Evaluation of forest conservation programs in Amazonia Rainforest in Brazil

Duke MIDS Capstone Project in collaboration with Conservation International

Vivek Sahukar & Zhenhua Wang

Conservation International:	Dr. Rachel Golden Kroner & Dr. Sebastien Costedoat
Faculty Lead:	Dr. Alex Pfaff
Capstone Director:	Dr. Gregory Herschlag
Project Manager:	Dr. Heather Huntington & Dr. Ryan Huang

April 14, 2020

CONSERVATION INTERNATIONAL

Duke | SANFORD SCHOOL of PUBLIC POLICY

DEVLAB @ Duke

Duke MIDS Master in Interdisciplinary Data Science

PA & PES are two major forest conservation programs

PA - Protected Area: Government restricts the amount of human activity

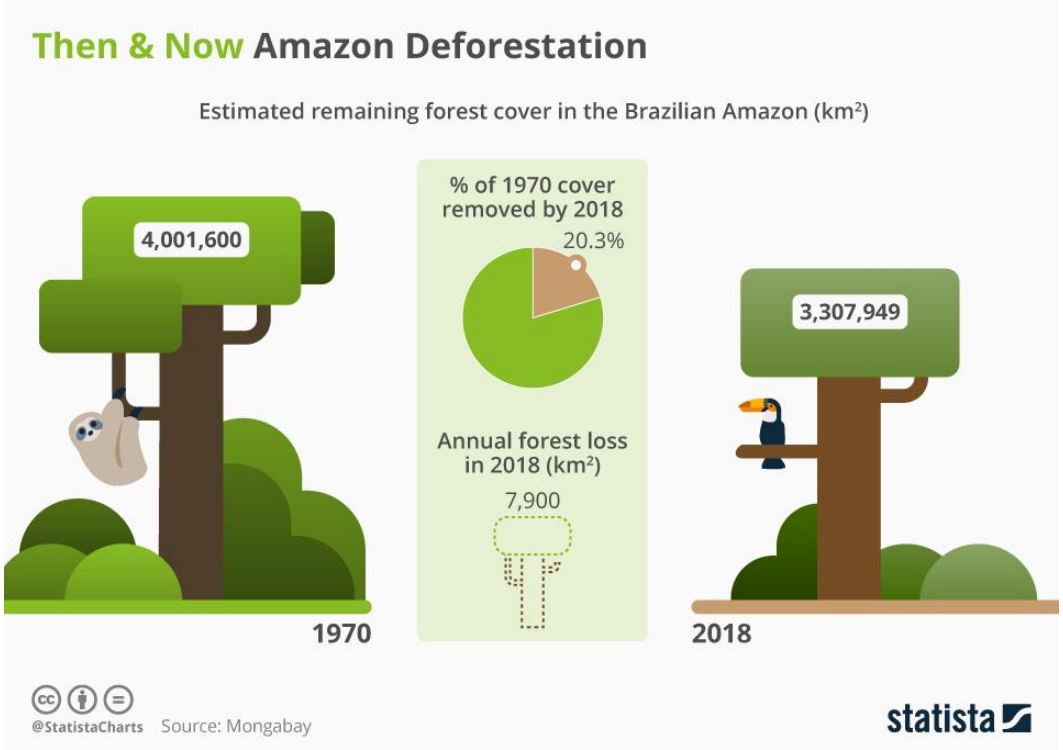
PES - Payment for Ecosystem Services: Government pays local community in exchange for conservation measures

Brazil’s Bolsa Verde (BV), a type of PES

- implemented from 2011 - 2018 in both PAs & rural settlements
- cash transfer program for extremely poor rural populations instead of paying landowners
- Condition: maintain 80% of the original forest cover in the area.

We have considered BV implemented in PAs only to study the interaction effect between PA & PES.

Forest loss in Amazonia is alarming and detrimental to climate change



- Rainforest deforestation is second largest cause of climate change.
- Brazil has 60% of Amazonia - largest tropical rainforest in the world
- Annual forest loss area in Brazil was more than 12 times the size of New York City for 2018.
- This is the highest deforestation rate in Brazil since 2008 and is 30% increase over the rate during the previous year-long span.

Data Source: Brazil's National Institute for Space Research

Project Goal: Evaluate forest conservation programs in Amazonia Rainforest in Brazil

- Conservation International (CI): American nonprofit environmental conservation organization
- CI's Goal for Amazonia is *"**to achieve zero net deforestation in Amazonia"** to protect essential resources, mitigate climate change and increase prosperity for people.*"
- CI is making **Global Conservation Atlas** which is a database of area-based conservation systems, to identify and map natural capital.
- **Evaluation studies** are important to judge the success of forest conservation programs and improve upon in the future.
- **Value for stakeholders** is to understand spatial trends in forest cover loss in Amazonia rainforest in Brazil

Our project answers the following questions:

1. Are PA & PES successful in reducing deforestation in Amazonia Rainforest in Brazil?
2. Which variables affect the forest cover loss?
3. When is the Bolsa Verde eco-payments program most effective within protected area？

Data Variables

	Variables	**Data Source**
Outcome	Forest cover loss	Hansen's Analysis Results of Landsat Images
Treatment	Presence of Bolsa Verde (0 / 1)	Conservation Governance Atlas
Confounders	Elevation (m)	The NASA Shuttle Radar Topographic Mission
	Slope (degree)	Calculated from elevation and is in degree.
	Annual air temperature (℃)	The Climate Data Guide: global precipitation and temperature
	Annual precipitation (mm)	The Climate Data Guide: global precipitation and temperature
	Distance to road (decimal degree)	Calculated from global roads open access data set, Socioeconomic Data and Applications Center
	Estimated travel time to the nearest city of 50,000 or more people (minutes)	Joint Research Center of the European Commission

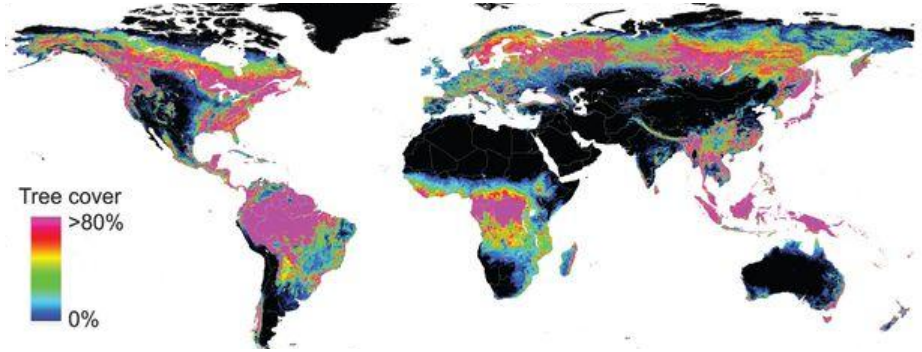
Data Source: Raster

Forest cover: Hansen's analysis results of landsat images

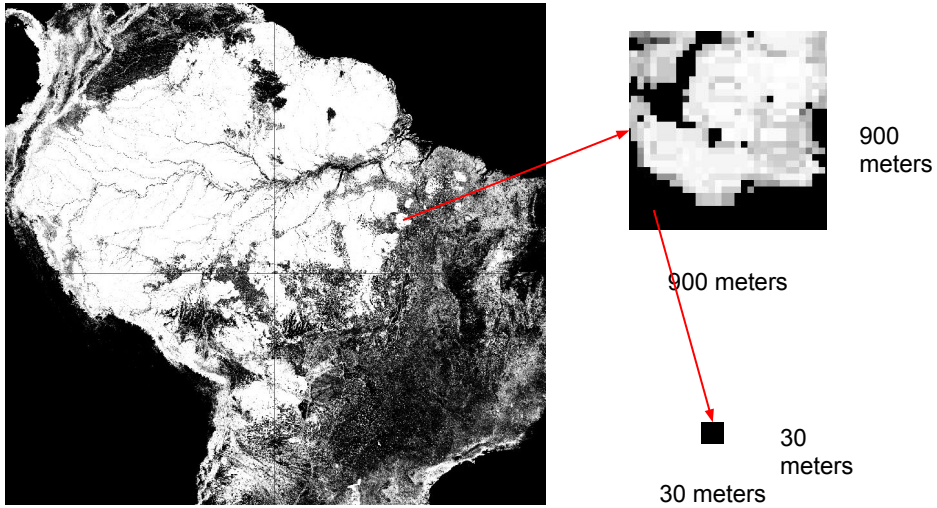
- Pixel: annual forest cover
- Spatial resolution: 30x30 sq m.

Due to limited computation resources, Combined 900 cells into a larger pixel.

- Spatial resolution: 900x900 sq m
- Each new pixel represents count of forest cover plots (30x30 square metres) out of total 900x900 square meters region.



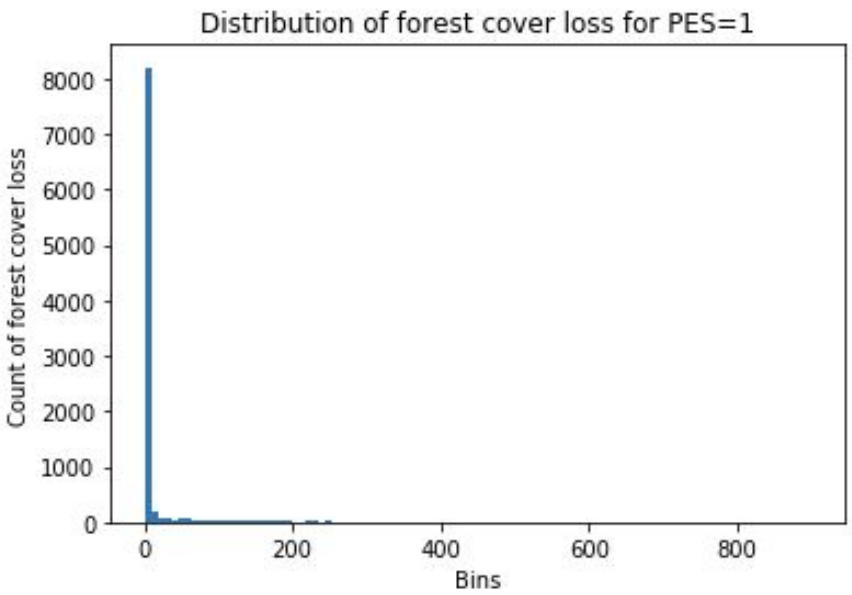
Example of Hansen's data set

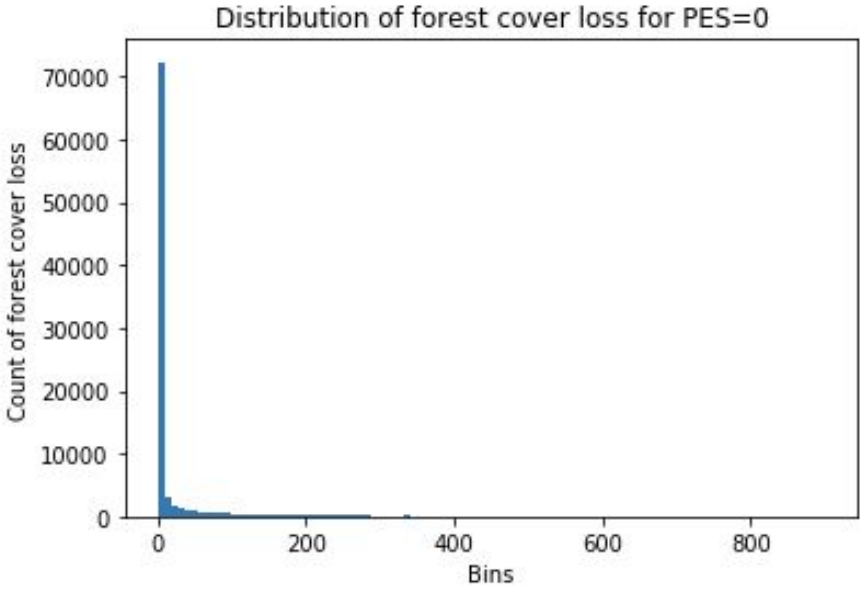


Example of aggregated forest loss

Data Source: Hansen/UMD/Google/USGS/NASA

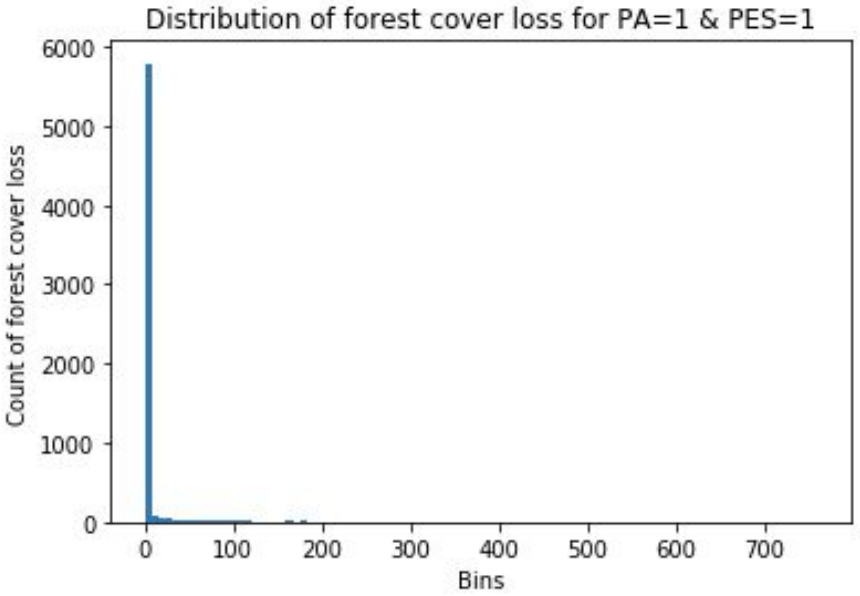
Distribution of forest cover loss is not normal

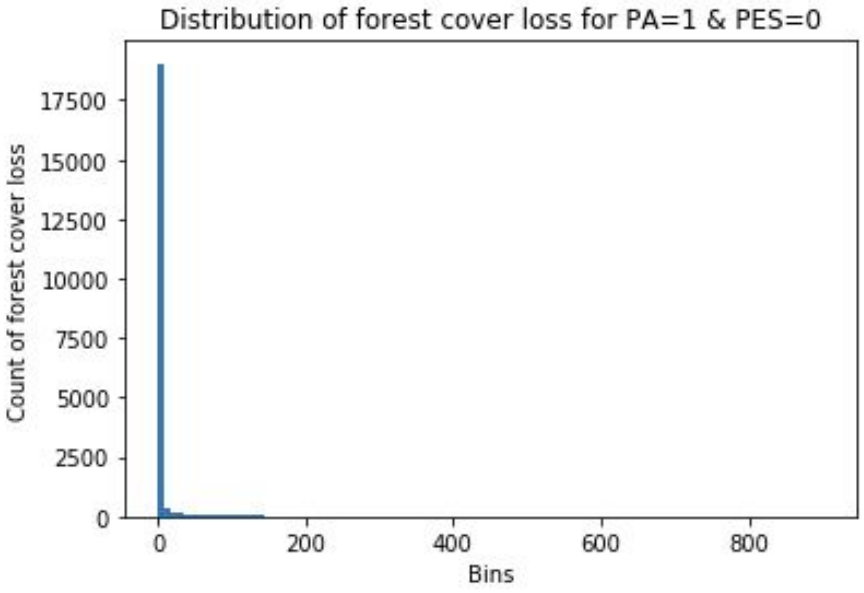




- Hence the assumptions for the linear regression model are not satisfied.
- Therefore, logistic and poisson regression models are used.

Distribution of forest cover loss is not normal





- Hence the assumptions for the linear regression model are not satisfied.
- Therefore, logistic and poisson regression models are used.

Study design - Treatment vs Control Group

- 2 groups studied to study effect of policy on forest cover loss:
  - Treatment group: where policy (PES) has been implemented
  - Control group: where policy (PES) has not been implemented

- Balance Tables used to see whether treatment and control groups have similar distribution of covariates.

- 2 subgroups studied:
  - PES alone (without considering effect of PA)
  - Subset PA alone → Study interaction of PES with PA

Study design - Propensity Score Matching (PSM) with Regression

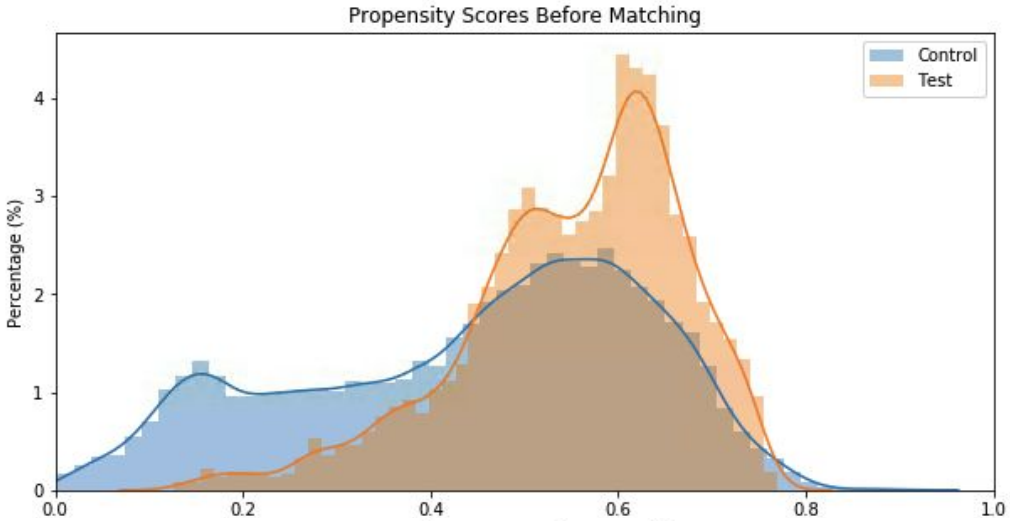
- Imbalance in treatment vs control group
- Randomized study not possible since policy already implemented
- Causal inference method - Matching with regression is used
- Matching → Propensity Score Matching with k-nearest (k=1) neighbor → to achieve balance
- Regression: Logistic & Poisson

Propensity Score Matching

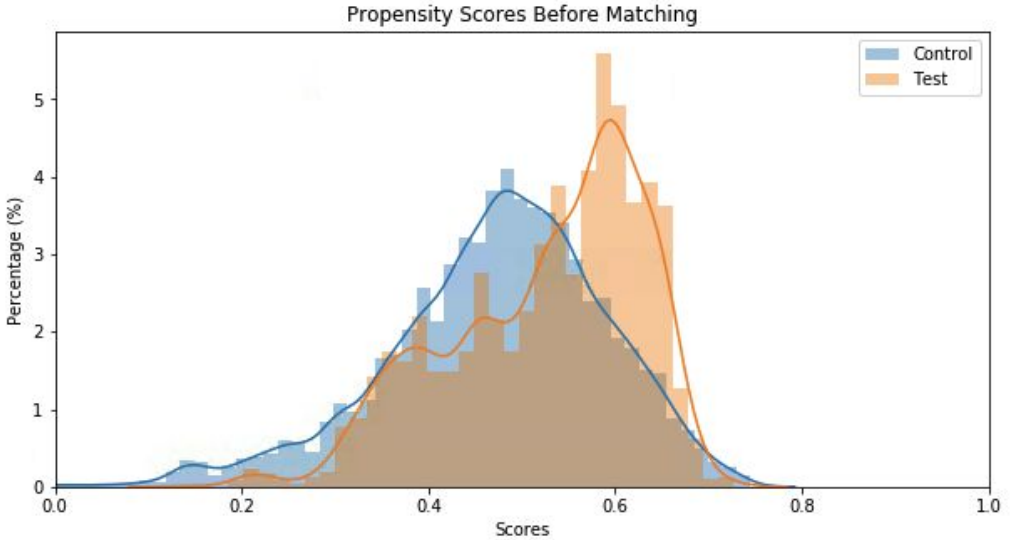
Logistic Regression	Propensity Score	Region of Common Support	Matching
Outcome variable: PES (0 / 1) - Treatment variable Dependent variable: Covariates	Probability of being assigned to treatment or control group given all the values of the covariates	Distribution of propensity score across the treatment and control group: More the overlap, greater the balance	Matched each observation in treatment group to the observation in the control group that has the closest propensity score (k-nearest neighbor matching, k=1) Matching without replacement

Region of Common Support - PSM

PES (0 / 1)



PES (0 / 1) with PA = 1



Balance tables - Distribution of mean of covariates

PES vs non-PES - before matching

Variable	PES	Non-PES	P-value for difference
Slope	0.25	0.32	0.00
Elevation	130.02	169.08	0.00
Distance to road	0.62	0.95	0.00
Access to cities	1743.60	1734.75	0.59
Precipitation 2011	20.00	19.53	0.00
Air temperature 2011	26.87	26.38	0.00

PES vs non-PES - after matching

Variable	PES	Non-PES	P-value for difference
Slope	0.25	0.24	0.11
Elevation	130.02	123.29	0.00
Distance to road	0.62	0.64	0.00
Access to cities	1743.60	1775.13	0.13
Precipitation 2011	20.00	19.95	0.35
Air temperature 2011	26.87	26.85	0.09

Balance tables - Distribution of mean of covariates

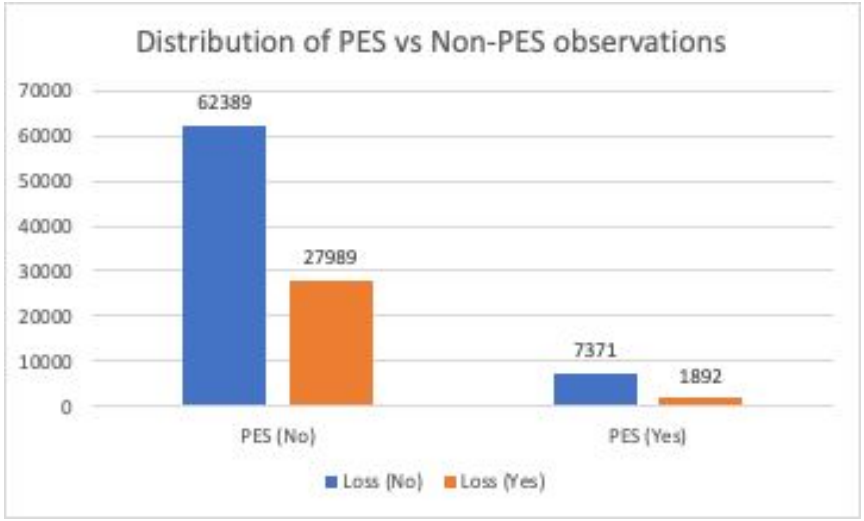
PES vs non-PES (PA=1) - before matching

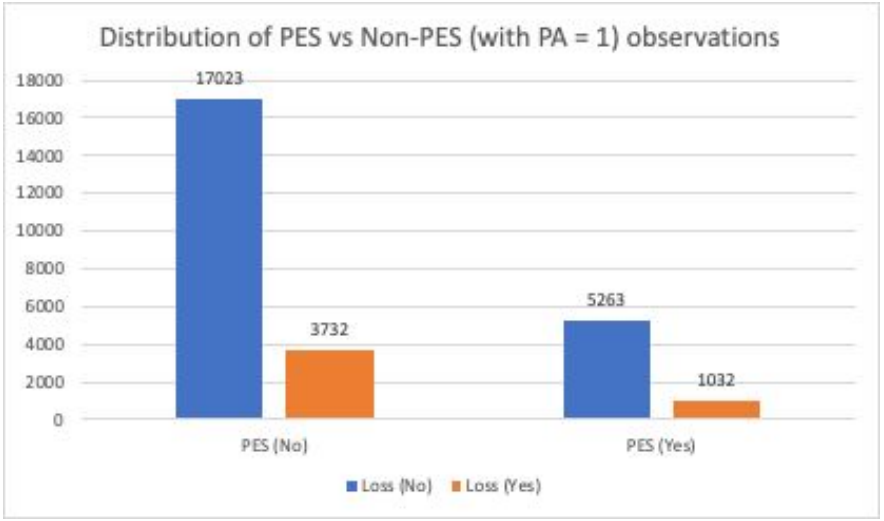
Variable	PES	Non-PES	P-value for difference
Slope	0.24	0.34	0.00
Elevation	135.20	151.24	0.00
Distance to road	0.73	0.98	0.00
Access to cities	1947.46	2134.80	0.00
Precipitation 2011	19.55	20.61	0.00
Air temperature 2011	26.78	26.74	0.00

PES vs non-PES (PA=1) - after matching

Variable	PES	Non-PES	P-value for difference
Slope	0.24	0.24	0.41
Elevation	135.20	133.84	0.41
Distance to road	0.73	0.73	0.90
Access to cities	1947.46	1898.79	0.04
Precipitation 2011	19.55	19.58	0.65
Air temperature 2011	26.78	26.78	0.96

Distribution for Binary Loss variable is skewed





Therefore, PSM was used before applying logistic regression to achieve balance in terms of covariates in both the treatment and control groups.

Model: PSM with Logistic Regression

- Distribution of data is not normal and does not satisfy assumptions of normal distribution
- Hence, linear regression with forest loss as continuous variable is not used
- Instead, Logistic regression is used, where outcome variable: forest cover loss → binary variable

Modeling was done in two steps:

1. PES (0 / 1)
2. PA (=1 only) → PES (0 / 1)

Results

- All confounding variables are significant at explaining forest cover loss.
- PA & PES are effective in decreasing forest cover loss.
- Forest loss decreases as elevation, distance to road and access to cities increase.
- Matching improved the balance in covariates across the treatment & control group.
- Then studied trends in each cluster separately to understand spatial effects of covariates.

Key Question: When is the Bolsa Verde eco-payments program most effective within protected area?

We need:

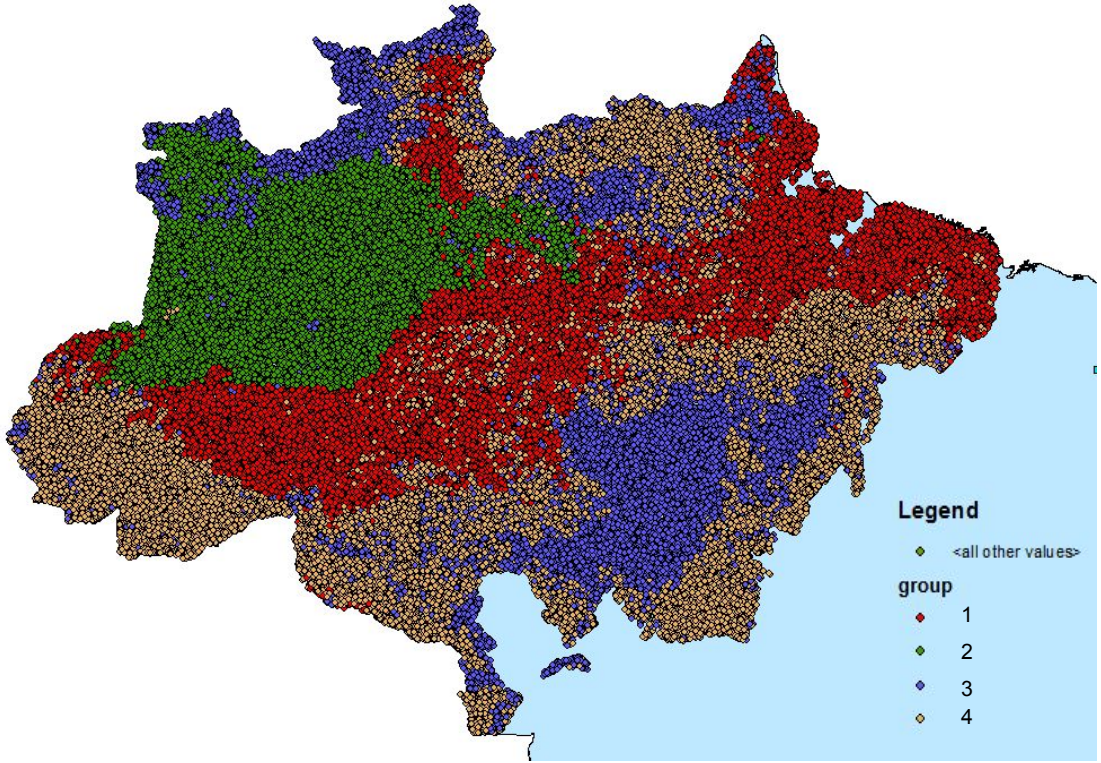
- Regions with different physical covariates

Porto Velho, Rondonia State, Brazil,

Amazon River in Brazil

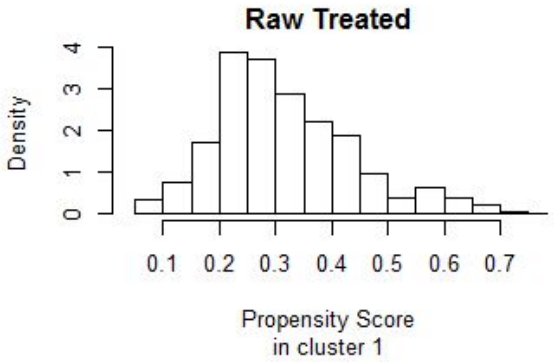
KMeans Clustering used to generate regions with heterogeneous settings

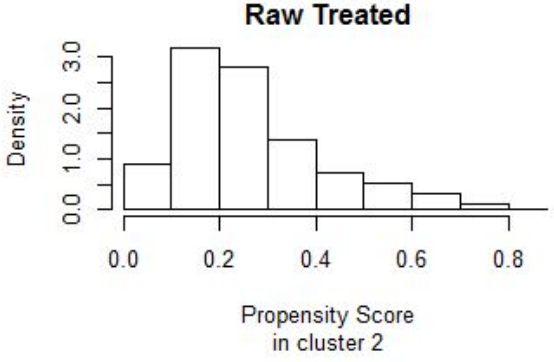
- Cluster 1 is closest to roads and cites.
- Cluster 2 has the largest distance to road, accessibility to cites.
- Cluster 3 has the largest slope and elevation.
- Cluster 4 has the lowest precipitation and is close to cities and roads.

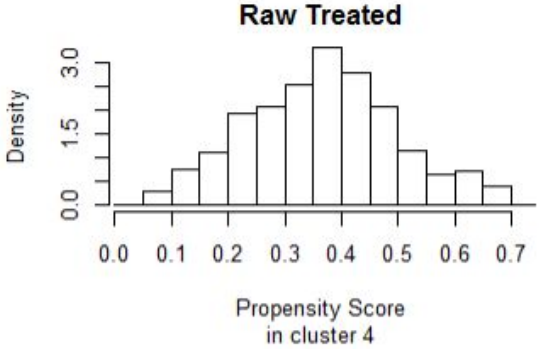


Regions with heterogeneous settings implies Bolsa Verde's enrollment criteria

- Logistic regression were used to calculate propensity score
- The probability of enrollment in Bolsa Verde is different for each cluster
- Cluster 3 and 4 has higher probability of enrollments
  - Higher elevation
  - Higher slope

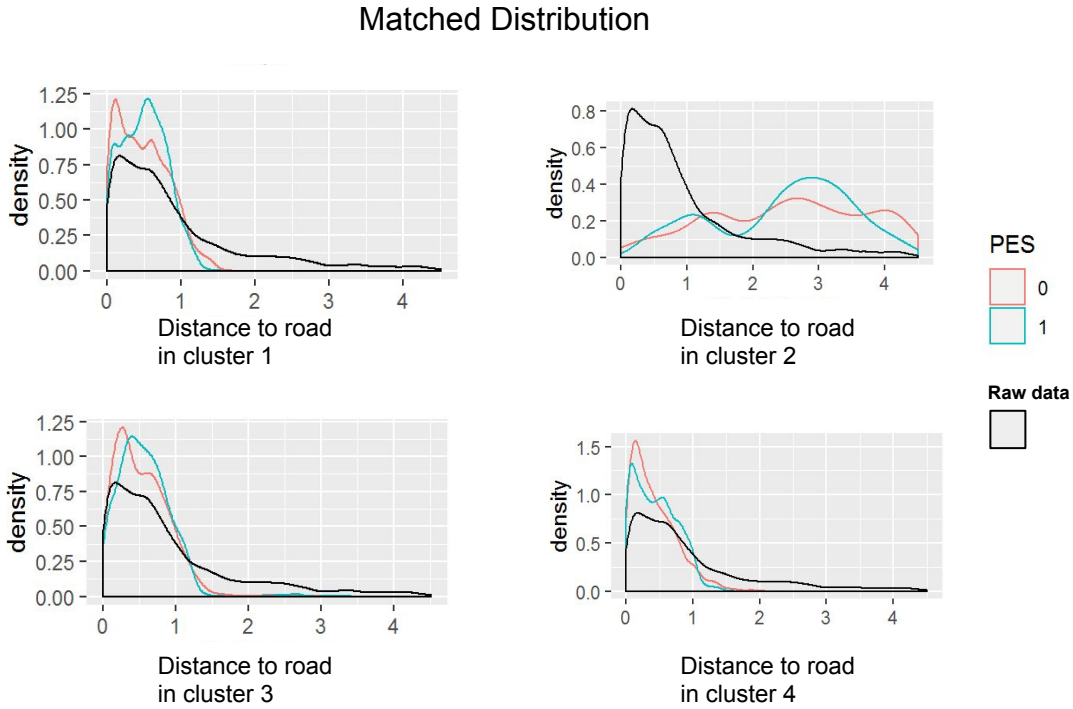






Balance were achieved in each cluster

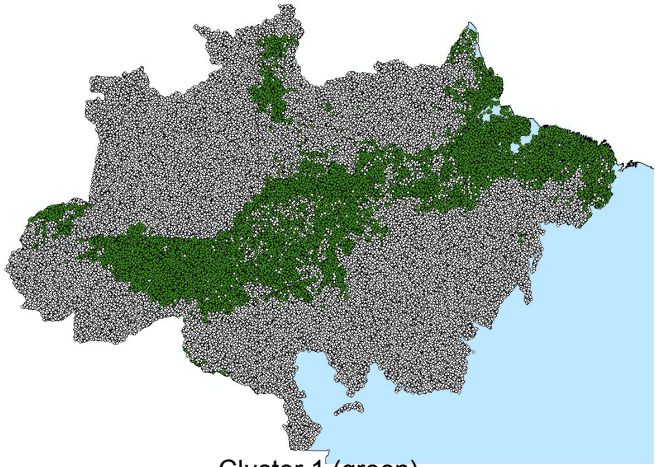
- PSM was used to achieve balance separately in each cluster
- Distributions of confounders are different in clusters



Bolsa Verde tends to be more effective when it is closer to cities and has lower elevation

- Poisson Regression were used to estimates the effects of Bolsa Verde
- In cluster 1 (marked in green), Bolsa Verde reduces forest loss by 18.8%
- In cluster 3, Bolsa Verde reduces forest loss by 2.1%
- In cluster 2 and 4, Bolsa Verde is not effective.

Confounding Variables		*Slope & Elevation*	
	Cluster Numbers	*Low*	*High*
Distance to roads & Access to cities	*Low*	1	4
	High	2	3



Cluster 1 (green)

Conclusion

1. Are PA & PES successful in reducing deforestation in Amazonia Rainforest in Brazil?
   Yes, but it is necessary to run propensity score matching in each cluster to achieve balance

2. Which variables affect the forest cover loss?
- Loss decreases as elevation, distance to road and access to cities increase.
- Loss increases as slope increases.
- The effect of variables differs by clusters.

3. When is the Bolsa Verde eco-payments program most effective within protected area?
- Carrying out deforestation activities in areas having higher slope and elevation is tough even when those areas are closer to cities and accessible by roads.
- Areas that are far away from roads and cities are less influenced by human settlement activities, and hence there is less deforestation.
- Therefore, implementing BV in clusters 2 & 4 is not effective and very less effective in cluster 3; BV is most effective in cluster 1, which is closer to roads and cities and has lower slope and elevation, because those areas are more prone to deforestation by human activities.

Thank you!

Questions?