Practical guidance for conservation impact evaluation using remotely sensed data

Alberto Garcia and Robert Heilmayr

July 20, 2020

Abstract

Conservation practitioners need rigorous evidence measuring the effectiveness of proposed policy interventions. In response, scientists are increasingly combining methods of impact evaluation with remotely sensed data on land use change to assess conservation effectiveness. Here we review this burgeoning literature to develop practical guidance for the design of econometric models quantifying conservation policy effectiveness. Using Monte Carlo simulations and analytical proofs, we demonstrate that many of the models employed for conservation impact evaluation suffer from significant bias - the significance, magnitude and even direction of estimated effects from many studies may be incorrect. These errors threaten to undermine the evidence base that is increasingly used to inform conservation policy adoption. To address this concern, we provide clear guidance to help scientists minimize the bias of their impact evaluations by carefully designing the structure of their econometric model, their unit of observation and their method and scale of data aggregation.

Introduction

The founding goal of conservation biology is to provide principles and tools to preserve biological diversity (Soulé 1985). To live up to this goal, scientists must generate causal evidence detailing the effectiveness of conservation interventions (Williams, Balmford, and Wilcove 2020). Such evidence is critical for practitioners who grapple with challenging questions of cause and effect. Do marine protected areas stop unsustainable harvesting of fish? Can payments for ecosystem services encourage lasting reforestation? When successful, conservation science provides answers that improve the way society confronts environmental challenges. However, inappropriate methods can yield misleading conclusions and, as a result, risk diverting scarce financial and political resources from the most effective conservation strategies.

Increasingly, conservation science has turned to econometric methods of impact evaluation to disentangle causal relationships (Butsic et al. 2017; Baylis et al. 2016). While randomized experiments are the gold standard for scientific discovery in both the natural and social sciences (Edwards et al. 2020; Jayachandran et al. 2017), conservation often poses questions that are prohibitively expensive, unethical or impossible to pursue through experimentation. In such settings, a growing portfolio of statistical techniques enable researchers to draw causal conclusions using observational data (Larsen, Meng, and Kendall 2019). These methods generate accurate estimates of an intervention's impact by comparing observed outcomes to a rigorous counterfactual of what would have happened in the absence of an intervention (Ferraro 2009; Meyfroidt 2016; Ferraro, Sanchirico, and Smith 2019; Ribas et al. 2020). Importantly, these methods account for the non-random assignment of interventions that often confound identification of causal relationships; a low rate of deforestation within a remote protected area may reflect the protected area's effectiveness, or it may be indicative of the remote location's poor suitability for agricultural development (Andam et al. 2008; Pfaff et al. 2009). A diversity of research designs such as regression discontinuity, synthetic control, matching, instrumental variables and event studies have been used to generate statistical counterfactuals of conservation interventions (Butsic et al. 2017). These approaches frequently build upon panel data settings in which units are observed repeatedly through time, enabling researchers to observe changes in outcomes after the adoption of an intervention (Blackman 2013). When a rigorous research design is applied to panel data,

1

observational studies can yield conclusions that are comparable to what a researcher would discover if they were able to run a randomized experiment (Ferraro and Miranda 2017).

Conservation impact evaluation increasingly makes use of panel data thanks, in large part, to the proliferation of remote sensing (Blackman 2013; Jones and Lewis 2015). For example, NASA's landsat missions provide detailed and consistent information on land use spanning the entirety of the world since the 1970s (Hansen and Loveland 2012). As a result, a scientist hoping to quantify the impacts of a land use policy adopted decades ago can assemble data for treated and control units that span both pre- and post-implementation periods (Jain 2020). New sensors and methods allow for similar time series tracking changes in air pollution, fires, or animal movement [XX].

However, many of these remotely sensed metrics have structural differences from the data used in more traditional applications of causal inference. These differences include the underlying error structure of generated data products and unique properties of the ecological processes being studied. Here we demonstrate that, as a result of these differences, many of the econometric models used in this growing literature are likely biased - significance, magnitude and even direction of estimated effects might be incorrect. These biases arise even when researchers follow common guidance to adopt "rigorous" research designs with valid counterfactuals (Blackman 2013; Jones and Lewis 2015). Based on a review of the existing literature, we identify key model design decisions that researchers need to make. Using Monte Carlo simulations and analytical proofs, we show how these design decisions affect the validity of causal inference in these settings. Lastly, we illustrate ways in which scientists can tailor these design decisions to minimize bias and better understand the implications of their impact evaluations.

Key considerations for impact evaluations of deforestation

We focus on the case in which a researcher would like to quantify the impact that an intervention has had on deforestation rates. We assume that the intervention has clearly defined boundaries (e.g. a protected area, certified concession, or indigenous territory), and that the researcher has access to spatially explicit observations of forest cover and forest loss spanning the periods before and after the intervention was adopted. The fundamental problem of causal inference is that, for every treated unit, we fail to observe the value that the outcome would have taken in the absence of treatment (Holland 1986). Figure 1 displays this problem in the context of our simulated conservation intervention — Panel A depicts the landscape as observed by the researcher at the end of the observation period, while Panel B depicts the unobservable counterfactual of what would have happened if the conservation intervention had not been adopted. The researcher's goal is to measure the Average Treatment Effect on the Treated (ATT), which quantifies the avoided deforestation occurring inside treated units. This general setting describes a broad array of research studies that apply panel methods to remotely sensed data (Table 1).
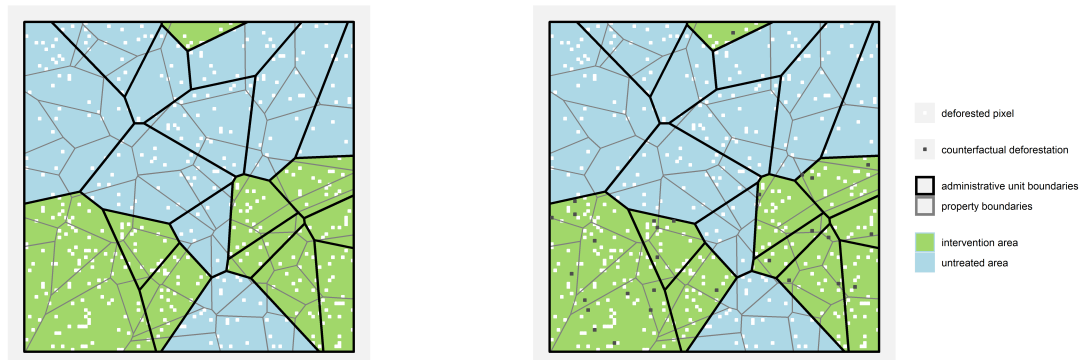

Figure 1: Left: deforestation observed in the period after the policy has been implemented in the intervention area; Right: deforestation that would have ocurred in the absence of the conservation intervention at the same point in time. Note that deforestation in the control area is the same in both cathe realized and counterfactual cases, since no intervention occurred.

In panel data settings, two broad methods are often used to measure the impact of conservation interventions: Difference-in-Differences (DID) and Two-way Fixed Effects (TWFE) regression models (Blackman 2013; Jones and Lewis 2015). The typical DID regression model includes a dummy variable equal to one for units in the treatment group, a dummy variable equal to one for observations in the period after the intervention, and their interaction. Conceptually, the DiD estimator calculates the treatment effect as the difference between the differences of the treated and untreated observations before and after treatment (Butsic et al. 2017). The DID estimator identifies the ATT under one main assumption, known as the common trends assumption. It amounts to assuming that both units in the intervention area and untreated units would have experienced the same average change in the outcome in the absence of the intervention. While fundamentally untestable, ensuring that deforestation rates in the intervention area and the control area followed parallel trajectories prior to the date of the intervention can give creedence to this assumption (Butsic et al. 2017).

Figure 2 shows the intuition behind the DID estimator in two time periods. Deforestation rates in the control area changed between the first and second period. The common trends assumption amounts to assuming that the deforestation rate in the intervention area would have changed by the same amount in the absence of the intervention. In our example, the DID estimator is the difference between post and pre-treatment deforestation rates in units with the intervention, minus the difference between post and pre-treatment deforestation rates in units without.
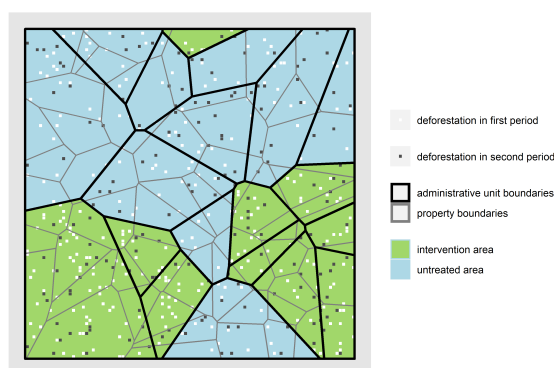

Figure 2: Left: deforestation observed in the first period; Right: deforestation observed in the second period. Note that pixels deforested in the first period are still observed as deforested in the second

TWFE regression models are often used to apply DID methods to multiple groups or treatment periods. This amounts to estimating a regression that controls for unit and time fixed effects. Intuitively, this can be thought of as including a dummy variable for each unit of analysis and each time period. The fixed effects account for any unobservable confounding variables that may vary across units or through time. When the treatment effect is constant across groups and over time, TWFE regressions estimate the *ATT* under the standard common trends assumption (**???**). Because TWFE regression models are often used to generalize the DID method, they are used in a wider variety of settings. Settings in which units undergo treatment in more than two distinct time periods may be amenable to TWFE regression, but not the standard DID method. For example, a researcher may use an TWFE regression model to examine the effectiveness of a network of protected areas, where the protected areas where created at different times or a payment for ecosystem services (PES) program, which enrolls properties in annual cohorts. In the case of two groups and two time periods, the TWFE regression should give an estimate equivalent to that of the DID model (Wooldridge, n.d.). This flexibility has led to TWFE regression models becoming commonplace in the literature.

Key model parameterizations in the literature

Unit of analysis Analyses using a binary pixel or plot are common in the literature (Jones and Lewis 2015). One benefit of a pixel level analysis is that results can be interpreted directly as the average effect across the landscape (Alix-Garcia and Gibbs 2017). This may be preferable in contexts where there is no clear

alternative unit of interest. For example, Anderson et al. (2018) explore how overlapping land allocations impact deforestation. Here, the estimated effect from a pixel-level analysis is more intuitive than would be generated from an "overlap" level analysis, where the coefficient could be interpreted as the average effect on a representative "overlap". Pixel level analyses often times requires the researcher to sample points in the interest of computational feasibility (e.g. Alix-Garcia et al. 2018; Anderson et al. 2018).

Sampling pixels from the landscape may ignore important spatial dependencies or relationships that would be captured with aggregation. Some papers, as a result, choose to aggregate pixels to the grid level. This maintains the landscape-scale interpretation provided by pixel level analyses, while allowing the author to include every pixel in the analysis within a manageable number of grid cells. In exchange for fine-scale spatial specificity, using coarser-resolution cells rather than pixels has the benefits of diluting the effects of possible spatial misalignments between datasets, enabling easier interpolation of missing data within cells, and subsuming localized spatial correlation (**???**). Although not necessarily an issue, one consideration of using grid cell aggregation is the potential for the grid cell to overlap two or more administrative units. In this case, the treatment variable will be continuous, or the researcher must make a decision as to how treatment should be assigned.

Several other papers choose to conduct analysis at the level of the property or comparable decision making unit. This level of analysis matches the scale at which management decisions occur (Carlson et al. 2018), and has been promoted in the context of PES programs that enroll at the property level. Using the entire area owned by an individual landowner may provide the secondary advantage of indirectly addressing slippage within the property. Spatial spillover effects that may occur when the intervention spurs deforestation on other parts of a management unit or landowner property are accounted for (Blackman, Goff, and Rivera Planter 2018; Alix-Garcia and Wolff 2014; Arriagada et al. 2012). Although this does not apply to general equilibrium leakage due to price responses or changes in relative profitability (Arriagada et al. 2012), this offers one advantage relative to a pixel or grid level analysis.

Property level analyses have been promoted in the context of PES programs, but some researchers have avoided them due to interpretability in certain situations. The interpretation on the treatment effect coefficient is the average effect on a property with average characteristics (Alix-Garcia and Gibbs 2017). The effect of the intervention on a landowner's land use decisions is likely what the researcher is after when evaluating PES program, however, this measure may be less directly related to the success of a landscape-scale intervention. If a property does not have an easily attributable first date of treatment because of shifts or overlaps in boundaries over time, it may also not be an easy unit of analysis to work with.

Some researchers choose to use a larger administrative boundary as the unit of analysis such as the state, municipality, or county in order to address the level at which the intervention is applied. Again, this may mitigate concerns over bias emerging from local spillovers within an administrative unit (**???**).

Calculating deforestation rates Upon choosing an aggregated unit of analysis, the researcher must compute the deforestaion rate. This varies thoughout the literature, and many authors do not explicitly define the formula used. Different names are used to describe the calculation of the annual deforestation rate, which generates further confusion (Puyravaud 2003).

A common formula to calculate annual deforestation rates is

$$r_1 = \frac{F_{i,t-1} - F_{it}}{F_{i,t-1}} \tag{1}$$

, where F_{it} and $F_{i,t-1}$ are the forest cover at times t and $t-1$, respectively. This calculation is used consistently in the literature (e.g. Carlson et al. 2018; **???**) and has been referred to as the rate of deforestation, and annual deforestation percentage.

Some authors have also calculated the deforestation rate in relation to the initial observed level of forest cover, presumably replacing A_1 with A_0, the baseline forest cover, in equation (1)

$$r_2 = \frac{F_{i0} - F_{it}}{F_{i0}} \tag{2}$$

Puyravaud (2003) suggest the rate of annual forest change be calculated instead using

$$r_3 = ln(F_{it}/F_{i,t-1}) \tag{3}$$

The deforestation rate can be obtained by computing $-r_3$. This formula is derived from the Compound Interest Law and has also been used in recent studies (e.g. Ruggiero et al. 2019). It has been suggested, because r_1 may underestimate deforestation, relying arbitrarily on the period's initial forest cover (F_{it}) as a baseline. As deforestation rates grow larger, the difference between r_1 and $-r_3$ grows.

Functional form The decision of functional form is another often overlooked decison, particularly with a binary measure of forest cover. The vast majority of papers use OLS, maintaining a linear functional form. Few authors explain their reasoning for choosing the functional form of the econometric model, whether linear or non-linear. In many cases, the distribution of observed data is nonnegative and concentrated near zero deforestation (Carlson et al. 2018; **???**). (**???**) argues a poisson model is theoretically consistent with forest cover loss within an aggregated unit being the count of many independent, discrete binary observations of forest cover loss at the pixel level.

While non-linear functional forms may be theoretically preferable to OLS, this is often times more complicated to implement in practice. In order to obtain an interpretable estimate of the ATT, researchers using non-linear functional forms generally need to compute marginal or average partial effects from the estimated coefficients. It has been noted in the impact evaluation literature that fixed effects cannot typically be used in most non-linear methods due to the incidental parameters problem (e.g. Jones and Lewis 2015; Wendland et al. 2015), favoring the use of OLS in regression models containing fixed effects. Fernández-Val and Weidner (2016) show that a bias correction is needed to get asymptotically unbiased estimates logit, probit, and poisson TWFE models. These bias corrections are available in various statistical packages and are necessary in order to get unbiased estimates when using fixed effects with non-linear functional forms.

Table 1: Table of common methods in the literature

Paper	Panel Method	Unit of analysis	Functional form
Alix-Garcia and Gibbs 2017	two-way FE	binary point/pixel	OLS
Alix-Garcia et al 2018	two-way FE	binary point/pixel	OLS
Anderson et al. 2018	matched DID	binary point/pixel	OLS
Araujo et al. 2009	two-way FE using instrument	state	OLS
Arriagada et al. 2012	matched DID	farm	OLS
Baylis et al. 2012	DID	grid cell	OLS
BenYishay et al. 2017	two-way FE	grid cell	OLS
Blackman 2015	matched regression	binary point/pixel	probit
Blackman et al 2017	two-way FE	community	OLS
Blackman et al. 2018	matched two-way FE	forest magagement unit	OLS
Busch et al. 2015	matched two-way FE	grid cell	poisson
Carlson et al. 2018	matched two-way FE	plantation	poisson
Heilmayr and Lambin 2016	matched DID	property	OLS
Herrera et al. 2019	matched regression	binary point/pixel	OLS
Holland et al. 2017	matched two-way FE	landowner parcel/predio	OLS
Jones and Lewis 2015 (1)	matched two-way FE	binary point/pixel	OLS
(2)	matched two-way FE	household parcel	OLS
Jones et al. 2017	matched two-way FE	household	OLS
Koch et al. 2018	matched DID	municipality	OLS
Oliviera Fiorini et al. 2020	matched regression	binary point/pixel	WLS
Pfaff 1999	regression	county	OLS
Sanchez-Azofeifa et al 2007	regression	grid cells	OLS
Shah and Baylis 2015	DID	grid cell	OLS
Sims and Alix-Garcia 2017	two-way FE	locality	OLS
Tabor et al. 2017	two-way FE	fokontany	OLS
Wendland et al. 2015	matched two-way FE	binary point/pixel	OLS

Methods

Monte Carlo simulations

We employ a series of Monte Carlo simulations to (1) generate synthetic landscapes with known policy effectiveness and (2) analyze the performance of different econometric models in estimating the policy's known impact. Our landscape consists of administrative units that are either untreated ($D = 0$) or are assigned to a conservation treatment ($D = 1$). We observe deforestation in two, even-length periods, a pre-treatment ($t < t_0$) and a post-treatment ($t \geq t_0$) period. The data generating process underlying our

Monte Carlo simulations begins with the assignment of four parameters: The mean, pre-treatment period deforestation rate for untreated units, $baseline_0$; the mean, pre-treatment period deforestation rate for treated units, $baseline_1$; untreated units' mean change in deforestation rates occurring between the pre- and post-treatment periods, $trend$; and lastly, the average treatment effect of the policy on the treated units, ATT. The ATT is the primary parameter the researcher is interested in estimating.

Rather than directly observing each unit's deforestation rate in each time period, the researcher observes annualized maps depicting pixel-level, binary deforestation ($y_{it} \in 0, 1$). We follow XX and assume that these binary observations of deforestation reflect each pixel's unobservable value along a continuous, latent variable indicating the pixel's suitability for deforestation. We define this latent variable as:

$$y_{it}^* = \beta_0 + \beta_1 \mathbb{1}\{D_i = 1\} + \beta_2 \mathbb{1}\{t \geq t_0\} + \beta_3 \mathbb{1}\{D_i = 1\}\mathbb{1}\{t \geq t_0\} + \alpha_i + u_{it} \tag{4}$$

The β coefficients can be defined as functions of the four parameters assigned by the researcher (Appendix 1). A pixel-specific random disturbance is generated according to $\alpha_i \sim N(0, \sigma_a^2)$ and the error term is generated according to $u_{it} \sim N(0, \sigma_u^2)$.

The mapping from the latent to observed variable y_{it} is

$$y_{it} = \begin{cases} 1 & y_{it}^* > 0 \\ 0 & otherwise \end{cases} \tag{5}$$

Here, the observed outcome variable, y_{it}, is equal to 1 if pixel i is observed as deforested in time t and 0 otherwise. The observed variable is the binary outcome visible to the researcher, which represents the result of a more complex data generating process that reflects the pre-defined effectiveness of the policy (ATT) as well as temporal trends in deforestation ($trend$), pre-treatment differences in deforestation probabilities ($baseline_1$), unobservable, pixel-specific determinants of deforestation (α_i) and time-varying, pixel-level disturbances (u_{it}). The area represented by a pixel may not be entirely deforested but may still be classified as deforested using the binary outcome metric.

Set up and evaluation

It is not uncommon to see annual treatment effects amounting to less than a 1% reduction in the annual deforestation rate (e.g. Robalino and Pfaff 2013; Jones et al. 2017). These modest reductions in the annual deforestation rate, however, can amount to large effects over the course of the study period. For example, Alix-Garcia et al. (2018) find that environmental land registration in Brazil's Amazonian states of Mato Grosso and Para reduced the annual deforestation rate by an average 0.5 percentage points, which has amounted to an overall deforestation reduction of 10%. We have parameterized a guiding example, representative of an impactful intervention, to explore for the remainder of the paper.

The primary criteria we use to compare econometric models are the bias and coverage probability of their estimate of the ATT parameter. Using our Monte Carlo simulations, we determine bias by computing the difference between each model's estimate of the ATT and the known ATT parameter. Coverage probability is defined as the proportion of simulations in which the true ATT lies within the simulation's 95% confidence interval (CI). As such, we would expect the ATT to lie within this CI 95% of the time, however, factors such as the bias of the estimates, their distribution, and treatment of standard errors may impact coverage. If the estimator is biased, for example, it is ex-ante less likely that the true parameter falls within the CI.

Results

Bias inherent to binary deforestation DGP

Bias often arises in the DID and TWFE models because of the data generating process inherent to remotely sensed metrics of deforestation. In the case of a binary pixel, the researcher does not observe the underlying spatial process of deforestation. The pixel is realized as deforested only when clearing exceeds some specific threshold (expressed in equation (5)). The use of NDVI or canopy cover thresholds to define a binary forest cover variable are prevalent in the literature (e.g. Holland et al. 2017, @honey–roses2011, @bruggeman2015).

For example, Takahashi and Todo (2013) and Southworth, Munroe, and Nagendra (2004) explicitly determine a threshold value of NDVI and define pixels with NDVI values that are greater than the threshold as forested. Here, a pixel may be subject to continuous clearing pressure but only be classified as deforested when the NDVI value drops below the threshold value. Alternatively, discontinuous clearing pressure may exist but will only be realized when a deforestation event causes the pixel to drop below the threshold. Even popular forest loss datasets (e.g. Hansen et al. 2013) that detect deforestation by stand level disturbance may suffer from similar bias, since a clearing event must exceed some threshold to be detected. Therefore, small-scale logging and deforestation events are less likely to be detected, potentially leading to an underestimate of forest loss (Burivalova et al. 2015). Some level of misclassification error, and resulting bias, is extremely likely in the classification of pixels into a binary forest cover outcome.

Mathematically, this bias will arise any time the latent variable has error generated according to a distribution with a nonlinear cumulative distribution function and is subsequently mapped to a binary observable outcome subject to a threshold. In Figure 3, we can see this bias in the context of our guiding example. Allowing the outcome to vary between 0 and 1 across time periods allows us to see the bias due only to the disparity between the underlying spatial data generating process and the observed binary outcome. We do not exolore the sensitivity of this bias to various thresholds, but bring attention to its existence. For the remainder of the paper, we net out this bias in order to focus on bias that arises from various model selection decisions.
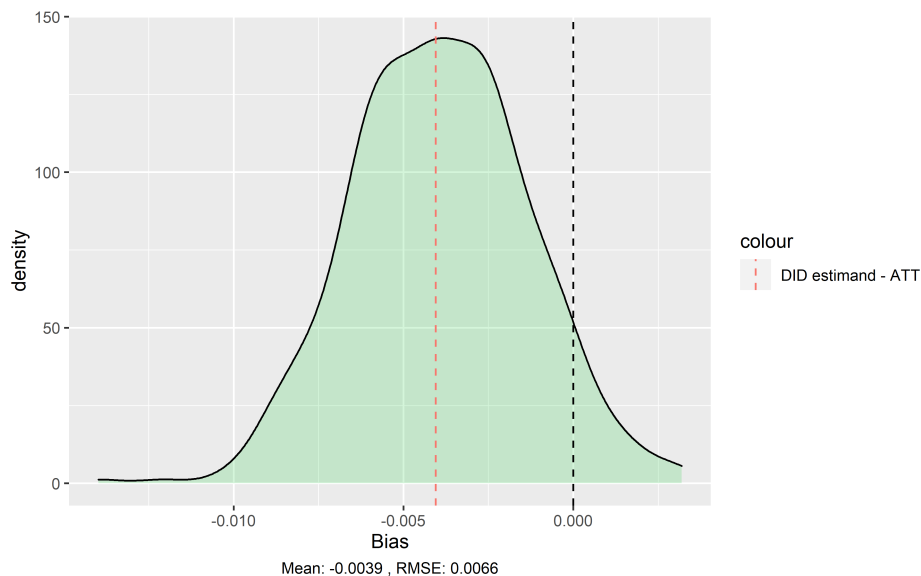

Figure 3: Inherently biased estimates due to nature of observed data

Unit of analysis

Treatment of binary outcome   Analyses at the pixel level are prevalent in the literature, as seen in Table 1. Further, it has often been promoted as the preferred unit of analysis in certain cases. The pixel is generally the level at which the researcher is able to observe the data and is assigned a binary outcome. Remotely sensed metrics of deforestation at the pixel level are often subject to the dynamics of forest disturbance and regrowth. After a deforestation event occurs, the deforested area is unlikely to revert to forest cover within the study period, as it takes several years for forest to regenerate to a detectable level. In the panel therefore, it is probable that in the periods after a pixel is first realized as deforested, subsequent observations of the pixel will also observe the pixel as deforested.

In order to account for these dynamics in the context of deforestation, it has been advised to drop deforested pixels in the periods after they first become deforested (Jones and Lewis 2015; Alix-Garcia and Gibbs 2017). The logic for doing so is as follows. A forested pixel switches from its assigned value of 0 to a value of 1 following a discrete deforestation event. Keeping the deforested pixel in the panel beyond the first period in

which it was observed as deforested may imply that it has actively been deforested in each subsequent time period. In fact, no new deforestation event has ocurred, but the area simply remains deforested from the prior event. These pixels, therefore, contribute positively towards the deforestation rate in each period they are left in the panel. Indeed, we see in our Monte Carlo simulations that DID regression models failing to drop deforested pixels in subsequent periods incur severe bias.

Figure 4 demonstrates the magnitude and direction of the bias incurred from keeping deforested pixels in the panel after they are first realized as deforested in the context of our guiding example. Pixels that were deforested prior to the implementation of the policy continued to contribute to the deforestation rate in the post period in both the treatment and control groups. Dropping the pixels in the periods after they are first observed as deforested mitigates this bias in the DID model, as seen in Figure 4.
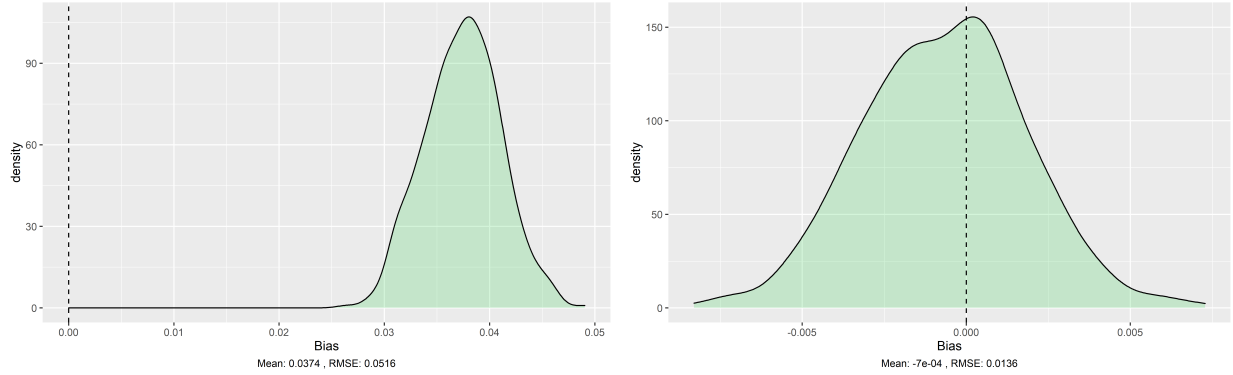

Figure 4: Left: distribution of DID estimates leaving deforested pixels in the panel; Right: DID estimates dropping deforested pixels

Issue with TWFE using pixel as unit of analysis Despite widespread use of pixel level analyses, they are problematic in the context of TWFE regression models. In fact, the TWFE model yields the post-treatment difference in outcomes (single difference), rather than the desired ATT. We provide two forms of evidence to support this claim: (1) evidence from our Monte Carlo simulations and (2) an analytical proof found in the appendix. The result stems from the fact that the TWFE regression is only able to identify off of pixels that are not dropped over the course of the observation period. Thus the pre-treatment period deforestation rates are not accounted for in the TWFE estimates.
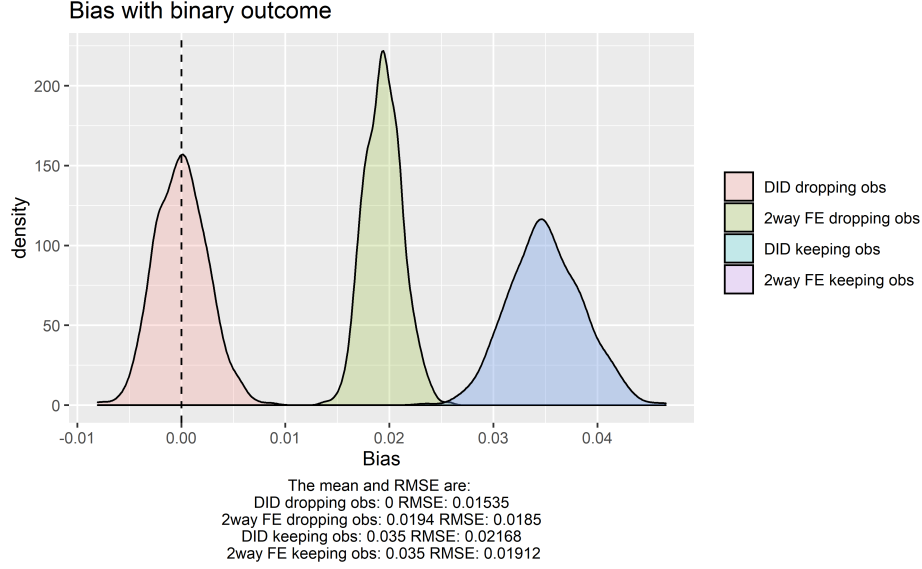
Figure 5: Distribution of estimates produced depending on pixel level regression model

In Figure 5, we see Monte Carlo Outcomes for four econometric model specifications with a binary outcome: (1) DID dropping deforested pixels from the panel for the periods after they are first realized as deforested; (2) TWFE dropping deforested pixels from the panel for the periods after they are first realized as deforested; (3) DID keeping deforested pixels in the panel for the length of the study period; and (4) TWFE keeping deforested pixels in the panel for the length of the study period. We see that estimate distributions from specifications (3) and (4) are identical, showing that DID and TWFE regression models are generally identical in the two-group, two-period case. As described above, we again see the bias resulting from leaving deforested pixels in the panel for the duration of the study period in specifications (3) and (4).

We now bring attention to the distinction between specifications (1) and (2). In both specifications, observations are dropped from the panel in the periods after they are first realized as deforested. As discussed and seen in Figure 4, this is preferable in terms of bias when using the pixel as the unit of analysis. The figure shows, however, that the TWFE model returns a biased measure of the ATT, and in fact, estimates an ex-post single difference. Table 3 shows the resulting bias in the context of the guiding example.

Table 2: ATT versus ex-post single difference (difference in treated and untreated deforestation rates after intervention) in guiding example

treated ex-post	untreated ex-post	ex-post single difference	true ATT
0.025	0.015	0.01	-0.01

In the guiding example, the positive bias also leads to a change in the direction of the estimated *ATT*, which could influence policymakers to avoid an effective policy design. This shows that TWFE models using the pixel as the unit of analysis are not a viable approach to estimate the *ATT* in deforestation impact evaluations. In contrast, the DID regression model does not suffer from this severe bias, meaning that a pixel level analysis is feasible in the two-period, two-group case.

Calculating deforestation rates

Upon aggregating data, the researcher must determine how to calculate deforestation rates in the outcome. With no clear guidance on how this deforestation rate should be computed, there have been a variety of

10

techniques used. We examine the three methods outlined in section 2.

	Outcome
(1)	$\frac{F_{i,t-1}-F_{it}}{F_{i,t-1}}$
(2)	$\frac{F_{i0}-F_{it}}{F_{i0}}$
(3)	$log(F_{i,t-1}/F_{it})$

Outcome 1 Mean: 1e-04 , RMSE: 0.0147
Outcome 2 Mean: -4e-04 , RMSE: 0.0142
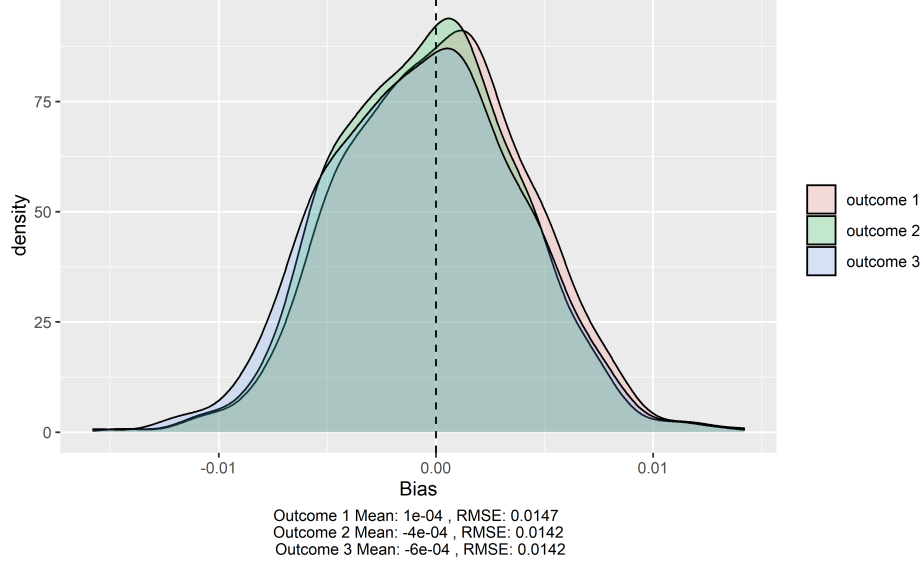Outcome 3 Mean: -6e-04 , RMSE: 0.0142

Figure 6: Distribution of estimates produced by different outcome variable formulae

As seen in Figure 7, outcome 1 results in the least bias in our guiding example. The other outcomes result in relatively greater bias, although this difference seems minimal. We express concern surrounding the use of the initial observed baseline forest cover in the calculation of deforestation rates, as in formula (2). In settings with high deforestation rates, this outcome may overestimate the deforestation rate. As the rate is calculated in periods far away from the baseline, the deforestation rate will be increasingly high. This may impact the viability of the analysis, as the rate becomes artificially inflated in later years. As seen in Appendix T2, outcome two is the least robust to alternate landscape parameterizations.

Regardless of authors' choice of formula, we advise that this formula be explicitly stated in the paper. This will help to avoid confusion as to which formula was used and help researchers understand which methods are the standard within the literature. Moving forward in our paper, all specifications using aggregated data use outcome 1. In our guiding example, it resulted in the least bias, and it is currently the most common deforestation rate calculation used in the literature.

Aggregated Outcomes Since pixel level analyses are not feasible in the context of TWFE regressions, researchers should be aware of the tradeoffs using aggregated units of analysis. The following results apply to both DID and TWFE regression models, as both are equivalent in the two-period, two-group example. We now opt to use TWFE regression models when aggregating data, and DID regression models for pixel level analyses due to the aforementioned issues with TWFE regressions at the pixel level.

For simplicity, we assume the researcher can choose between three levels at which to aggregate the data: grid cell, county, and property. Grid cells are uniform grids layered over the study area and may have a treatment

11

value between 0 and 1 following aggregation of pixels. Counties are heterogeneous administrative units at which we now assign the treatment. Lastly, properties are smaller administrative units within a county.
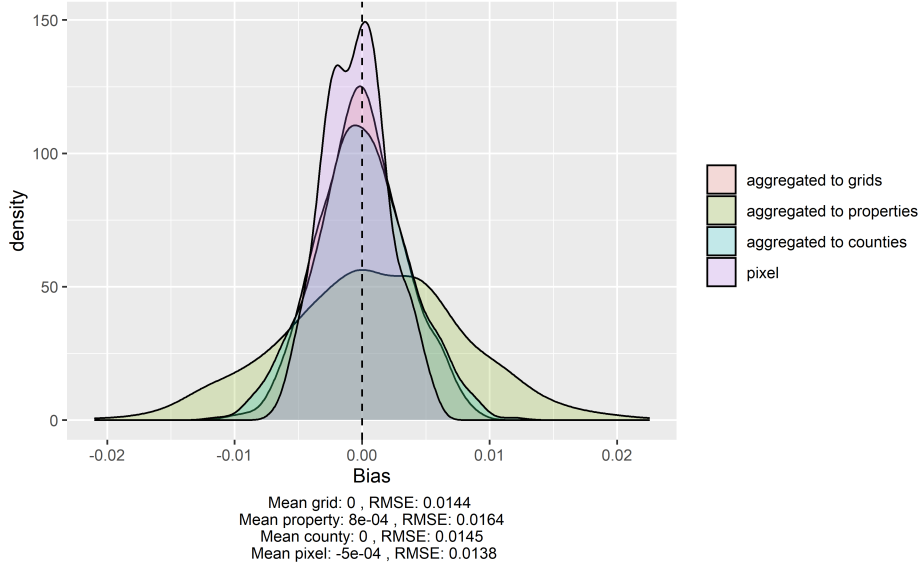

Figure 7: Distribution of estimates produced by level of aggregation

We find little evidence that any one level of aggregation is consistently preferrable in terms of bias, meaning that coverage probability plays a larger role as an evaluation metric. We explore the bias and coverage probabilities corresponding to different levels of aggregation in the context of our guiding example. Figure 8 shows that the bias of the estimates is not critically different depending on level of aggregation, however, the distributions are varied. Notice that the distribution of estimates when aggregating to the property level is wider than the others. This is likely due to heterogeneity in property size. The distribution of the estimates, bias, and standard error structure will affect the coverage probabilities. We compare coverage at each level of aggregation in Table 2 using classical standard errors. We have not yet introduced any structure into the DGP that would warrant clustering. Despite the clear differences in the distributions of estimates, both pixel and property-level analyses yield the desired 95% coverage probabilities.

Table 3: Table of coverage based on unit of analysis

unit of analysis	standard error structure	coverage probability
grid	classical	0.932
property	classical	0.950
county	classical	0.914
pixel	classical	0.950

Property level unobservables   Property level unobservables may impact both treatment effect estimates and coverage probabilities. This is likely to be a factor when land use decisions are made at the property level or the intervention seeks to alter underlying landowner incentives. We introduce an additional error term to the initial DGP that varies at the property level in order to account for these unobservables. This property level error term is generated according to $\rho_{ip} \sim N(0, .15^2)$. Table 3 shows that property level unobservables negatively impact coverage at every level without clustering standard errors. The property level analyses suffer the least from this reduced coverage probability.

12

Table 4: Table of coverage in presence of property level unobservables

unit of analysis	standard error structure	coverage probability
grid	classical	0.740
property	classical	0.875
county	classical	0.755
pixel	classical	0.725

As property level unobservables play a larger role, the treatment of standard errors also becomes more important. Almost all studies in the literature use clustered standard errors for inference. The clustering problem is casued by the presence of a common unobserved ramdom shock at the group level that will lead to correlation between all observations within a group (Hansen 2007). We have introduced the random shock at the property level in our guiding example to represent the decision-making unit. As such, correlation within grids and counties will also be introduced, since they contain multiple pixels within a property. Clustering standard errors relaxes the assumption of no correlation across observations within the spatial unit used for clustering (Jones and Lewis 2015). Examining the below table shows that clustering to the respective level of aggregation does indeed improve the coverage probability. Property level aggregation leads to relatively preferred coverage probabilities, regardless of the standard error structure. These results highlight the benefit of using the relevant decision-making unit as the unit of analysis as well as clustering standard errors at that level.

Table 5: Table of coverage in presence of property level unobservables

unit of analysis	standard error structure	coverage probability
grid	clustered at grid	0.85
property	clustered at property	0.91
county	clustered at county	0.83
pixel	clustered at property	0.89

Functional form

We explore the effect of using non-linear functional forms. We use DID regression models at the pixel level in this section to avoid the aforementioned issues with pixel level TWFE models . We also reintroduce the consideration of bias due to classification error in the dgp in order to explore whether alternative functional forms may help to mitigate this bias. Coefficient estimates are converted into the *ATT* using the **mfx** package in R, which takes into account the binary nature of the treatment variable.

The underlying DGP and error structure play a role in determining which functional form leads to the least bias in a given scenario. We find that alternative parameterizations do affect which functional form is optimal in a given circumstance, as seen in Table XX. Therefore, authors, regardless of the preferred specification, should justify their choice. As is common in the literature already, we advise that authors ensure the result is robust to different choices of functional form. We leave a thorough discussion of the choice of functional form in the presence of classification error to future work, but highlight the importance of this area for exploration.

Table 6: functional form across landscape parameterizations

| model | 500 datasets of 10,000 pixels, $\sigma_a = 0.1$, $\sigma_v = 0.25$ | | | | | |
| | $ATT = -0.01$ $baseline_0 = 0.02$ $baseline_1 = 0.04$ $trend = -0.005$ | | $ATT = 0.01$ $baseline_0 = 0.05$ $baseline_1 = 0.02$ $trend = 0.005$ | | $ATT = -0.03$ $baseline_0 = 0.02$ $baseline_1 = 0.05$ $trend = 0.01$ | |
	bias	RMSE	bias	RMSE	bias	RMSE
DID	-0.0044238	0.0109512	0.0007987	0.0059909	0.0090085	0.0444103
Probit	0.0017479	0.0055368	0.0052713	0.0124090	0.0123822	0.0370975
Logit	0.0025495	0.0037282	0.0061994	0.0137431	0.0128176	0.0346648
Poisson	0.0026877	0.0045031	0.0064181	0.0070264	0.0129215	0.0364135

Discussion

This paper seeks to leave researchers with a clear sense of the decisions that play a role in their choice of econometric specification. We discuss and show the benefits of conducting the analysis at the level of the relevant decision making unit when it is of importance. Choice of functional form should be justified in the context of each institutional setting, taking into account the underlying DGP. Robustness checks should account for the possibility of alternative functional forms. Formulas in the calculation of the deforestation rate should be made explicit to avoid confusion. Ultimately, context plays a role in all of these decisions. There is no one size fits all econometric specification and researchers should make clear which decisions they have made as well as their reasoning. We have developed a shiny app available at .. so that researchers can explore the potential tradeoffs with choices of econometric model design.

Several econometric considerations relevant to model design and interpretation were not explored in this paper. FE regression models have received substantial attention in the econometrics literature recently regarding concerns surrounding their viability and interpretability (e.g. Kropko and Kubinec 2020). For example, properties of FE regression models that arise when the treatment effect is heterogeneous across groups or over time may lead to erroneous results (**???**). We largely abstract away from choices of scale with regards to grid size or unit of analysis, but biased estimates may result from scale choices that are too large or too small relative to the data generating process or decision unit (Avelino, Baylis, and Honey-Rosés 2016). This makes choice of grid cell size a relevant but unexplored aspect of our study. We do not address more general considerations surrounding the use of satellite data including satellite sensor characteristics, atmospheric conditions, or the error structure of certain remotely sensed data products (Jain 2020). These considerations all play a role in the choice of econometric method, but are left for future exploration. Lastly, the benefits of pre-matching control and treatment units is not addressed, but they are well understood to be substantial (e.g. Jones and Lewis 2015; Blackman 2013).

Without clear guidance on key econometric model design decisions, a wide variety of methods have been used in the literature using quasi-experimental methods for conservation impact evaluation. We show that a number of studies likely provide estimates that suffer from bias or improper inference. In particular, our result about the use of FE models with the pixel as the relevant unit of analysis casts doubt as to the reliability of a significant number of estimates. This said, a significant number of studies have used proper intuition in the design of their econometric models. The increasingly widespread awareness of methods of causal inference is a further positive step. The observations made here may apply to a wider audience beyond the set of researchers investigating the impacts of conservation interventions on deforestation. Any evaluation of interventions implemented at a spatial scale may benefit from this discussion. These results further apply to many settings in which the outcome represents an irreversible binary event. For example,

studies addressing technology adoption as an outcome may fall prey to the same issues we describe with a pixel level analysis. Moving forward, it is imperative that researchers use methods that minimize bias and allow inference at expected levels of confidence. Misleading results may lead policymakers to avoid impactful policy designs or adopt policy that worsens environmental damages.

Acknowledgements and data

References

Alix-Garcia, Jennifer, and Holly K. Gibbs. 2017. "Forest Conservation Effects of Brazil's Zero Deforestation Cattle Agreements Undermined by Leakage." *Global Environmental Change* 47 (November): 201–17. https://doi.org/10.1016/j.gloenvcha.2017.08.009.

Alix-Garcia, Jennifer, Lisa L. Rausch, Jessica L'Roe, Holly K. Gibbs, and Jacob Munger. 2018. "Avoided Deforestation Linked to Environmental Registration of Properties in the Brazilian Amazon: Environmental Registration in the Amazon." *Conservation Letters* 11 (3): e12414. https://doi.org/10.1111/conl.12414.

Alix-Garcia, Jennifer, and Hendrik Wolff. 2014. "Payment for Ecosystem Services from Forests." *Annual Review of Resource Economics* 6 (1): 361–80. https://doi.org/10.1146/annurev-resource-100913-012524.

Andam, K. S, P. J Ferraro, A. Pfaff, G. A Sanchez-Azofeifa, and J. A Robalino. 2008. "Measuring the Effectiveness of Protected Area Networks in Reducing Deforestation." *Proceedings of the National Academy of Sciences* 105 (42): 16089.

Anderson, Christa M., Gregory P. Asner, William Llactayo, and Eric F. Lambin. 2018. "Overlapping Land Allocations Reduce Deforestation in Peru." *Land Use Policy* 79 (December): 174–78. https://doi.org/10.1016/j.landusepol.2018.08.002.

Arriagada, R. A., P. J. Ferraro, E. O. Sills, S. K. Pattanayak, and S. Cordero-Sancho. 2012. "Do Payments for Environmental Services Affect Forest Cover? A Farm-Level Evaluation from Costa Rica." *Land Economics* 88 (2): 382–99. https://doi.org/10.3368/le.88.2.382.

Avelino, Andre Fernandes Tomon, Kathy Baylis, and Jordi Honey-Rosés. 2016. "Goldilocks and the Raster Grid: Selecting Scale When Evaluating Conservation Programs." *PLOS ONE* 11 (12): e0167945. https://doi.org/10.1371/journal.pone.0167945.

Baylis, Kathy, Jordi Honey-Rosés, Jan Börner, Esteve Corbera, Driss Ezzine-de-Blas, Paul J. Ferraro, Renaud Lapeyre, U. Martin Persson, Alex Pfaff, and Sven Wunder. 2016. "Mainstreaming Impact Evaluation in Nature Conservation." *Conservation Letters* 9 (1): 58–64. https://doi.org/10.1111/conl.12180.

Blackman, Allen. 2013. "Evaluating Forest Conservation Policies in Developing Countries Using Remote Sensing Data: An Introduction and Practical Guide." *Forest Policy and Economics* 34 (September): 1–16. https://doi.org/10.1016/j.forpol.2013.04.006.

Blackman, Allen, Leonard Goff, and Marisol Rivera Planter. 2018. "Does Eco-Certification Stem Tropical Deforestation? Forest Stewardship Council Certification in Mexico." *Journal of Environmental Economics and Management* 89 (May): 306–33. https://doi.org/10.1016/j.jeem.2018.04.005.

Bruggeman, Derek, Patrick Meyfroidt, and Eric F. Lambin. 2015. "Production Forests as a Conservation Tool: Effectiveness of Cameroon's Land Use Zoning Policy." *Land Use Policy* 42 (January): 151–64. https://doi.org/10.1016/j.landusepol.2014.07.012.

Burivalova, Zuzana, Martin R. Bauert, Sonja Hassold, Nandinanjakana T. Fatroandrianjafinonjasolomiovazo, and Lian Pin Koh. 2015. "Relevance of Global Forest Change Data Set to Local Conservation: Case Study of Forest Degradation in Masoala National Park, Madagascar." *Biotropica* 47 (2): 267–74. https://doi.org/10.1111/btp.12194.

Butsic, Van, David J. Lewis, Volker C. Radeloff, Matthias Baumann, and Tobias Kuemmerle. 2017. "Quasi-Experimental Methods Enable Stronger Inferences from Observational Data in Ecology." *Basic and Applied Ecology* 19 (March): 1–10. https://doi.org/10.1016/j.baae.2017.01.005.

Carlson, Kimberly M., Robert Heilmayr, Holly K. Gibbs, Praveen Noojipady, David N. Burns, Douglas C. Morton, Nathalie F. Walker, Gary D. Paoli, and Claire Kremen. 2018. "Effect of Oil Palm Sustainability Certification on Deforestation and Fire in Indonesia." *Proceedings of the National Academy of Sciences* 115 (1): 121–26. https://doi.org/10.1073/pnas.1704728114.

Edwards, Ryan B, Walter P Falcon, Gracia Hadiwidjaja, Matthew M Higgins, Rosamond L Naylor, and Sudarno Sumarto. 2020. "Fight Fire with Finance: A Randomized Field Experiment to Curtail Land-Clearing Fire in Indonesia," 62.

Fernández-Val, Iván, and Martin Weidner. 2016. "Individual and Time Effects in Nonlinear Panel Models with Large N , T." *Journal of Econometrics* 192 (1): 291–312. https://doi.org/10.1016/j.jeconom.2015.12.014.

Ferraro, Paul J. 2009. "Counterfactual Thinking and Impact Evaluation in Environmental Policy." *New Directions for Evaluation* 2009 (122): 75–84. https://doi.org/10.1002/ev.297.

Ferraro, Paul J., and Juan José Miranda. 2017. "Panel Data Designs and Estimators as Substitutes for Randomized Controlled Trials in the Evaluation of Public Programs." *Journal of the Association of Environmental and Resource Economists* 4 (1): 281–317. https://doi.org/10.1086/689868.

Ferraro, Paul J., James N. Sanchirico, and Martin D. Smith. 2019. "Causal Inference in Coupled Human and Natural Systems." *Proceedings of the National Academy of Sciences* 116 (12): 5311–8. https://doi.org/10.1073/pnas.1805563115.

Hansen, Christian B. 2007. "Generalized Least Squares Inference in Panel and Multilevel Models with Serial Correlation and Fixed Effects." *Journal of Econometrics* 140 (2): 670–94. https://doi.org/10.1016/j.jeconom.2006.07.011.

Hansen, Matthew C., and Thomas R. Loveland. 2012. "A Review of Large Area Monitoring of Land Cover Change Using Landsat Data." *Remote Sensing of Environment*, Landsat Legacy Special Issue, 122 (July): 66–74. https://doi.org/10.1016/j.rse.2011.08.024.

Hansen, M. C., P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, et al. 2013. "High-Resolution Global Maps of 21st-Century Forest Cover Change." *Science* 342 (6160): 850–53. https://doi.org/10.1126/science.1244693.

Holland, Margaret B., Kelly W. Jones, Lisa Naughton-Treves, José-Luis Freire, Manuel Morales, and Luis Suárez. 2017. "Titling Land to Conserve Forests: The Case of Cuyabeno Reserve in Ecuador." *Global Environmental Change* 44 (May): 27–38. https://doi.org/10.1016/j.gloenvcha.2017.02.004.

Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81 (396): 945–60. https://doi.org/10.1080/01621459.1986.10478354.

Honey-Rosés, Jordi, Kathy Baylis, and M. Isabel Ramírez. 2011. "A Spatially Explicit Estimate of Avoided Forest Loss: Spatial Estimate of Avoided Forest Loss." *Conservation Biology* 25 (5): 1032–43. https://doi.org/10.1111/j.1523-1739.2011.01729.x.

Jain, Meha. 2020. "The Benefits and Pitfalls of Using Satellite Data for Causal Inference." *Review of Environmental Economics and Policy* 14 (1): 157–69. https://doi.org/10.1093/reep/rez023.

Jayachandran, Seema, Joost de Laat, Eric F. Lambin, Charlotte Y. Stanton, Robin Audy, and Nancy E. Thomas. 2017. "Cash for Carbon: A Randomized Trial of Payments for Ecosystem Services to Reduce Deforestation." *Science* 357 (6348): 267–73. https://doi.org/10.1126/science.aan0568.

Jones, Kelly W., Margaret B. Holland, Lisa Naughton-Treves, Manuel Morales, Luis Suarez, and Kayla Keenan. 2017. "Forest Conservation Incentives and Deforestation in the Ecuadorian Amazon." *Environmental Conservation* 44 (1): 56–65. https://doi.org/10.1017/S0376892916000308.

Jones, Kelly W., and David J. Lewis. 2015. "Estimating the Counterfactual Impact of Conservation Programs on Land Cover Outcomes: The Role of Matching and Panel Regression Techniques." *PLOS ONE* 10 (10): e0141380. https://doi.org/10.1371/journal.pone.0141380.

Kropko, Jonathan, and Robert Kubinec. 2020. "Interpretation and Identification of Within-Unit and Cross-Sectional Variation in Panel Data Models." Edited by Talib Al-Ameri. *PLOS ONE* 15 (4): e0231349. https://doi.org/10.1371/journal.pone.0231349.

Larsen, Ashley E., Kyle Meng, and Bruce E. Kendall. 2019. "Causal Analysis in ControlImpact Ecological Studies with Observational Data." *Methods in Ecology and Evolution* 10 (7): 924–34. https://doi.org/10.1111/2041-210X.13190.

Meyfroidt, Patrick. 2016. "Approaches and Terminology for Causal Analysis in Land Systems Science." *Journal of Land Use Science* 11 (5): 501–22. https://doi.org/10.1080/1747423X.2015.1117530.

Pfaff, Alexander, Juan Robalino, G. Arturo Sanchez-Azofeifa, Kwaw S. Andam, and Paul J. Ferraro. 2009. "Park Location Affects Forest Protection: Land Characteristics Cause Differences in Park Impacts Across Costa Rica." *The B.E. Journal of Economic Analysis & Policy* 9 (2).

Puyravaud, Jean-Philippe. 2003. "Standardizing the Calculation of the Annual Rate of Deforestation." *Forest Ecology and Management* 177 (1-3): 593–96. https://doi.org/10.1016/S0378-1127(02)00335-3.

Ribas, Luiz Guilherme dos Santos, Robert L. Pressey, Rafael Loyola, and Luis Mauricio Bini. 2020. "A Global Comparative Analysis of Impact Evaluation Methods in Estimating the Effectiveness of Protected Areas." *Biological Conservation* 246 (June): 108595. https://doi.org/10.1016/j.biocon.2020.108595.

Robalino, J., and A. Pfaff. 2013. "Ecopayments and Deforestation in Costa Rica: A Nationwide Analysis of PSA's Initial Years." *Land Economics* 89 (3): 432–48. https://doi.org/10.3368/le.89.3.432.

Ruggiero, Patricia G. C., Jean Paul Metzger, Leandro Reverberi Tambosi, and Elizabeth Nichols. 2019. "Payment for Ecosystem Services Programs in the Brazilian Atlantic Forest: Effective but Not Enough." *Land Use Policy* 82 (March): 283–91. https://doi.org/10.1016/j.landusepol.2018.11.054.

Soulé, Michael E. 1985. "What Is Conservation Biology?" *BioScience* 35 (11): 727–34. https://doi.org/10.2307/1310054.

Southworth, Jane, Darla Munroe, and Harini Nagendra. 2004. "Land Cover Change and Landscape FragmentationComparing the Utility of Continuous and Discrete Analyses for a Western Honduras Region." *Agriculture, Ecosystems & Environment* 101 (2-3): 185–205. https://doi.org/10.1016/j.agee.2003.09.011.

Takahashi, Ryo, and Yasuyuki Todo. 2013. "The Impact of a Shade Coffee Certification Program on Forest Conservation: A Case Study from a Wild Coffee Forest in Ethiopia." *Journal of Environmental Management* 130 (November): 48–54. https://doi.org/10.1016/j.jenvman.2013.08.025.

Wendland, K. J., M. Baumann, D. J. Lewis, A. Sieber, and V. C. Radeloff. 2015. "Protected Area Effectiveness in European Russia: A Postmatching Panel Data Analysis." *Land Economics* 91 (1): 149–68. https://doi.org/10.3368/le.91.1.149.

Williams, David R., Andrew Balmford, and David S. Wilcove. 2020. "The Past and Future Role of Conservation Science in Saving Biodiversity." *Conservation Letters* n/a (n/a): e12720. https://doi.org/10.1111/conl.12720.

Wooldridge, Jeffrey M. n.d. "Clusterextended.Pdf."

Appendix

Two-way fixed effects proof

We can show that in the case where the binary outcome is dropped in periods after the outcome is realized as a 1, two-way fixed effects regressions typically do not identify the ATT, but the ATT + the group difference.

proof:

deforestation rate calculation robustness to alternate parameterizations

Table 7: outcome calculation across landscape parameterizations

	500 datasets of 10,000 pixels, $\sigma_a = 0.1$, $\sigma_v = 0.25$					
	$ATT = -0.01$		$ATT = 0.01$		$ATT = -0.03$	
	$baseline_0 = 0.02$		$baseline_0 = 0.05$		$baseline_0 = 0.02$	
	$baseline_1 = 0.04$		$baseline_1 = 0.02$		$baseline_1 = 0.05$	
	$trend = -0.005$		$trend = 0.005$		$trend = 0.01$	
Outcome	**bias**	**RMSE**	**bias**	**RMSE**	**bias**	**RMSE**
1	-0.0000303	0.0144711	0.0024299	0.0072458	-0.0006179	0.0193108
2	-0.0004783	0.0140086	0.0060103	0.0052083	-0.0022662	0.0176621
3	-0.0006658	0.0138960	0.0027099	0.0072738	-0.0015277	0.0184756

Initial parameter to β coefficient mapping

The researcher sets the following four parameters:

$$baseline_0 = E[y_{it}(0)|t < t_0, D_i = 0]$$
$$baseline_1 = E[y_{it}(0)|t < t_0, D_i = 1]$$
$$trend = E[y_{it}(0)|t \geq t_0, D_i = 0] - E[y_{it}(0)|t < t_0, D_i = 0]$$
$$ATT = E[y_{it}(1) - y_{it}(0)|t \geq t_0, D_i = 1]$$

Note the following constraints on the parameters:

$$E[y_{it}(0)|t \geq t_0, D_i = 0] \geq 0$$
$$E[y_{it}(1)|t \geq t_0, D_i = 1] \geq 0$$

The parameters can be expressed as follows:

$$
\begin{aligned}
ATT =& E[y_{it}(1) - y_{it}(0)|t \geq t_0, D_i = 1] \\
=& E[y_{it}(1)|t \geq t_0, D_i = 1] - E[y_{it}(0)|t \geq t_0, D_i = 1] \\
=& P(y_{it}(1) = 1|t \geq t_0, D_i = 1) - P(y_{it}(0) = 1|t \geq t_0, D_i = 1) \\
=& P(y_{it}^*(1) > 0|t \geq t_0, D_i = 1) - P(y_{it}^*(0) > 0|t \geq t_0, D_i = 1) \\
=& P(\beta_0 + \beta_1 + \beta_2 + \beta_3 + \alpha_i + u_{it} > 0) - P(\beta_0 + \beta_1 + \beta_2 + \alpha_i + u_{it} > 0) \\
=& P(-\alpha_i - u_{it} < \beta_0 + \beta_1 + \beta_2 + \beta_3) - P(-\alpha_i - u_{it} < \beta_0 + \beta_1 + \beta_2) \\
=& F(\beta_0 + \beta_1 + \beta_2 + \beta_3) - F(\beta_0 + \beta_1 + \beta_2)
\end{aligned}
$$

$$
\begin{aligned}
trend =& E[y_{it}(0)|t \geq t_0, D_i = 0] - E[y_{it}(0)|t < t_0, D_i = 0]\\
=& P(y_{it}(0) = 1|t \geq t_0, D_i = 0) - P(y_{it}(0) = 1|t < t_0, D_i = 0)\\
=& P(y_{it}^*(0) > 0|t \geq t_0, D_i = 0) - P(y_{it}^*(0) > 0|t < t_0, D_i = 0)\\
=& P(-\alpha_i - u_{it} < \beta_0 + \beta_2) - P(-\alpha_i - u_{it} < \beta_0)\\
=& F(\beta_0 + \beta_2) - F(\beta_0)
\end{aligned}
$$

$$
\begin{aligned}
baseline_0 =& E[y_{it}(0)|t < t_0, D_i = 0]\\
=& P(y_{it}(0) = 1|t < t_0, D_i = 0)\\
=& P(y_{it}^*(0) > 0|t < t_0, D_i = 0)\\
=& P(-\alpha_i - u_{it} < \beta_0)\\
=& F(\beta_0)
\end{aligned}
$$

$$
\begin{aligned}
baseline_1 =& E[y_{it}(0)|t < t_0, D_i = 1]\\
=& P(y_{it}(0) = 1|t < t_0, D_i = 1)\\
=& P(y_{it}^*(0) > 0|t < t_0, D_i = 1)\\
=& P(-\alpha_i - u_{it} < \beta_0 + \beta_1)\\
=& F(\beta_0 + \beta_1)
\end{aligned}
$$

, Where $F()$ is the CDF of a $N(0, \sigma_a^2 + \sigma_u^2)$

Now solving for the β coefficients:

solving for β_0

$$
baseline_0 = F(\beta_0)
$$
$$
\Leftrightarrow
$$
$$
\beta_0 = F^{-1}(baseline_0)
$$

solving for β_1

$$
baseline_1 = F(\beta_0 + \beta_1)
$$
$$
\Leftrightarrow
$$
$$
\beta_1 = F^{-1}(baseline_1) - \beta_0
$$

solving for β_2

$$
trend = F(\beta_0 + \beta_2) - F(\beta_0)
$$
$$
\Leftrightarrow
$$
$$
trend + baseline_0 = F(\beta_0 + \beta_2)
$$
$$
\Leftrightarrow
$$
$$
F^{-1}(trend + baseline_0) = \beta_0 + \beta_2
$$
$$
\Leftrightarrow
$$
$$
\beta_2 = F^{-1}(trend + baseline_0) - \beta_0
$$

solving for β_3

$$ATT = F(\beta_0 + \beta_1 + \beta_2 + \beta_3) - F(\beta_0 + \beta_1 + \beta_2)$$
$$\Leftrightarrow$$
$$ATT + F(\beta_0 + \beta_1 + \beta_2) = F(\beta_0 + \beta_1 + \beta_2 + \beta_3)$$
$$\Leftrightarrow$$
$$F^{-1}(ATT + F(\beta_0 + \beta_1 + \beta_2)) = \beta_0 + \beta_1 + \beta_2 + \beta_3$$
$$\Leftrightarrow$$
$$\beta_3 = F^{-1}(ATT + F(\beta_0 + \beta_1 + \beta_2)) - (\beta_0 + \beta_1 + \beta_2)$$

Bias inherent to binary deforestation DGP

This bias can be represented by the difference between the DID estimand and the ATT parameter of interest:

$$DID_{estimand} - ATT$$
$$\Leftrightarrow$$
$$E[y_{it}(1)|t \geq t_0, D_i = 1] - E[y_{it}(0)|t < t_0, D_i = 1] - (E[y_{it}(0)|t \geq t_0, D_i = 0] - E[y_{it}(0)|t < t_0, D_i = 0])$$
$$- E[y_{it}(1) - y_{it}(0)|t \geq t_0, D_i = 1]$$
$$\Leftrightarrow$$
$$F(\beta_0 + \beta_1 + \beta_2 + \beta_3) - F(\beta_0 + \beta_2) - (F(\beta_0 + \beta_1) - F(\beta_0))$$
$$- (F(\beta_0 + \beta_1 + \beta_2 + \beta_3) - F(\beta_0 + \beta_1 + \beta_2))$$
$$\Leftrightarrow$$
$$F(\beta_0 + \beta_1 + \beta_2) + F(\beta_0)) - (F(\beta_0 + \beta_2) + F(\beta_0 + \beta_1))$$