

# Conservation impact evaluation using remotely sensed data

Alberto Garcia and Robert Heilmayr

June 1, 2022

## Abstract

Conservation scientists are increasingly measuring the impacts of conservation interventions by applying quasiexperimental impact evaluation to remotely sensed panel data on land use change. However, these applications come with new challenges. Using Monte Carlo simulations and analytical proofs, we demonstrate that many of the panel econometric models employed for conservation impact evaluation are biased - the significance, magnitude and even direction of estimated effects from many studies are likely incorrect. These errors threaten to undermine the evidence base that underpins conservation policy adoption and design. We review the methods in this burgeoning literature and develop guidance for the design of econometric models quantifying conservation policy effectiveness.

## 1 Introduction

Policymakers often need to understand the causal impacts of conservation interventions. Do marine protected areas stop unsustainable harvesting of fish? Can payments for ecosystem services encourage lasting reforestation? While randomized experiments are the gold standard for the identification of causal relationships (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). Increasingly, these econometric approaches to impact evaluation are being used to disentangle the causal relationships that underpin conservation decisionmaking (Butsic et al. 2017; Baylis et al. 2016; Williams, Balmford, and Wilcove 2020). These econometric methods can generate estimates of an intervention’s impact by comparing observed outcomes to a statistical 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 econometric methods account for the non-random assignment of interventions that often confound identification of causal relationships in conservation. For example, 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). To build valid counterfactuals in the face of non-random selection, researchers frequently build upon spatial, panel data to observe changes in outcomes across treated and control units after the adoption of an intervention (Blackman 2013). When a rigorous research design is applied to panel data, 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).

One important development that has enabled the proliferation of conservation impact evaluation is the increasing prevalence of remotely sensed datasets detailing conservation outcomes through time (Blackman 2013; Jones and Lewis 2015). For example, NASA’s Landsat archive can be used to generate 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).

However, many commonly used measures of land use change have structural differences from the data used

in traditional, linear panel models. For example, deforestation is often measured using data with a similar structure to the Global Forest Change product produced by Hansen et al. (2013). When converted to a panel structure, these data yield binary observations detailing the first year in which each 30 by 30m pixel was deforested. Importantly, the data do not allow for the detection of reforestation timing and, as a result, are unable to detect repeated deforestation events in the same location. To date, insufficient attention has been paid to how such binary, irreversible outcome data may affect the performance of standard econometric tools.

Here, we use a combination of analytical proofs and Monte Carlo simulations to demonstrate that many econometric analyses of deforestation are likely biased - significance, magnitude and even direction of estimated effects might be incorrect. While we focus on deforestation, many of our results apply to any setting with binary, irreversible data including mortality (e.g., Friedman and Schady 2013), technology adoption (e.g., Bollinger et al., n.d.), and reincarceration (e.g., Mastrobuoni and Pinotti 2015). The resulting biases arise even when researchers follow common guidance to adopt “rigorous” research designs with valid counterfactuals (Blackman 2013; Jones and Lewis 2015). Our main result shows that two-way fixed effects regressions with pixel unit fixed effects cannot identify the desired treatment effect parameter. Papers published in both conservation science and economics journals as recently as 2021 use this problematic specification to recover treatment effect estimates. To help guide future impact evaluations, we identify multiple ways in which this bias can be reduced or even eliminated. We then explore non-random selection that arises due to irreversibility in the deforestation setting and how this feature of the data may lead to bias. Finally, we reflect on the econometric benefits that emerge when researchers are able to match their model structure to the relevant scale of the deforestation process.

## 2 Measuring avoided deforestation using panel methods

We focus on the case in which a researcher would like to quantify the impact an intervention has had on deforestation. 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. This general setting describes a broad array of studies that apply panel methods to remotely sensed data (Table 1).

Table 1: Econometric model structures used in avoided deforestation impact evaluations. All of the included studies use data accessed as pixelated binary deforestation.

<b>Paper</b>	<b>Panel Method</b>	<b>Unit of analysis</b>	<b>Unit fixed effects level</b>
Alix-Garcia and Gibbs 2017	TWFE	binary point/pixel	pixel
Alix-Garcia et al 2018	TWFE	binary point/pixel	pixel
Anderson et al. 2018	matched DID	binary point/pixel	NA
Araujo et al. 2009	TWFE using instrument	state	state
Arriagada et al. 2012	matched DID	farm	NA
Baehr et al. 2021	TWFE	binary pixel/grid cell	pixel
Baylis et al. 2012	DID	grid cell	NA
BenYishay et al. 2017	TWFE	grid cell	grid cell
Blackman 2015	matched unit FE model	binary point/pixel	county
Blackman and Villalobos 2020	matched TWFE	management unit	management unit
Blackman et al. 2017	TWFE	community	community
Blackman et al. 2018	matched TWFE	management unit	management unit
Busch et al. 2015	matched TWFE	grid cell	grid cell
Bustic et al. 2017	TWFE	binary point/pixel	pixel
Carlson et al. 2018	matched TWFE	plantation	plantation
Carlson et al. 2018 (robustness check)	Cox PH DID	pixel	NA
Heilmayr and Lambin 2016	matched DID	property	NA
Herrera et al. 2019	matched regression	binary point/pixel	NA
Holland et al. 2017	matched TWFE	landowner parcel	landowner parcel
Jones and Lewis 2015 (1)	matched TWFE	binary point/pixel	pixel
Jones and Lewis 2015 (2)	matched TWFE	household parcel	household parcel
Jones et al. 2017	matched TWFE	household	household
Koch et al. 2018	matched DID	municipality	NA
Nolte et al. 2017	DID	NA	deforestation
Panlasigui et al. 2018	TWFE	binary point/pixel	pixel
Pfaff 1999	regression	county	NA
Sales et al. 2022	Cox PH DID	pixel	NA
Sanchez-Azofeifa et al. 2007	regression	grid cells	NA
Shah and Baylis 2015	DID	grid cell	NA
Sims and Alix-Garcia 2017	TWFE	locality	locality
Tabor et al. 2017	TWFE	fokontany	fokontany
Wendland et al. 2015	matched TWFE	binary point/pixel	pixel

In each of the studies detailed in Table 1, the researcher’s goal is to measure the impact that a specific policy had on deforestation within treated units, also known as the *average treatment effect on the treated (ATT)*. The *ATT* estimates the difference between the average deforestation rate of treated units with treatment, and the average deforestation rate of treated units without treatment. The fundamental challenge is that, for every treated unit, the researcher is unable to observe the value that the outcome would have taken in the absence of treatment (Holland 1986). In our case, this means that the researcher cannot observe the deforestation that would have occurred in treated units had they not received treatment. Figure 1 displays this problem in the context of a simulated conservation intervention that reduced deforestation rates in treated areas — the landscape is depicted as observed by the researcher at the end of the observation period, including the unobservable counterfactual of what would have happened if the conservation intervention had not been adopted. Note that in untreated areas, there is no counterfactual deforestation, since no intervention ever took place.

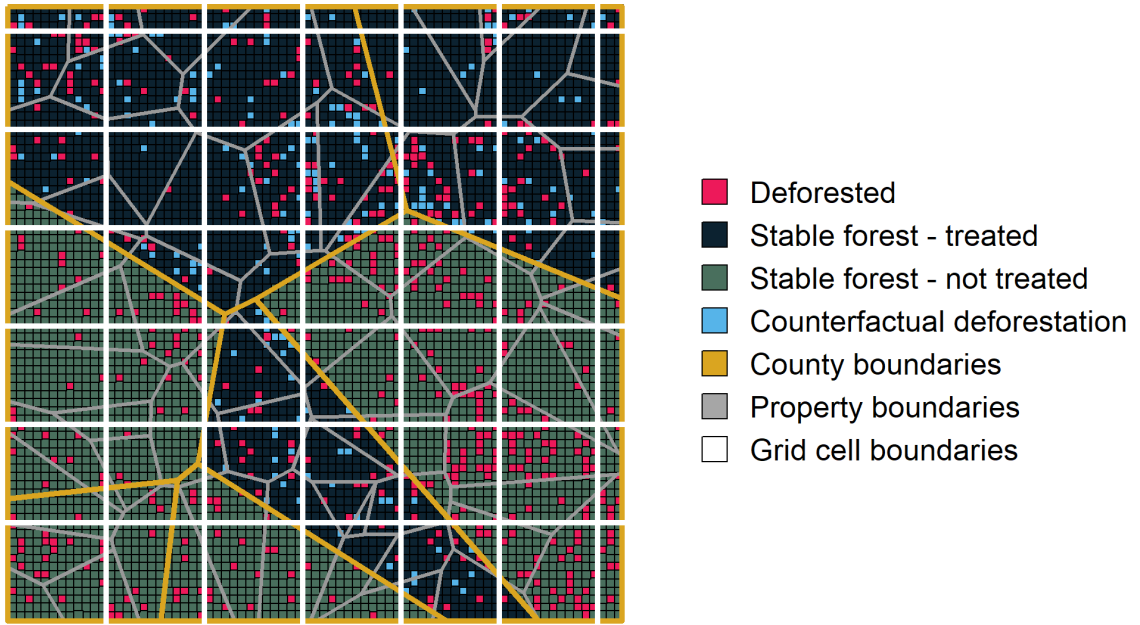


Figure 1: A map of a simulated landscape depicting patterns of deforestation under an effective conservation intervention, as well as counterfactual deforestation illustrating what would have happened in the absence of the intervention.

## 2.1 Modeling the decision to deforest

We model deforestation ( $y_{ivt}$ ) as a binary choice by a landowner to clear a small plot of land  $i$  within their larger property  $v$  at time  $t$ . The decision to deforest depends upon a latent variable ( $y_{ivt}^*$ ) that represents the returns from the plot of land in its cleared state ( $V_{ivt}^{cleared}$ ) relative to the returns from its forested state ( $V_{ivt}^{uncleared}$ ), such that:

$$y_{ivt}^* = V_{ivt}^{cleared} - V_{ivt}^{uncleared} \quad (1)$$

$$y_{ivt} = \begin{cases} 1 & \text{if } y_{ivt}^* > 0 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

This generic clearing rule underpins a broad class of more specific static and dynamic models that have

been used to explore the determinants of deforestation (e.g., Pfaff 1999; Kerr et al. 2003; Pfaff and Sanchez-Azofeifa 2004).

However, this basic model makes an assumption that the decision to deforest is reversible. In reality, a number of characteristics of both the process of deforestation, as well as the methods used to detect deforestation in individual plots, complicate this assumption. First, the goal of many conservation interventions is to prevent the loss of mature forests that may take decades, if not centuries, to regrow. In such cases, deforestation itself may be considered irreversible in human time scales, focusing the researchers' attention upon the first instance in which a plot is deforested. Even when deforestation of secondary forests is an object of interest, constraints imposed by remotely sensed datasets may force empirical researchers to treat deforestation as irreversible. Gradual processes of reforestation are inherently harder to identify than abrupt losses of forest cover (Hansen et al. 2013). In addition, determining the precise year in which the extended process of forest regrowth began is currently an active area of research for the remote sensing community, and often requires many years of post-regrowth observations. As a result, commonly used deforestation datasets such as the Global Forest Change product often only identify the first year in which a pixel was cleared (Hansen et al. 2013). Whether desired, or due to technical limitations, the resulting inability to observe repeated deforestation means that deforestation is, in effect, an irreversible process in most conservation impact evaluations. In response to this irreversibility, Jones and Lewis (2015) and Alix-Garcia and Gibbs (2017) have suggested that deforested pixels should be dropped in the periods after they are first cleared. We follow this guidance, further modifying our binary deforestation variable:

$$y_{ivt}^o = \begin{cases} 1 & y_{ivt}^* > 0 \text{ and } y_{iv\tau}^* \leq 0 \text{ for all } \tau < t \\ 0 & y_{ivt}^* \leq 0 \text{ and } y_{iv\tau}^* \leq 0 \text{ for all } \tau < t \\ \text{NAN} & \text{otherwise} \end{cases} \quad (3)$$

## 2.2 Estimating the *ATT* using difference-in-differences and two-way fixed effects estimators

Two methods commonly used to estimate the *ATT* in conservation intervention settings with panel data are Difference-in-Differences (DID) and Two-way Fixed Effects (TWFE) regression (Blackman (2013); Jones and Lewis (2015)). Our parameter of interest, the *ATT*, is the average effect of the conservation intervention on treated pixels. Let  $y_{ivt}(1)$  and  $y_{ivt}(0)$  denote the potential outcomes of pixel  $i$  in property  $v$  in time  $t$  with and without the treatment, respectively. In addition, let  $t_0$  denote the year that the intervention was implemented and let  $D_i$  represent a dummy indicating whether pixel  $i$  is ever treated. The *ATT* can now be expressed as:

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

DID and TWFE methods have become popular in part, because the researcher does not need random assignment of treatment to generate convincing estimates of a program's impact on avoided deforestation. Instead, the researcher must make a common trends assumption, under which we evaluate each method.

**Assumption 1:** (Common trends)

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

Assumption 1 requires that pixels in treated and untreated areas would have experienced the same change in their probability of deforestation across the two periods had no intervention occurred. 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 lend credibility to this assumption (Butsic et al. 2017).

We also make the following stable unit treatment value assumption (SUTVA)

**Assumption 2:** (SUTVA)

$$\forall d \in \{0, 1\} : \text{ if } D_i = d \text{ and } t \geq t_0, \text{ then } y_{ivt}(d) = y_{ivt}$$

Assumption 2 requires that the potential outcomes for pixel  $i$ ,  $y_{ivt}(1)$  and  $y_{ivt}(0)$ , do not depend on the treatment status of any other pixel. There also cannot exist unobserved versions of treatment that may affect the potential outcomes.

**Regression 1:** (DID regression) Under these two assumptions, the researcher can estimate the  $ATT$  using a traditional DID regression. Let  $\beta_{DID}$  denote the coefficient of the interaction between  $D_i$  and an indicator for whether the intervention has been implemented in time  $t$ ,  $\mathbb{1}\{t \geq t_0\}$ , in the following (population) OLS regression:

$$y_{ivt}^o = \alpha_0 + \alpha_1 D_i + \alpha_2 \mathbb{1}\{t \geq t_0\} + \beta_{DID} \times D_i \mathbb{1}\{t \geq t_0\} + \epsilon_{it} \quad (4)$$

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).

$$\beta_{DID} = E[y_{ivt}^o | t \geq t_0, D_i = 1] - E[y_{ivt}^o | t < t_0, D_i = 1] - (E[y_{ivt}^o | t \geq t_0, D_i = 0] - E[y_{ivt}^o | t < t_0, D_i = 0])$$

When the  $y_{ivt}^o$ s are i.i.d. and Assumptions 1 and 2 hold, it is straightforward to show that

$$\beta_{DID} = ATT$$

Often, however, the researcher wants to estimate the  $ATT$  in a setting that does not fit the two-group, two-period case covered by the standard DID model. In such cases, TWFE regressions are frequently used to apply DID methods to multiple groups or treatment periods. This amounts to estimating a regression that includes unit and time fixed effects to control for unobservable confounding variables that vary across units or through time.

**Regression 2:** (TWFE regression) Let  $\beta_{TWFE}$  denote the coefficient of the interaction between  $D_i$  and  $\mathbb{1}\{t \geq t_0\}$  in the following (population) OLS regression:

$$y_{ivt}^o = \alpha + \beta_{TWFE} \times D_i \mathbb{1}\{t \geq t_0\} + \lambda_t + \gamma_i + \epsilon_{it} \quad (5)$$

Here  $\lambda_t$  and  $\gamma_i$  represent the year and pixel fixed effects, respectively.

In the case of two groups and two time periods, the TWFE regression typically yields an estimate of the  $ATT$  that is equivalent to the estimate generated by the DID model (Wooldridge 2010). With this in mind, many researchers have used the TWFE model as a “generalized DID” that can be estimated in settings where different units are exposed to treatment in more than two distinct time periods (Table 1). For example, a researcher may use a TWFE regression model to examine the effectiveness of a network of protected areas where the protected areas were created at different times, or a payment for ecosystem services (PES) program that enrolled properties in annual cohorts.

### 2.2.1 Pixel level, TWFE models fail to estimate the $ATT$

Despite widespread use of pixel level analyses of deforestation, the application of TWFE models to a binary, irreversible process such as deforestation yields a biased estimate of the  $ATT$ . Specifically, the proof contained in Appendix 10.1 shows that, in the two-group, two-period case, the coefficient of interest from the TWFE model ( $\beta_{TWFE}$ ) estimates the post-treatment difference in outcomes (single difference), rather than the desired  $ATT$ :

$$\beta_{TWFE} = ATT + \underbrace{E[y_{it}(0) | t < t_0, D_i = 1] - E[y_{it}(0) | t < t_0, D_i = 0]}_{\text{pre-treatment difference in deforestation rates}}$$

Regression 2 thus forgoes the benefits that panel methods provide, and if the treated area has a different baseline deforestation rate than the control, will generate a biased estimate of the intervention’s impact. Many conservation interventions are specifically targeted towards locations with either low opportunity costs for conservation or high threats of conversion. As a result, it is likely that many conservation impact evaluations will have treatment and control units that experienced different pre-treatment deforestation rates. It is important to note that this bias could even lead to changes in the estimated treatment effect’s sign, in addition to errors in the effect’s magnitude and significance.

This result stems from researchers’ use of pixel fixed effects in the regression specification. By including pixel fixed effects in TWFE regressions, researchers hope to control for local confounders, including pre-treatment differences in the outcome. However, when following common guidance to drop pixels in the periods after they are first deforested, these fixed effects do not behave as the researcher expects. Because the pixels that are deforested in the first period are not observed for the entire panel, these observations go unaccounted for in  $\beta_{TWFE}$ . Implicitly, the pre-treatment deforestation rates get assigned as 0 for both the treatment and control groups. In Appendix 10.2, we show that  $\hat{\beta}_{TWFE}$  is equivalent to the coefficient from a regression on a dataset where all pixels deforested in the pre-treatment period are simply dropped from the dataset completely.

### 2.2.2 Alternative construction of $y_{i,t}^o$ cannot recover the $ATT$

Although dropping previously deforested pixels from the panel introduces bias into the TWFE estimate of the  $ATT$ , keeping observations in the panel after initial deforestation introduces its own problems. The  $ATT$  as defined in Section 2.2 is an estimate of the impact of an intervention on the frequency of deforestation events (i.e. the decision to clear). Keeping the deforested pixel in the panel beyond the first period in which it was observed as deforested would incorrectly imply that it has actively been deforested in each subsequent time period, when in fact, no new deforestation event or clearing decision has occurred. This is intuitively problematic, because the deforestation rate in each period would be monotonically increasing by construction, which is not necessarily the case. We show that pixel-level regression models that do not drop deforested pixels in subsequent periods do indeed incur severe bias if used to estimate the  $ATT$  as it is defined here (Appendix 10.3).

## 2.3 Estimating the $ATT$ using survival analysis

### 2.3.1 Hazard rate ratios from a single survival model do not estimate the $ATT$ under common trends

Survival analysis has emerged as a common approach to modeling irreversible changes such as mortality, technology adoption, or recidivism (Friedman and Schady 2013; Mastrobuoni and Pinotti 2015). Survival models, such as the Cox Proportional Hazards model, quantify how covariates relate to changes in the length of time that a unit remains in a sample. In the case of deforestation, survival analyses can be used to explore how policy adoption changes the duration that treated, forested pixels survive until they are first cleared.

Despite the theoretical appeal of using survival models to study deforestation, they are still relatively uncommon in conservation impact evaluation. One emerging approach introduces the intuition of a difference in differences research design into a Cox Proportional Hazards model (e.g. Heilmayr et al. 2020; Sales, Strobl, and Elliott 2022). Specifically, researchers estimate a Cox proportional hazards model of the following general form:

**Regression 3:** (Cox DID regression) Let  $\beta_{coxDID}$  denote the coefficient of the interaction between  $D_i$  and  $\mathbb{1}\{t \geq t_0\}$  in the following (population) OLS regression:

$$h(t) = \delta_0(t) \exp(\alpha_0 + \alpha_1 D_i + \alpha_2 \mathbb{1}\{t \geq t_0\} + \beta_{coxDID} \times D_i \mathbb{1}\{t \geq t_0\} + \epsilon_{it}) \quad (6)$$

, where  $h(t)$  is the hazard rate of deforestation,  $t$  years into the study period; and  $\delta_0(t)$  is the baseline hazard function.

Multiple studies across a wide variety of settings have interpreted the resulting exponentiated coefficient  $\exp(\beta_{\text{coxDID}})$  as a hazard ratio that is indicative of the impact that treatment has had on the relative likelihood of survival. Specifically, based on the way this hazard ratio is interpreted in multiple papers, it appears that many researchers expect this hazard ratio to represent the ratio of the hazard rates in the treatment group post-treatment, relative to the counterfactual in that group had treatment not occurred. This desired hazard ratio measuring the relative impact of treatment on the treated ( $HRTT$ ) can be considered a reframing of the traditional  $ATT$  as a ratio rather than a difference:

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

Both in conservation and alternative settings, researchers using Regression 3 make Assumption 1, and evaluate whether it is plausible in their setting. However, Appendix 10.4 shows that  $\exp(\beta_{\text{coxDID}})$  only identifies the  $HRTT$  under an alternative assumption:

**Assumption 3:** (Proportional trends)

$$\frac{E[y_{ivt}(0)|t \geq t_0, D_i = 1]}{E[y_{ivt}(0)|t < t_0, D_i = 1]} = \frac{E[y_{ivt}(0)|t \geq t_0, D_i = 0]}{E[y_{ivt}(0)|t < t_0, D_i = 0]} \quad (7)$$

Assumption 3 requires that pixels in treated and untreated areas would have experienced the same ratio of change in their probability of deforestation across the two periods had no intervention occurred. Note that in general, Assumption 1 and Assumption 3 cannot simultaneously hold. This means that researchers estimating Regression 3 under the traditional common trends assumption (Assumption 1) will not recover the  $HRTT$ , the relevant treatment effect parameter.

### 2.3.2 Proposing a new survival analysis-based estimator of the $ATT$

To the best of our knowledge, no prior studies have successfully combined the Cox Proportional Hazards model and the difference in differences research design to recover an unbiased estimate of the  $ATT$  under the traditional common trends assumption (Assumption 1). Here we outline a new estimation approach that first recovers an unbiased estimate of the  $HRTT$  and then translates this into an estimate of the  $ATT$  that holds under Assumption 1. First, we note that the desired  $HRTT$  can be re-written as a combination of three different hazard ratios:

$$\begin{aligned} HR_1 &= \frac{E[y_{ivt}(1)|t \geq t_0, D_i = 1]}{E[y_{ivt}(0)|t < t_0, D_i = 1]} \\ HR_2 &= \frac{E[y_{ivt}(1)|t \geq t_0, D_i = 1]}{E[y_{ivt}(0)|t \geq t_0, D_i = 0]} \\ HR_3 &= \frac{E[y_{ivt}(0)|t \geq t_0, D_i = 0]}{E[y_{ivt}(0)|t < t_0, D_i = 0]} \\ HRTT &= \frac{E[y_{ivt}(1)|t \geq t_0, D_i = 1]}{E[y_{ivt}(0)|t \geq t_0, D_i = 1]} = \frac{1}{1/HR_1 + 1/HR_2 - 1/(HR_2 * HR_3)} \end{aligned}$$

Each of the three hazard ratios,  $HR_1$ ,  $HR_2$ , and  $HR_3$ , can be estimated through separate Cox Proportional Hazards models estimated on subsets of the larger dataset. Specifically:

- $HR_1 = \exp(\alpha)$ , where  $\alpha$  is estimated by subsetting to observations from the treated group ( $D_i = 1$ ), and estimating the hazard rate of deforestation at time  $t$  as  $h(t) = \lambda_0(t)\exp(\alpha 1\{t \geq t_0\})$ ; and
- $HR_2 = \exp(\beta)$ , where  $\beta$  is estimated by subsetting to observations from the post-treatment period ( $t \geq t_0$ ), and estimating the hazard rate of deforestation at time  $t$  as  $h(t) = \gamma_0(t)\exp(\beta 1\{D_i = 1\})$ ;



- $HR_3 = \exp(\delta)$ , where  $\delta$  is estimated by subsetting to observations from the untreated group ( $D_i = 0$ ), and estimating the hazard rate of deforestation at time  $t$  as  $h(t) = \psi_0(t)\exp(\delta 1\{t \geq t_0\})$ .

Because the numerator of  $HRTT$ ,  $E[y_{ivt}(1)|t \geq t_0, D_i = 1]$ , can be estimated as the mean of post-treatment deforestation rates in the treated group (denoted  $\widehat{defor}_{D_i=1, t \geq t_0}$ ), we can estimate the  $ATT$  using this estimated deforestation rate and our estimate of  $HRTT$ :

$$\widehat{ATT}_{Cox} = \widehat{defor}_{D_i=1, t \geq t_0} - \frac{\widehat{defor}_{D_i=1, t \geq t_0}}{\widehat{HRTT}}$$

We have shown that the simple extension of the traditional DID to the survival setting only recovers an easily interpretable measure of a policy’s impact under an assumption that cannot simultaneously hold with the traditional common trends assumption, the “Proportional Trends” assumption. In contrast, our proposed estimator, which relies on separate estimation of relevant hazard ratios, does recover the relevant analog of the  $ATT$  under common trends. We explore the performance of  $\widehat{ATT}_{Cox}$  relative to the proposed OLS regressions under various circumstances likely to arise in the deforestation setting in the next sections. If a researcher opts to use survival analysis to recover an intervention’s impact, their choice of estimator should depend on which trends assumption is plausible in their specific setting.

### 3 Monte Carlo simulations to compare alternative model performance

The rapid growth of the conservation impact evaluation literature has resulted in a diversity of model structures that all attempt to estimate the effectiveness of conservation interventions (Table 1). To explore the relative performance of these different models, we employ a series of Monte Carlo simulations. Specifically, we randomly generate synthetic landscapes with known policy effectiveness and analyze the performance of different econometric models in estimating the policy’s known impact.

#### 3.1 Landscape configuration

We begin each Monte Carlo simulation by creating a synthetic landscape consisting of 150 rows and 150 columns of square pixels, equivalent to a raster that is 4 times larger than what is illustrated in Figure 1. We assume that each pixel has a resolution of 30 meters, comparable to the resolution of many Landsat-based, remote sensing analyses. The landscape thus represents an area of approximately 20.25 km<sup>2</sup>. We then divide this landscape into a variety of spatial units, composed of either uniform aggregations of pixels (i.e. large or small “grid cells”), or randomly spaced Thiessen polygons (i.e. “counties” or “properties”). Grid cells are intended to represent arbitrary units of spatial aggregation imposed by the researcher. In contrast, counties and properties are intended to represent simulated administrative units over which policy or land use decisions are made. Table 2 summarizes the relative scale of each of these spatial units under our baseline specifications.

Table 2: Spatial unit structure and size

Spatial unit	Spatial structure	Avg. number of pixels	Area (hectares)
Property	Thiessen polygons	100	9
County	Thiessen polygons	900	81
Large grid	Uniform square	900	81
Small grid	Uniform square	100	9

### 3.2 Data generating process

Each of our simulated landscapes consists of administrative units that are either untreated ( $D_i = 0$ ) or are assigned to a conservation treatment ( $D_i = 1$ ). We observe deforestation in two, even-length periods, a pre-treatment ( $t < t_0$ ) and a post-treatment ( $t \geq t_0$ ) period.

We follow Equation (1) and model these binary deforestation events as a function of each pixel’s unobservable value along the continuous, latent variable ( $y_{ivt}^*$ ) indicating the return to clearing pixel  $i$ , in property  $v$ , in year  $t$ :

$$\begin{aligned} y_{ivt}^* &= V_{ivt}^{cleared} - V_{ivt}^{uncleared} \\ &= \beta_0 + \beta_1 D_i + \beta_{2,0}(1 - D_i)\mathbb{1}\{t \geq t_0\} + (\beta_{2,1} + \beta_3)D_i\mathbb{1}\{t \geq t_0\} + \alpha_i + u_{it} + \rho_v \end{aligned}$$

That is, the returns to deforestation evolve over the two time periods ( $\mathbb{1}\{t \geq t_0\}$ ), and differ across the control ( $D_i = 0$ ) and treated pixels ( $D_i = 1$ ). In addition, we assume that the value of deforestation is influenced by time-invariant random disturbances at the scale of individual pixels ( $\alpha_i \sim N(0, \sigma_a^2)$ ) or properties ( $\rho_v \sim N(0, \sigma_p^2)$ ), as well as time-varying, pixel-scale disturbances ( $u_{it} \sim N(0, \sigma_u^2)$ ). These disturbances can represent a variety of spatial and temporal processes including, for example, the biophysical characteristics of a location, or the preferences of a property owner.

### 3.3 Assumed parameter values

For the remainder of the paper, we explore a guiding example that has been parameterized to represent an impactful intervention in a high deforestation setting. Conservation interventions often have annual treatment effects smaller than a 1 percentage point 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 landscape-scale effects. 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 of 0.5 percentage points, which has resulted in an overall reduction in deforestation of 10%.

Our initial simulated landscape has the following characteristics: a pre-treatment deforestation rate of 2% in the control area; a pre-treatment deforestation rate of 5% in the intervention area; a decrease in the deforestation rate of 0.5 percentage points between the first and second period in the absence of treatment; and an average reduction of 1 percentage point in the deforestation rate in treated units due to the intervention ( $ATT = -0.01$ ). We assume that  $\sigma_u = 0.5$ . Finally, we begin by assuming away time invariant pixel ( $\sigma_a = 0$ ) and property-level disturbances ( $\sigma_p = 0$ ) but relax this assumption in Section 5. Note that Assumptions 1 and 2 are satisfied by construction. The derivations detailing the mapping from the landscape characteristics to the corresponding parameters in  $y_{ivt}^*$  can be found in Appendix 10.5.

### 3.4 Evaluation criteria

We compare econometric models using a combination of estimate bias, root mean squared error (RMSE), and coverage probability. Using our Monte Carlo simulations, we calculate estimate bias as the difference between each model’s mean estimate of the  $ATT$  and the known  $ATT$  parameter. RMSE describes the distribution of estimates around the  $ATT$ . 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.

## 4 Alternative model specifications can yield unbiased estimates of the $ATT$

TWFE models have risen to prominence due to their flexibility in applying DID methods to settings with multiple groups and variation in treatment timing. However, TWFE models with pixel fixed effects are not a viable approach to estimate the  $ATT$  in deforestation impact evaluations. Column 1 of Figure 2 shows the bias associated with Regression 2, a TWFE regression with pixel unit fixed effects. In our guiding example, the ex-post single difference is 0.02 (the  $ATT$  plus the post-treatment group difference in deforestation

rates), when the true  $ATT$  is equal to -0.01. This means that a positive bias of 0.03 results from the use of this regression model. However, we show that multiple alternate model specifications enable researchers to generate unbiased estimates of the  $ATT$ . We describe several straightforward solutions to this challenge below.

#### 4.1 Traditional difference-in-differences model

In the two-group, two-period case the traditional DID (Regression 1) is an unbiased estimator of the  $ATT$ , as shown in column 3 of Figure 2 (The typical DID is equivalent to including treatment fixed effects). However, researchers often want to use TWFE models because of their flexibility in situations that do not fall under the simplest DID setting. Therefore, researchers should be aware of the trade-offs using aggregated units of analysis and fixed effects when using TWFE models for deforestation impact evaluation.

#### 4.2 Spatially aggregated fixed effects

One can also use fixed effects at the level of a spatially aggregated unit rather than the pixel to resolve bias associated with TWFE regressions with pixel fixed effects. 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. Counties are heterogeneous administrative units at which we assign the treatment. Lastly, properties are smaller spatial units that lie within one or more counties. Pixel level TWFE models with spatially aggregated unit fixed effects are all in the form of Regression 4.

**Regression 4:** (pixel level TWFE regression with spatially aggregated unit fixed effects) Let  $\beta_{FE,j}$  denote the coefficient of the interaction between  $D_i$  and  $\mathbb{1}\{t \geq t_0\}$  in the following (population) OLS regression:

$$y_{i,t}^o = \alpha + \beta_{FE,j} \times D_i \mathbb{1}\{t \geq t_0\} + \lambda_t + \gamma_j + \psi_k + \epsilon_{it} \quad (8)$$

, where  $\lambda_t$  denotes year fixed effects and  $\gamma_j$  and  $\psi_k$  denote one of grid ( $g$ ), county ( $c$ ), or property ( $v$ ) fixed effects ( $j, k = \{g, c, v\}$ ). We include both  $\gamma_j$  and  $\psi_k$  in order to account for fixed effects at both the level of treatment assignment,  $k$  (In this case,  $k = c$ ), and another spatial unit,  $j$ .

Columns 4-6 of Figure 2 show that pixel level TWFE regressions are unbiased estimators of the  $ATT$  when grid, county, or property fixed effects are used rather than pixel fixed effects. We also see that, in the absence of property level perturbations (i.e.  $\sigma_p = 0$  in the DGP), all three models provide similar estimates and estimate distributions.

#### 4.3 Spatially aggregated units of analysis

Another potential solution to the bias associated with TWFE models is for researchers to aggregate multiple pixel level observations into larger units of analysis. The researcher must now calculate the deforestation rate in each time period. While we detail the tradeoffs of various deforestation rate calculations in the appendix (Appendix 10.10), the following results are based on Equation (9) which is arguably the most commonly used formula in the literature (e.g. Carlson et al. 2018; Busch et al. 2015).

$$z_{jt} = \frac{F_{j,t-1} - F_{j,t}}{F_{j,t-1}}$$

, where  $F_{j,t}$  and  $F_{j,t-1}$  are the forest cover in unit  $j$  at times  $t$  and  $t - 1$ , respectively. Unit  $j$  represents one of the three aggregated units, the grid cell ( $g$ ), county ( $c$ ), or property ( $v$ ).

**Regression 5:** (grid/county/property level regression) Let  $\beta_j$  denote the coefficient of the interaction between  $D_j$  and  $\mathbb{1}\{t \geq t_0\}$  in the following (population) OLS regression, where  $j = \{g, c, v\}$ :

$$z_{jt} = \alpha + \beta_j \times D_j \mathbb{1}\{t \geq t_0\} + \lambda_t + \gamma_j + \epsilon_{jt} \quad (9)$$

, where  $\lambda_t$  denotes year fixed effects and  $\gamma_j$  denotes one of grid, county, or property fixed effects. The treatment variable  $D_j = \frac{1}{N_j} \sum_{i=1}^{N_j} D_i$ , is the average treatment value amongst all pixels in unit  $j$ . Note that in these regressions, the level of unit fixed effects matches the unit of analysis.

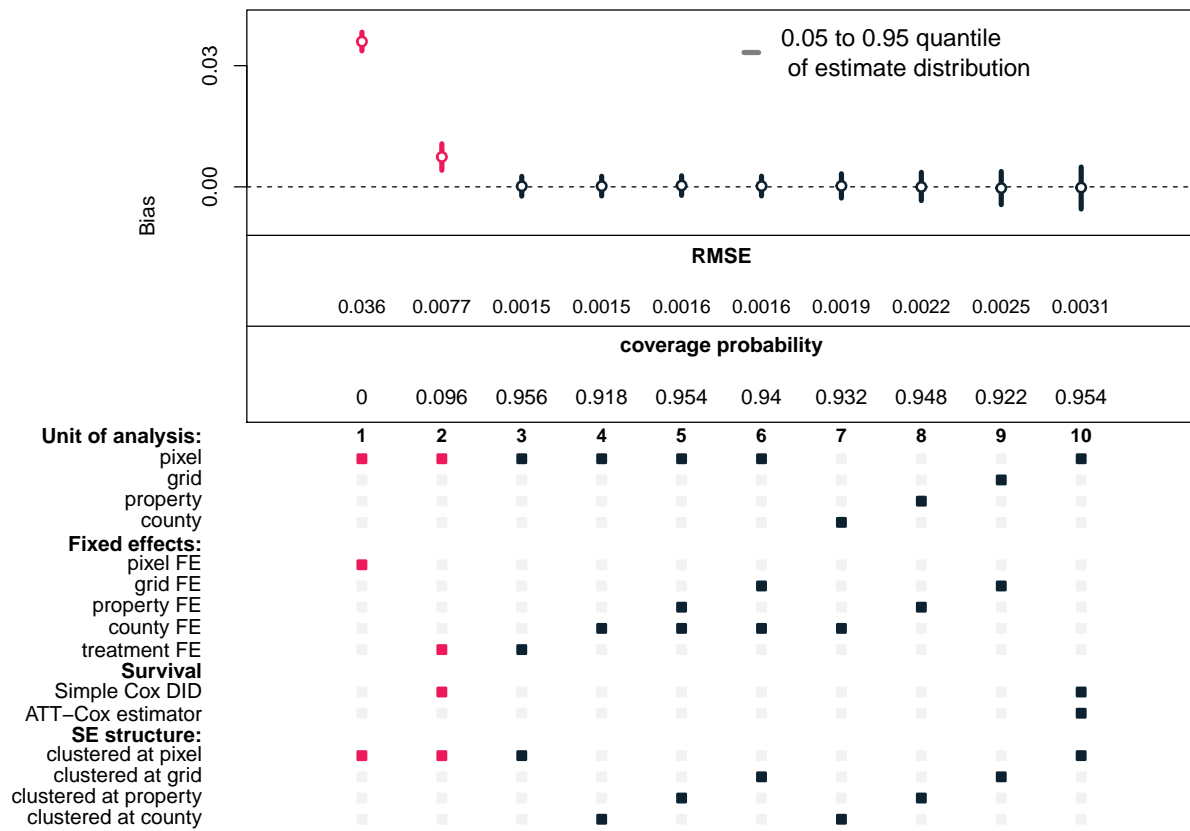


Figure 2: bias, distribution, and coverage of all models with clustered standard errors.

Columns 7-9 of Figure 2 show that the bias of the estimates does not vary dramatically across different levels of aggregation, however, the distributions are varied. In particular, the property-level model had the highest RMSE of unbiased specifications due to the heterogeneity in property sizes across the landscape.

#### 4.4 Survival analysis

Survival analysis provides an appealing alternative to traditional linear estimators when studying irreversible changes such as deforestation. The simple, single-regression DID framing of the Cox Proportional hazards model (Regression 3), however, is not a viable solution (Section 2.3) under the typical common trends assumption (Assumption 1). Column 2 of Figure 2 shows the bias associated with this model in our parameterization where Assumption 1 holds but Assumption 3 does not. In light of the fact that the simple Cox DID is not a viable analog to the traditional DID, we explore whether our proposed estimator,  $\widehat{ATT}_{Cox}$ , recovers relevant treatment effect parameters in the deforestation setting. We see in column 10 that  $\widehat{ATT}_{Cox}$  indeed recovers our parameter of interest, the  $ATT$ .

Although our proposed, survival analysis-based approach to estimation yields a good estimate of the true  $ATT$  in this simple setting, multiple considerations raise questions about the utility of this non-linear model in more complex settings. One of the primary reasons for the use of survival analysis is censoring. This occurs when the researcher has partial information about the subjects' survival times but does not have access to precise event times. While the researcher may not observe all pixels until they are deforested, many other common forms of censoring are rarely a concern in the context of deforestation since remote sensing typically enables the creation of balanced panels. Further, the proposed strategy to drop pixels in the periods after they are first deforested successfully addresses irreversibility in deforestation events. Finally, as we show in the following sections, survival analysis is likely to suffer from bias when unobservables enter the landscape at a scale different than the unit of analysis. While researchers can account for this in OLS specifications, there is no clear solution in survival analysis.

### 5 Non-random sample selection can generate bias

Irreversibility in observed deforestation also creates the potential for non-random sample selection. Specifically, deforested pixels are no longer at risk of clearing in the periods after they are first deforested. This means the "at risk" set of pixels changes through time as more pixels become deforested. As such, the distribution that describes the returns to clearing the at-risk pixels may change through time as well, leading to non-random selection of the sample through time. For example, pixels with extremely high returns to clearing are more likely to be cleared early on, regardless of treatment status. In subsequent periods, therefore, these high return pixels are less likely to be present in the sample at all. In the context of two-groups and two-periods, only the second period suffers from this non-random sample selection. We express the bias introduced from non-random sample selection below.

**Result:** Under Assumptions 1 and 2, in the two-group, two-period case,  $\beta_{DID}$  suffers from non-random sample selection bias when the  $y_{ivt}^o$ s are not i.i.d.

$$\beta_{DID} = ATT + \underbrace{E[y_{ivt}^o | t \geq t_0, D_i = 1] - E[y_{ivt} | t \geq t_0, D_i = 1] - (E[y_{ivt}^o | t \geq t_0, D_i = 0] - E[y_{ivt} | t \geq t_0, D_i = 0])}_{\text{bias emerging from non-random sample selection}} \quad (10)$$

**Proof:** Appendix 10.6 Q.E.D.

In essence, the first and third expectations in the bias term are conditional on the pixel remaining forested after the first period.

Non-random sample selection did not bias our initial simulations as presented in Figure 2 because we assumed away time-invariant pixel and property-level disturbances. The sample of at risk pixels in each time period did not depend on the deforestation that occurred the previous period, since the  $y_{ivt}^o$ s were i.i.d.. However, once time-invariant disturbances enter the DGP, the distribution of the  $y_{iv2}^o$ s is potentially different from

that of the  $y_{iv1}^o$ s. This is likely to be the case in reality, since each plot of land will have time-invariant characteristics such as slope, elevation, market access, agricultural suitability, etc. that impact its expected returns to clearing.

In order to see how this non-random selection influences estimates in our simulated setting, we set  $\sigma_a$ , the standard error of the time-invariant pixel-level disturbances equal to 0.1. Figure 3 shows that non-random selection introduces a slight downward bias across every specification.

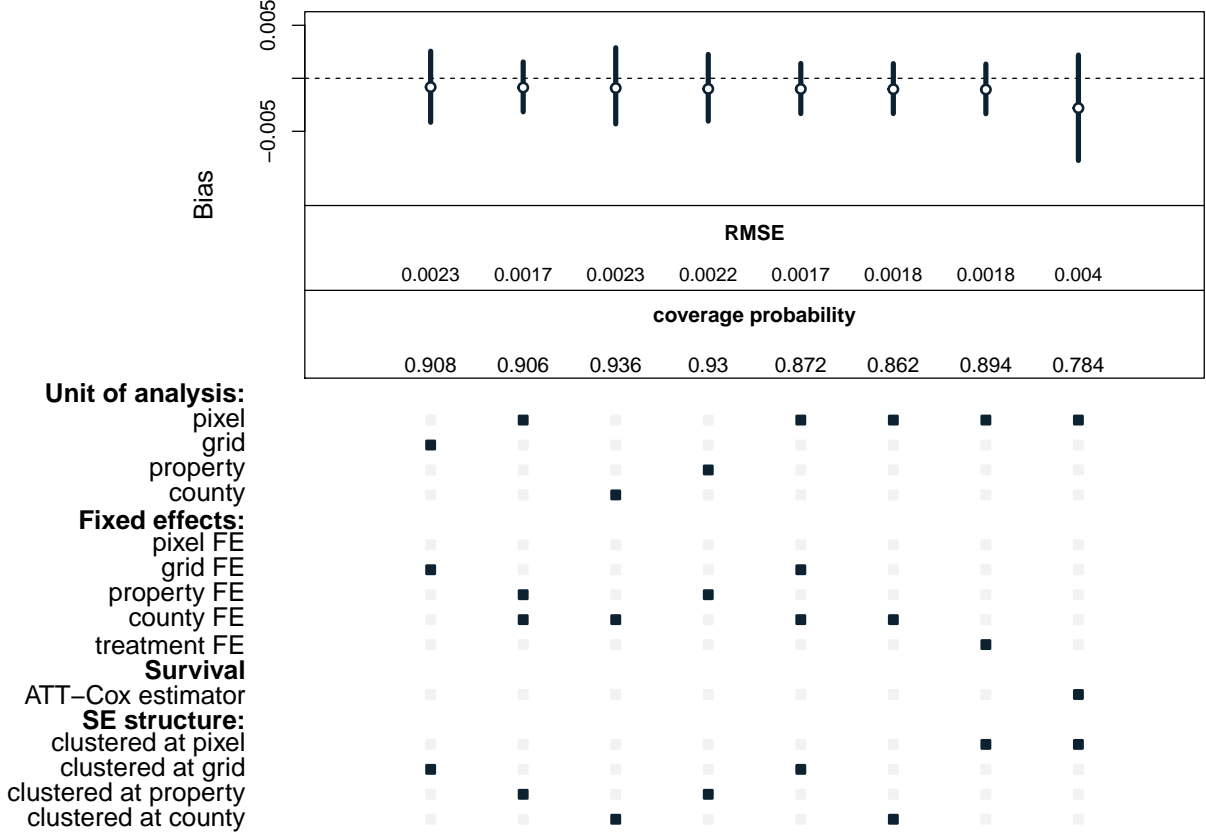


Figure 3: Non-random selection introduces bias into estimates. Specifications are ordered by absolute value of mean bias (least to most bias)

## 6 Selecting the appropriate spatial structure

### 6.1 Model structures that match the spatial process deforestation can reduce bias

Connecting the econometric model to the process by which land use change occurs on the ground has clear benefits for estimation and inference in deforestation impact evaluation. Table 1 shows that researchers often use an arbitrary spatial unit such as a point, pixel, or grid cell as the unit of analysis. While this may be a useful way of structuring data, it can lead to biased results if land use change is determined through a process that is mediated by other spatial structures.

In reality, property level unobservables such as the preferences and resources of a landowner may drive significant variation in land use across a landscape. These differences will impact both treatment effect estimates and coverage probabilities. To illustrate this effect, we introduce property-level perturbations to the returns from forest clearing by varying  $\sigma_p$ , the standard deviation of time-invariant property level disturbances in the DGP.

The introduction of  $\sigma_p$  changes the relative performance of each specification. The traditional DID does not account for the spatial nature of the deforestation process, and in Figure 4, we see that the pixel-level DID begins to suffer in terms of bias, RMSE, and coverage as these property-level unobservables play a larger role in the data generating process.

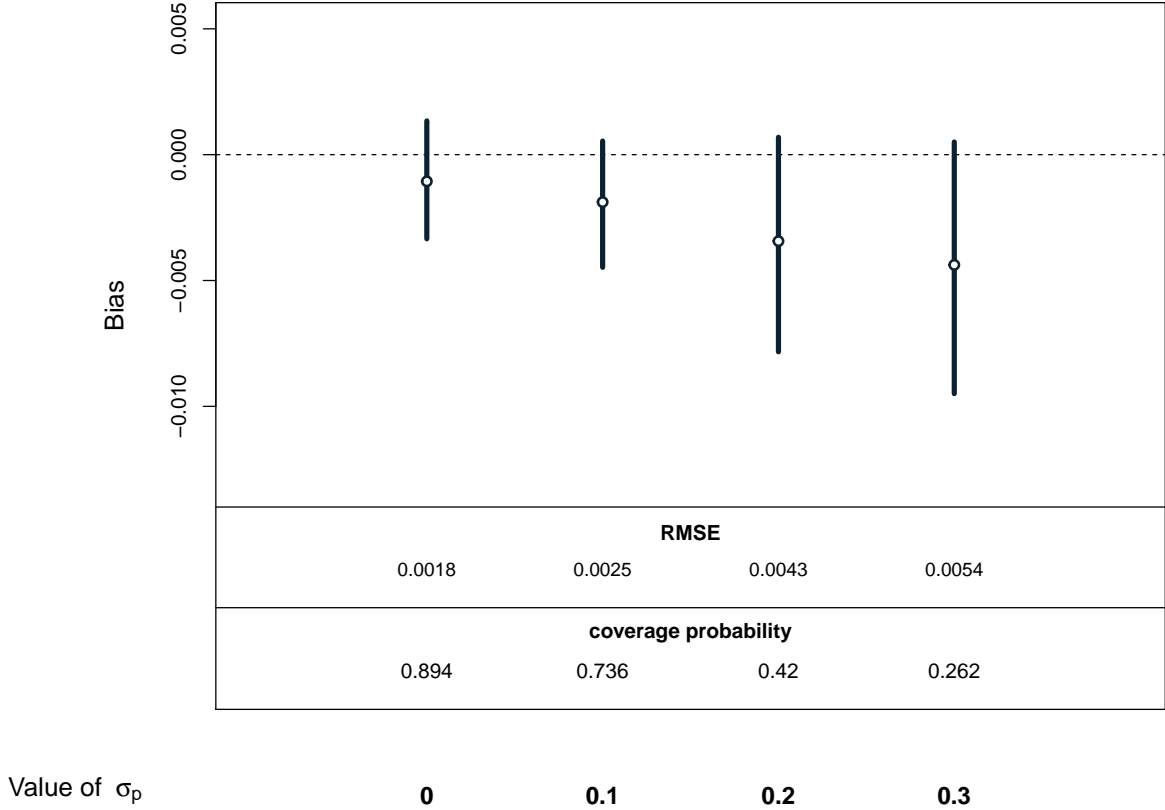


Figure 4: Bias, RMSE, and coverage of the pixel-level DID model as the relative scale of property level disturbances increase

In Figure 5 we see that, by incorporating spatially aggregated units into the model structure, the researcher can reduce bias relative to the simple pixel-level DID in settings where property-level perturbations are relatively large ( $\sigma_p = 0.3$ ). This improvement is apparent across specifications that either control for spatially aggregated fixed effects (Regression 4; left panel) or use a spatially aggregated unit of analysis (Regression 5; right panel). That said, specifications with a spatially aggregated unit of analysis consistently outperform their counterpart with spatially aggregated fixed effects. Further, we see that the scale of aggregation plays a role. Analyses incorporating the property or grid cells near the size of the average property outperform models using larger or smaller scales. Ultimately, all models that incorporate spatial aggregation suffer from relatively little bias and incur less bias than the simple pixel-level DID. In Appendix 10.8, we see that specifications incorporating the property as the unit of analysis continue to outperform other models in alternate parameterizations.

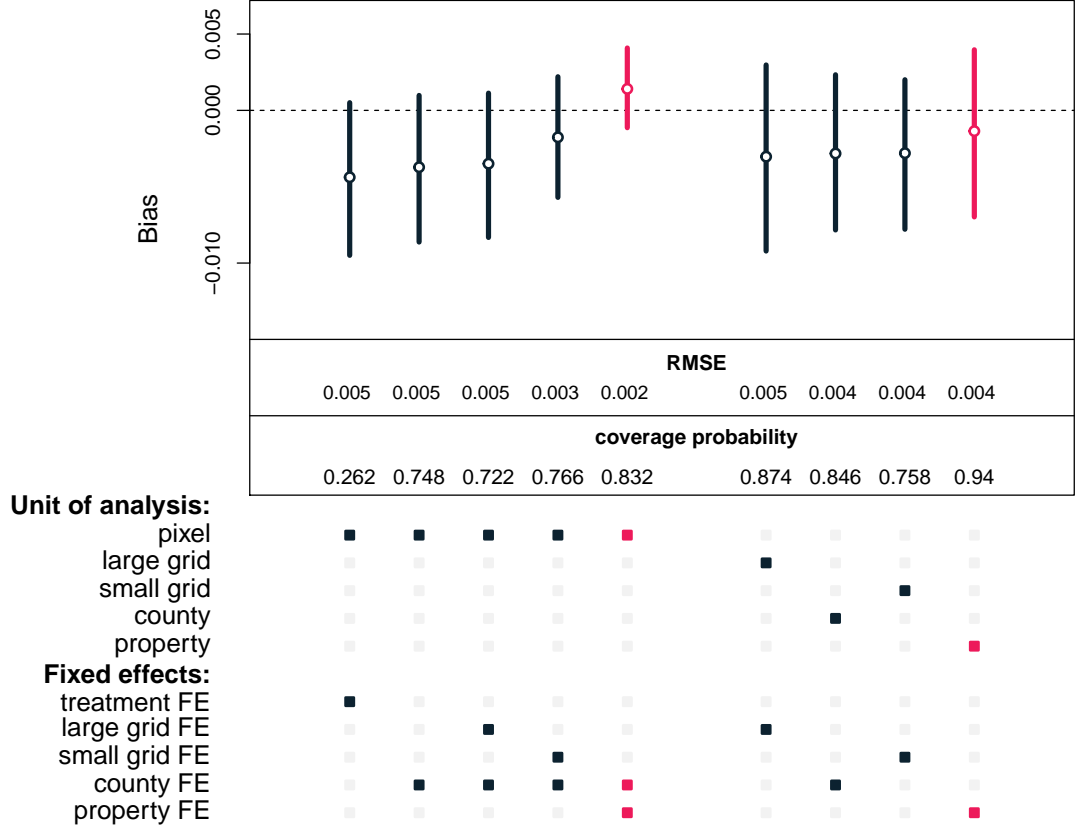


Figure 5: Bias, distribution, and coverage with  $\sigma_p = 0.3$  for specifications with aggregated unit fixed effects (left panel), and specifications with aggregated units of analysis (right panel)

Although spatial aggregation can improve the performance of OLS-based model specifications, there is no clear analog for survival models. Figure 12 in the appendix shows that the performance of  $\widehat{ATT}_{Cox}$  suffers as  $\sigma_p$  increases. When unobserved, spatial processes contribute to the underlying DGP, linear models that effectively control for these processes are likely to outperform survival analysis-based estimates of the impact of conservation interventions.

## 6.2 Weighting by area recovers landscape scale estimates

As researchers transition towards spatially aggregated units of analysis, interpretation of the estimated  $ATT$  can become more complicated. Authors frequently choose to use a set of evenly-sized pixels or grid cells as their preferred units of analysis in order to simplify the interpretation of their estimated  $ATT$  (Alix-Garcia and Gibbs 2017). For example, when researchers estimate a model with pixel-level units of analysis, the coefficient of interest can be interpreted as a population average for all treated, forested pixels. In contrast, if a property is used as the unit of analysis, the coefficient should be interpreted as the effect of the intervention on the characteristic property in the sample. In order to obtain a landscape-scale interpretation, one must weight the regression by the area of each unit of analysis (i.e. property).

Weighting does not have a large impact on bias, RMSE, or coverage probability when the treatment effect is constant across properties (even with property-level unobservables). The use of area weights is likely to be most useful when the treatment effect in the characteristic property differs from the landscape's  $ATT$ . To illustrate this effect, we consider a landscape in which treatment effects are correlated with property size. The full DGP for this case can be found in Appendix 10.5.1.



The treatment effect now varies across properties, and properties with greater areas experience treatment effects of a lower magnitude than smaller properties. For clarity of definitions, we assign treatment at the property level in this subsection. We consider two sample *ATT*s: the landscape *ATT* and the property-level *ATT*. They can be defined as follows:

- $ATT_{ls} = \frac{1}{n_{i:D_i=1}} \sum_{i:D_i=1} (y_{iv2}(1) - y_{iv2}(0))$ , where  $n_{i:D_i=1}$  is the number of treated pixels in the simulated landscape; and
- $ATT_{property} = \frac{1}{n_{v:D_v=1}} \sum_{v:D_v=1} (\frac{1}{n_{iv}} \sum_{i=1}^{n_{iv}} (y_{iv2}(1) - y_{iv2}(0)))$ , where  $n_{v:D_v=1}$  is the number of treated properties in the simulated landscape; and  $n_{iv}$  is the number of pixels in property  $v$ .

Note that neither  $ATT_{ls}$  nor  $ATT_{property}$  can be estimated directly, because  $y(0)_{iv2}$  is not observed for treated units.

Because the treatment is more effective in properties of a smaller size, the treatment effect for the average property is greater than the average treatment effect experienced across the landscape. Figure 6 shows the sample *ATT*s for both the property and landscape. In our simulation,  $ATT_{property} = -0.0136$ , and  $ATT_{landscape} = -0.0092$ . The property-level TWFE regression recovers the *ATT* relative to the characteristic property when area weights are not used and the landscape scale *ATT* when they are used. Researchers should use these area weights when they are interested in the impact of the intervention across the landscape. In cases where the researcher is interested in how an intervention affects incentives at the property level, using these weights may not be necessary.

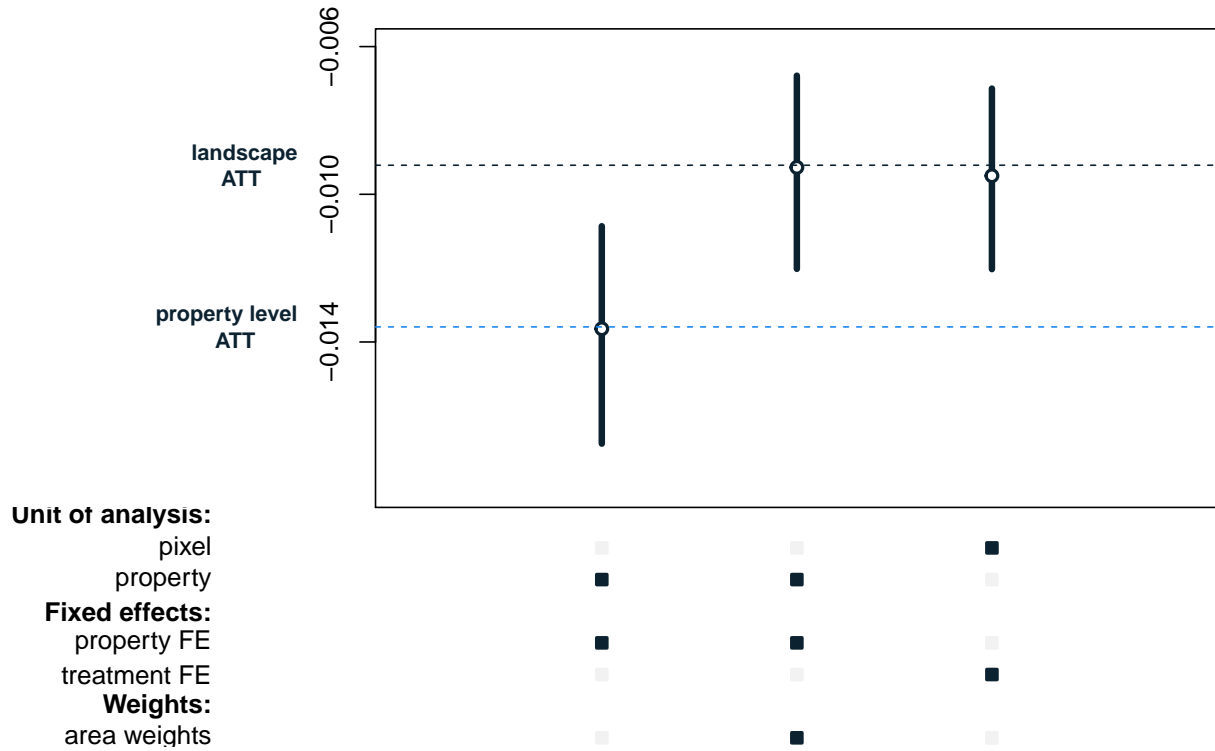


Figure 6: Weighting recovers landscape scale interpretation

## 7 Estimating the *ATT* under staggered treatment

The traditional DID regression applies to settings with two groups and two time periods. However, researchers often use TWFE regressions to exploit variation across groups of units that receive treatment at different

times. Recent work has shown that, in these staggered treatment settings, TWFE regressions identify a weighted average of all possible two-group/two-period DID estimators in the data (Goodman-Bacon 2021).

Further, when estimating the *ATT*, some weights on each group-time treatment effect parameter may actually be negative (de Chaisemartin and D’Haultfoeulle 2020). Newly developed DID estimators seek produce unbiased estimates of the *ATT* in settings with multiple groups and time periods. These estimators do so through a variety of strategies including imputation (e.g., Borusyak, Jaravel, and Spiess 2021), two-stage least squares (e.g., Gardner 2021), and the re-weighting of group-time *ATT*s (e.g., Callaway and Sant’Anna 2020). Some researchers might hope that these new estimators would solve the bias detailed in Section 2.2.

### 7.1 New DID estimators are biased when applied to binary, irreversible outcomes

Although the new class of DID estimators effectively address concerns about staggered treatment timing and heterogeneous treatment effects, they continue to yield biased treatment effect estimates when applied to binary, irreversible, outcomes. To illustrate this, we introduce a setting in which groups of units receive treatment at different times (full DGP can be found in Appendix 10.9). We consider three groups: an early group, a late group, and a never-treated group, where the early and late groups undergo treatment in years three and four, respectively. Each group experiences differing pre-treatment deforestation rates (7%, 4%, and 2% for the early, late, and never-treated groups, respectively) and no time trend. The *ATT* for both treated groups is  $-0.02$ . Common trends is satisfied by construction, and we do not introduce any dynamic effects. Figure 7 shows the observed deforestation rates ( $E[y_{it}^o]$ ) from one iteration of our simulation in this setting.

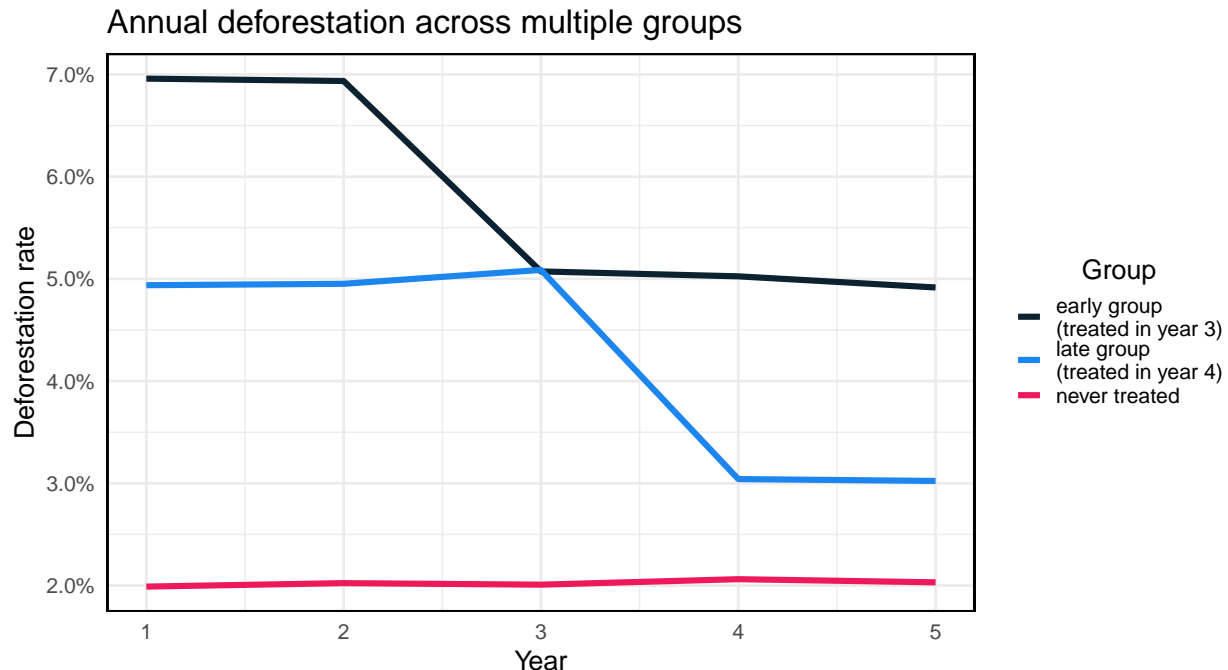


Figure 7: Observed deforestation in simulated example with multiple groups and periods.

The left panel of Figure 8 shows that the estimators developed in Callaway and Sant’Anna (2020), Gardner (2021), and Borusyak, Jaravel, and Spiess (2021) suffer from similar bias to TWFE regressions with pixel unit fixed effects if the pixel is used as the unit of analysis. All methods yield a treatment effect greater than or equal to 0 in all post-treatment periods, reflecting the fact that pre-treatment period deforestation rates are unaccounted for by the estimators. This is particularly clear in the Callaway and Sant’Anna (2020) estimator in which pre-treatment periods are all precisely zero, indicating that the estimator could only compute treatment effects using pixels that survived until the end of the observation period. The right panel

of Figure 8 shows that this bias is eliminated when one uses an aggregated unit of analysis with binary treatment (e.g., county). We do not include pixel-level TWFE regressions with spatially aggregated fixed effects, because most recently developed estimators do not allow for a comparable implementation. Therefore if the researcher wants to obtain a landscape-scale interpretation, they must use an aggregated grid cell or weight the regression by the unit area.

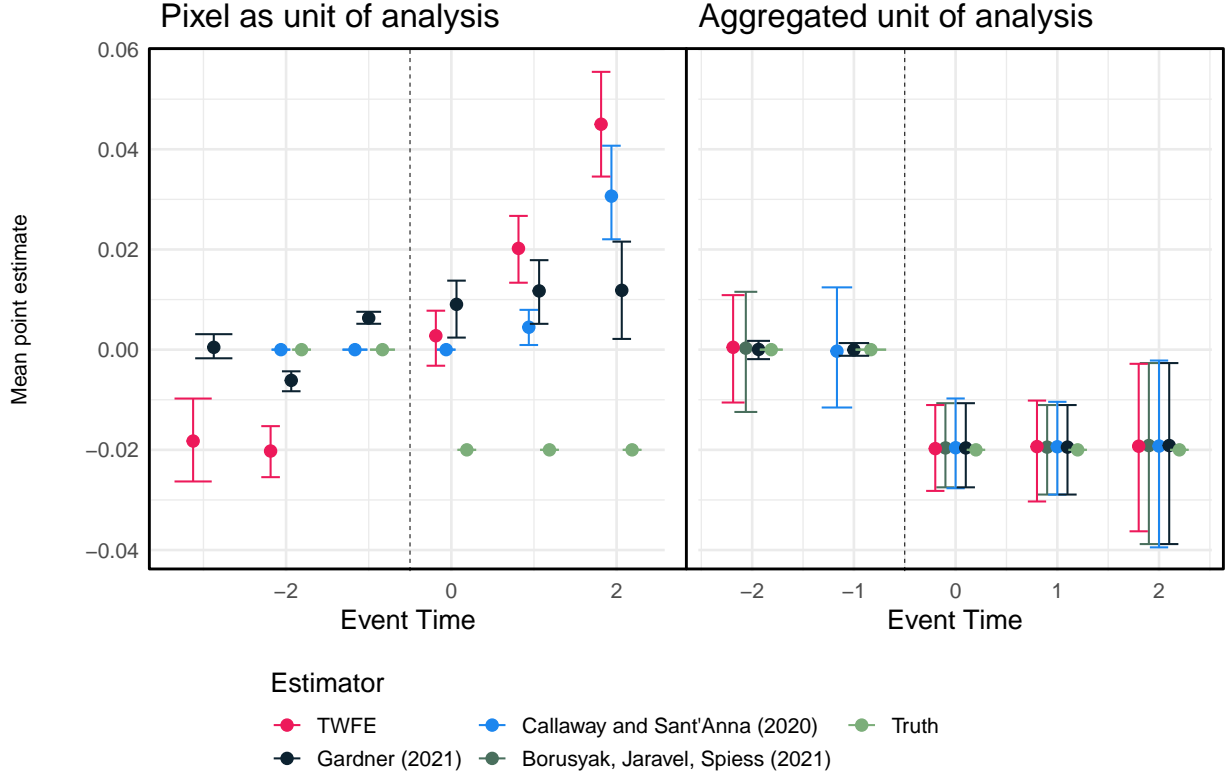


Figure 8: New estimators, similar to TWFE regressions with pixel unit fixed effects, cannot identify ATT with pixel as unit of analysis

## 7.2 New DID estimators can yield unbiased estimates of heterogeneous treatment effects

Finally, we examine the performance of the new DID estimators relative to a traditional TWFE regression when treatment effects vary across time and across groups. We again work with an early, late and untreated group. The full parameterization and DGP can be found in Appendix 10.9.1. Figure 9 shows deforestation rates in each of the three groups through time.

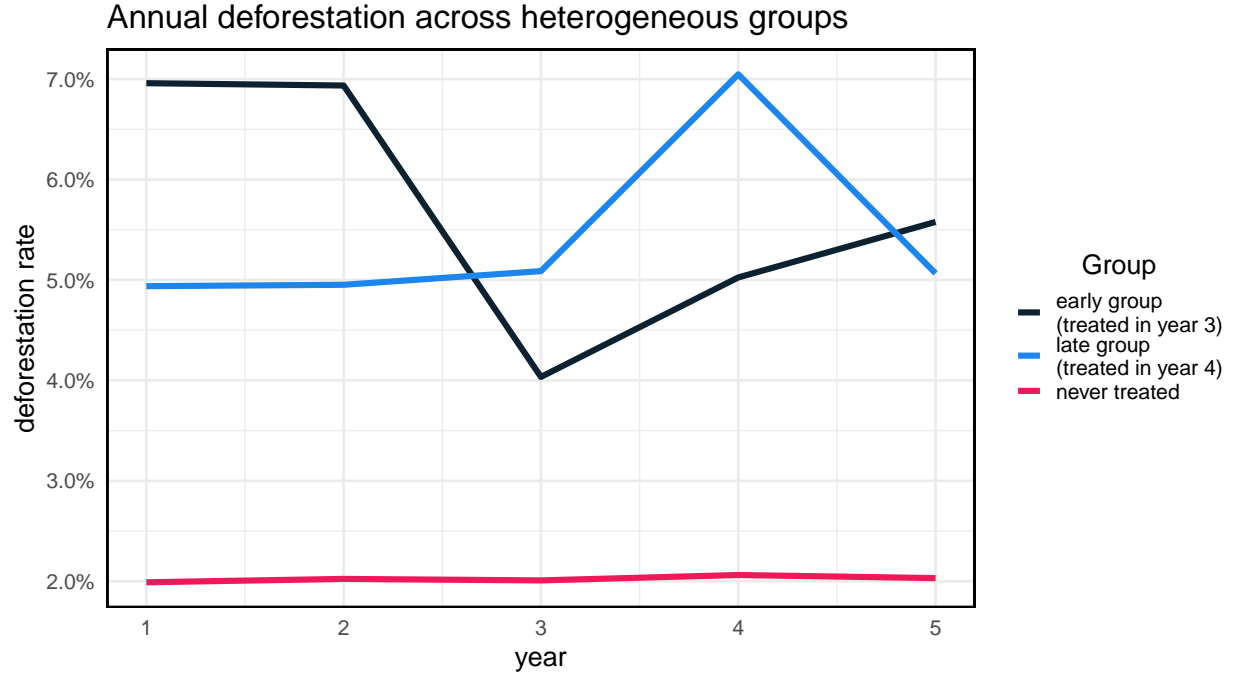


Figure 9: Observed deforestation in simulated example when treatment effects vary across groups and through time

Figure 10 shows the event study estimates produced by each of the three estimators as well as the “truth” for both pixel and county-level analyses. Again, none of the estimators yield the *ATT* with pixel-level analyses. In the county-level estimates, we see that the newer estimators slightly outperform the TWFE estimator. This is evidence of the weighting that has become a concern with TWFE estimators in these type of settings. While TWFE estimates represent a weighted average of all possible 2x2 DID estimates, the weights may not always be intuitive (Goodman-Bacon 2021). In contrast, newer estimators do not suffer from this concern.

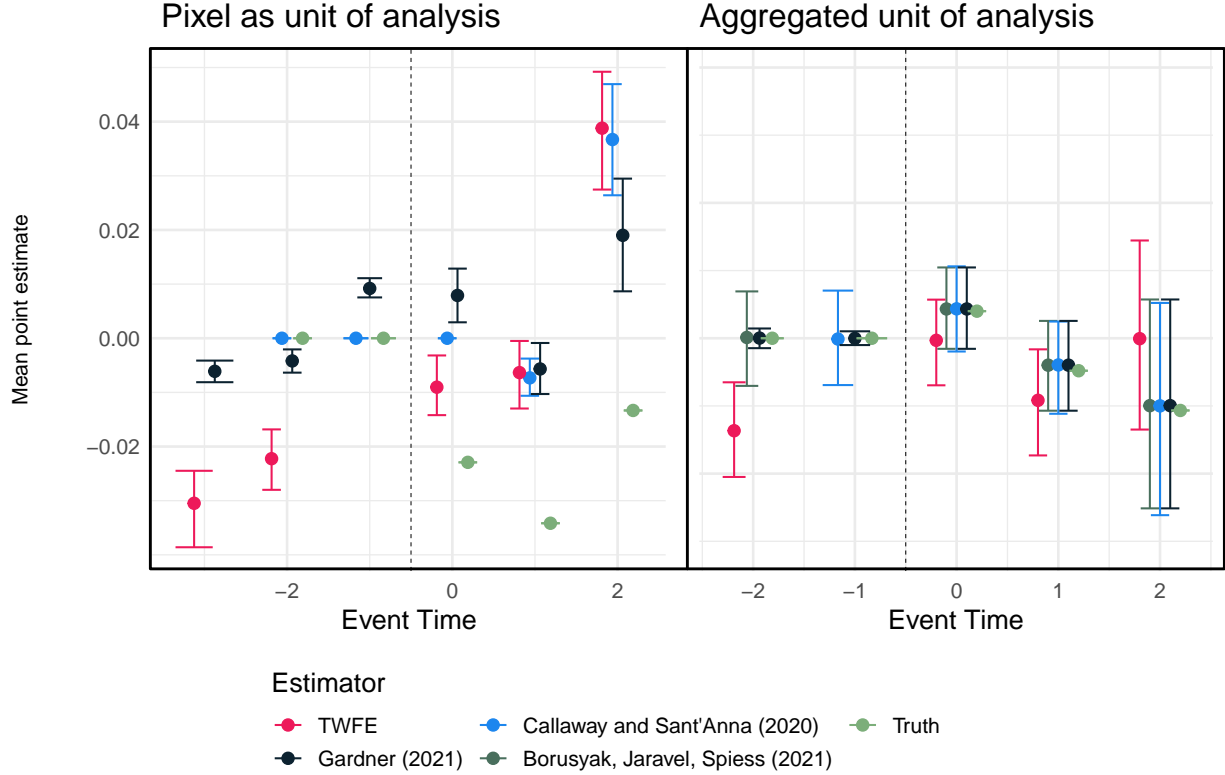


Figure 10: TWFE regressions suffer from weighting concerns when treatment effects vary across groups and through time

## 8 Conclusions

By applying econometric methods of causal inference to remotely-sensed measurements of land use change, researchers have advanced society’s understanding of the impacts of conservation interventions. However, this interdisciplinary research community has insufficiently considered how the data generating processes underpinning land use change and its measurement might affect the performance of standard econometric models. The analytical proofs and simulations presented in this paper highlight that the conclusions made in many prior studies may be biased.

Researchers can take several practical steps in the design of their econometric models to more accurately measure the impacts of conservation policies. First, despite past guidance to the contrary, researchers should recognize that pixel-level, TWFE models are unable to yield unbiased estimates of a policy’s impact when applied to irreversible, binary outcomes. Researchers can easily avoid this bias by aggregating either the units of observation, or the scale at which fixed effects are estimated. Second, while survival models provide an appealing empirical framework with which to study deforestation, past studies have typically overlooked implicit assumptions made when applying survival models to the difference in differences research design. To resolve this challenge, we propose a new, survival-based estimation procedure that enables researchers to recover an unbiased estimate of the *ATT* under the traditional parallel trends assumption. Finally, we provide evidence suggesting that researchers should seek to align the structure of their econometric models to match the real-world units at which land use decisions are being made. For example, if unobservable, property-level characteristics are thought to be an important driver of deforestation, the inclusion of property-level fixed effects can improve the accuracy of model estimates and inference. Ultimately, context plays a role in what is feasible, and researchers should make clear the limits to their impact evaluation strategy.

Our research contributes to an emerging literature calling for a deeper understanding of the interdependencies

between the creation of remotely sensed data, and the interpretation of that data through econometric models (Jain 2020; Alix-Garcia and Millimet, n.d.). However, we have largely abstracted away from prior concerns that characteristics of the remote sensing data collection process, including sensor properties, atmospheric conditions, and image processing methods, may influence the structure of output data products. Of particular concern is the potential for these processes to give rise to non-classical measurement error, which can lead to biased estimates of the *ATT* (Wooldridge 2010). Importantly, our study implicitly assumes that pixel-level outcomes are measured without any non-classical measurement error. Fortunately, Alix-Garcia and Millimet (n.d.) propose a solution for unbiased causal inference in the context of a remotely sensed binary outcome in which misclassification is non-classical.

Although this paper focuses upon efforts to identify the impacts of conservation policies on deforestation, the lessons we highlight are relevant to a wider audience. Specifically, many of our key findings apply to diverse settings in which the outcome of interest represents an irreversible, binary event. For example, studies exploring the drivers of unidirectional technology adoption may fall prey to the same issues we have identified in the context of deforestation. Moving forward, researchers should carefully consider the underlying structure of their data, and ensure that their chosen models minimize bias and allow inference at expected levels of confidence. Misleading causal inference may lead policymakers to avoid effective policies, or to adopt interventions that worsen environmental damages.

## Acknowledgements

We thank Kelsey Jack, Andrew Plantinga, and Jennifer Alix-Garcia for useful comments on early versions of this paper. We are grateful for feedback received at the BIOECON XXII conference. We thank the University of California, Santa Barbara’s Academic Senate for a Faculty Research Grant that supported this work. This paper is based upon work supported by the National Aeronautics and Space Administration under Grant No. 80NSSC20K1489 issued through the Land Cover and Land Use Change Program. This paper contributes to the Global Land Programme.

## 9 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, and Daniel L. Millimet. n.d. “Remotely Incorrect? Accounting for Nonclassical Measurement in Satellite Data on Deforestation,” 56.
- 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>.
- 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 Llacayo, 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>.
- 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>.
- Bollinger, Bryan, Kenneth Gillingham, Steven Sexton, and A Justin Kirkpatrick. n.d. “Visibility and Peer in Uence in Durable Good Adoption,” 65.
- Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. 2021. “Revisiting Event Study Designs: Robust and Efficient Estimation.” *arXiv:2108.12419 [Econ]*, August. <http://arxiv.org/abs/2108.12419>.
- Busch, Jonah, Kalifi Ferretti-Gallon, Jens Engelmann, Max Wright, Kemen G. Austin, Fred Stolle, Svetlana Turubanova, et al. 2015. “Reductions in Emissions from Deforestation from Indonesia’s Moratorium on New Oil Palm, Timber, and Logging Concessions.” *Proceedings of the National Academy of Sciences* 112 (5): 1328–33. <https://doi.org/10.1073/pnas.1412514112>.
- 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>.
- Callaway, Brantly, and Pedro H. C. Sant’Anna. 2020. “Difference-in-Differences with Multiple Time Periods.” *Journal of Econometrics*, December, S0304407620303948. <https://doi.org/10.1016/j.jeconom.2020.12.001>.
- 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>.
- de Chaisemartin, Clément, and Xavier D’Haultfœuille. 2020. “Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects.” *American Economic Review* 110 (9): 2964–96. <https://doi.org/10.1257/aer.20181169>.
- 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.
- 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>.
- Friedman, Jed, and Norbert Schady. 2013. “How Many Infants Likely Died in Africa as A Result of the 2008-2009 Global Financial Crisis?: Excess Infant Mortality in Africa Due to the Global Financial Crisis.” *Health Economics* 22 (5): 611–22. <https://doi.org/10.1002/hec.2818>.
- Gardner, John. 2021. “Two-Stage Differences in Differences.” *Working Paper*, April, 34.
- Goodman-Bacon, Andrew. 2021. “Difference-in-Differences with Variation in Treatment Timing.” *Journal of Econometrics*, June, S0304407621001445. <https://doi.org/10.1016/j.jeconom.2021.03.014>.
- 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>.

- Heilmayr, Robert, Lisa L. Rausch, Jacob Munger, and Holly K. Gibbs. 2020. "Brazil's Amazon Soy Moratorium Reduced Deforestation." *Nature Food* 1 (12): 801–10. <https://doi.org/10.1038/s43016-020-00194-5>.
- 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>.
- 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>.
- Kerr, Suzi, Shuguang Liu, Alexander S. P. Pfaff, and R. Flint Hughes. 2003. "Carbon Dynamics and Land-Use Choices: Building a Regional-Scale Multidisciplinary Model." *Journal of Environmental Management* 69 (1): 25–37. [https://doi.org/10.1016/S0301-4797\(03\)00106-3](https://doi.org/10.1016/S0301-4797(03)00106-3).
- Koch, Nicolas. n.d. "Agricultural Productivity and Forest Conservation: Evidence from the Brazilian Amazon," 22.
- Larsen, Ashley E., Kyle Meng, and Bruce E. Kendall. 2019. "Causal Analysis in Control Impact Ecological Studies with Observational Data." *Methods in Ecology and Evolution* 10 (7): 924–34. <https://doi.org/10.1111/2041-210X.13190>.
- Mastrobuoni, Giovanni, and Paolo Pinotti. 2015. "Legal Status and the Criminal Activity of Immigrants." *American Economic Journal: Applied Economics* 7 (2): 175–206. <https://doi.org/10.1257/app.20140039>.
- 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>.
- Panlasigui, Stephanie. 2018. "Impacts of Certification, Uncertified Concessions, and Protected Areas on Forest Loss in Cameroon, 2000 to 2013." *Biological Conservation*, 7.
- 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).
- Pfaff, Alexander S. P. 1999. "What Drives Deforestation in the Brazilian Amazon?" *Journal of Environmental Economics and Management* 37 (1): 26–43. <https://doi.org/10.1006/jeem.1998.1056>.
- Pfaff, Alexander S. P., and G. Arturo Sanchez-Azofeifa. 2004. "Deforestation Pressure and Biological Reserve Planning: A Conceptual Approach and an Illustrative Application for Costa Rica." *Resource and Energy Economics* 26 (2): 237–54. <https://doi.org/10.1016/j.reseneeco.2003.11.009>.
- 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](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>.

Sales, Vilane G., Eric Strobl, and Robert J. R. Elliott. 2022. “Cloud Cover and Its Impact on Brazils Deforestation Satellite Monitoring Program: Evidence from the Cerrado Biome of the Brazilian Legal Amazon.” *Applied Geography* 140 (March): 102651. <https://doi.org/10.1016/j.apgeog.2022.102651>.

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/consnl.12720>.

Wooldridge, Jeffrey M. 2010. “Econometric Analysis of Cross Section and Panel Data, Second Edition.” The MIT Press.

## 10 Appendix

### 10.1 proof showing pixel level TWFE regression models with pixel fixed effects do not identify $ATT$

In settings with a binary and unrepeatable outcome variable, the commonly used pixel level TWFE model yields the post-treatment difference in outcomes (single difference), rather than the desired  $ATT$ .

proof:

Consider a two-period setting ( $t = 1, 2$ ) with multiple pixels indexed by  $i$ . We observe  $y_{it}$ , the realized deforestation occurring in each pixel in each time period. Some units are exposed to a policy treatment ( $D_i = 1$ ) in the second time period ( $t_0 = 2$ ). Using the potential outcomes framework, we consider the potential outcomes for each observation as  $y_{i,2}(D_i)$ . In this notation, the treatment effect for unit  $i$  can be defined as

$$\tau_i = y_{i,2}(1) - y_{i,2}(0)$$

In many deforestation maps generated through remote sensing, deforestation is represented as a binary indicator. Furthermore, deforestation is only observed once for a given location since these data products typically do not monitor the timing of reforestation. Given these constraints, deforestation is typically (Alix-Garcia and Gibbs (2017); Jones and Lewis (2015)) represented as a binary, unrepeatable variable taking the following values:

$$y_{it} = \begin{cases} 0 & \text{the pixel has never been deforested} \\ 1 & \text{the pixel was deforested in year } t \\ NAN & \text{the pixel was deforested in a year } < t \end{cases}$$

The traditional two-way fixed effects model seeks to estimate this effect using the following regression specification:

$$y_{it} = \alpha + \beta_{TWFE} \times D_i \mathbb{1}\{t \geq t_0\} + \gamma_i + \eta_{t=2} + u_{it}, \text{ for } t = 1, 2$$

Taking into account the data structure of  $y_{it}$  and our two-period case, we are left with:

$$y_{i1} = \alpha + \gamma_i + u_{i1}$$

and

$$y_{i2} = \begin{cases} \alpha + \beta_{TWFE} \times D_i + \gamma_i + \eta_{t=2} + u_{i2} & y_{i1} = 0 \\ NAN & y_{i1} \neq 0 \end{cases}$$

First differencing,

$$y_{i2} - y_{i1} = \begin{cases} (\alpha + \beta_{TWFE} \times D_i + \gamma_i + \eta_{t=2} + u_{i2}) - (\alpha + \gamma_i + u_{i1}) & y_{i1} = 0 \\ NAN & y_{i1} \neq 0 \end{cases}$$

Focusing on the first case, where  $y_{i1} = 0$

$$\begin{aligned} y_{i2} - y_{i1} &= (\alpha + \beta_{TWFE} \times D_i + \gamma_i + \eta_{t=2} + u_{i2}) - (\alpha + \gamma_i + u_{i1}) \\ &= \beta_{TWFE} \times D_i + \eta_{t=2} + \Delta u_i \end{aligned}$$

The general expression can be restated as:

$$y_{i2} - y_{i1} = \begin{cases} \beta_{TWFE} \times D_i + \eta_{t=2} + \Delta u_i & y_{i1} = 0 \\ NAN & y_{i1} \neq 0 \end{cases}$$

With binary treatment ( $D_i$ ),  $\hat{\beta}_{TWFE}$ , the regression's estimate of  $\beta_{TWFE}$  can be expressed as the double difference in mean outcomes across treated / untreated units, and across the two time periods:

$$\hat{\beta} = \frac{1}{n_{i:D_i=1}} \sum_{i:D_i=1} y_{i2} - \frac{1}{n_{i:D_i=1}} \sum_{i:D_i=1} y_{i1} - \left( \frac{1}{n_{i:D_i=0}} \sum_{i:D_i=0} y_{i2} - \frac{1}{n_{i:D_i=0}} \sum_{i:D_i=0} y_{i1} \right)$$

However, this is only valid when  $y_{i1} = 0$ . As a result, we can restate as:

$$\hat{\beta}_{TWFE} = \frac{1}{n_{i:D_i=1}} \sum_{i:D_i=1} y_{i2} - 0 - \left( \frac{1}{n_{i:D_i=0}} \sum_{i:D_i=0} y_{i2} - 0 \right)$$

Applying the potential outcomes notation to indicate whether we see the treated or untreated outcome:

$$\hat{\beta}_{TWFE} = \frac{1}{n_{i:D_i=1}} \sum_{i:D_i=1} y_{i2}(1) - \frac{1}{n_{i:D_i=0}} \sum_{i:D_i=0} y_{i2}(0)$$

Adding and subtracting  $\frac{1}{n_{i:D_i=1}} \sum_{i:D_i=1} y_{i2}(0)$  gives:

$$\begin{aligned} \hat{\beta}_{TWFE} &= \frac{1}{n_{i:D_i=1}} \sum_{i:D_i=1} y_{i2}(1) - y_{i2}(0) \\ &\quad + \frac{1}{n_{i:D_i=0}} \sum_{i:D_i=1} y_{i2}(0) - \frac{1}{n_{i:D_i=0}} \sum_{i:D_i=0} y_{i2}(0) \end{aligned}$$

And finally, taking the expectation gives:

$$\begin{aligned} E[\hat{\beta}_{TWFE}] &= ATT + Diff \\ \beta_{TWFE} &= ATT + Diff \end{aligned}$$

■

## 10.2 TWFE with pixel fixed effects is equivalent to running TWFE on dataset after dropping all pixels that were deforested in the first period

Table 3 shows that the coefficient of interest from the TWFE regression with pixel fixed effects is equivalent to the same regression on a dataset where all pixels deforested in the first period are dropped. We generate a panel of two years pre-and post treatment in order to show that the estimated coefficient is not equivalent to the ex-post difference in means, although both should be estimating the same parameter. This exercise provides further evidence that this commonly used TWFE regression is not adequate for recovering the *ATT*.

Table 3: TWFE with pixel fixed effects is equivalent to TWFE on dataset with all first-period deforested pixels dropped completely from the dataset

model	Bias	RMSE	0.25 to 0.75 quantile
DID	0.0001034	0.0015268	-0.00088 , 0.0011
ex-post difference in means	0.0300441	0.0300609	0.02935 , 0.03074
TWFE	0.0359502	0.0359744	0.03501 , 0.03684
TWFE on dataset dropping deforested pixels prior to treatment	0.0359502	0.0359744	0.03501 , 0.03684

## 10.3 Keeping pixels in periods after they are first deforested is not a viable solution

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. Further, many data products do not allow for the monitoring of forest regrowth. In the panel therefore, it is likely that in the periods after a pixel is first realized as deforested, subsequent observations of the pixel will also observe the pixel as deforested.

The logic for dropping binary pixels after they first become deforested 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 occurred, 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. As such, the coefficient cannot recover the *ATT*.

### 10.3.1 Analytical expression of bias in two-period two-group case when pixels are not dropped from the panel

The DID estimand is

$$\begin{aligned}
& E[y_{it}|t \geq t_0, D_i = 1] - E[y_{it}|t < t_0, D_i = 1] - (E[y_{it}|t \geq t_0, D_i = 0] - E[y_{it}|t < t_0, D_i = 0]) \\
& = P(y_{it} = 1|t \geq t_0, D_i = 1) - P(y_{it} = 1|t < t_0, D_i = 1) - \\
& \quad (P(y_{it} = 1|t \geq t_0, D_i = 0) - P(y_{it} = 1|t < t_0, D_i = 0))
\end{aligned}$$

In the case of an irreversible binary outcome, the first and third terms can be reexpressed, giving

$$\begin{aligned}
& P(y_{it} = 1|t \geq t_0, D_i = 1) \cup P(y_{it} = 1|t < t_0, D_i = 1) - P(y_{it} = 1|t < t_0, D_i = 1) - \\
& \quad (P(y_{it} = 1|t \geq t_0, D_i = 0) \cup P(y_{it} = 1|t < t_0, D_i = 0) - P(y_{it} = 1|t < t_0, D_i = 0)) \\
& = P(y_{it} = 1|t \geq t_0, D_i = 1) + P(y_{it} = 1|t < t_0, D_i = 1) - P(y_{it} = 1|t \geq t_0, D_i = 1) \cap P(y_{it} = 1|t < t_0, D_i = 1) - \\
& \quad P(y_{it} = 1|t < t_0, D_i = 1) - \\
& \quad (P(y_{it} = 1|t \geq t_0, D_i = 0) + P(y_{it} = 1|t < t_0, D_i = 0) - P(y_{it} = 1|t \geq t_0, D_i = 0) \cap P(y_{it} = 1|t < t_0, D_i = 0))
\end{aligned}$$

### 10.3.2 Monte Carlo evidence

Figure 11 demonstrates 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 eliminates this bias in the DID model, as seen in Figure 11.

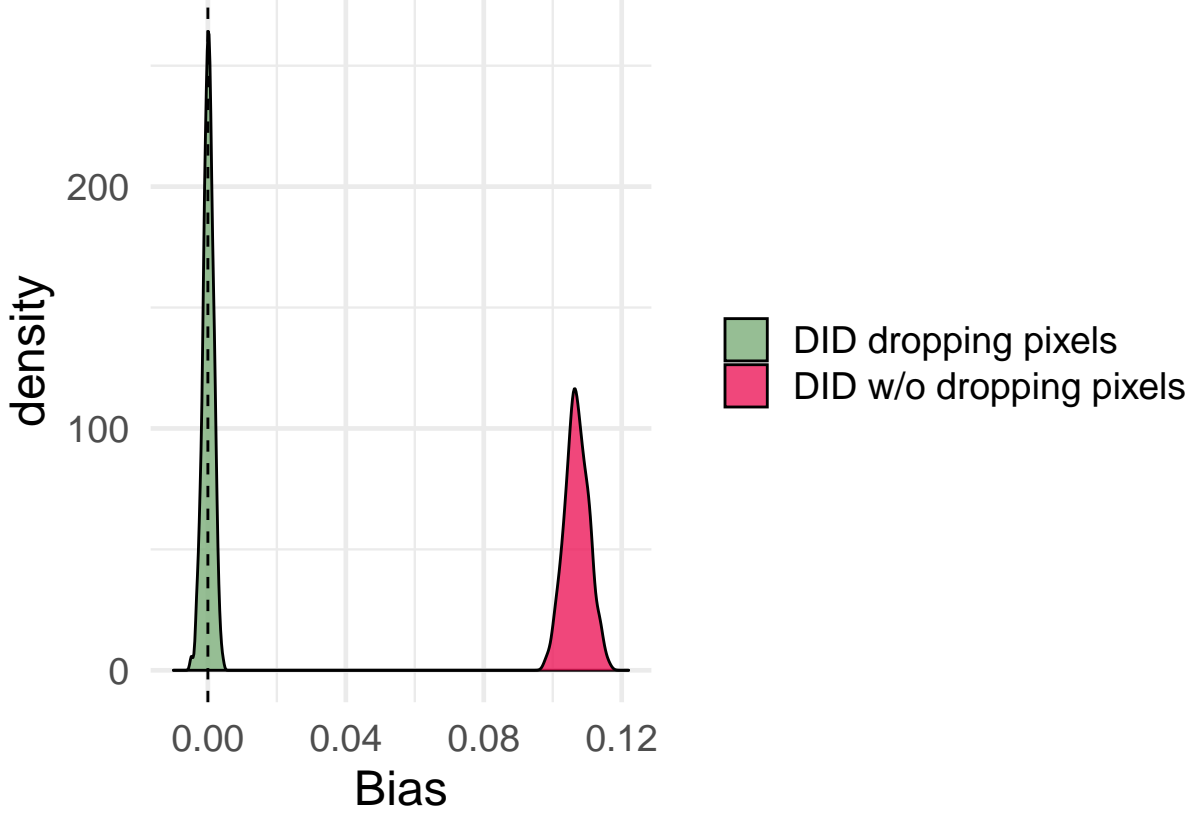


Figure 11: Distribution of DID estimates leaving deforested pixels in the panel and of DID estimates dropping deforested pixels. Note that leaving deforested pixels in the panel incurs severe bias.

### 10.4 Cox PH DID identifies $HRTT$ when proportional trends assumption holds

The exponentiated coefficient on the interaction between two binary variables,  $D_i$  and  $\mathbb{1}\{t \geq t_0\}$ ,  $\exp(\beta_{coxDID})$ , is expressed as the ratio of the two pre-post hazard rate ratios across the two groups:

$$\exp(\beta_{coxDID}) = \frac{E[y_{ivt}|t \geq t_0, D_i = 1]/E[y_{ivt}|t < t_0, D_i = 1]}{E[y_{ivt}|t \geq t_0, D_i = 0]/E[y_{ivt}|t < t_0, D_i = 0]}$$

Introducing potential outcomes and simplifying:

$$\exp(\beta_{coxDID}) = \frac{E[y_{ivt}(1)|t \geq t_0, D_i = 1]E[y_{ivt}(0)|t < t_0, D_i = 0]}{E[y_{ivt}(0)|t \geq t_0, D_i = 0]E[y_{ivt}(0)|t < t_0, D_i = 1]}$$

Now, operating under Assumption 1 (Proportional Trends) and substituting for the right-hand side of (7):

$$\begin{aligned} \exp(\beta_{coxDID}) &= \frac{E[y_{ivt}(1)|t \geq t_0, D_i = 1]E[y_{ivt}(0)|t < t_0, D_i = 1]}{E[y_{ivt}(0)|t \geq t_0, D_i = 1]E[y_{ivt}(0)|t < t_0, D_i = 1]} \\ &= HRRT \end{aligned}$$

■

## 10.5 Initial Monte Carlo parameter to $\beta$ coefficient mapping

The following five parameters and their definitions inform the simulation parameterizations.

$$\begin{aligned} baseline_0 &= E[y_{it}(0)|t < t_0, D_i = 0] \\ baseline_1 &= E[y_{it}(0)|t < t_0, D_i = 1] \\ trend_0 &= E[y_{it}(0)|t \geq t_0, D_i = 0] - E[y_{it}(0)|t < t_0, D_i = 0] \\ trend_1 &= E[y_{it}(0)|t \geq t_0, D_i = 1] - E[y_{it}(0)|t < t_0, D_i = 1] \\ ATT &= E[y_{it}(1) - y_{it}(0)|t \geq t_0, D_i = 1] \end{aligned}$$

Note the following constraints on the parameters:

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

The parameters can be expressed as follows:

$$\begin{aligned} ATT &= E[y_{it}(1) - y_{it}(0)|t \geq t_0, D_i = 1] \\ &= E[y_{it}(1)|t \geq t_0, D_i = 1] - E[y_{it}(0)|t \geq t_0, D_i = 1] \\ &= P(y_{it}(1) = 1|t \geq t_0, D_i = 1) - P(y_{it}(0) = 1|t \geq t_0, D_i = 1) \\ &= P(y_{it}^*(1) > 0|t \geq t_0, D_i = 1) - P(y_{it}^*(0) > 0|t \geq t_0, D_i = 1) \\ &= P(\beta_0 + \beta_1 + \beta_{2,1} + \beta_3 + \alpha_i + u_{it} > 0) - P(\beta_0 + \beta_1 + \beta_{2,1} + \alpha_i + u_{it} > 0) \\ &= P(-\alpha_i - u_{it} < \beta_0 + \beta_1 + \beta_{2,1} + \beta_3) - P(-\alpha_i - u_{it} < \beta_0 + \beta_1 + \beta_{2,1}) \\ &= F(\beta_0 + \beta_1 + \beta_{2,1} + \beta_3) - F(\beta_0 + \beta_1 + \beta_{2,1}) \end{aligned}$$

$$\begin{aligned} trend_0 &= E[y_{it}(0)|t \geq t_0, D_i = 0] - E[y_{it}(0)|t < t_0, D_i = 0] \\ &= P(y_{it}(0) = 1|t \geq t_0, D_i = 0) - P(y_{it}(0) = 1|t < t_0, D_i = 0) \\ &= P(y_{it}^*(0) > 0|t \geq t_0, D_i = 0) - P(y_{it}^*(0) > 0|t < t_0, D_i = 0) \\ &= \frac{(1 - P(y_{it}^*(0) > 0|t < t_0, D_i = 0))P(y_{it}^*(0) > 0|t \geq t_0, D_i = 0)}{(1 - P(y_{it}^*(0) > 0|t < t_0, D_i = 0))} - P(y_{it}^*(0) > 0|t < t_0, D_i = 0) \\ &= P(-\alpha_i - u_{it} < \beta_0 + \beta_{2,0}) - P(-\alpha_i - u_{it} < \beta_0) \\ &= F(\beta_0 + \beta_{2,0}) - F(\beta_0) \end{aligned}$$

$$\begin{aligned}
trend_1 &= E[y_{it}(0)|t \geq t_0, D_i = 1] - E[y_{it}(0)|t < t_0, D_i = 1] \\
&= P(y_{it}(0) = 1|t \geq t_0, D_i = 1) - P(y_{it}(0) = 1|t < t_0, D_i = 1) \\
&= P(y_{it}^*(0) > 0|t \geq t_0, D_i = 1 \cap y_{it}^*(0) < 0|t < t_0, D_i = 1) - P(y_{it}^*(0) > 0|t < t_0, D_i = 1) \\
&= P(-\alpha_i - u_{it} < \beta_0 + \beta_1 + \beta_{2,1}) - P(-\alpha_i - u_{it} < \beta_0 + \beta_1) \\
&= F(\beta_0 + \beta_1 + \beta_{2,1}) - F(\beta_0 + \beta_1)
\end{aligned}$$

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

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

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

Now solving for the  $\beta$  coefficients:

solving for  $\beta_0$

$$\begin{aligned}
baseline_0 &= F(\beta_0) \\
\Leftrightarrow \\
\beta_0 &= F^{-1}(baseline_0)
\end{aligned}$$

solving for  $\beta_1$

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

solving for  $\beta_{2,0}$

$$\begin{aligned}
trend &= F(\beta_0 + \beta_{2,0}) - F(\beta_0) \\
\Leftrightarrow \\
trend + baseline_0 &= F(\beta_0 + \beta_{2,0}) \\
\Leftrightarrow \\
F^{-1}(trend + baseline_0) &= \beta_0 + \beta_{2,0} \\
\Leftrightarrow \\
\beta_{2,0} &= F^{-1}(trend + baseline_0) - \beta_0
\end{aligned}$$

solving for  $\beta_{2,1}$

$$\begin{aligned}
& trend = F(\beta_0 + \beta_1 + \beta_{2,1}) - F(\beta_0 + \beta_1) \\
& \Leftrightarrow \\
& trend + baseline_1 = F(\beta_0 + \beta_1 + \beta_{2,1}) \\
& \Leftrightarrow \\
& F^{-1}(trend + baseline_1) = \beta_0 + \beta_1 + \beta_{2,1} \\
& \Leftrightarrow \\
& \beta_{2,1} = F^{-1}(trend + baseline_1) - \beta_0 - \beta_1
\end{aligned}$$

solving for  $\beta_3$

$$\begin{aligned}
& ATT = F(\beta_0 + \beta_1 + \beta_{2,1} + \beta_3) - F(\beta_0 + \beta_1 + \beta_{2,1}) \\
& \Leftrightarrow \\
& ATT + F(\beta_0 + \beta_1 + \beta_{2,1}) = F(\beta_0 + \beta_1 + \beta_{2,1} + \beta_3) \\
& \Leftrightarrow \\
& F^{-1}(ATT + F(\beta_0 + \beta_1 + \beta_{2,1})) = \beta_0 + \beta_1 + \beta_{2,1} + \beta_3 \\
& \Leftrightarrow \\
& \beta_3 = F^{-1}(ATT + F(\beta_0 + \beta_1 + \beta_{2,1})) - (\beta_0 + \beta_1 + \beta_{2,1})
\end{aligned}$$

### 10.5.1 when treatment effects are correlated with property size

$$\begin{aligned}
ATT &= E(\beta_0 + \beta_1 + \beta_{2,1} + \beta_3) - E(\beta_0 + \beta_1 + \beta_{2,1}) \\
&= P(-\alpha_i - u_{it} - \beta_3 < \beta_0 + \beta_1 + \beta_{2,1} + \mu) - P(-\alpha_i - u_{it} < \beta_0 + \beta_1 + \beta_{2,1}) \\
&= G(\beta_0 + \beta_1 + \beta_{2,1} + \mu) - F(\beta_0 + \beta_1 + \beta_{2,1})
\end{aligned}$$

, where  $\beta_3 \sim N(\mu, \sigma_{te}^2)$  and  $G()$  is the CDF of a  $N(0, \sigma_a^2 + \sigma_u^2 + \sigma_p^2 + \sigma_{te}^2)$  and

$$\begin{aligned}
& ATT = G(\beta_0 + \beta_1 + \beta_{2,1}) - F(\beta_0 + \beta_1 + \beta_{2,1}) \\
& \Leftrightarrow \\
& ATT + F(\beta_0 + \beta_1 + \beta_{2,1}) = G(\beta_0 + \beta_1 + \beta_{2,1} + \mu) \\
& \Leftrightarrow \\
& G^{-1}(ATT + F(\beta_0 + \beta_1 + \beta_{2,1})) = \beta_0 + \beta_1 + \beta_{2,1} + \mu \\
& \Leftrightarrow \\
& \mu = G^{-1}(ATT + F(\beta_0 + \beta_1 + \beta_{2,1})) - (\beta_0 + \beta_1 + \beta_{2,1})
\end{aligned}$$

## 10.6 Analytical expression of non-random sample selection bias in two-period two-group setting

In the two-group, two-period case, the bias can be expressed by the difference between the *DID* estimand and the *ATT*:

$$\begin{aligned}
\beta_{DID} - ATT &= E[y_{ivt}^o | t \geq t_0, D_i = 1] - E[y_{ivt}^o | t < t_0, D_i = 1] - (E[y_{ivt}^o | t \geq t_0, D_i = 0] - E[y_{ivt}^o | t < t_0, D_i = 0]) \\
&\quad - (E[y_{ivt}(1) | t \geq t_0, D_i = 1] - E[y_{ivt}(0) | t \geq t_0, D_i = 1])
\end{aligned}$$

In the first period, the expectation of  $y_{ivt}^o$  is the same as that of  $y_{ivt}$ , giving:

$$\begin{aligned}\beta_{DID} - ATT = & E[y_{ivt}^o|t \geq t_0, D_i = 1] - E[y_{ivt}|t < t_0, D_i = 1] - (E[y_{ivt}^o|t \geq t_0, D_i = 0] - E[y_{ivt}|t < t_0, D_i = 0]) \\ & - (E[y_{ivt}(1)|t \geq t_0, D_i = 1] - E[y_{ivt}(0)|t \geq t_0, D_i = 1])\end{aligned}$$

Applying potential outcomes:

$$\begin{aligned}\beta_{DID} - ATT = & E[y_{ivt}^o(1)|t \geq t_0, D_i = 1] - E[y_{ivt}(0)|t < t_0, D_i = 1] - (E[y_{ivt}^o(0)|t \geq t_0, D_i = 0] - E[y_{ivt}(0)|t < t_0, D_i = 0]) \\ & - (E[y_{ivt}(1)|t \geq t_0, D_i = 1] - E[y_{ivt}(0)|t \geq t_0, D_i = 1])\end{aligned}$$

Applying our common trends assumption:

$$\begin{aligned}\beta_{DID} - ATT = & E[y_{ivt}^o(1)|t \geq t_0, D_i = 1] - E[y_{ivt}(0)|t < t_0, D_i = 0] - (E[y_{ivt}^o(0)|t \geq t_0, D_i = 0] - E[y_{ivt}(0)|t < t_0, D_i = 0]) \\ & - (E[y_{ivt}(1)|t \geq t_0, D_i = 1] - E[y_{ivt}(0)|t \geq t_0, D_i = 0])\end{aligned}$$

3 Simplifying:

$$\begin{aligned}\beta_{DID} - ATT = & E[y_{ivt}^o(1)|t \geq t_0, D_i = 1] - E[y_{ivt}^o(0)|t \geq t_0, D_i = 0] \\ & - (E[y_{ivt}(1)|t \geq t_0, D_i = 1] - E[y_{ivt}(0)|t \geq t_0, D_i = 0])\end{aligned}$$

**Extending to our simulations:**

In the context of our monte carlo simulations, this can be extended:

$$\begin{aligned}\beta_{DID} - ATT = & P[y_{ivt}^o(1) = 1|t \geq t_0, D_i = 1] - P[y_{ivt}^o(0) = 1|t \geq t_0, D_i = 0] \\ & - (P[y_{ivt}(1) = 1|t \geq t_0, D_i = 1] - P[y_{ivt}(0) = 1|t \geq t_0, D_i = 0])\end{aligned}$$

$$\begin{aligned}\beta_{DID} - ATT = & P[(y_{ivt}(1) = 1|t \geq t_0, D_i = 1)|(y_{ivt}(0) = 0|t < t_0, D_i = 1)] \\ & - P[(y_{ivt}(0) = 1|t \geq t_0, D_i = 0)|(y_{ivt}(0) = 0|t < t_0, D_i = 0)] \\ & - (P[y_{ivt}(1) = 1|t \geq t_0, D_i = 1] - P[y_{ivt}(0) = 1|t \geq t_0, D_i = 0])\end{aligned}$$

$$\begin{aligned}\beta_{DID} - ATT = & P[(y_{ivt}^*(1) > 0|t \geq t_0, D_i = 1)|(y_{ivt}^*(0) \leq 0|t < t_0, D_i = 1)] \\ & - P[(y_{ivt}^*(0) > 0|t \geq t_0, D_i = 0)|(y_{ivt}^*(0) \leq 0|t < t_0, D_i = 0)] \\ & - (P[y_{ivt}^*(1) > 0|t \geq t_0, D_i = 1] - P[y_{ivt}^*(0) > 0|t \geq t_0, D_i = 0])\end{aligned}$$

Here, we let  $t \in \{1, 2\}$  denote the first and second periods, respectively:

$$\begin{aligned}\beta_{DID} - ATT = & P[(\beta_0 + \beta_1 + \beta_{2,1} + \beta_3 + \alpha_i + u_{i2} + \rho_v > 0|t \geq t_0, D_i = 1)|(\beta_0 + \beta_1 + \alpha_i + u_{i1} + \rho_v \leq 0|t < t_0, D_i = 1)] \\ & - P[(\beta_0 + \beta_{2,0} + \alpha_i + u_{i2} + \rho_v > 0|t \geq t_0, D_i = 0)|(\beta_0 + \alpha_i + u_{i1} + \rho_v \leq 0|t < t_0, D_i = 0)] \\ & - (P[\beta_0 + \beta_1 + \beta_{2,1} + \beta_3 + \alpha_i + u_{i2} + \rho_v > 0|t \geq t_0, D_i = 1] - P[\beta_0 + \beta_{2,0} + \alpha_i + u_{i2} + \rho_v > 0|t \geq t_0, D_i = 0])\end{aligned}$$



## 10.7 Full summary figure from all specifications and values of $\sigma_p$

Using Figure 12 to compare across all specifications and varying  $\sigma_p$ , we see that RMSE tends to increase across all specifications as  $\sigma_p$  increases. The pixel-level TWFE specifications with spatially aggregated unit fixed effects tend to have the lowest RMSE whenever  $\sigma_p$  is nonzero. In contrast, the specification with the property as the unit of analysis and pixel-level DID tend to have the highest RMSE.

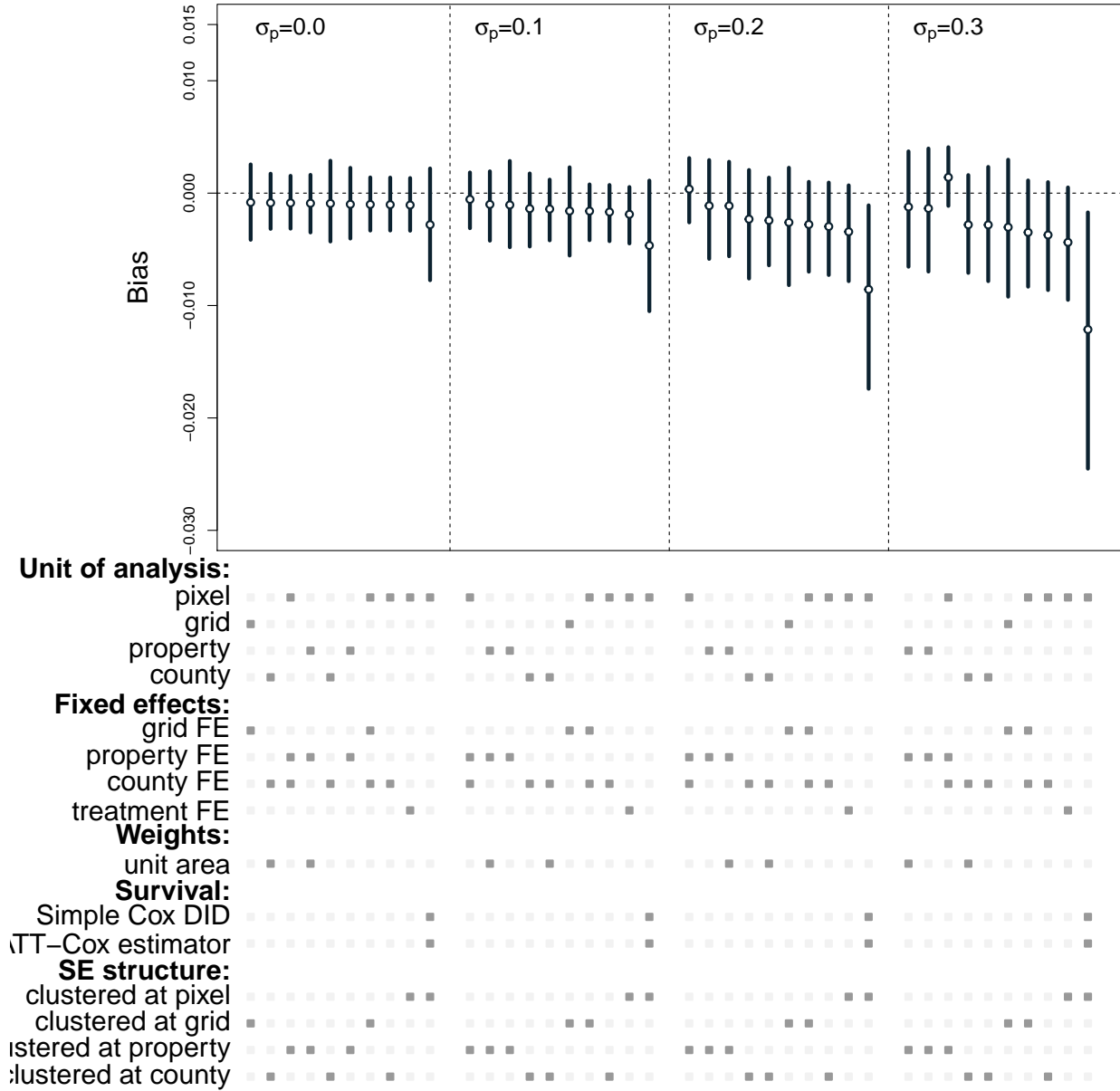


Figure 12: bias, distribution, and coverage of all models with clustered standard errors. Each subplot represents simulated landscapes with varying  $\sigma_p$ . specifications are ordered within subplots by absolute value of mean bias (by least to most).

## 10.8 Property-level models still outperform in alternative parameterizations

Figure 13 shows model performance for two alternate landscape parameterizations in the presence of property unobservables ( $\sigma_p$ ), ordered from least to most biased. The top panel considers our initial parameterization, but switches the pre-treatment deforestation rates for the two groups. This leaves a pre-treatment deforestation rate of 5% in the control area and 2% in the intervention area. The bottom panel considers the initial parameterization, but instead makes the *ATT* positive (i.e.  $ATT = 0.01$  rather than  $-0.01$ ).

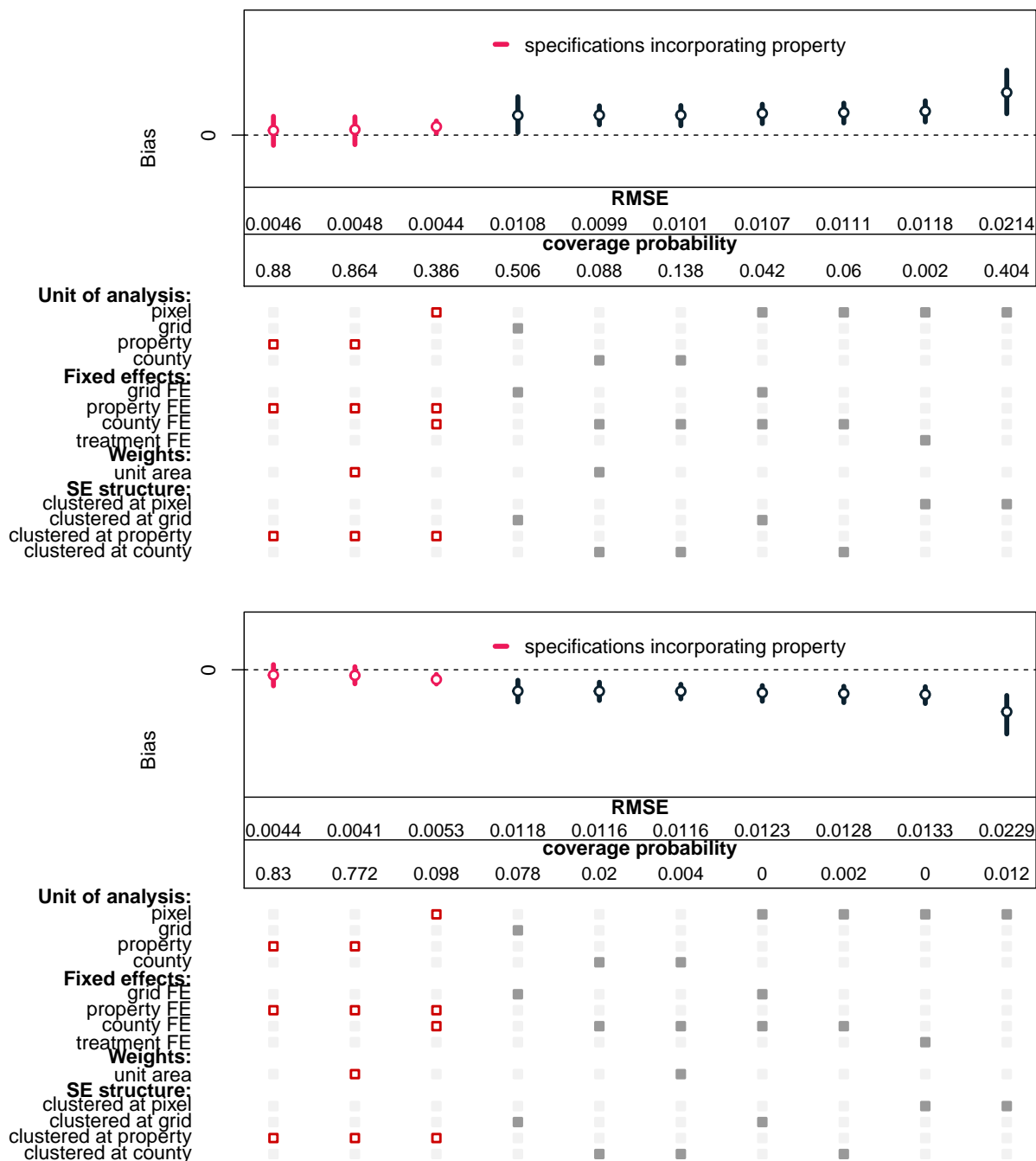


Figure 13: Property-level specification still outperform others in presence of property unobservables with alternative landscape parameterizations

## 10.9 DGP for multiple groups and variation in treatment timing

The following parameters and their definitions inform the simulation parameterizations.

$$\begin{aligned}
baseline_a &= E[y_{it}(0)|t < t_0, G_i = a] \\
baseline_b &= E[y_{it}(0)|t < t_0, G_i = b] \\
baseline_c &= E[y_{it}(0)|t < t_0, G_i = c] \\
trend_1 &= E[y_{it}(0)|t = 1, G_i = g] - E[y_{it}(0)|t = 0, G_i = g] \\
trend_2 &= E[y_{it}(0)|t = 2, G_i = g] - E[y_{it}(0)|t = 1, G_i = g] \\
trend_3 &= E[y_{it}(0)|t = 3, G_i = g] - E[y_{it}(0)|t = 2, G_i = g] \\
trend_4 &= E[y_{it}(0)|t = 4, G_i = g] - E[y_{it}(0)|t = 3, G_i = g] \\
ATT &= E[y_{it}(1) - y_{it}(0)|t \geq t_0, G_i = g]
\end{aligned}$$

Here, three groups,  $g \in \{a, b, c\}$  have different baseline deforestation rates, and all three groups would experience the same trends in the absence of treatment. Group  $a$  experiences treatment in time 2, group  $b$  experiences treatment in time 3, and group  $c$  is never treated. The  $ATT$  is equal across the two treated groups and there are no dynamic effects.

The DGP for each observation can be written as follows:

Group  $a$ :

$$y_{it}^* = \beta_{0,a}1\{t = 0\} + \beta_{1,a}1\{t = 1\} + \beta_{2,a}1\{t = 2\} + \beta_{3,a}1\{t = 3\} + \beta_{4,a}1\{t = 4\} + \tau_a1\{t \geq 2\} + \alpha_i + u_{it}$$

Group  $b$ :

$$y_{it}^* = \beta_{0,b}1\{t = 0\} + \beta_{1,b}1\{t = 1\} + \beta_{2,b}1\{t = 2\} + \beta_{3,b}1\{t = 3\} + \beta_{4,b}1\{t = 4\} + \tau_b1\{t \geq 3\} + \alpha_i + u_{it}$$

Group  $c$ :

$$y_{it}^* = \beta_{0,c}1\{t = 0\} + \beta_{1,c}1\{t = 1\} + \beta_{2,c}1\{t = 2\} + \beta_{3,c}1\{t = 3\} + \beta_{4,c}1\{t = 4\} + \alpha_i + u_{it}$$

, where the  $\beta$  and  $\tau$  coefficients are calculated as follows:

$$\begin{aligned}
\beta_{0,a} &= F^{-1}(baseline_a) \\
\beta_{1,a} &= F^{-1}(trend_1 + baseline_a) - \beta_{0,a} \\
\beta_{2,a} &= F^{-1}(trend_2 + F(\beta_{0,a} + \beta_{1,a})) - \beta_{0,a} - \beta_{1,a} \\
\beta_{3,a} &= F^{-1}(trend_3 + F(\beta_{0,a} + \beta_{1,a} + \beta_{2,a})) - \beta_{0,a} - \beta_{1,a} - \beta_{2,a} \\
\beta_{4,a} &= F^{-1}(trend_4 + F(\beta_{0,a} + \beta_{1,a} + \beta_{2,a} + \beta_{3,a})) - \beta_{0,a} - \beta_{1,a} - \beta_{2,a} - \beta_{3,a} \\
\tau_a &= F^{-1}(ATT + F(\beta_{0,a} + \beta_{1,a} + \beta_{2,a})) - \beta_{0,a} - \beta_{1,a} - \beta_{2,a}
\end{aligned}$$

$$\begin{aligned}
\beta_{0,b} &= F^{-1}(\text{baseline}_b) \\
\beta_{1,b} &= F^{-1}(\text{trend}_1 + \text{baseline}_b) - \beta_{0,b} \\
\beta_{2,b} &= F^{-1}(\text{trend}_2 + F(\beta_{0,b} + \beta_{1,b})) - \beta_{0,b} - \beta_{1,b} \\
\beta_{3,b} &= F^{-1}(\text{trend}_3 + F(\beta_{0,b} + \beta_{1,b} + \beta_{2,b})) - \beta_{0,b} - \beta_{1,b} - \beta_{2,b} \\
\beta_{4,b} &= F^{-1}(\text{trend}_4 + F(\beta_{0,b} + \beta_{1,b} + \beta_{2,b} + \beta_{3,b})) - \beta_{0,b} - \beta_{1,b} - \beta_{2,b} - \beta_{3,b} \\
\tau_b &= F^{-1}(\text{ATT} + F(\beta_{0,b} + \beta_{1,b} + \beta_{2,b} + \beta_{3,b})) - \beta_{0,b} - \beta_{1,b} - \beta_{2,b} - \beta_{3,b}
\end{aligned}$$

$$\begin{aligned}
\beta_{0,c} &= F^{-1}(\text{baseline}_c) \\
\beta_{1,c} &= F^{-1}(\text{trend}_1 + \text{baseline}_c) - \beta_{0,c} \\
\beta_{2,c} &= F^{-1}(\text{trend}_2 + F(\beta_{0,c} + \beta_{1,c})) - \beta_{0,c} - \beta_{1,c} \\
\beta_{3,c} &= F^{-1}(\text{trend}_3 + F(\beta_{0,c} + \beta_{1,c} + \beta_{2,c})) - \beta_{0,c} - \beta_{1,c} - \beta_{2,c} \\
\beta_{4,c} &= F^{-1}(\text{trend}_4 + F(\beta_{0,c} + \beta_{1,c} + \beta_{2,c} + \beta_{3,c})) - \beta_{0,c} - \beta_{1,c} - \beta_{2,c} - \beta_{3,c}
\end{aligned}$$

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

### 10.9.1 parameterization for heterogeneous treatment effects example

The following parameters and their definitions inform the simulation parameterizations.

$$\begin{aligned}
\text{baseline}_a &= E[y_{it}(0)|t < t_0, G_i = a] \\
\text{baseline}_b &= E[y_{it}(0)|t < t_0, G_i = b] \\
\text{baseline}_c &= E[y_{it}(0)|t < t_0, G_i = c] \\
\text{trend}_1 &= E[y_{it}(0)|t = 1, G_i = g] - E[y_{it}(0)|t = 0, G_i = g] \\
\text{trend}_2 &= E[y_{it}(0)|t = 2, G_i = g] - E[y_{it}(0)|t = 1, G_i = g] \\
\text{trend}_3 &= E[y_{it}(0)|t = 3, G_i = g] - E[y_{it}(0)|t = 2, G_i = g] \\
\text{trend}_4 &= E[y_{it}(0)|t = 4, G_i = g] - E[y_{it}(0)|t = 3, G_i = g] \\
\text{ATT}_{0,a} &= E[y_{it}(1) - y_{it}(0)|t = 2, G_i = a] \\
\text{ATT}_{1,a} &= E[y_{it}(1) - y_{it}(0)|t = 3, G_i = a] \\
\text{ATT}_{2,a} &= E[y_{it}(1) - y_{it}(0)|t = 4, G_i = a] \\
\text{ATT}_{0,b} &= E[y_{it}(1) - y_{it}(0)|t = 3, G_i = b] \\
\text{ATT}_{1,b} &= E[y_{it}(1) - y_{it}(0)|t = 4, G_i = b]
\end{aligned}$$

Here, three groups,  $g \in \{a, b, c\}$  have different baseline deforestation rates, and all three groups would experience the same trends in the absence of treatment. Group  $a$  experiences treatment in time 2, group  $b$  experiences treatment in time 3, and group  $c$  is never treated. The  $ATT$  is equal across the two treated groups and there are no dynamic effects.

The DGP for each observation can be written as follows:

Group  $a$ :

$$y_{it} = \beta_{0,a}1\{t=0\} + \beta_{1,a}1\{t=1\} + (\beta_{2,a} + \tau_{0,a})1\{t=2\} + (\beta_{3,a} + \tau_{1,a})1\{t=3\} + (\beta_{4,a} + \tau_{2,a})1\{t=4\} + \alpha_i + u_{it}$$

Group  $b$ :

$$y_{it} = \beta_{0,b}1\{t=0\} + \beta_{1,b}1\{t=1\} + \beta_{2,b}1\{t=2\} + (\beta_{3,b} + \tau_{0,b})1\{t=3\} + (\beta_{4,b} + \tau_{1,b})1\{t=4\} + \alpha_i + u_{it}$$

Group  $c$ :

$$y_{it} = \beta_{0,c}1\{t=0\} + \beta_{1,c}1\{t=1\} + \beta_{2,c}1\{t=2\} + \beta_{3,c}1\{t=3\} + \beta_{4,c}1\{t=4\} + \alpha_i + u_{it}$$

, where the  $\beta$  and  $\tau$  coefficients are calculated as follows:

$$\begin{aligned}\beta_{0,a} &= F^{-1}(\text{baseline}_a) \\ \beta_{1,a} &= F^{-1}(\text{trend}_1 + \text{baseline}_a) - \beta_{0,a} \\ \beta_{2,a} &= F^{-1}(\text{trend}_2 + F(\beta_{0,a} + \beta_{1,a})) - \beta_{0,a} - \beta_{1,a} \\ \beta_{3,a} &= F^{-1}(\text{trend}_3 + F(\beta_{0,a} + \beta_{1,a} + \beta_{2,a})) - \beta_{0,a} - \beta_{1,a} - \beta_{2,a} \\ \beta_{4,a} &= F^{-1}(\text{trend}_4 + F(\beta_{0,a} + \beta_{1,a} + \beta_{2,a} + \beta_{3,a})) - \beta_{0,a} - \beta_{1,a} - \beta_{2,a} - \beta_{3,a} \\ \tau_{0,a} &= F^{-1}(\text{ATT}_{0,a} + F(\beta_{0,a} + \beta_{1,a} + \beta_{2,a})) - \beta_{0,a} - \beta_{1,a} - \beta_{2,a} \\ \tau_{1,a} &= F^{-1}(\text{ATT}_{1,a} + F(\beta_{0,a} + \beta_{1,a} + \beta_{2,a} + \beta_{3,a} + \tau_{0,a})) - \beta_{0,a} - \beta_{1,a} - \beta_{2,a} - \beta_{3,a} - \tau_{0,a} \\ \tau_{2,a} &= F^{-1}(\text{ATT}_{2,a} + F(\beta_{0,a} + \beta_{1,a} + \beta_{2,a} + \beta_{3,a} + \tau_{0,a} + \tau_{1,a})) - \beta_{0,a} - \beta_{1,a} - \beta_{2,a} - \beta_{3,a} - \tau_{0,a} - \tau_{1,a}\end{aligned}$$

$$\begin{aligned}\beta_{0,b} &= F^{-1}(\text{baseline}_b) \\ \beta_{1,b} &= F^{-1}(\text{trend}_1 + \text{baseline}_b) - \beta_{0,b} \\ \beta_{2,b} &= F^{-1}(\text{trend}_2 + F(\beta_{0,b} + \beta_{1,b})) - \beta_{0,b} - \beta_{1,b} \\ \beta_{3,b} &= F^{-1}(\text{trend}_3 + F(\beta_{0,b} + \beta_{1,b} + \beta_{2,b})) - \beta_{0,b} - \beta_{1,b} - \beta_{2,b} \\ \beta_{4,b} &= F^{-1}(\text{trend}_4 + F(\beta_{0,b} + \beta_{1,b} + \beta_{2,b} + \beta_{3,b})) - \beta_{0,b} - \beta_{1,b} - \beta_{2,b} - \beta_{3,b} \\ \tau_{b,0} &= F^{-1}(\text{ATT}_{0,b} + F(\beta_{0,b} + \beta_{1,b} + \beta_{2,b} + \beta_{3,b})) - \beta_{0,b} - \beta_{1,b} - \beta_{2,b} - \beta_{3,b} \\ \tau_{b,1} &= F^{-1}(\text{ATT}_{1,b} + F(\beta_{0,b} + \beta_{1,b} + \beta_{2,b} + \beta_{3,b} + \tau_{b,0})) - \beta_{0,b} - \beta_{1,b} - \beta_{2,b} - \beta_{3,b} - \tau_{b,0}\end{aligned}$$

$$\begin{aligned}\beta_{0,c} &= F^{-1}(\text{baseline}_c) \\ \beta_{1,c} &= F^{-1}(\text{trend}_1 + \text{baseline}_c) - \beta_{0,c} \\ \beta_{2,c} &= F^{-1}(\text{trend}_2 + F(\beta_{0,c} + \beta_{1,c})) - \beta_{0,c} - \beta_{1,c} \\ \beta_{3,c} &= F^{-1}(\text{trend}_3 + F(\beta_{0,c} + \beta_{1,c} + \beta_{2,c})) - \beta_{0,c} - \beta_{1,c} - \beta_{2,c} \\ \beta_{4,c} &= F^{-1}(\text{trend}_4 + F(\beta_{0,c} + \beta_{1,c} + \beta_{2,c} + \beta_{3,c})) - \beta_{0,c} - \beta_{1,c} - \beta_{2,c} - \beta_{3,c}\end{aligned}$$

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

## 10.10 Calculating deforestation rates

Upon choosing an aggregated unit of analysis, the researcher must compute the deforestation rate. This varies throughout 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). We test the performance of three common deforestation rate formulas in the literature.

One common formula to calculate annual deforestation rates is

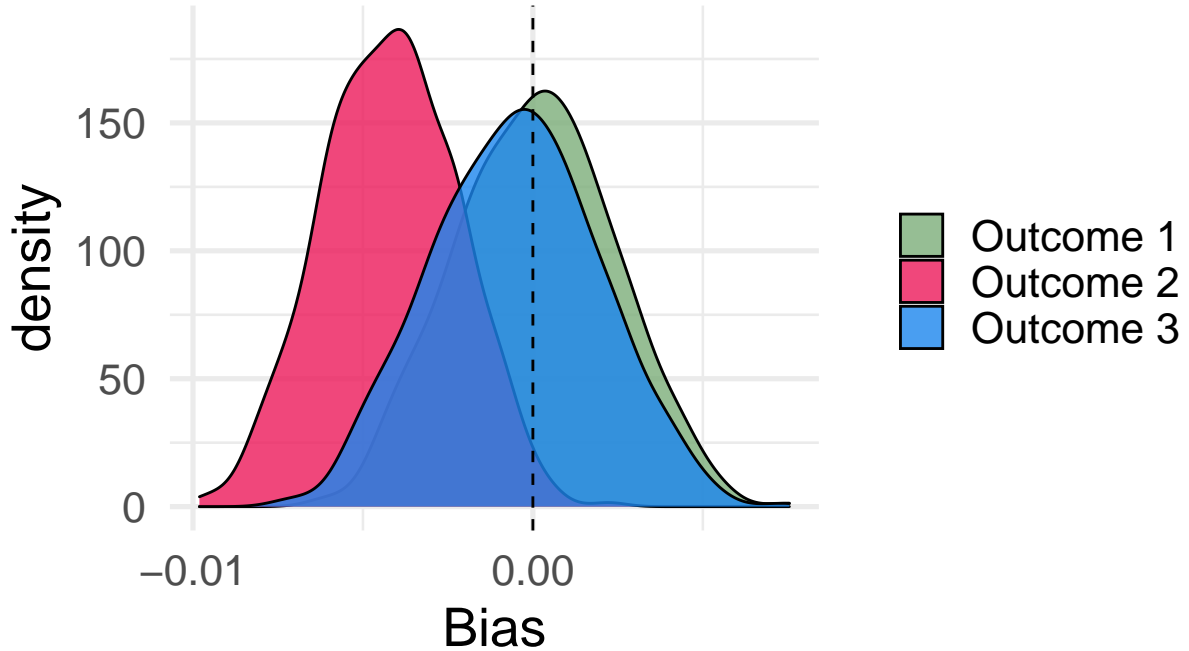
$$\text{Outcome 1) } = \frac{F_{i,t-1} - F_{it}}{F_{i,t-1}} \quad (11)$$

, 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; Busch et al. 2015), and is arguably the most widely used formula. Some authors have also calculated the deforestation rate in relation to the initial observed level of forest cover, replacing  $F_{it}$  with  $F_{i0}$ , the baseline forest cover, in equation (1). This gives outcome 2):

$$\text{Outcome 2) } = \frac{F_{i0} - F_{it}}{F_{i0}} \quad (12)$$

Lastly, we consider a formula derived from the Compound Interest Law that has also been used in recent studies (e.g. Ruggiero et al. 2019; Puyravaud 2003). Outcome 3) is given by:

$$\text{Outcome 3) } = \ln(F_{i,t-1}/F_{it}) \quad (13)$$



Outcome 1: 1e-04 RMSE: 0.00231  
Outcome 2: -0.0042 RMSE: 0.00465  
Outcome 3: -5e-04 RMSE: 0.00248

Figure 14: Distribution of estimates produced by different outcome variable formulae

Figure 14 demonstrates that outcome 1 results in the least bias in our guiding example. The other outcomes result in relatively greater bias, although this difference seems minimal in our setting. RMSE is also comparable across the three outcomes. We express concern surrounding the use of the initial observed baseline forest cover in the calculation of deforestation rates, as in formula (2). As the rate is calculated in periods further away from the baseline, the deforestation rate will be increasingly high, even if the rate has not changed over time. This concern is echoed by the fact that this outcome results in the most biased estimates of the *ATT*.

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 lowest RMSE, and it is currently the most common deforestation rate calculation used in the literature.