Payment for Ecosystem Services in Costa Rica: Evaluation of a Country-wide Program

Pablo Ordóñez †
with Juan Manuel Murguia†, Leonardo Corral†, and Gilmar Navarrete-Chacón*

JOB MARKET PAPER

Abstract

Our study evaluates the effect on deforestation and land cover of Costa Rica's payment-for-ecosystem-services (PES) program, one of the oldest country-wide PES programs in the world. Using property level data from over 2,600 landowners who applied to participate in the program between 2016 and 2019, we employ an event study design using modern methods that account for rollout under treatment heterogeneity and find a statistically significant decrease in the deforestation rate. The estimated effect represents an 87% decrease with respect to the pre-2016 average deforestation rate and is equivalent to 0.09 hectares of avoided deforestation per property (a small total effect given the low baseline deforestation). We find no significant effect on forest cover, but we find suggestive evidence that there is a shift from annual to perennial crops. Given that the lack of additionality is one of the main critiques of PES programs, we explore whether the program could increase its additionality by targeting properties with higher ex-ante deforestation risk. For this, we train a machine learning model to predict which properties have a higher risk of deforestation and find that the program is currently not enrolling disproportionately more high-risk properties. Limiting our focus to these properties, we find that the reduction in the deforestation rate is 27-73% larger than what we find for the whole sample of participants. Risk-based targeting could reduce the cost of avoided CO_2 emissions by 42%, from \$71 for the current program to \$41 per ton, well under current estimates of the social cost of carbon.

---

† Pablo Ordóñez, Juan Manuel Murguia, and Leonardo Corral are at the Inter-American Development Bank.

* Gilmar Navarrete-Chacón is the Director of the Environmental Services Program of the National Forest Financing Fund (FONAFIFO) of Costa Rica.

Acknowledgements: We are thankful to the Ministry of the Environment and Energy of Costa Rica (MINAE), in particular to FONAFIFO for their support in the design and implementation of this study, and to the National System of Conservation Areas (SINAC) for their technical support and discussion on MINAE prioritization criteria for PES. Funding has been provided by the Inter-American Development Bank; Economics Sector Work "Costa Rica: Impact Evaluation of Payment for Ecosystem Services Program 2010-2020" (RG-E1851). We wish to thank Allen Blackman, Kathy Baylis and Mateus Souza for their comments on earlier drafts of this paper.

1 Introduction

To decrease deforestation, reduce carbon emissions and preserve biodiversity, many countries have implemented payment for ecosystem services (PES), buoyed by the promise of these programs as a win-win strategy that would allow both the conservation of natural resources and the reduction of poverty for rural households and communities. However, the evaluations of these programs have found only modest effects, with small reductions in deforestation and increases in wellbeing (Ferraro, 2017). Across different systematic reviews and studies, the average estimated effects on deforestation are small, with an approximate reduction of 0.2% per year in the deforestation rate (Ferraro, 2017; Snilsveit et al., 2019; Wunder et al., 2020), modest increases in income ranging from 2% to 14%, and with no effects on other dimensions of human welfare (Ferraro, 2017). These modest results highlight the low additionality these programs can have, and the challenges associated with using them to reduce deforestation. This is especially true in contexts like Costa Rica, where the deforestation rate is relatively low and the total forest area has been increasing since 1990 (Ritchie & Roser, 2021).

Here, we evaluate the effect of Costa Rica's PES program on deforestation and land cover. Given that the limited additionality of PES programs is one of the main criticisms against them, we directly tackle this problem by developing a novel machine learning model of future deforestation rates within properties to evaluate whether improving the targeting of the program could increase the program's effects. Additionally, to consider the broader effects of the program on carbon and landscapes, we measure the program's effects not only on deforestation, but also on total forest cover and on annual and perennial crops. Finally, we also focus on newer cohorts of participants who have not been previously studied.

For this study we have a unique dataset that includes all eligible applicants (enrollees and non-enrollees) for all the cohorts between 2016 and 2020 (2,619 landowners). We combine these data with deforestation data from Hansen et al., (2013), with Costa Rica's official land cover data, and with satellite weather (rainfall and temperature) and elevation data to create a panel dataset of properties from 2010 to 2019. We use these data in an event study design with staggered entry into treatment, where the main identifying assumption is that of parallel trends between treated and untreated units. This means that in the absence of the PES program, enrolled properties would have followed a similar trend in deforestation and land cover as unenrolled properties. We believe this is a reasonable assumption in this setting because our control group consists only of property owners who applied and were eligible to enroll but did not enroll in the program. In fact, we find that few of the differences between treated and untreated properties are statistically significant.

Recent work has shown that in an event study design with staggered entry into treatment, the estimation of treatment effects using OLS can be problematic in settings where there is heterogeneity in treatment effects across cohorts. The two main issues are that, first, the estimated effects for some units will have a negative weight on the treatment effect, and second, the OLS estimates can suffer from cross-lag contamination (Roth et al., 2023). Thus, we use the methods developed by Borusyak et al. (2021) and Callaway & Sant'Anna (2021), which are robust to treatment effect heterogeneity across cohorts and time.

The results of the estimated effect of the program show a statistically significant decrease in the deforestation rate for properties who enroll in the PES program. This effect is relatively large: between an 81% and 87% decrease with respect to the baseline deforestation rate. However, the absolute magnitude of the avoided deforestation is small (0.09 hectares per property) because the baseline deforestation rate in the country is small. We find no effect of the program on forest cover, but we find suggestive evidence of a small reduction in annual crops accompanied by a proportional increase in perennial crops, which has been shown to be associated with more biomass, soil organic carbon and total nitrogen stock (Chen et al., 2022; Means et al., 2022)

To evaluate whether the program's additionality could be increased by enrolling a larger share of properties with higher ex-ante risk of deforestation, we use machine learning to predict which properties, in the year of enrollment, will have high deforestation rates in the future. To do this, we use pre-2016 data (before any of the properties were enrolled) to train a machine learning model to predict the properties for which the future two-year deforestation rate will be above the 90th percentile of all the properties in a given year. We call these 'high-risk' properties. Then, we use this model to classify all the eligible properties for each PES cohort (2016 to 2019) as either high-risk or low-risk. We find that the share of high-risk properties in enrolled and unenrolled groups is similar for almost all cohorts. This suggests that the program is not successfully targeting and enrolling high-risk properties. When we limit the treatment group to only include high-risk properties, the estimated effects are between 27% and 73% higher (although they are not statistically significant). Thus, we believe that targeting high-risk properties to enroll in the program is an opportunity for the program to increase its additionality. We estimate the cost per ton of avoided CO_2 emissions by the program through the avoided deforestation to be equal to \$71 USD per ton of CO_2. If the program exclusively enrolled high-risk properties, the increase in the avoided deforestation would decrease the cost of each ton of avoided CO2 emissions by 42%, to \$41 USD per ton of CO_2.

Being one of the oldest PES programs in the world, Costa Rica's PES program has been extensively studied. However, the conditions of both the country and the program have changed since its inception, and the current the context of this program is by no means unique. The lessons learned from it could be beneficial to countries at a similar income level and with comparable rates of deforestation (Table 11 in the Appendix shows that these countries represent 22% of the world's population and 11.4% of the world's forest area). In this sense, we believe we advance the current understanding of the effectiveness of this program in at least three dimensions.

First, given that the lack of additionality of PES programs is currently one of the main concerns about the effectiveness of these programs for forest conservation (Pattanayak et al., 2010), our approach using machine learning to predict which are the properties with the highest ex-ante risk of deforestation is a valuable contribution that has tangible policy implications. It points to areas in which the implementation of the program could be modified to increase its additionality and highlights the importance of a more risk-oriented targeting for other PES programs. To the best of our knowledge, ours is the first paper to do this.

Second, the data we use allows us to construct a control group using data on actual applicants to the PES program, whereas in previous studies, the construction of the control group was done by choosing areas not enrolled in the program (Arriagada et al., 2012; Arturo et al., 2007; Robalino et al., 2021). The advantage of our data and approach is that the control group we use in our analysis is directly comparable to the group of enrolled properties, which in turn requires us to make less strong assumptions for our estimated effects to be considered as causal. We believe this is an improvement on the existing studies that use quasi-experimental methods. The gold standard set by the only RCT that has been done for a PES program cannot be reached in this setting (Jayachandran, de Laat, et al., 2017), but nonetheless we believe that our methodological approach is an improvement with respect to other studies.

Relatedly, the unit of observation in our analysis is the property. This is an important distinction from previous work, where the units were either grid cells (Arturo et al., 2007) or randomly drawn points (Robalino et al., 2021; Robalino & Pfaff, 2013). To the best of our knowledge, only one study used properties as the unit of observation, and it focused on only one region and not the whole country (Arriagada et al., 2012). Given that the property is the level at which decisions about land use are made, estimating effects at this level can better capture the total effects of the program (i.e., the effect net of any leakage from areas under contract to no-contract areas within the property).

Finally, this is the first study of the effects of the PES program in Costa Rica that has focused on the latest cohort of participants and on other outcomes besides deforestation (forest and agricultural land cover). By studying the most recent cohorts that have not yet been evaluated, we have a more comprehensive understanding of the effects of the program since we can also focus on the stock of land cover within the properties, which allows us to assess the net changes on forest cover. This is important given that different land covers provide different environmental services, and we believe this is the direction in which the conservation literature has started to move (Ordóñez et al., 2023).

The remainder of this paper is organized as follows: section 2 describes the program background and data we use, and section 3 focuses on our estimation strategy. Section 4 presents the results and discusses their significance, and section 5 presents our analysis of the program's additionality based on the predicted deforestation. Finally, section 6 presents a cost-benefit analysis of the program and section 7 presents our conclusions.

2 Background and Data

In 1996 Costa Rica approved the current Forest Law (Law No. 7575), creating a payment-for-ecosystem-services (PES) program that compensates landowners for forest conservation. This law banned the clearing of mature forests in the country and established the National Forestry Finance Fund (FONAFIFO), a semi-autonomous body responsible for managing the PES program. The program is financed by a tax on fossil fuels and its current source of funding might be curtailed in the future as the country moves towards the decarbonization of its economy (IDB, 2020).

Participation in the program is voluntary and each year FONAFIFO opens a call for applications to participate in the program; applications are rated based on an existing set of criteria that focus mostly on the location of the land (indigenous land, protected areas, environmentally strategic land, level of development of the district) and the size of the property. Each application's score must reach a minimum eligibility threshold and this threshold changes every year depending on the available funding. Enrollees sign a 5-year contract with FONAFIFO that establishes the total area to be protected and the payment enrollees will receive in return.

We use the data from all the properties that have applied to the PES program in Costa Rica between 2016 and 2020 to construct a panel dataset of properties. Using the polygons of each property, we overlay the polygons with publicly available spatial data on deforestation, weather, terrain characteristics and accessibility, such that we can know what happened inside each one of the polygons, for each year between 2010 and 2019.

For deforestation, we use Hansen et al.'s (2013) deforestation dataset. This dataset is based on Landsat images and have a resolution of 30m, which means that each pixel covers an area of approximately 0.09 hectares. We not only include data on deforestation but also on actual forest cover. These were developed by the Costa Rican National Meteorological Institute (INM per its acronym in Spanish), based on Landsat images, which were classified based on a Random Forest Model. The resulting rasters have data on land cover for 2013, 2015, 2017, and 2019, with a 30m resolution and 17 types of land covers classified.

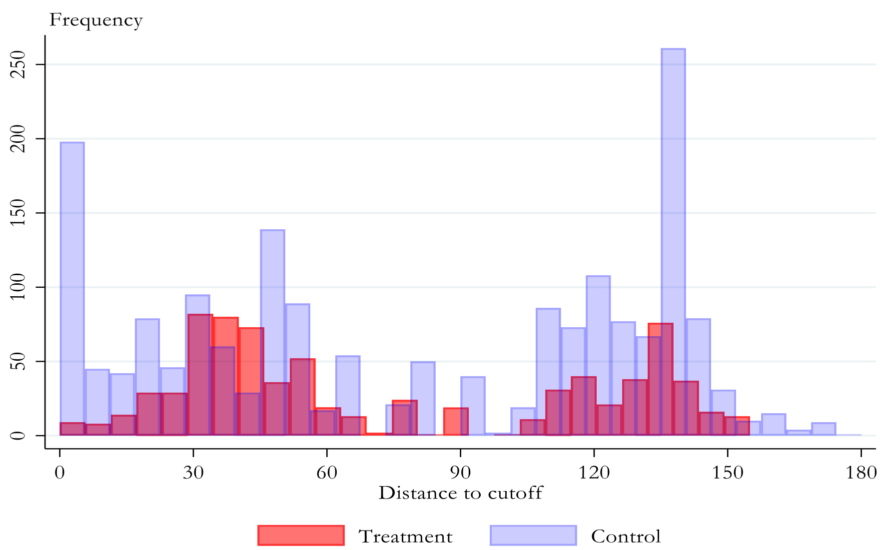
We complement these data with satellite weather data, with the monthly rainfall data from the Climate Hazards Group Infrared Precipitation with Stations (CHIRPS) dataset (Funk et al., 2015), and the monthly mean temperature from MODIS (Wan et al., 2015). We also include data on the characteristics of the terrain, with data on the elevation and slope from the CGIAR STM DEM data and the potential yield for different crops from the FAO global agroecological zones (IIASA & FAO, 2012). Finally, we include a measure of proximity to markets, by including data on travel time to major cities (Nelson, 2008).

We limit our sample to landowners who have never participated in the program for any year between 2010 and 2015 and who have a total area under 300 hectares, which effectively excludes very large indigenous territories. We focus only on those landowners who participate in the forest protection category of the PES program ("Protección de bosques"), which accounts for 74% of all the applications submitted to FONAFIFO and 69% of all the contracts signed. This excludes other categories of the PES such as agroforestry systems, reforestation, and natural regeneration. Given that the forest protection category is the one with the highest demand and that this demand exceeds the available funds every year, the applications are scored using prioritization criteria from the Ministry of the Environment and Energy (MINAE)[2]. These criteria assign a higher score for forests that are in areas considered to have a high environmental value: protected areas, indigenous territories, biological corridors, and forests that provide hydrological services. Additionally, previous program participants,

[2] Set by decree in "Decreto Ejecutivo N 39871-MINAE"

smallholders and those from poorer districts also receive additional points in their applications. The applications are then ranked, from the highest to the lowest score. The funds available each year allow the enrollment of a limited number of hectares and so the applications are selected from highest to lowest score until they reach the maximum number of hectares allowed for that year. This effectively establishes a score cutoff below which none of the applications are accepted. However, some of the applicants that have a score above the cutoff do not always end up enrolling in the program even though they are given the option to do so. Thus, we limit our sample to only those applications for which the score received was above the score cutoff for the year in which they applied (see Figure 1) and that as such had the option to enroll in the program. We believe this allows us to have a more comparable control group, given that they will have similar characteristics.

Figure 1. Histogram of the distance of the application score to the cutoff



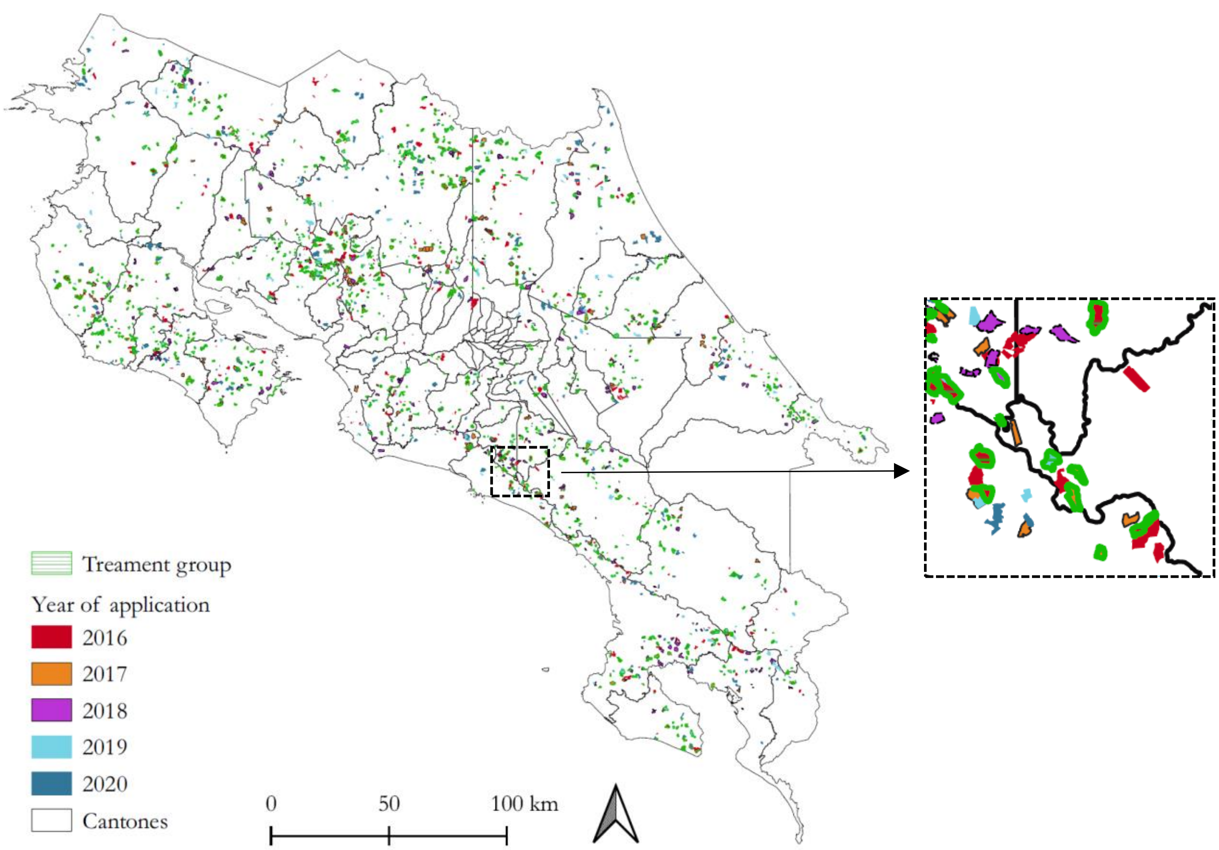
The number of properties that apply to the program changes for every cohort, as does the enrollment rate (Table 1). The 2016 cohort is the one with highest number of applications and enrolled properties, and 2017 is the year with the lowest number of applications. The total enrollment rate is of 42% for the whole set of cohorts in our sample (including 2020), and the average enrollment rate per cohort is 37%. Since we focus only on those that are eligible to enroll in the program, this means that the control group for a given cohort consists of 58% of all the applicants who were eligible to enroll in the program but ended up not enrolling.

Table 1. Applications that are eligible for enrollment per cohort from 2016 to 2020

Cohort	Non-participants	Participants	Enrollment rate	Total per cohort
2016	458	269	37%	727
2017	219	149	40%	368
2018	356	165	32%	521
2019	285	191	40%	476
2020	527	0	0%	527
Total sample	1,846	774	42%	2,619

Finally, we also observe that the properties are all evenly distributed throughout the country (see Figure 2).

Figure 2. Geographical distribution of properties



3 Empirical Strategy

Our approach to estimate the effect that the enrollment in the PES program had on deforestation is based on an event study design, where we estimate the effect of enrollment in the PES program on deforestation for the years before and after enrollment in the program. For this approach, we use OLS to estimate the following model:

$$FL_{it} = \sum_{j=-q}^{-1} \kappa_j PES_j + \sum_{j=0}^{m} \lambda_j PES_j + X_{it}\beta_2 + \alpha_t + \alpha_i + \varepsilon_{it} \quad (1)$$

Where FL_{it} is the deforestation rate inside the polygon of property i in year t (measured as the proportion of the deforested land in a year over the total land area of the property). The model includes leads, which capture the effect of enrollment j years before the actual enrollment in the PES program, such that κ_{-2} would capture the effect of enrolling in the program, 2 years before enrollment. It also includes lags, which allow us to estimate the effect on deforestation j years after enrollment, such that λ_3 for example, would be the effect from enrolling in the program, 3 years after enrollment. We include a set of control variables X_{it} that includes rainfall, temperature and their interactions with crop suitability for different crops, elevation, slope and travel time to cities. We also include year and property fixed effects.

The main identifying assumption for the estimation of a causal effect from participation in the PES program is that of the parallel trends in deforestation between treated and untreated properties. In this case, this implies that the trends in deforestation that treated properties would have followed had they not been treated, would have been similar to the trend followed by the untreated properties. While this is an untestable assumption, we provide evidence that shows that the trends followed by treated properties prior to enrollment, were similar to those followed by the untreated properties during that same period. In our setting, this would imply that the κ_{-j} coefficients are statistically equal to zero and that they are jointly not significant.

Additionally, given that in this setting we have properties that enroll in the program at different points in time, we need to assume that there is no heterogeneity in the treatment effects for the different cohorts, to have unbiased estimates of λ_j. With treatment heterogeneity across program cohorts, there are two issues that arise (Roth et al., 2023). First, the coefficient λ_j may put a negative weight on the treatment effect j periods after treatment, so that for example, the treatment effect for some properties three years after entering the program may enter λ_3 negatively. Second, the coefficient λ_j may put non-zero weight for treatment effects at lags $j' \neq j$, creating cross-lag contamination. This could lead to a situation where λ_3 is affected by the treatment effect for some properties that have been in the program for four years. Importantly, (Sun & Abraham, 2021) show that if there are heterogeneous treatment effects, the κ_{-j} coefficients from (1) can be equal to zero even in there are pre-trends or can also be different from zero even if there are no pre-trends. Thus, if there is treatment heterogeneity, these coefficients are not informative about the validity of the parallel trends assumption.

We believe that in our setting, the presence of heterogeneous dynamic treatment effects is highly likely, given that new properties with different characteristics apply to the program every year (see Table 13 in the Appendix), which together with changes in the external conditions that might affect land use decisions, could change the effects that the program has for different groups at different points in time. Thus, for example, the assumption that the average treatment effect of the program 2 years after enrollment for properties who enrolled in year t is the same as what it is for properties who enrolled in $t+1$, might not be a reasonable assumption.

Thus, we use two recently developed methods that are robust to the presence of heterogeneous treatment effects. First, we use the method developed by Callaway & Sant'Anna (2021)[3], which is based on a doubly-robust approach from Sant'Anna & Zhao (2020), under which the estimated effects are consistent if either the propensity score model or the outcome model are correctly specified. The group-average treatment effect is estimated from two elements: (1) the estimated probability of a property being part of a treated cohort based on their pre-treatment characteristics; (2) the expected difference between the outcome in the period before treatment and the expected outcome for the never treated group.

Second, we also use Borusyak et al.'s (2021) method[4], which is an imputation estimator that uses the not-yet-treated properties and time periods in a regression with property and year fixed effects to generate the predicted potential outcome for the treated properties after treatment. These can be used to estimate the treatment effect for each property and year, and they can be aggregated to estimate the average treatment effect at different lags. The main difference between the CS and BJS methods is that the base period in BJS is the average of all the pre-treatment, whereas the base period for CS is the last pre-treatment period.

4 Results

We compare the characteristics of both participants and non-participants in our sample, to assess if on average, these two groups are systematically different. We find that both groups have similar characteristics, but there are statistically significant differences in the means of both groups for certain variables (Table 2).

[3] Henceforth, we will refer to this method as CS.

[4] We will refer to this method as BJS.

Table 2. Summary statistics of main characteristics by group

	All sample	Control	Treatment	Diff.	t-statistic
Application score	116.87	113.16	125.72	-12.56	(-10.90)***
Deforestation in ha (pre-2016)	0.10	0.12	0.07	0.05	(4.65)***
Deforestation Rate (pre-2016)	0.152%	0.159%	0.134%	0.026%	(1.49)
Forest Area 2015 (ha)	56.14	60.20	46.48	13.72	(6.02)***
Forest Area 2015 (% of total area)	0.82	0.81	0.84	-0.03	(-3.20)**
Total crops area 2015 (ha)	1.83	2.03	1.34	0.70	(1.66)
Annual crops 2015 (ha)	0.52	0.58	0.39	0.19	(1.27)
Perennial crops 2015 (ha)	1.31	1.46	0.95	0.51	(1.43)
Share annual crops 2015 (% total)	0.81%	0.83%	0.76%	0.07%	(0.42)
Share perennial crops 2015 (% total)	1.93%	2.23%	1.21%	1.02%	(3.06)**
Area (ha)	68.82	74.36	55.61	18.75	(7.24)***
Proposed area under PES (ha)	38.90	41.90	31.72	10.23	(6.15)***
Elevation (m.a.s.l)	608.69	617.35	588.04	29.31	(1.13)
Slope (degrees)	12.33	12.27	12.47	-0.20	(-0.67)
Potential yield – Maize (kg/ha)	701.89	699.14	708.45	-9.31	(-0.36)
Potential yield – Sugarcane (kg/ha)	1246.05	1258.31	1216.82	41.50	(0.90)
Potential yield – Wheat (kg/ha)	15.02	15.50	13.85	1.65	(0.69)
Potential yield – Citrus (kg/ha)	261.76	272.44	236.31	36.13	(2.31)*
Potential yield – Coffee (kg/ha)	321.92	324.55	315.65	8.91	(0.71)
Travel time (min)	340.77	338.57	346.03	-7.46	(-0.95)
Mean annual rainfall (mm)	2969.79	2980.43	2944.44	35.99	(1.03)
Mean temperature (Celsius)	26.23	26.23	26.24	-0.01	(-0.10)
Obs.	2,619	1,845	774		

As expected, we find that there are significant differences in the application score, with participants having a higher score than non-participants. In terms of the outcome variables, we find that non-participants have higher total deforestation (significant difference), a higher deforestation rate (non-significant difference), more forest area in hectares but less forest area as a share of the property's land area (significant difference), which is explained by the fact that non-participants have larger properties (significant difference). In terms of environmental characteristics, the only significant difference between the two groups is that non-participants have a slightly higher potential yield for citrus. Importantly, none of the other characteristics that could influence land use decisions, such as elevation, slope, potential yields, and weather are different between the two groups.

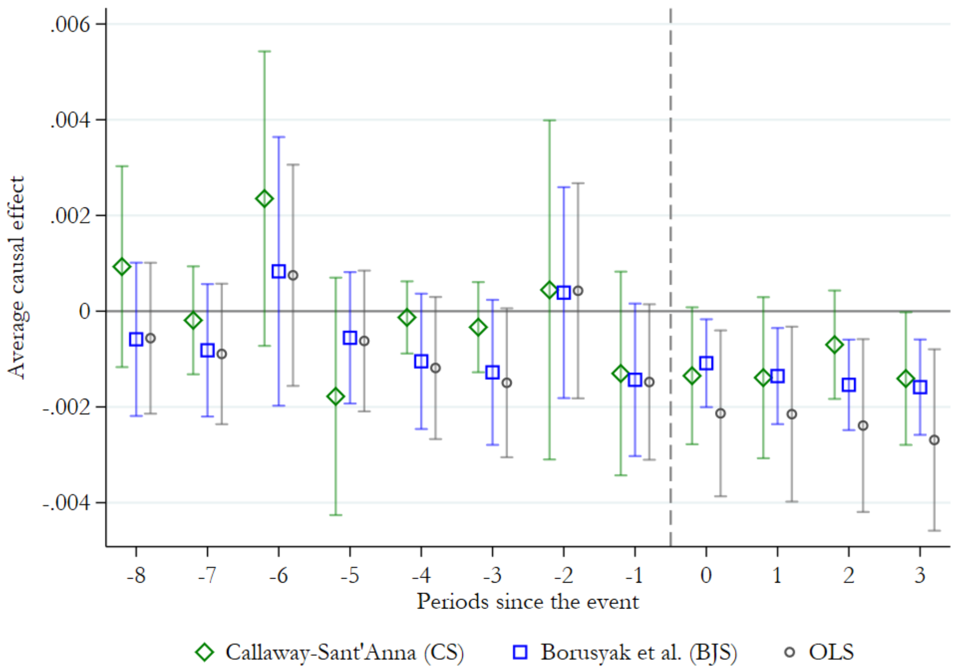
4.1 Deforestation

Our main outcome variable is the deforestation rate inside each of the properties that applied to the program. Given that enrollment in the program is staggered and that we believe that effects of the program are likely to be heterogeneous in time, our two estimates of the effects of the program on deforestation come from CS and BJS (we also include the estimated results from OLS for comparison). The main assumption for the unbiasedness of the estimated effects in this case is that of the parallel trends, which means that the trend in the deforestation rate within each property after enrolling in the PES program, would have been the same as the trend followed by the control group, if the enrolled properties had not participated in the program. Our belief that this is a plausible assumption in this case, comes from the fact that both groups are statistically similar in their observable environmental characteristics (Table 2) and that both groups applied to the program and were eligible to enroll. This makes us confident that the scope of the potential selection bias is limited, especially for the selection into participating in the program. However, it is possible that there is some selection into enrollment, since all the properties in our sample were eligible to enroll, but not all ended enrolling into the program. Thus, our main assumption is that the decision to not enroll in the program is on average uncorrelated with any time varying external factor that affects the deforestation rate and land use within the properties. From anecdotal evidence from the program administrators, both participants and not participants are equally interested in the program and nonenrolment is ultimately the result of unforeseen circumstances (for example, the applicant takes longer than anticipated gathering all the necessary documents or cannot find one of these documents). In fact, we observe that the average date at which the applications are submitted are similar for both participants and nonparticipants (Table 12 in the Appendix).

Furthermore, given our event study specification, we can test whether the trends in deforestation between participants and nonparticipants before enrollment are the same, which is in fact what we find (Figure 3). This does not mean that the parallel trends assumption is true in our case, since this assumption is ultimately untestable, but it provides evidence in favor of the validity of the assumption in this setting.

We find that after enrollment in the program there is a decrease in the deforestation rate. All the effects estimated using BJ and OLS are statistically significant, while those based on CS are non-significant for all the post-treatment period, although they are significant at the 10% significance level. The magnitude of the estimated effects with both methods are similar for the first two post-treatment years, after which the estimated effects from BJS are larger. The largest effect happens three years after enrollment, with an estimated reduction in the deforestation rate of -0.0016.

Figure 3. Event study results for the deforestation rate



The average treatment effect on the treated provides a more informative measure of the magnitude of the program's effects on deforestation. The estimated effects from both methods are very similar (also similar to the OLS estimate), and they are all statistically significant (Table 3). Given that the baseline level of deforestation in our sample is low (which is also true for the whole country), the relative magnitude of the effect is large. When compared to the pre-2016 average level of within farm deforestation rate of 0.00152 (see Table 2), the estimated effects would imply an 87% reduction in the deforestation rate based on BJS and an 81% based on CS. With the average area of 69 hectares, this represents between 0.085 and 0.09 hectares of avoided deforestation per property per year.

Table 3. Deforestation Rate - Average Treatment Effects on the Treated

	Estimated coefficient	Std. error	p-value
OLS	-0.00122	0.00037	0.001
BJS - Post-treatment ATT	-0.00132	0.00034	0.000
CS - Post-treatment ATT	-0.00123	0.00045	0.024

To compare our results with the results from previous studies, we calculate the standardized effect based on the results from the BJS estimation, by dividing the estimated effect by the standard deviation of the deforestation rate for the comparison group. Our standardized effect is -0.12, which is within the range of all the values reported by Ferraro (2017) for all the studies on the effect of PES on deforestation in Costa Rica (ranging from -0.07 from Robalino & Pfaff [2013] to -0.18 from Robalino et al. [2008]). Importantly, Robalino & Pfaff (2013) focus on the first cohorts of the program (1997 to 2000), while Robalino et al. (2008) focus on the cohorts between 2000 to 2005. Robalino et al. (2008) find that there was an increase in the effect of the program when compared to the results for earlier cohorts, since for the earlier cohorts, properties were enrolled on a "first come first serve" basis and enrolled properties tended to be the ones where deforestation would have been low in absence of the program. Based on our results, it seems like the effect of the program has not increased after 2005 (although there are no studies for the cohorts between 2005 and 2016) and the changes to the prioritization criteria in 2016 do not seem to have led to a larger effect from the program on deforestation.

4.2 Forest Cover

To complement our analysis of the effects of the PES program on deforestation (the flow variable), we examine whether the program has any detectable effects on forest cover (the stock). Given that the magnitude of the effect of the program on deforestation is very small on absolute terms, we would not expect this to be picked up as an effect of the program. This is exactly what we find: the estimated effect of enrollment in the PES program on forest cover is zero for all the years after enrollment in the program (Figure 4). Furthermore, the average treatment effect on the treated, using both CS and BJS is also effectively equal to zero (Table 4). This does not mean that the program has no effect on forest cover, only that with our current data and given the magnitude of the effect of the program on deforestation, we cannot detect any effect on forest cover.

Figure 4. Event study results for forest cover

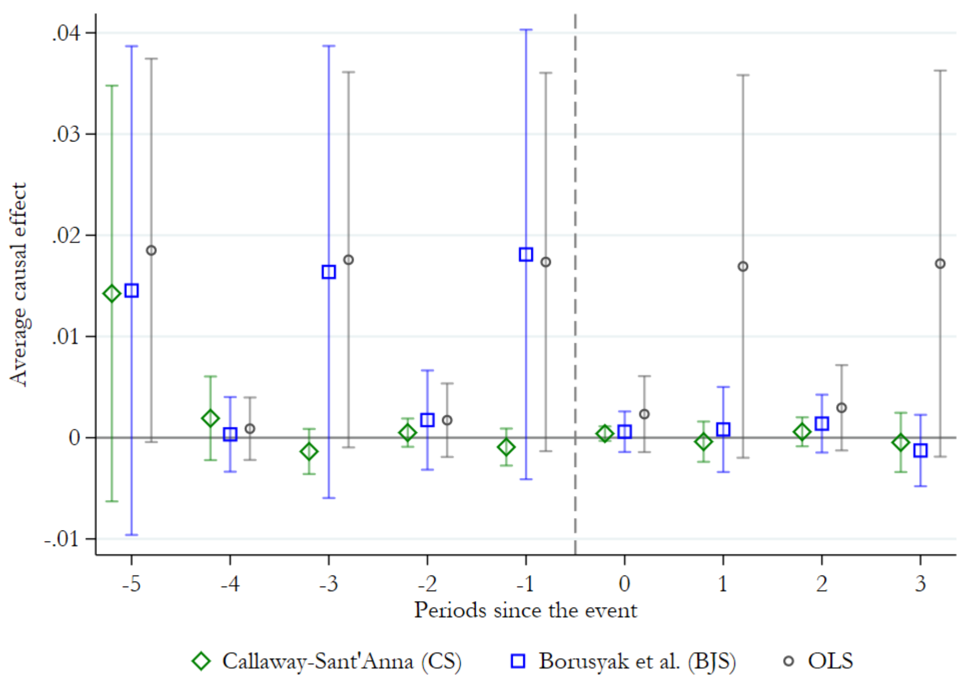


Table 4. Forest Cover - Average Treatment Effects on the Treated

	Estimated coefficient	Std. error	p-value
OLS	-0.00002	0.00085	0.981
BJS - Post-treatment ATT	0.00037	0.00091	0.689
CS - Post-treatment ATT	-0.00004	0.00072	0.959

4.3 Agricultural land uses

We also explore the effects from enrolling in the program on land with annual and perennial crops within the property. A common concern related to PES programs is that the new source of income (i.e., the payments) could relax a credit constraint and could be used to expand the agricultural frontier, by converting forestland to cropland. Conversely, the increased income could also be used to intensify agricultural production in the existing cropland, through investments in the current crops or through the shift to more economically valuable crops (such as perennial crops).

Related to the conversion of forestland to cropland, we have no evidence that this is the case in Costa Rica (Figure 5 and Table 5). However, we do find suggestive evidence that there is a shift from annual to perennial crops, given that the share of annual crops over the total land in the property decreases while there is a proportional increase in land with perennial crops. This is relatively small increase in the share of the total land dedicated to perennial crops (7% increase with respect to the mean), and evaluated at the mean property size, represents a shift of 0.1 hectares from annual to perennial crops.

Figure 5. Event study results for forest cover

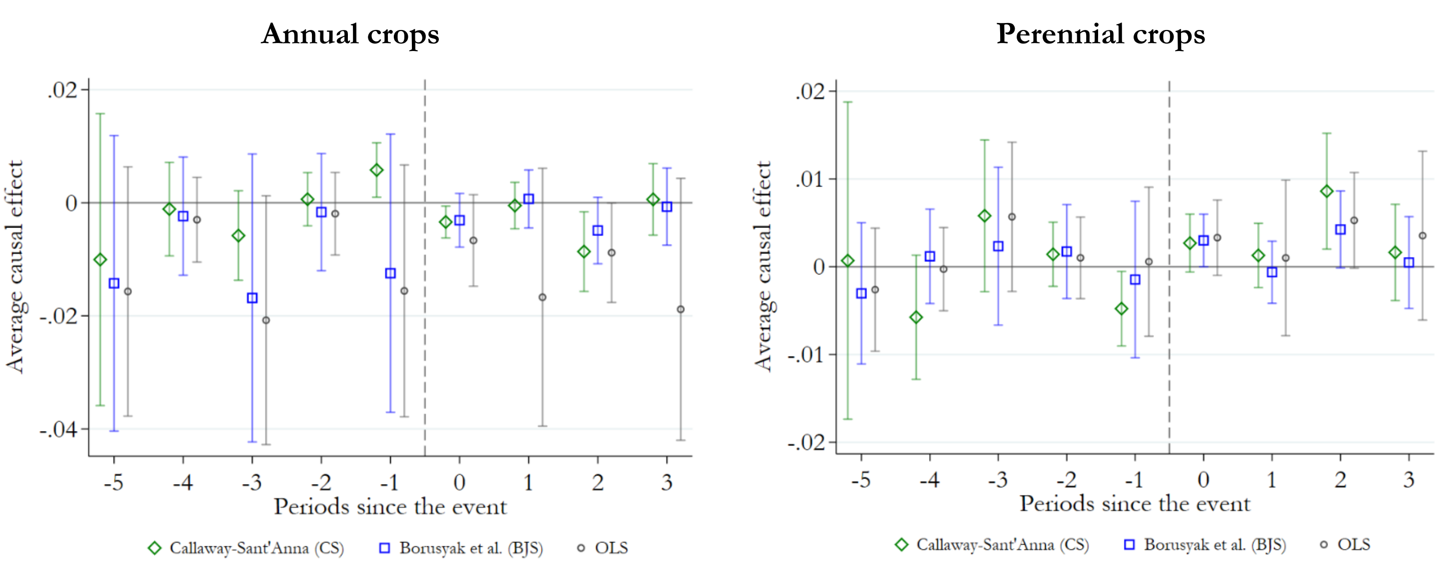


Table 5. Agricultural land uses - Average Treatment Effects on the Treated

Panel A - Annual crops

	Estimated coefficient	Std. error	p-value
OLS	-0.00083	0.00177	0.640
BJS - Post-treatment ATT	-0.00147	0.00174	0.396
CS - Post-treatment ATT	-0.00212	0.00180	0.239

Panel B - Perennial crops

	Estimated coefficient	Std. error	p-value
OLS	0.00033	0.00139	0.809
BJS - Post-treatment ATT	0.00134	0.00136	0.326
CS - Post-treatment ATT	0.00267	0.00163	0.103

5 Deforestation risk and heterogeneous effects

The effectiveness of the program depends on the additionality that it has: how many hectares of forest would have been lost in the absence of the program. If the program targets areas with a higher risk of deforestation, where presumably more hectares of forest would be lost in absence of the program, then the program will be much more effective (higher additionality).

As a thought exercise, imagine the program administrators had complete information and perfect foresight, such that based on the location of a property and its characteristics, they could perfectly predict which properties will have higher deforestation in the future in the absence of the program. If this was the case, then with this information, they would target those properties where deforestation will be relatively higher (relative to the pool of applicants), given that for every dollar spent, they would get more hectares of avoided deforestation. From this simple thought experiment, we can see that increasing the additionality of the program is to a large extent a prediction problem. We want to be able to know which properties are the ones where future deforestation will be highest, and we then want to select those properties to enroll in the program.

Given the predictive nature of this problem, we can use tools that are specifically designed to generate accurate predictions (such as machine learning models), to examine whether the program currently targets properties with a higher risk of deforestation and whether there are differential effects when focusing on the properties that the model would classify to be as high-risk. Thus, first we train a machine learning model to predict which properties are the most likely to have high levels of deforestation in the future, trying to replicate as closely as possible the thought experiment we proposed, and then estimate the share of participants and non-participants that are classified as high-risk properties. Secondly, we estimate the effects of the program focusing only on the treated properties that are classified as high-risk by the model, so that we can compare the estimated effects for these properties, to the effects we estimate for the whole sample.

To train the model, we take all the properties in our sample, and for each year, calculate the future two-year deforestation rate. For example, for each property in 2012, we sum the total deforestation in 2013 and 2014, and then divide by the area of the property to get the future two-year deforestation rate. We do this for every property and every year up to 2015. We then create a dummy variable that takes the value of 1 if the property has a two-year deforestation rate above the 90th percentile of all the properties in a given year (we call this variable "high deforestation risk"). We choose this cutoff since the two-year deforestation rate is zero for most properties, which means that for most years, the 75th percentile of the two-year deforestation rate is zero.

We train a gradient boosting model to classify properties as being either high-risk (1) or low-risk (0) of deforestation, using data from properties in 2012. We chose this model given that it frequently outperforms other models such as Random Forest and XGBoost (Bentéjac et al., 2021). We use 10-fold cross-validation to tune the three most important hyperparameters in the model: the learning rate, the optimal number of trees and the maximum tree depth (see section 8.1 in the Appendix for a brief description of the model and the gradient boosting algorithm). In terms of the predictors, we include variables that are associated with the potential suitability of the forest land for agriculture, given that the existing evidence shows that in general agriculture is the main driver of deforestation (Busch & Ferretti-Gallon, 2017; Pendrill et al., 2022). This is also true in Costa Rica, where between 2001 and 2022, shifting agriculture was the main driver for 93% of the hectares of forest loss, according to data from Global Forest Watch[5]. As such, we include data at the property level on elevation, slope, potential yields for 14 crops, rainfall, and temperature (both in the present and lagged up to two periods), travel time to the closest city with a population of 50 thousand residents, and then also Canton level population variables (total population, male and female).

With the hyperparameters from the cross-validation exercise, we create a model that predicts which properties could be classified as being of high deforestation risk. To assess the model's out of sample performance, we compare the predicted classification with the real classification for 2013 and 2014. We then calculate the recall rate, which is the ratio between the properties that are correctly classified as high deforestation risk and the total number of properties that are actually high deforestation (classified as 1 when they are actually 1), and the precision rate (also called true positive rate) as the ratio between the properties that are correctly classified as high risk and the total number properties classified as high risk (both the true and false positives). The recall rate then allows us to know what percentage of the high-risk properties the model correctly identified, whereas the precision rate allows us to know what percentage of the properties classified as high-risk, are actually high-risk. We find that for 2013, 46% of the properties are incorrectly classified as high deforestation risk and the value goes down to 36% in 2014 (Table 6). This means that if we based the targeting of the program solely on

[5] Global Forest Watch, 2014, World Resources Institute. Accessed on July 4, 2023.

the predictions generated by our model, we would get 45.6% of the high-risk properties in 2013 and 36% in 2014. Additionally, of the properties that the model classifies as high deforestation risk in both 2013 and 2014, 26% of these would actually be high-risk.

Table 6. Recall and precision rates of the classification model

	Recall	Precision
2013	45.6%	26.0%
2014	36.0%	26.0%

When evaluating the model's performance, it is important to keep two things in mind. The first is that our true outcome variable is by definition a rare event, since for a given year, only 10% of the properties are considered to be high risk. The rarer the event to be predicted, the harder it will be for the model to have higher recall and precision rates. Second, the objective is not to perfectly predict all the properties that we classified as being high-risk, based on an arbitrary cutoff, but to predict those properties where for a given year, future deforestation will be higher than what it is for the average property in our sample. In that regard, we believe that our model performs very well. To evaluate this, we calculate the average future two-year deforestation rate for all the properties that our model classifies as high-risk, and we compare it to that of the properties classified as low-risk by our model (Table 7). We see that for both years, the properties that are classified as high-risk have significantly higher future two-year deforestation rates (4x times the rate of the low-risk properties in 2013 and 3.3x times for 2014). These differences are statistically significant for both years.

Table 7. Future two-year deforestation rates based on the model's classification

	Low-risk	High-risk	Difference	t-statistic
2013	0.135%	0.547%	-0.412%	7.64***
2014	0.162%	0.530%	-0.368%	4.17***

Given the model's performance at identifying high-risk properties, we can use the model's prediction to examine whether the program is currently being successful at enrolling high-risk properties. To do this, we use the model to classify all the properties for every cohort as being high-risk or low-risk, and then calculate the share of properties for each cohort that are classified as high-risk. If the program is being successful at enrolling high-risk properties, then the share of high-risk properties would be higher in the participants' than in the non-participants group. We find that this is not the case (Table 8). For most of the cohorts, the share of properties that are classified as high-risk is the same for both groups, with the only exception being the cohort of 2019, when the share of high-risk properties in the participant's group is double that of the non-participants. Additionally, we also find that the average of the application scores of low and high-risk properties are not statistically different, as would be expected if the prioritization criteria were correlated with a measure of ex-ante deforestation risk, in which case the application scores of the high-risk properties would be higher (see Table 14 in the Appendix). Thus, we believe that currently the program is not successfully enrolling the properties that have the highest risk of deforestation.

Table 8. Share of participants by cohort that are classified as high-risk

Cohort	Non-participants	Participants
2016	19%	20%
2017	13%	9%
2018	10%	11%
2019	7%	14%

Finally, we can also explore whether the program's effects are different when only focusing on those properties that we classify as high-risk. If the effects are higher, then it would provide evidence of how targeting and successfully enrolling properties that are considered to be high-risk, can increase the program's additionality. We find that when limiting the treatment group only to the properties that are classified as high-risk by our model, and using the same control group as in our original analysis, the estimated effects increase when compared to the whole treatment group (Table 9), even though the results are no longer statistically significant (possibly because the treatment group is much smaller). Compared to our original results (Table 3), we see that when focusing only on those that we predict would have had higher deforestation, the estimated effects from BJS are 27% larger and the ones from CS are 73% larger. Given that average area of these sample of properties is larger (76.6 ha), this represents between 0.12 ha and 0.16 ha of avoided deforestation, an increase of between 33% and 78% with respect to the avoided deforestation for the whole sample.

Table 9. Estimated effects on deforestation with high-risk properties as treatment group

	Estimated coefficient	Std. error	p-value
BJS - Post-treatment ATT	-0.001608	0.00124	0.195
CS - Post-treatment ATT	-0.00213	0.00149	0.154

6 Cost-benefit analysis from avoided CO_2 emissions

The main benefit of the program is that it protects existing forestlands that would have otherwise been lost. These forestlands provide several environmental services, but the main one is that by keeping these trees standing, the carbon that is contained in their biomass, both below and above ground, is prevented from going into the atmosphere when these trees are cut down. We can then calculate the approximate amount of avoided carbon emissions and use this to evaluate whether the benefits from these avoided emissions exceed the total payments required to achieve this result. To do this, we break down the analysis into two different steps.

First, we define the amount of avoided carbon emissions associated with the avoided deforestation. To define the carbon content per hectare of forest in Costa Rica, we use the values from Saatchi et al., (2011), which provide the estimated carbon content per hectare of forest for several countries and have estimated the mean value for Costa Rica to be 108 tons of carbon per hectare of forest[6]. Given that a molecule of CO_2 is 3.67 times heavier than a carbon atom, this is equivalent to 396 tons of CO_2 per hectare of forest (Table 15 in the Appendix).

Our estimates show that the reduction in deforestation caused by the program is equivalent to 0.09 hectares of forest per enrolled property, which implies that on average every enrolled property represents 36 tons of avoided CO_2 emissions. However, these emissions do not occur instantaneously, and so we assume that there is an average lag of 10 years between the time when the trees are felled and the carbon in them is released into atmosphere, which comes from assuming that to 45% of the biomass is burned in the first year, 45% decays over 15 years (Hérault et al., 2010) and the remaining 10% is stored in wood for 30 years (this is a similar assumption as the one made by Jayachandran, De Laat, et al., 2017). Given that the value of avoided present emissions is higher than that of future emissions, we use a 2% discount rate to bring that future stream of emissions into the present (Nesje et al., 2023; Sarofim & Giordano, 2018).

Second, we calculate the costs associated with preventing these emissions. We have that on average each enrolled property has an area under contract of 39 hectares (the size of the area that is legally being protected under the contract signed between the beneficiary and FONAFIFO). With an average payment of $54 USD/ha, the payments made to each beneficiary amount to $2,106 USD/year.

Thus, the estimated cost from avoided CO_2 emissions is $71 USD/MT CO_2 (Table 10, Panel A). This cost can then be compared to the benefits from the avoided CO_2 emissions, which have been defined in three different ways (Table 10, Panel B). One is based on the social cost of carbon (SCC), which is an estimate of the economic damages done by the emission of an additional unit of carbon. These are estimated using Integrated Assessment Models (IAMs), which use climatic models to estimate the climate's response to changes in emissions, which are in turn used as inputs in an economic damages

[6] We use the data provided in Table 3 from the "Supplementary Material" from Saatchi et al., (2011), which can be found in https://www.pnas.org/doi/full/10.1073/pnas.1019576108#supplementary-materials.

function, to estimate the damages from emissions. Given all the modelling choices, the estimated SCCs can vary greatly, although there has been a documented upward trend in the published estimates (Rennert et al., 2022; Tol, 2023).

Another way in which the benefits from avoided CO_2 emissions can be valued is based on a target-consistent pricing approach, in which an implicit carbon price is estimated at any point in time, based on an emissions reduction target (such as the one set by the Paris Agreement). These are based on estimated marginal abatement cost curves, which are used to estimate the implicit price of CO_2 at any point in time that will lead to a reduction in CO_2 emissions that is consistent with a given target (High-Level Commission on Carbon Prices, 2017; Stern et al., 2022). The third one is based on survey-elicited experts' opinions on the probabilities of extreme outcomes associated with climate change and the reductions in emissions needed to avert them (Pindyck, 2019). The SCC is then estimated as the ratio of the losses to the reduction in CO_2 emissions needed to avert them.

We calculate the benefit-cost ratios based on what we believe are the best estimates of the benefits from the avoided CO_2 emissions (Table 10, Panel C). We find that in all cases the benefit-cost ratio is greater than one (the benefit from the avoided emissions is higher than the costs of achieving them). To the best of our knowledge, this is the first time that a cost-benefit analysis has been performed for this program. Jayachandran, De Laat, et al., (2017) performed a similar analysis for a PES program in Uganda and found that the estimated cost per ton of avoided CO_2 was much lower ($0.46 per averted MT of CO_2), which then implies that the benefit-cost ratios are much higher for their program.

Table 10. Cost-benefit analysis per ton of avoided CO_2

Panel A. Cost of avoided CO_2

	Value	**Measurement Units**
Avoided deforestation per property	0.09	ha
Present equivalent of avoided CO2 emissions per property	29.5	MT
Total average payment per contract/year (USD)	$2,106	USD$
Amount paid per ton of CO2 avoided	$71	USD/MT CO_2

Panel B. Benefits of avoided CO_2

Rennert et al. (2023) - 2% discount rate	$185	USD (2020)/MT CO_2
Pindyck, 2019	$90	USD (2020)/MT CO_2
High-Level Commission on Carbon Prices, 2017	$80	USD (2020)/MT CO_2

Panel C. Benefit-cost ratios

Rennert et al. (2023) - 2% discount rate	2.59
Pindyck, 2019	1.26
High-Level Commission on Carbon Prices, 2017	1.12

7 Conclusions

Costa Rica's PES program is one of the oldest country wide PES programs in the world. The evaluation of its effectiveness in the early years of the program showed that the program had little to no effect on deforestation (Arturo et al., 2007; Robalino & Pfaff, 2013). However, in regions where the threat of deforestation was higher and there was active targeting of participants, evidence suggests that the program increased forest cover (Arriagada et al., 2012). There is also evidence showing that the effect of the program on deforestation increased for the later cohorts (Robalino et al., 2021).

Ours is the first study of the latest cohort of participants of the PES program (2016-2020) and is also the first one to use actual applicants to the program as a control group. Using an event study design, we find that there is a statistically significant decrease in the deforestation rate after enrollment in the program and the relative magnitude of the effect is large compared to the baseline level of deforestation, with the estimated effect representing an 87% decrease in the deforestation rate. However, in terms of the total area of avoided deforestation, this amounts to 0.09 hectares per property per year. Given this small effect, we do not find that there is a detectable effect of the program on forest cover, but we do find suggestive evidence that there is a shift from annual to perennial crops.

We find that there is a potential to leverage the predictive power from machine learning models to increase the additionality of the program. We train a model to predict which are the properties with relatively higher deforestation risk and find that currently the program is not disproportionally enrolling these high-risk properties, which is what one would expect to see if the program was successfully targeting properties with an ex-ante higher risk of deforestation. We find that if the program focused on the properties that are predicted to be high-risk, the additionality of the program could be higher, given that the effect of the program is between 27% and 73% higher when only focusing on these predicted high-risk properties, which would represent between 0.12 ha and 0.16 ha of avoided deforestation.

Finally, we evaluated whether the benefits from the avoided CO_2 emissions exceed the costs of the program. We find that that this is indeed the case and that the benefit-cost ratio could be higher if the program targeted more high-risk properties (increasing its additionality). Given that the avoided deforestation could increase between 0.12 ha up and 0.16 ha when focusing only on high-risk properties, the expected cost per ton of avoided CO_2 could then decrease to $54-$41 (Table 16 in the Appendix), which represents a reduction of up to 42% in the cost per ton of avoided CO_2. This is highly relevant in this case given the possibility that the source of funding for the program will decrease in the future and that the cost per avoided metric ton of CO_2 is much higher than the only other available evaluation of a PES program (in Uganda).

Given these results, we believe that future work should aim to understand how the program can increase the enrollment of more high-risk properties, since this will be critical to increasing the additionality and cost-effectiveness of the program in the future. Additionally, the program presents an opportunity to evaluate the other environmental benefits provided by forests in Costa Rica, such as its effects on air and water quality.

References

Arriagada, R. A., Ferraro, P. J., Sills, E. O., Pattanayak, S. K., & Cordero-Sancho, S. (2012). Do Payments for Environmental Services Affect Forest Cover? A Farm-Level Evaluation from Costa Rica. *Land Economics*, *88*(2), 382–399. https://doi.org/10.3368/le.88.2.382

Arturo, G., Sanchez-Azofeifa, G. A., Pfaff, A., Robalino, J. A., & Boomhower, J. P. (2007). Costa Rica's Payment for Environmental Services Program: Intention, Implementation, and Impact. *Conservation Biology*, *21*(5), 1165–1173. https://doi.org/10.1111/j.1523-1739.2007.00751.x

Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, *54*(3), 1937–1967. https://doi.org/10.1007/S10462-020-09896-5/TABLES/12

Borusyak, K., Xavier, J., & Spiess, J. (2021). *Revisiting Event Study Designs: Robust and Efficient Estimation*. https://arxiv.org/abs/2108.12419v4

Busch, J., & Ferretti-Gallon, K. (2017). What Drives Deforestation and What Stops It? A Meta-Analysis. *Review of Environmental Economics and Policy*, *11*(1), 3–23. https://doi.org/10.1093/reep/rew013

Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-Differences with multiple time periods. *Journal of Econometrics*, *225*(2), 200–230. https://doi.org/10.1016/J.JECONOM.2020.12.001

Chen, J., Lærke, P. E., & Jørgensen, U. (2022). Land conversion from annual to perennial crops: A win-win strategy for biomass yield and soil organic carbon and total nitrogen sequestration. *Agriculture, Ecosystems & Environment*, *330*, 107907. https://doi.org/10.1016/J.AGEE.2022.107907

Ferraro, P. J. (2017). Are payments for ecosystem services benefiting ecosystems and people? In *Effective Conservation Science: Data Not Dogma* (pp. 159–166). Oxford University Press. https://doi.org/10.1093/oso/9780198808978.003.0025

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O., & Townshend, J. R. G. (2013). High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science*, *342*(6160), 850–853. https://doi.org/10.1126/science.1244693

Hérault, B., Beauchêne, J., Muller, F., Wagner, F., Baraloto, C., Blanc, L., & Martin, J. M. (2010). Modeling decay rates of dead wood in a neotropical forest. *Oecologia*, *164*(1), 243–251. https://doi.org/10.1007/S00442-010-1602-8/FIGURES/3

High-Level Commission on Carbon Prices. (2017). *Report of the High-Level Commission on Carbon Prices I*. www.carbonpricingleadership.org.

IDB. (2020). *Hacia una Economía Verde: Apoyo al Plan de Descarbonización de Costa Rica*.

IIASA, & FAO. (2012). *Global Agro-ecological Zones Model Documentation*.

Jayachandran, S., De Laat, J., Lambin, E. F., Stanton, C. Y., Audy, R., & Thomas, N. E. (2017). Cash for carbon: A randomized trial of payments for ecosystem services to reduce deforestation. *Science*, *357*(6348), 267–273. http://science.sciencemag.org/

Jayachandran, S., de Laat, J., Lambin, E. F., Stanton, C. Y., Audy, R., & Thomas, N. E. (2017). Supplementary Material - Cash for carbon: A randomized trial of payments for ecosystem services to reduce deforestation. *Science (New York, N.Y.)*, *357*(6348), 267–273. https://doi.org/10.1126/science.aan0568

Means, M., Crews, T., & Souza, L. (2022). Annual and perennial crop composition impacts on soil carbon and nitrogen dynamics at two different depths. *Renewable Agriculture and Food Systems*, *37*(5), 437–444. https://doi.org/10.1017/S1742170522000084

Nelson, A. (2008). *Estimated travel time to the nearest city of 50,000 or more people in year 2000*. Global Environment Monitoring Unit - Joint Research Centre of the European Commission. https://forobs.jrc.ec.europa.eu/products/gam/

Nesje, F., Drupp, M. A., Freeman, M. C., & Groom, B. (2023). Philosophers and economists agree on climate policy paths but for different reasons. *Nature Climate Change*, *13*(6), 515–522. https://doi.org/10.1038/s41558-023-01681-w

Ordóñez, P. J., Baylis, K., & Ramírez, M. I. (2023). Land cover change effects from community forest management in Michoacán, Mexico. *Environmental Research Letters*, *18*(6), 065008. https://doi.org/10.1088/1748-9326/ACCDEF

Pattanayak, S. K., Wunder, S., & Ferraro, P. J. (2010). Show Me the Money: Do Payments Supply Environmental Services in Developing Countries? *Review of Environmental Economics and Policy*, *4*(2), 254–274. https://doi.org/10.1093/reep/req006

Pendrill, F., Gardner, T. A., Meyfroidt, P., Persson, U. M., Adams, J., Azevedo, T., Lima, M. G. B., Baumann, M., Curtis, P. G., De Sy, V., Garrett, R., Godar, J., Goldman, E. D., Hansen, M. C., Heilmayr, R., Herold, M., Kuemmerle, T., Lathuillière, M. J., Ribeiro, V., … West, C. (2022). Disentangling the numbers behind agriculture-driven tropical deforestation. *Science*, *377*(6611). https://doi.org/10.1126/SCIENCE.ABM9267/SUPPL_FILE/SCIENCE.ABM9267_SM.PDF

Pindyck, R. S. (2019). The social cost of carbon revisited. *Journal of Environmental Economics and Management*, *94*, 140–160. https://doi.org/10.1016/J.JEEM.2019.02.003

Rennert, K., Errickson, F., Prest, B. C., Rennels, L., Newell, R. G., Pizer, W., Kingdon, C., Wingenroth, J., Cooke, R., Parthum, B., Smith, D., Cromar, K., Diaz, D., Moore, F. C., Müller, U. K., Plevin, R. J., Raftery, A. E., Ševčíková, H., Sheets, H., … Anthoff, D. (2022). Comprehensive evidence implies a higher social cost of CO2. *Nature*, *610*(7933), 687–692. https://doi.org/10.1038/s41586-022-05224-9

Ritchie, H., & Roser, M. (2021). *Forests and Deforestation*. Published Online at OurWorldInData.Org.

Robalino, J., & Pfaff, A. (2013). Ecopayments and deforestation in Costa Rica: A nationwide analysis of PSA's initial years. *Land Economics*, *89*(3), 432–448. https://doi.org/10.3368/le.89.3.432

Robalino, J., Pfaff, A., Sánchez, G., Alpízar, F., León, C., & Rodríguez, C. (2008). Deforestation impacts of environmental services payments: Costa Rica's PSA Program 200-2005. In *Discussion paper series (EfD)* (EfD DP 08-24; Environment for Development Discussion Paper Series). JSTOR.

Robalino, J., Pfaff, A., Sandoval, C., & Sanchez-Azofeifa, G. A. (2021). Can we increase the impacts from payments for ecosystem services? Impact rose over time in Costa Rica, yet spatial variation indicates more potential. *Forest Policy and Economics*, *132*, 102577. https://doi.org/10.1016/J.FORPOL.2021.102577

Roth, J., Sant'Anna, P. H. C., Bilinski, A., & Poe, J. (2023). What's trending in difference-in-differences? A synthesis of the recent econometrics literature. *Journal of Econometrics*, *235*(2), 2218–2244. https://doi.org/10.1016/J.JECONOM.2023.03.008

Saatchi, S. S., Harris, N. L., Brown, S., Lefsky, M., Mitchard, E. T. A., Salas, W., Zutta, B. R., Buermann, W., Lewis, S. L., Hagen, S., Petrova, S., White, L., Silman, M., & Morel, A. (2011). Benchmark map of forest carbon stocks in tropical regions across three continents. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(24), 9899–9904. https://doi.org/10.1073/PNAS.1019576108/-/DCSUPPLEMENTAL

Sant'Anna, P. H. C., & Zhao, J. (2020). Doubly robust difference-in-differences estimators. *Journal of Econometrics*, *219*(1), 101–122. https://doi.org/10.1016/J.JECONOM.2020.06.003

Sarofim, M. C., & Giordano, M. R. (2018). A quantitative approach to evaluating the GWP timescale through implicit discount rates. *Earth System Dynamics*, *9*(3), 1013–1024. https://doi.org/10.5194/ESD-9-1013-2018

Snilsveit, B., Stevenson, J., Langer, L., Tannous, N., Ravat, Z., Nduku, P., Polanin, J., Shemilt, I., Eyers, J., & Ferraro, P. J. (2019). Incentives for climate mitigation in the land use sector—the effects of payment for environmental services on environmental and socioeconomic outcomes in low- and middle-income countries: A mixed-methods systematic review. *Campbell Systematic Reviews*, *15*(3), e1045. https://doi.org/10.1002/CL2.1045

Stern, N., Stiglitz, J., Karlsson, K., & Taylor, C. (2022). *A Social Cost of Carbon Consistent with a Net-Zero Climate Goal*.

Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, *225*(2), 175–199. https://doi.org/10.1016/J.JECONOM.2020.09.006

Tol, R. S. J. (2023). Social cost of carbon estimates have increased over time. *Nature Climate Change*, *13*(6), 532–536. https://doi.org/10.1038/s41558-023-01680-x

Wunder, S., Börner, J., Ezzine-de-Blas, D., Feder, S., & Pagiola, S. (2020). Payments for Environmental Services: Past Performance and Pending Potentials. *Annual Reviews*, *12*, 209–234. https://doi.org/10.1146/ANNUREV-RESOURCE-100518-094206

8 Appendix

Table 11. Countries with similar per capita GDP and deforestation rates*

Country	GDP per capita, PPP (2021 - current USD $)	Total Population - 2021	Forest area (2021 - ha)	Forest Loss (2021 - ha)	Forest Loss Rate (2021 - based on forest area)
Albania	15,533	2,811,666	788,900	1,686	0.21%
Antigua and Barbuda	21,959	93,219	8,053	39	0.49%
Barbados	15,178	281,200	6,300	17	0.27%
China	19,484	1,412,360,000	221,857,750	536,656	0.24%
Colombia	17,105	51,516,562	58,942,627	265,168	0.45%
Costa Rica	22,643	5,153,957	3,051,240	7,848	0.26%
Ecuador	11,773	17,797,737	12,433,560	30,273	0.24%
El Salvador	10,143	6,314,167	579,380	1,489	0.26%
Equatorial Guinea	16,151	1,634,466	2,440,060	8,906	0.36%
Jamaica	10,601	2,827,695	600,773	1,485	0.25%
Mexico	19,578	126,705,138	65,564,317	189,079	0.29%
Moldova	15,010	2,615,199	386,500	1,131	0.29%
North Macedonia	18,344	2,065,092	1,001,490	2,304	0.23%
Peru	13,831	33,715,471	72,157,533	225,204	0.31%
Serbia	21,647	6,834,326	2,722,940	4,072	0.15%
South Africa	14,689	59,392,255	17,013,690	52,067	0.31%
St. Lucia	14,396	179,651	20,770	35	0.17%
Trinidad and Tobago	25,421	1,525,663	227,770	316	0.14%
Tunisia	11,471	12,262,946	704,260	2,079	0.30%
Turks and Caicos Islands	21,803	45,114	10,520	43	0.41%
Total		1,746,131,524	460,518,433	1,329,898	0.29%
Share of World total		22%	11%	5.3%	

* Lighter shading denotes countries in Latin America.

Table 12. Time from beginning of year to application date

Cohort	Number of days between beginning of the year and date of application			
	Nonparticipants	**Participants**	**Difference**	**t-statistic**
2016	45.4	40.5	5.0	3.15***
2017	37.7	39.6	-1.9	0.91
2018	49.6	53.4	-3.8	1.68
2019	57.4	54.5	2.9	1.45
2020	54.1	-	-	-
Total	49.6	46.5	3.0	3.02***

Table 13. Characteristics by cohort

	Cohorts				
	2016	2017	2018	2019	2020
Application score	128.15	126.17	103.83	106.46	117.05
Deforestation in ha (pre-2016)	0.11	0.09	0.07	0.12	0.11
Deforestation Rate (pre-2016)	0.163%	0.136%	0.096%	0.200%	0.160%
Forest Area 2015 (ha)	60.11	65.60	54.14	49.70	51.77
Forest Area 2015 (% of total area)	0.84	0.85	0.82	0.79	0.78
Total crops area 2015 (ha)	1.65	1.40	1.99	1.44	2.56
Annual crops 2015 (ha)	0.53	0.46	0.55	0.39	0.65
Perennial crops 2015 (ha)	1.12	0.94	1.44	1.05	1.92
Share annual crops 2015 (% total)	0.79%	0.69%	0.55%	0.89%	1.10%
Share perennial crops 2015 (% total)	1.57%	1.15%	2.15%	2.00%	2.66%
Area (ha)	70.72	77.34	66.36	63.49	67.40
Proposed area under PES (ha)	39.38	45.33	39.50	36.13	35.75
Elevation (m.a.s.l)	668.87	617.10	616.05	626.85	496.12
Slope (degrees)	13.06	12.46	12.76	11.88	11.21
Potential yield – Maize (kg/ha)	639.53	711.04	635.30	717.09	833.62
Potential yield – Sugarcane (kg/ha)	1,105.72	1,244.00	1,175.98	1,313.92	1,449.04
Potential yield – Wheat (kg/ha)	16.81	14.18	14.56	12.97	15.43
Potential yield – Citrus (kg/ha)	254.13	242.15	297.31	302.26	214.25
Potential yield – Coffee (kg/ha)	296.71	313.60	297.61	335.22	374.54
Travel time (min)	347.76	346.74	330.12	331.71	345.69
Mean annual rainfall (mm)	2,938.64	3,041.14	2,995.22	2,931.78	2,966.51
Mean temperature (Celsius)	25.96	26.06	26.19	26.20	26.79

Table 14. Application score by classification of deforestation risk

Cohort	Low-risk	High-risk	Difference	t-statistic
2016	129.0	123.8	5.2	3.51***
2017	126.2	124.3	1.8	0.86
2018	108.6	114.9	-6.3	-0.98
2019	106.1	107.9	-1.8	-0.29
All cohorts	117.9	119.6	-1.7	-0.81

8.1 Tuning the Hyperparameters for the Gradient Boosting Model

The model is based on weak learners (decision trees in this case), that are recursively trained to reduce the prediction error. The Gradient Boosting algorithm proceeds as follows:

1. A weak learner creates a prediction that defines the starting point for the optimization. For a classifier, this will typically be the ratio of positive to negative values (high-risk to low-risk, in our case).
2. Based on this prediction, it will calculate the residuals (the difference between the ground truth and the predicted value → $R_1 = Y - \lambda Pred_1$). This is equivalent to minimizing a log-loss function.
3. Using the residuals from the previous step as the outcome variable, it will then fit a new classifier tree, using the same features (independent variables).
4. The prediction is updated as the sum of the original prediction and the new prediction, which is adjusted by a hyperparameter called the learning rate (λ). This controls the weight given to each new prediction as is added to generate the overall prediction.
5. The residuals are then recalculated, based on the updated prediction ($R_2 = Y - \lambda Pred_1 - \lambda Pred_2$).
6. The process is then repeated *N* times, each time adding the new prediction (adjusted by the learning rate) to the previous prediction and then updating the residuals.

Thus, there are three hyperparameters that are usually tuned when using Gradient Boosting Models for classification. First, since the weak learners are decision trees, we need to tune one of the hyperparameters that controls one of the main characteristics of the estimated trees: the maximum tree depth. This hyperparameter controls the maximum number of nodes in each tree[7] and thus controls the complexity of the tree. Second, we also tune the total number of trees that will be trained by the model, which determines how many times the predictions are updated and added to the initial prediction (the *N* from step 6 above). Third, we tune the learning rate (λ from step 4 above), which is the weight given to each new prediction. To tune these three hyperparameters, we did a 10-fold cross-validation in a grid search approach, where we evaluated a total of 120 combinations of these 3 hyperparameters.

[7] The maximum number of nodes in the tree is equal to 2 to the power of the depth. So, for a tree with a maximum depth of 5, the maximum number of terminal (i.e., leaf) nodes is $2^5 = 32$.

8.2 Cost-benefit analysis

Table 15. Main values used in CBA

	Value	Measurement Units
Carbon content per hectare of forest	108	Metric tons (MT)/ha
Carbon to CO_2 conversion factor	3.67	
CO_2 emissions per hectare of forest	396	MT/ha
Average property size	69	ha
Avoided deforestation per property	0.09	ha
Time from tree loss to CO_2 emissions	10.2	Years
Discount rate for future emissions	2%	%/year
Present Avoided CO_2 emissions per property	29.5	MT
Average contract size per property (ha)	39	ha
Payment per ha in PES (USD)	$54	USD$
Total average payment per contract/year (USD)	$2,106	USD$

Table 16. Sensitivity analysis of the cost of avoided CO2 emissions

Parameter	Alternative parameter value	Cost per ton of CO2 avoided (USD $)
Avoided deforestation per enrolled property (ha)	0.09	71
	0.12	54
	0.16	41
Discount rate	0%	58
	3%	79
	5%	96
Time from tree loss to CO_2 emissions (years)	0	58
	5	64
	15	79
Carbon content of forest (MT/ha)	96	80
	108	71
	119	65