

Env 327 Problem Set 2

Experiencing significant deforestation and biodiversity loss because of domestic and international pressures for development, Indonesia is a biodiversity hotspot (Myers et al, 2000). Establishing protected areas (PAs) is widely seen as the primary intervention for halting habitat degradation and preserving ecosystems. PAs are also likely to affect livelihoods by restricting extractive activities within their boundaries and by generating tourist revenue. However, we still have a poor understanding of the effectiveness of PAs at protecting ecosystems and of their impact on livelihoods (Ferraro and Pattanayak, 2006; Ferraro et al, 2012).

Focusing on propensity score matching, this assignment aims to walk you through the steps of evaluating the effectiveness of the protected areas on stalling deforestation and preventing poverty on the island of Sumatra, Indonesia. To this end, you will work with actual data spanning the subdistricts of Sumatra from 2000 to 2006. **Please do not use these data for purposes other than this class.**

The main data for this assignment are in env327_ps2.dta posted on Sakai. The spatial coordinates for the maps are in coord.dta. The treatment here is protection: a subdistrict is “treated” if it contains a protected area within its boundaries and in the “control” group if it doesn’t. The variable definitions appear on the back of this assignment as well as in the env327_ps2.dta file.

Part I: Getting to know your data

- a. Generate three maps showing the deforestation and poverty levels within the subdistricts as well as the location of protected subdistricts. Make sure each graph is properly labeled. By just looking at the maps, can you identify patterns in space? For deforestation and poverty, where are the highest values concentrated? Do the protected subdistricts seem evenly spaced? Describe briefly what you observe in each map
- b. The goal of matching is to compare treated subdistricts with observationally similar control subdistricts that are not affected by the treated subdistricts. Looking at the map of the protected areas you generated in (a), what observations would you include in your control group? Which ones would you omit?
- c. Generate box plots for the two outcome variables (poverty and deforestation) for the treatment and control groups. (Note, you need two plots-one for each outcome variable.) Based on your graphs, what can you conclude about the effectiveness of protected areas? Describe briefly.
- d. Are the treatment and control groups similar in their characteristics? Generate table with the mean, standard deviation and the number of observations of the covariates for each group separately. Based on this table, are the two groups similar in terms of their characteristics?

- e. Briefly summarize your conclusions from the descriptive analysis and point out any potential issues that may arise if you attempt inference with traditional regression methods.

Part II. Conducting a simple propensity score matching (PSM) analysis

Like other matching methods, the PSM technique compares treated with observationally similar observations to identify the causal impact of the protected areas on the outcome variables (deforestation and poverty incidence in Sumatra). The method collapses the multiple observed characteristics of the treatment and control observations in a single variable—the propensity score, which is the likelihood that an observation is protected given the values of its covariates. The causal effect of protection, the average treatment effect on the treated (ATT), is obtained by the difference between the treated and the observationally similar control subdistricts, matched based on similarities in the propensity score. In other words, the matching procedure allows you to tell what the outcomes of the protected subdistricts would have been, had they not been protected. The ATT gives you the magnitude of the average effect of the program for the protected subdistricts.

- a. The first step is to generate the propensity score. Because the treatment is binary (i.e. a subdistrict is either protected or not), the propensity score can be obtained in a logit. Run a logit with the dependent variable being whether or not a subdistrict is protected. For the covariates, pick the ones that that theory tells you influence **both** the placement of protected areas **and** the outcomes. In this step, you may want to do model specification checks (i.e. check for outliers, multicollinearity etc. Guidance on how to do this for a logit can be found here: <http://www.ats.ucla.edu/stat/stata/webbooks/logistic/chapter3/statalog3.htm>.) Also, you may want to include interaction terms.
Once you are happy with the model specification, use the *predict* option to generate *spa_protected_hat*, which is propensity score.
Interpret your results from the logit: Do your data suggest that protected areas are randomly scattered across Sumatra? What variables seem to be associated with the establishment of protected areas?
- b. Run the propensity score matching command (*psmatch2*) using the predicted propensity score. What is the causal impact of protected areas? Note, you have to run this command twice for each outcome variable.
- c. Unbalanced covariates indicate that your results may be biased. Use the *pstest* command to check whether the covariates are balanced. Present your results in a neat table and briefly interpret.
- d. The PSM technique requires overlapping distributions of the propensity scores for the control and treatment observations. Using the *psgraph* command, check whether this requirement is met. Present and briefly discuss the graph.

- e. The assumption behind the matching techniques is that observations that are similar in their observed characteristics are also similar in their unobserved characteristics (this is the Conditional Independence Assumption (CIA)). Use Rosenbaum's test (the *rbounds* command) to assess the sensitivity of your estimates to hidden bias. Briefly interpret the results
- f. Using the unprotected neighbours as the treatment group and the unprotected non-neighbours as the control group, repeat steps *a through e* to test for the presence of spillover effects. Are there any?
- g. Briefly summarize your findings from Part II. Are the protected areas in Sumatra effective at stalling deforestation and reducing poverty?
- h. Would you recommend establishing protected areas as an effective biodiversity conservation tool in other developing countries? Why/why not? Justify your answer.

Variable code	Variable definition
kecsd_id	2003 subdistrict identifier code
prurban00	proportion villages in urban areas
mean_main_road_light00	proportion of villages with lit main roads in 2000
avgaccessibility00	proportion of villages accessible by car all year round in 2000
avgricearea00	avg rice area in 2000
avgdist2pmkr00	avg distance to mkt w/ permanent structures in 2000
avgdist2ports_km	avg distance to ports, in km
avgai	avg aridity index
avgtoc	avg organic content in the topsoil
avgtph	avg pH levels in the topsoil
avgriv_km	avg river km per subdist
avaveyrtemp	avg yearly temperature per subdist
aavgprecip	avg precipitation per subdist
aelev	avg elevation per subdist
avgslope	avg slope per subdist
avgdist2cap	avg distance to the capital in km
avgmilldist	avg distance to mills in km
avgdist2ports_depth	avg distance to port*depth of the ports
heavily_logged	indicates whether the subdistrict had multiple concessions in 1996; 1 if yes, 0 otherwise
voter_grp	binary indicator whether most of the pop 2000 voted (1 if yes; 0 otherwise)
pforest00	baseline forest as a fraction of total subdistrict area
protect	indicates protection status: 0 "unprotected non-neighbour" 1 "unprotected neighbour"
fraction	"neighbour" pertains to whether the subdistrict is adjacent to a protected one
poverty_change_fraction	proportion of forest lost between 2000 and 2006 as a fraction of the 2000 forest
id	change in poverty between 2000 and 2003 as a fraction of baseline poverty
	identifier field needed to generate the maps

Formulae for the dependent variables

Variable name	Formula	Notes
Forest lost/baseline forest	$fraction = \frac{forest00 - forest06}{forest00}$	No forest regrowth
Poverty change/baseline poverty	$poverty_change_fraction = \frac{poverty00 - poverty03}{poverty00}$	The poverty rate (<i>poverty</i>) the # poor households as total