Estimating Heterogeneous Effects of Cash Bonuses in Retaining Teachers in Remote Peru

Alexander Quispe, World Bank, alexander.quispe@pucp.edu.pe
Sibei Zhang, Stanford University, sibei@stanford.edu
Richard Tang, Stanford University, richtang@stanford.edu

December 2, 2022

Abstract

In 2015, as part of a larger education reform, the Peruvian government implemented a salary bonus scheme ($120 million, or 5% payroll of entire Peru) favoring regions far from urban centers.

In this paper, we build on Leon Perez (2019)'s results, who applied a matching estimator to a geographic boundary discontinuity design to estimate the average treatment effect (ATE) at the boundary, with 5-km half-bandwidth. In this context, we aim to i) replicate Leon Perez (2019)'s findings using a variety of machine learning techniques, ii) further identify heterogeneous treatment effects with additional covariates given unconfoundedness, and iii) produce a geographic visualization of the treatment and control schools with heterogeneous treatment effects.

The contribution of our study are i) validated Leon Perez (2019)'s findings, ii) identified additional heterogeneity in treatment effects and highly influential covariates using causal trees (Athey and Imbens (2016), and iii) created the first-of-its-kind heterogeneous regional treatment effects visualization based on the Peruvian geographies and school locations.

This study would offer valuable insights to Peruvian education policy makers and government officials on the heterogeneous effectiveness of the education reform and possible next steps to further incentivize teacher retention in remote schools.

1 Introduction

As part of a larger education reform in 2015, the Peruvian government implemented a salary bonus scheme ($120 million, or 5% payroll of entire Peru) favoring regions far from urban centers. Because the bonuses are added to base salaries that are uniform across regions, teachers face wage offers that change discontinuously across space.

Leon Perez (2019) applied a matching estimator to a geographic boundary discontinuity design to estimate the average treatment effect (ATE) at the boundary, with 5-km half-bandwidth. He found that the bonuses offered by the Peruvian government have a heterogeneous effect (see Figure 1: temporary teachers react to the bonuses with a sizable increase in their retention rate.

Authors of this paper acquired the dataset used in Leon Perez (2019) with consent from both the author and Peruvian government. The data generating process of the dataset qualifies a quasi-experimental study in which the treatment and control group were clearly identified and separated by geographic boundaries and different levels of cash bonuses (see Figure 1 and Figure 2). The treatment group is Isolated, in which teachers receive $2100, or 25% of annual salary, cash bonus, and the control group is Rural, in which teachers receive $420, or 5% of annual salary, cash bonus.

In this paper, we build on Leon Perez (2019)'s results, applied several machine learning techniques to i) replicate Leon Perez (2019)'s results, ii) further discover heterogeneous treatment effects with additional covariates given unconfoundedness, and iii) generate a geographic visualization of the treatment and control schools with heterogeneous treatment effects.

The contribution of our study are i) validated Leon Perez (2019)'s findings, ii) identified additional heterogeneity in treatment effects and highly influential covariates using causal trees (Athey and Imbens (2016), and iii) created the first-of-its-kind heterogeneous regional treatment effects visualization based on the Peruvian geographies and school locations.We believe our study offers valuable insights to Peruvian education policy makers and government officials on the heterogeneous effectiveness of the education reform and possible next steps to further incentivize teacher retention in remote schools.

2 Related Work

Leon Perez (2019) is the original research that this study was based on. He has made tremendous contributions including i) used a new approach to study regression discontinuity designs in spatial settings; ii) quantified the effect of monetary incentives and discovered heterogeneous treatment effects on the type of teacher contracts (permanent vs. temporary, and qualified vs. unqualified). Our study builds on top of its contribution via ways outlined in the Introduction.

Athey and Imbens (2016) adopted a tree-based partitioning model to estimate heterogeneity by covariates in causal effects in experimental or observational studies. For experiments or observational studies under unconfoundedness, their method allows researchers to identify heterogeneity in treatment effects that was not specified in a preanalysis plan. We adopted their causal tree (CT) method to further investigate the heterogeneity that wasn't specified in Leon Perez (2019) and conduct inference about the magnitude of the differences in treatment effects across subsets of the population.

3 Dataset

This study uses the dataset constructed byLeon Perez (2019), which contains a school panel dataset that provides information on schools as well as the town surrounding it and a teacher panel dataset which, coupled with the school data, allows researchers to follow teachers across schools and space. The school data contains school and surrounding town's characteristics as well as average standardized test scores; the teacher data contains information including teacher salary, bonuses, age, contract type, working location, qualification among others from 2013 to 2017.

4 Research Questions

Given the rich and comprehensive dataset, we expressed particular interest in the following three questions:

1. Can we replicate Leon Perez (2019)'s results and reaffirm significant treatment effects of cash bonuses on qualified, temporary teachers using machine learning techniques?

2. Can we discover more nuanced, multi-dimensional heterogeneous treatment effects along subgroups not specified in the original Leon Perez (2019) study?

3. Is there a more informative and effective visualization to communicate the heterogeneous treatment effects based on geographic and spatial setting in Peru?

Throughout the paper, we maintain the assumption of randomization conditional on the covariates, or "unconfoundedness", formalized as $W_i \perp (Y_i(0), Y_i(1)) | X_i$.

We believe that the dataset was constructed in a quasi-experiment setting induced by the 2015 policy implemented by the Peruvian government, as such implicit boundaries were created and, on one side of the boundary, teachers received an additional 20% salary bonus. Therefore, along with the comprehensive collection of both school and teacher characteristics, authors believe we have observed all the possible variables that affect the teachers' receipt of treatment and are associated with the potential outcomes.

5 Methods, Experiment and Analysis

5.1 Data cleaning and processing

According to the original paper, the only statistically significant results are for qualified, temporary teachers in 2015 and 2016, who as a result of the cash bonus are more likely to stay an extra year at the same school by 0.05 and 0.04 percentage points respectively (see Figure 3).

Therefore, we only keep qualified, temporary teachers in year 2016 (exclude 2015 for simplicity on subsequent HTE analyses). We create binary dummy variables for categorical covariates: sex, level, and classification. The detailed variable list are shown in Table 8.1.

5.2 Replicate the Treatment Effects

5.2.1 Using OLS

Our first task is to replicate the 2016 result using OLS. The results yield a similar point estimate on the treatment variable, 0.0477 (0.006), compared to 0.04 (0.01) in the original paper, with slightly higher point estimate but much smaller standard error (see Figure 3 and Table 8.2).

5.2.2 Using Double/Debiased Machine Learning

We extend the analysis for the treatment effects using the double / debiased machine learning framework of Chernozhukov and et.al (2018) using a Partially Linear Regression Model:

$$Y = D\theta_0 + g_0(X) + \zeta, \qquad E[\zeta \mid D, X] = 0,$$
$$D = m_0(X) + V, \qquad E[V \mid X] = 0$$

Where Y is our outcome variable and D is the treatment. To estimate the causal parameter θ_0, we use double machine learning with 3-fold cross-fitting. The estimation of the nuisance components g_0 and m_0 is based on five different methods (OLS, Lasso, Random Forest, Trees and Boosted Trees). Table 8.3 shows the $RMSE$ for Y and D across all methods. We choose the ones with the lowest values for each variable. Therefore, our **best combination** is predicting Y using trees and boosted trees for D. Figure 4 shows the coefficients of all the methods. Our preferred estimation method coefficient is 0.065 which is statistically significant and around 0.02 units higher than the value found in the original paper.

5.3 Estimate HTE via Causal Tree (CT)

In the following sections, we attempt to estimate HTE following the tutorial (ML-based Causal Inference Tutorial, Golub Capital Social Impact Lab, May 05, 2021). We first used causal trees (Athey and Wager (2019)) to discover subgroups with different treatment effects. In addition to covariates listed above, we added two additional variables, $time\_ie$ (time to the provincial capital) and $population\_ccpp$ (town population surrounding the school), to further distinguish different treatment effects among subgroups. The reason to select these two variables is that the treatment/control classification in the original study depends on these two variables. If our causal tree method is valid, we should discover the high relevance of these two variables in HTEs.

The output shows that leaves that have the top 3 highest treatment effects are (see Figure 5):

- Leaf 5, (17 percentage points, 4% of the sample)
- Leaf 6, (10 percentage points, 25% of the sample)
- Leaf 4, (10 percentage points, 5% of the sample).

These results are aligned with the criteria of the cash bonus policy in which the isolated (or treatment) region is defined by $population\_ccpp < 500$ and $time\_ie > 120$.

Our generation causal trees results have quantified a more nuanced heterogeneous treatment effect estimation: the subgroup with the highest treatment effect is the one with i) a surrounding town population's at least 268 people; ii) time to provincial capital no longer than 202 minutes; iii) teachers age below 36. As one may conclude by these results, it seems like younger teachers are more likely to stay for another year at a school that is **not extremely** isolated in terms of town population and time needed to travel to the provincial capital.

Besides, we look at the distribution of predicted CATE (see Figure 6). The majority of observations fall between 0.05 and 0.15 and behave very much like in a normal distribution. However, we cannot simply draw conclusions about the heterogeneity of the treatment effects solely based on this normal distribution because our method may be simply overfitting. By looking at the variable importance that indicates how often a variable was used in a tree split, we have discovered additional covariates that may be crucial in determining HTEs. However, it is only a rough diagnostic because two covariates might be highly correlated and the trees might only split on one of them.

5.4 Estimate HTE via Data-driven Subgroups

To get unbiased estimates of the CATE, we use the sample splitting method that breaks the data into 5 quartiles of estimated CATEs. We then look in held out data to see what is the estimated treatment effects for each quartile of CATE. Using those estimates, we define a mapping from covariates to the quartiles of the distributions. For every x, we assigned them $1, 2, 3, 4, 5$ based on where they are in the distribution of CATE. That way, we use the data outside of the fold k to build the model and assign k to corresponding quartile, making sure that the assignment does not depend on that unit outcome.

We see some evidence of heterogeneity (see Figure 7). The Q_3, Q_4 are clearly above the Q_1, Q_2, but it is very hard to tell the relationship between Q_5 and the others as Q_5 has a very wide confidence interval. Also, it is hard to tell the difference between Q_1 and Q_2, and Q_3 and Q_4. Next, we check if different groups have different average covariate levels across rankings (see Figure 8).

5.5 Assessing Heterogeneity via Linear Fit

We then run a regression, doing a best linear fit using forest predictions on held-out data (see Figure 9). The regressors are the estimated CATE. The differential forests prediction estimate is the $CATE - ATE$. Based on the result shown in Figure 9, it suggests that there is actual heterogeneity. The result 0.79837 means when we are estimating the CATE changes, it is positively correlated with actual treatment effects.

5.6 Examine CATE Estimates via Partial Dependence

In this section, we examine how our CATE estimates behave when we change a single covariate, while keeping all the other covariates at their median. In Figure 10, we have two major findings: i) the result is only statistically significant for the age 31 group as its confidence interval is not including 0; ii) the treatment effect is decreasing from age 28 to age 44, which aligns well with our findings that older teachers are less sensitive to the salary bonus. It might be the case that varying some particular variable while keeping others fixed may just not be very interesting. Therefore, we should try varying different variables and figure out the most salient ones to dive into.

5.7 Visualize Heterogeneous Regional Treatment Effects

Following 5.4, we also examine the predicted treatment effect by varying latitude and longitude (see Figure 11). Next, we create a map to show the heterogeneous regional treatment effects in different areas in Peru. We observe that in the northwest and southeast of Peru, there is a much larger treatment effect than the rest of regions in Peru. Also, the treatment effect in the northern part and middle part of Peru is subtle (see Figure 12).

6 Conclusion and Limitations

This study has successfully answered three research questions outlined in the beginning of the paper. Authors have replicated Leon Perez (2019)'s results with our choice of covariates and higher power, discovered more nuanced heterogeneity in treatment effects along the government's original treatment policy, in which younger teachers are more likely to stay for another year at a school that is **not extremely** isolated in terms of town population and time needed to travel to the provincial capital.

More importantly, at least from a qualitative and policy-making perspective, we have generated a geographic visualization of the heterogeneity of teacher retention as the result of cash bonus across different regions of Peru. This visualization is first-of-its-kind and may tremendously help policymakers understand the differential impact of their $120 million investment across different regions in Peru.

The main limitation of these results is that the stable unit treatment value assumption (SUTVA) may not hold. This assumption requires that control units are not affected by the bonus schedule at the boundary, so it could be a possible venue of investigation

Leon Perez (2019). In addition, we find that some estimated propensity scores are clustered at 0 and 1 (see figure 13), which will not produce good estimates by using IPW or AIPW methods. One improvement we could make is to chop off the two sides of the estimated propensity scores and re-run our methods.

7 Appendix

References

Athey, S. and Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *PNAS*.

Athey, S. and Wager, S. (2019). Estimating treatment efects with causal forests: An application.

Chernozhukov, V. and et.al (2018). Double/debiased machine learning for treatment and structural parameters.

Leon Perez, M. (2019). Inducing teacher retention in remote locations: Evidence from peru.

8 Tables

Table 8.1: Variables' explanation

Exogenous Variables	Variable	Name of Variable in dataset
Y	The dependent variable is an indicator variable equal to 1 if the teacher stayed an extra year at the same school next year.	$stay$
T	Treatment unit: teacher working at Isolated school (S/500 = \$131 Bonus per month) , Control unit: teacher works at Rural school (S/100 = \$26 Bonus per month).	$treat12$
X	Level of Education (Kindergarten, Elementary, High School)	$level$
	Sex of Teacher	sex
	Age of Teacher	age
	School Classification School only has one Teacher Multi-Teacher Complete: each grade level has designated teachers	$classification$
	Town Characteristics: electricity, water	$cpelec, cpwater$

Table 8.2: OLS Results

Coefficients	**Estimate**	**Std. Error**	**t-value**	**Pr(> $\|t\|$)**	
Intercept	0.125048	0.020723	6.034	1.62e-09	***
treat12	0.047713	0.006041	7.898	2.94e-15	***
sex	-0.007348	0.006219	-1.181	0.2374	
age	0.003071	0.000405	7.584	3.47e-14	***
factor(level)2	0.004622	0.008742	0.529	0.5970	
factor(level)3	0.009403	0.011835	0.794	0.4269	
factor(clasification)2	-0.008398	0.009140	-.919	0.3562	
factor(clasification)3	0.062801	0.010961	5.729	1.023-98	***
cpelec	-0.020036	0.010036	-1.996	0.0459	*
cpwater	0.009932	0.006043	1.644	0.1003	

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*'*

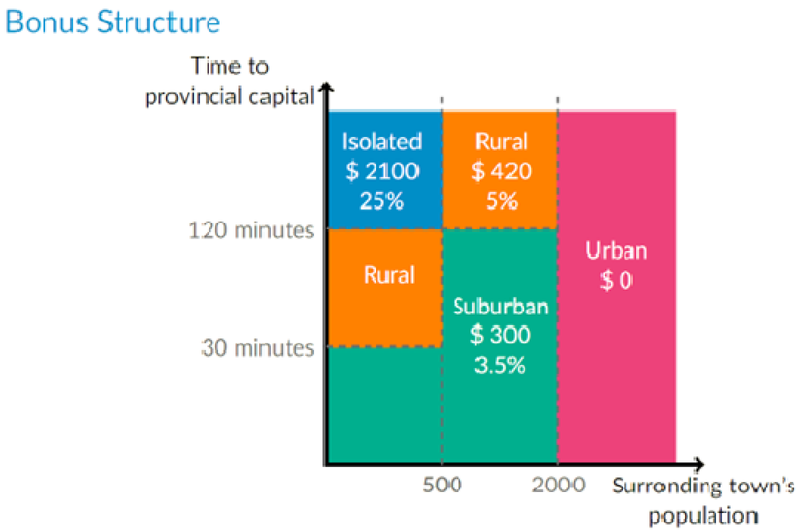
Table 8.3: RMSE for Y and D across all methods

	OLS	Lasso	Random_Forest	Tress	Boosted_Trees
RMSE D	0.281427	0.281411	0.281216	0.281139	0.287129
RMSE Y	0.622952	0.623220	0.622981	0.622667	0.619851

9 Figures

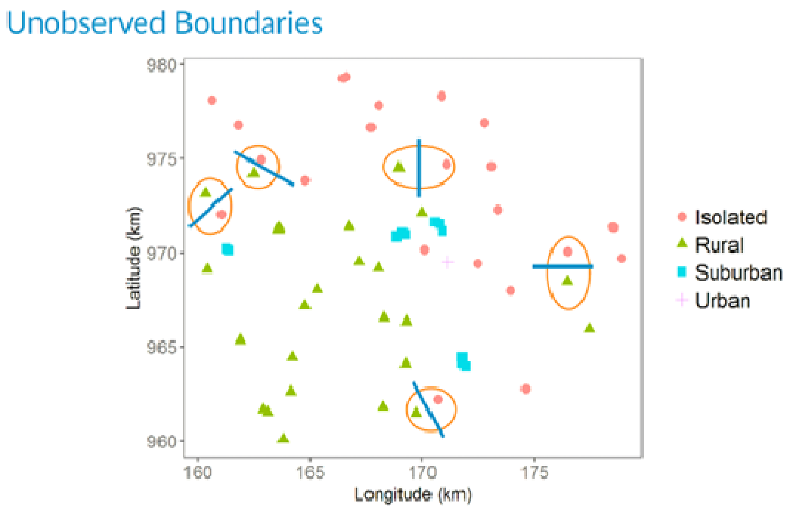
9.1 Introduction

Figure 1: Bonus Structure



Source : León (2019).

Figure 2: Unobserved Boundaries



Source : León (2019).

9.2 Data cleaning and processing

Figure 3: Treatment Effects on Retention for Teachers at Isolated/Rural Boundary

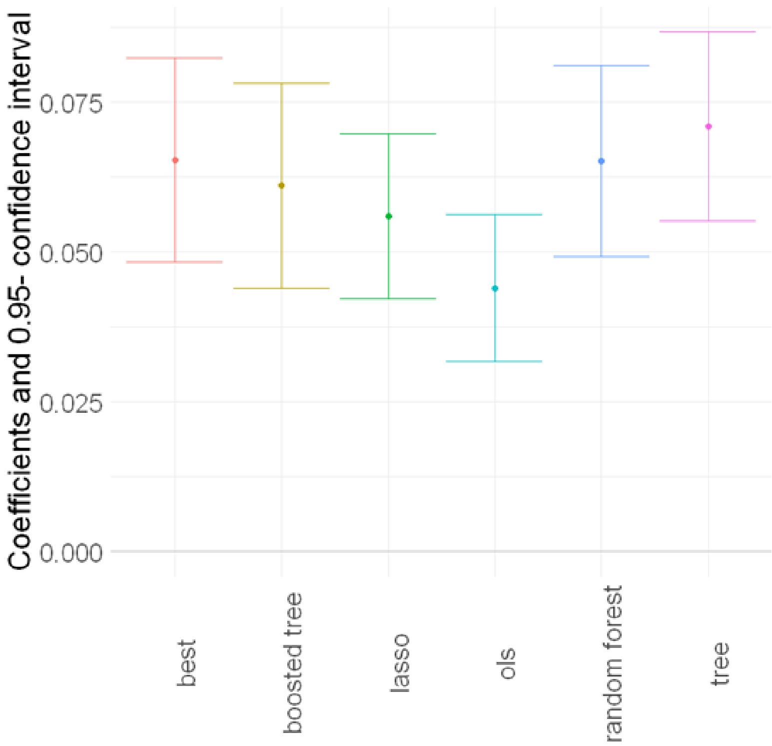
Table 8
Treatment Effect on Retention for Teachers at Isolated/Rural Boundary

	Pre-reform		Post-reform	
	(1) 2013	(2) 2014	(3) 2015	(4) 2016
Permanent	-0.01	0.00	0.02	0.03**
	(0.01)	(0.00)	(0.01)	(0.01)
Qualified Temporary	-0.00	0.02	0.05***	0.04***
	(0.02)	(0.02)	(0.01)	(0.01)
Unqualified Temporary	0.05	-0.04	-0.00	0.01
	(0.13)	(0.09)	(0.08)	(0.05)

Source : León (2019).

9.3 Replicate the Treatment Effects

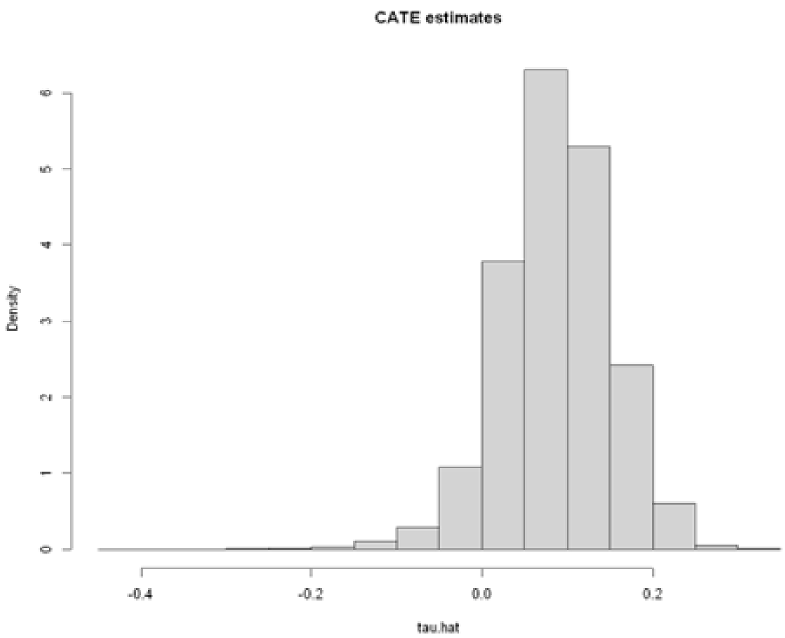
Figure 4: Coefficient Plot among all methods



9.4 Estimate HTE via Causal Tree (CT)

Figure 5: Causal Tree Output

Figure 6: Distribution of predicted CATE



9.5 Estimate HTE via Data-driven Subgroups

Figure 7: Average CATE within each ranking (as defined by predicted CATE)

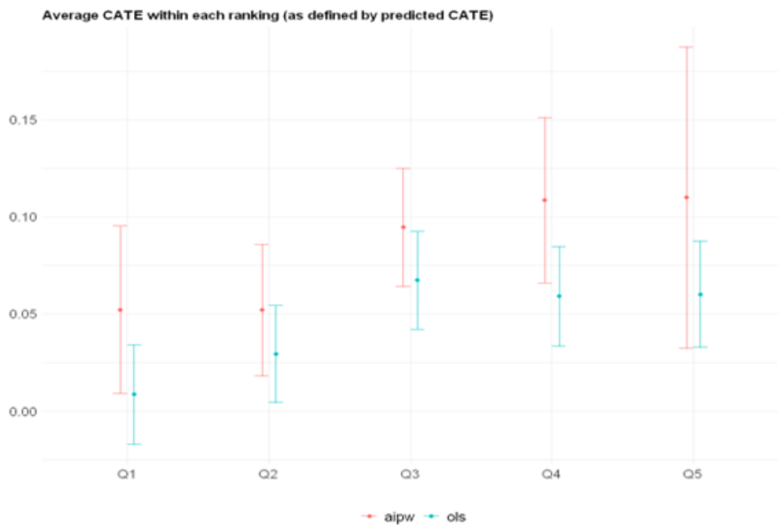
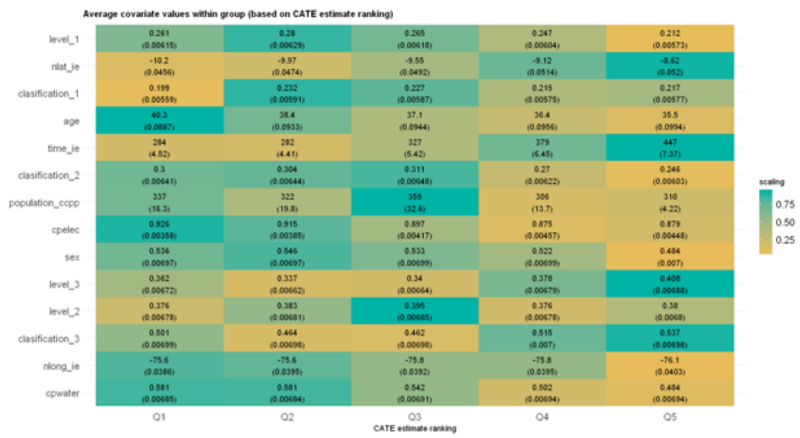


Figure 8: Average covariate values within group (based on CATE estimate ranking)



9.6 Assessing Heterogeneity via Linear Fit

Figure 9: Average CATE within each ranking (as defined by predicted CATE)

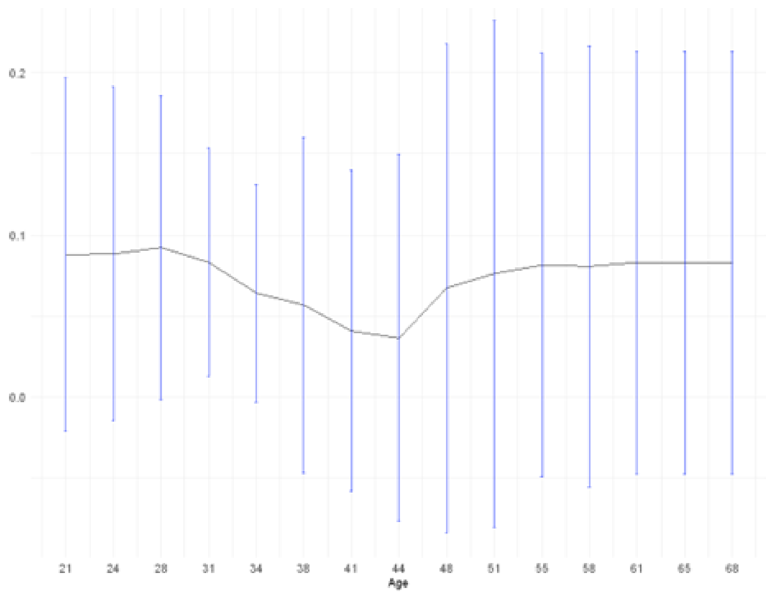
```
test_calibration(forest.tau)
```

```
Best linear fit using forest predictions (on held-out data)
as well as the mean forest prediction as regressors, along
with one-sided heteroskedasticity-robust (HC3) SEs:

                               Estimate Std. Error t value    Pr(>t)
mean.forest.prediction          1.00759    0.13663  7.3747 8.490e-14 ***
differential.forest.prediction  0.79837    0.19059  4.1890 1.406e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

9.7 Examine CATE Estimates via Partial Dependence

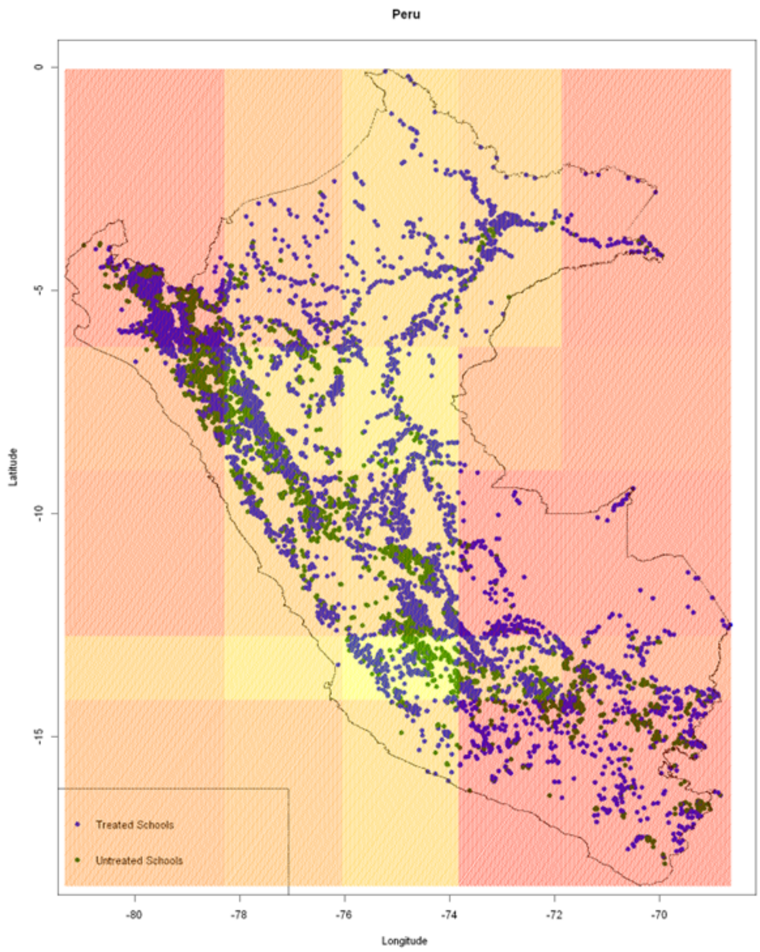
Figure 10: Average CATE within each ranking (as defined by predicted CATE)



9.8 Visualize Heterogeneous Regional Treatment Effects

Figure 11: Average CATE within each ranking (as defined by predicted CATE)

Figure 12: Average CATE within each ranking (as defined by predicted CATE)



9.9 Conclusion and Limitations

Figure 13: Average CATE within each ranking (as defined by predicted CATE)