream_forest	0.5869
elevation	0.4789
slope	0.3045
flow_accumulation	0.6576
drainage_density	0.2347
HSG	0.0949
CEC	0.0111
pH	0.0007

We also perform t-test on the covariate distributions (refer Table 1). We observe non-significant p-values for most of the covariates, which indicates that the two groups (treatment and control) are from the same distribution.

- **No interference:** We ensure that spill over effects from the treated sites to control sites are kept to a minimum by employing a masking strategy that masks out regions in a 500 meters buffer around all treated sites for sampling the control locations. Further, we also mask out all regions in a 500 meters buffer around water bodies to exclude their effect from the control locations.

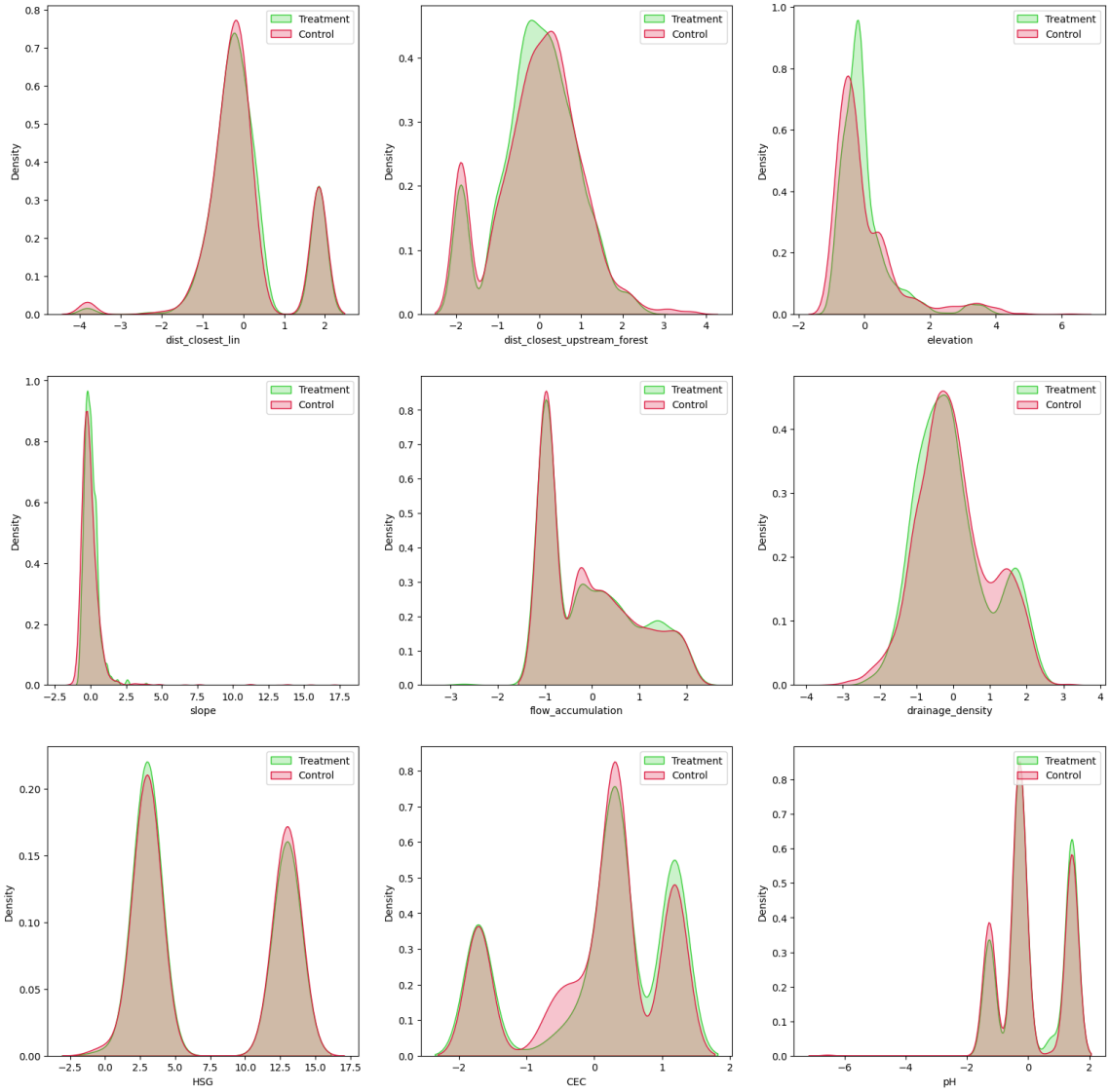

Figure 2: Distribution of covariates for the treatment and control groups