

# Conservation impact evaluation using remotely sensed data

Alberto Garcia\*

Robert Heilmayr†

June 27, 2023

## Abstract

The application of quasiexperimental impact evaluation to remotely sensed measures of deforestation has yielded important evidence detailing the effectiveness of conservation policies. However, researchers have paid insufficient attention to the binary and irreversible structure of most deforestation datasets. Using analytical proofs and simulations, we demonstrate that many commonly employed econometric models are biased when applied to binary and irreversible outcomes. The significance, magnitude and even direction of estimated effects from many studies are likely incorrect, threatening to undermine the evidence base that underpins conservation policy adoption and design. To address these concerns, we provide guidance and new strategies for the design of panel econometric models that yield more reliable estimates of the impacts of forest conservation policies.

## 1 Introduction

Policymakers often need to understand the causal impacts of conservation interventions. Can payments for ecosystem services encourage lasting reforestation? Does the allocation of land rights slow deforestation? 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; P. J. Ferraro and Hanauer 2014; Miteva, Pattanayak, and Ferraro 2012). Increasingly, these econometric approaches to impact evaluation are being used to disentangle the causal relationships that underpin conservation decisionmaking (Butsic, Lewis, et al. 2017; Baylis et al. 2016; Williams, Balmford, and Wilcove 2020).

The proliferation of panel, impact evaluations of conservation interventions has been enabled, in large part, by the increasing prevalence of remotely sensed datasets detailing deforestation over time (Blackman 2013; Jones and Lewis 2015). Decades of remotely sensed images taken across the entirety of Earth’s surface have yielded data that are incredibly well-suited for observational impact evaluations — a scientist hoping to quantify the impacts of a conservation policy adopted decades ago can observe deforestation in treated and untreated units spanning both pre- and post-implementation periods (Jain 2020). Responding to this opportunity, hundreds of studies have applied econometric techniques of impact evaluation to a single dataset, the Global Forest Change product produced by Hansen et al. (2013). However, this data product and others like it have a somewhat unusual structure in that they detail the first year in which a given pixel was deforested and, importantly, do not detect repeated deforestation events in the same location. As a result, empirical researchers are left with data describing deforestation as a binary and irreversible process.

---

\*University of California, Santa Barbara, Bren School of Environmental Science & Management; [albertogarcia@ucsb.edu](mailto:albertogarcia@ucsb.edu)

†University of California, Santa Barbara, Bren School of Environmental Science & Management and Environmental Studies Program; [rheilmayr@ucsb.edu](mailto:rheilmayr@ucsb.edu)

In this paper, we investigate whether the irreversible, binary data structure that characterizes many deforestation studies affects the performance of common panel econometric methods including difference-in-differences (DID), two-way fixed effects (TWFE) and survival models. We use a combination of analytical proofs and Monte Carlo simulations to demonstrate that many prior econometric analyses of deforestation are likely biased — the significance, magnitude and even direction of estimated effects might be incorrect. While we focus primarily upon impact evaluation of forest conservation policies, our findings are relevant to the broader set of research contexts in which the outcome of interest represents an irreversible, binary event, including recidivism (Agan and Makowsky 2018; Mastrobuoni and Pinotti 2015), individual mortality (Li et al. 2022; Friedman and Schady 2013; Dolan et al. 2019), technology adoption (Bollinger et al. 2022), and employee retention (Bueno and Sass 2018; Feng and Sass 2018).

One of our core results shows that TWFE regressions with individual unit fixed effects do not identify the desired treatment effect when applied to panel datasets with binary, irreversible outcomes. In more traditional data settings, the DID estimator is numerically equivalent to the linear TWFE estimator in the case of two-periods (each of which can consist of multiple years) and two-groups (Imai and Kim 2021; Goodman-Bacon 2021). However we demonstrate that, when applied to binary, irreversible panel data, the TWFE estimator is distinct from the standard DID estimator. Specifically, the coefficient of interest in TWFE specifications does not recover the *ATT*, but instead recovers an ex-post difference in deforestation rates between treatment and control areas. This is particularly worrisome in the context of forest conservation, where interventions often target areas with high deforestation rates (e.g., Brazil’s blacklisted Priority Municipalities) or low opportunity costs (e.g., protected areas). Papers published in both conservation science and economics journals frequently use this problematic specification to recover treatment effect estimates. We further show that this concern extends to many recently developed DID estimators that estimate treatment effects in staggered adoption settings (e.g., Sun and Abraham 2021; Gardner 2021; Borusyak, Jaravel, and Spiess 2021).

To help guide future impact evaluations, we identify multiple ways in which this bias can be reduced or even eliminated. In the land use context, one easily implemented solution is to aggregate the binary pixels spatially. Both pixel-level TWFE specifications with spatially aggregated unit fixed effects and TWFE specifications with spatially aggregated units of analysis recover the expected treatment effect parameter. Another potential solution is for researchers to adopt non-linear survival models that explicitly account for the binary, irreversible nature of the outcome variable. However, we show that prior efforts to use survival models in DID research designs require identifying assumptions that would violate the traditional linear common trends assumption. In response, we develop and test a new survival-based DID estimator that relies on the more traditional linear common trends assumption.

Finally, we reflect on the econometric benefits that emerge when researchers are able to incorporate institutional knowledge into their analyses, matching their model structure to the relevant scale of decisionmaking. For example, in settings where individual landowners may have strong, heterogeneous preferences affecting land-use change, we show that incorporating property-level fixed effects can reduce bias and improve statistical coverage over alternate model structures.

Our conclusions yield important insights for the community of environmental economists, conservation scientists, and policymakers that wish to understand the causal impacts of conservation policies. Causal impact evaluation in nature conservation has emerged relatively recently (Börner et al. 2020), and a wide array of papers has called for researchers to improve the rigor of their analyses in this space (Paul J. Ferraro, Sanchirico, and Smith 2019; Baylis et al. 2016; Miteva, Pattanayak, and Ferraro 2012). In this paper, we identify key challenges associated with the integration of popular deforestation datasets with panel econometric approaches frequently used in this literature. We also build on prior work by Avelino, Baylis, and Honey-Rosés (2016), showing the benefits to matching analyses to the scale at which heterogeneity operates in the context of inherently spatial processes such as land use change. Although recent studies in the forest conservation literature have begun to adopt alternative DID estimators that are robust to general treatment effect heterogeneity (e.g., Rico-Straffon et al. 2023), we demonstrate that many are not immune to the lessons identified in this paper. Ultimately, our analyses raise potential concerns about the conclusions in many past studies, while providing researchers with practical guidance to improve the reliability of future analyses.

Our research also contributes to an emerging literature proposing methodological improvements for the integration of remotely sensed data with econometric methods of causal inference. Review studies have identified the need for social scientists to better understand the structure and limitations of remotely sensed data products (Jain 2020; Donaldson and Storeygard 2016). Several studies have begun to document measurement error in satellite-based measurements and explore its implications for econometric analysis (Proctor, Carleton, and Sum 2023; Gibson et al. 2021). This includes work specific to the forest conservation context. For example, Alix-Garcia and Millimet (2022) address misclassification in the context of a remotely sensed binary forest cover outcome and propose a solution for unbiased causal inference. Torchiana et al. (2022) present a hidden Markov model that corrects for misclassification bias in land use change settings. Our paper implicitly assumes away issues of mismeasurement but illustrates that, even in settings without non-classical measurement error, the underlying structure and limitations of remotely sensed, deforestation data products may still lead researchers to recover biased treatment effect estimates.

The remainder of the paper proceeds as follows. In Section 2, we introduce our empirical context, focusing on the irreversibility inherent to forest conservation impact evaluation. We also discuss common panel econometric approaches used in the conservation literature and in other fields that often encounter binary, irreversible outcome data. Section 3 describes the setup for our analytical proofs and the Monte Carlo simulations we use to evaluate candidate econometric approaches in the context of a simulated conservation intervention. In Section 4, we present the results of our analytical proofs, including the identification of the *Irreversible unit fixed effects bias* associated with TWFE models incorporating irreversible, binary unit fixed effects and the non-random sample selection that may stem from irreversibility inherent to these settings. We also introduce our newly developed survival estimator. Next, in Section 5, we present results from our Monte Carlo simulations that illustrate the biases mentioned previously along with proposed solutions. We then explore the advantages to matching analyses to the scale of the decisionmaking unit and discuss the interpretation of coefficients from analyses incorporating area weighting. Additionally, we simulate an intervention with staggered adoption to show that many recently developed DID estimators fail to recover the expected treatment effect parameter when pixels are used as the unit of analysis, similar TWFE regressions with irreversible, binary unit fixed effects. Finally, Section 6 concludes.

## 2 Empirical context

### 2.1 Analysis setting

We focus on the case in which a researcher would like to quantify the impact an intervention has had on deforestation rates. We assume that the intervention has clearly defined boundaries (e.g., a protected area, certified concession, or indigenous territory), and that the researcher has access to spatially explicit observations of forest cover and forest loss spanning multiple years over 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 shows an unsystematic review of impact evaluations that apply panel econometric methods to data accessed as pixel-level observations of binary deforestation.

In each of the studies detailed in Table 1, the researcher’s goal is to measure the impact that a specific policy has 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 treated units would’ve experienced without treatment. The fundamental challenge is that the researcher is unable to observe what would have occurred in treated units had they not received treatment (P. W. Holland 1986).

Table 1: Panel econometric methods used in avoided deforestation impact evaluations. All included studies use data accessed as pixelated binary deforestation.

<b>Paper</b>	<b>Panel method</b>	<b>Unit of analysis</b>	<b>Unit FE 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	
Araujo et al. 2009	TWFE using instrument	state	state
Arriagada et al. 2012	matched DID	farm	
Baehr et al. 2021	TWFE	binary pixel/grid cell	pixel
Baylis et al. 2012	DID	grid cell	
BenYishay et al. 2017	TWFE	grid cell	grid cell
Blackman 2015	matched unit FE model	binary point/pixel	county
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
Butsic 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	
Heilmayr and Lambin 2016	matched DID	property	
Heilmayr et al. 2020	Triple DID/fixed effects	binary point/pixel	municipality
Herrera et al. 2019	matched regression	binary point/pixel	
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	
Nolte et al. 2017	DID	property	
Panlasigui et al. 2018	TWFE	binary point/pixel	pixel
Rico-Straffon et al. 2022	staggered DID	grid cell	
Ruggiero et al. 2022	TWFE	municipality	municipality
Sales et al. 2022	Cox PH DID	pixel	
Shah and Baylis 2015	DID	grid cell	
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

To further define the *ATT* in our research setting, we model deforestation ( $y_{ivt}$ ) as a binary choice by a landowner to clear a small plot of land  $i$  within their larger property  $v$  in year  $t$ , where  $t \in T$ . The decision to deforest depends upon a latent variable ( $y_{ivt}^*$ ) that represents the monetary 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 empirical reality of detecting deforestation in individual pixels, 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 remote sensing methods and datasets often force empirical researchers to treat deforestation as an absorbing state. Gradual processes of reforestation are inherently harder to identify than abrupt losses of forest cover. As a result, commonly used deforestation datasets such as the Global Forest Change product generally only identify the first year in which a pixel was cleared (Hansen et al. 2013). When converted to a panel dataset, deforestation must thus be treated as an absorbing state, meaning that once a pixel is deforested, it cannot revert to a forested state. Whether desired, or due to technical limitations, the resulting inability to observe repeated deforestation means that deforestation is, in effect, treated as an irreversible process in the vast majority of conservation impact evaluations. To incorporate this irreversibility into our model, we denote  $C_i$  as the first year in which  $y_{ivt}^* > 0$ , where  $y_{ivt}$  is not observed when  $t > C_i$ .

In response to this irreversibility, Jones and Lewis (2015) suggest that deforested pixels should be dropped in the periods after they are first cleared. We, alongside most studies in this literature, follow this guidance, further modifying our outcome,  $y_{ivt}^o$ , the observed binary deforestation variable:

$$y_{ivt}^o = \begin{cases} 1 & t = C_i \\ 0 & t < C_i \\ - & t > C_i \end{cases} \quad (3)$$

Here,  $-$  indicates that the outcome for pixel  $i$  in time  $t$  has been dropped from the panel in years where  $t > C_i$ . A simple illustrative panel dataset with this structure can be seen in A.1.

Our parameter of interest, the *ATT*, is the average effect of an intervention on treated pixels. Let  $y_{ivt}(1)$  and  $y_{ivt}(0)$  denote the potential outcomes of pixel  $i$  in property  $v$  in year  $t$  with and without the treatment, respectively. In addition, let  $t_0$  denote the first year in which the intervention of interest is implemented and let  $D_i$  represent a dummy indicating whether pixel  $i$  is ever treated. The *ATT* can now be expressed as:

$$ATT = E[y_{ivt}(1) - y_{ivt}(0) | t \geq t_0, D_i = 1] \quad (4)$$

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, researchers can take steps to address the plausibility of this assumption (Butsic, Lewis, et al. 2017; Roth 2022).

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  $i$ ,  $y_{it}(1)$  and  $y_{it}(0)$ , 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.

## 2.2 Candidate empirical models

We present five empirical model specifications that we will evaluate and refer to throughout the remainder of the paper: (1) the traditional difference-in-differences model; (2) the individual, unit-level two-way fixed effects model; (3) the two-way fixed effects model with aggregated unit fixed effects; (4) the two-way fixed effects model with aggregated units of analysis; and (5) the Cox proportional hazards difference-in-differences model. These models have all been used in the forest conservation literature to estimate the *ATT* of specific conservation interventions. While some approaches, such as the Cox proportional hazards model, are only beginning to emerge in the deforestation literature, they are popular in other areas of research that seek to estimate the impact of an intervention on the occurrence of binary, irreversible events.

### 2.2.1 Traditional difference-in-differences model

Under the above two assumptions, a common approach to estimating the *ATT* is the traditional DID regression:

**Regression 1:** (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:

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

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, Lewis, 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

$$\beta_{DID} = ATT$$

### 2.2.2 Individual unit-level TWFE model

Researchers often want 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 (Goodman-Bacon 2021). This amounts to estimating a regression that includes individual unit and time fixed effects to control for unobservable confounding variables that vary across units or through time.

**Regression 2:** (Individual unit 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:

$$y_{i,t}^o = \beta_{TWFE} \times D_i \mathbb{1}\{t \geq t_0\} + \lambda_t + \gamma_i + \epsilon_{i,t}$$

Here  $\lambda_t$  and  $\gamma_i$  represent the year and individual unit fixed effects, respectively. In the context of forest conservation, the individual unit  $i$  represents the pixel.

In the case of two groups and two time periods, the TWFE regression typically yields an estimate of the *ATT* that is numerically equivalent to the estimate generated by the DID model (Wooldridge 2010; Imai and Kim 2021). With this in mind, many researchers have used the TWFE model as a “generalized DID” that can be estimated not only in the 2x2 case, but also 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.

### 2.2.3 TWFE model with aggregated unit fixed effects

Some researchers have used unit fixed effects at the level of an aggregated unit rather than the individual binary unit. In the forest conservation context, aggregation is generally done spatially, aggregating pixels into larger units. For example, researchers using pixel-level TWFE regressions have used unit fixed effects at the level of a larger administrative unit such as the county (e.g., Blackman 2015) or municipality (e.g., Heilmayr et al. 2020). Regression 3 outlines the form of these individual unit-level TWFE models with aggregated unit fixed effects.

**Regression 3:** (Individual unit-level TWFE regression with 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:

$$y_{i,t}^o = \beta_{FE,j} \times D_i \mathbb{1}\{t \geq t_0\} + \lambda_t + \gamma_j + \epsilon_{i,t}$$

, where  $\gamma_j$  now denotes fixed effects at the level of an aggregated unit. If  $k$  differs from the level of treatment assignment, one must also include a treatment group indicator or fixed effects at the level of the unit at which treatment is assigned.

### 2.2.4 TWFE model with aggregated units of analysis

In addition to aggregating fixed effects, researchers can also transform their data by aggregating multiple pixel-level observations into larger units of analysis. For example, in the conservation case, researchers have often run a TWFE regression where the unit of analysis itself is the grid cell (e.g., Rico-Straffon et al. 2023), property (e.g., Heilmayr and Lambin 2016), or larger administrative unit (e.g., Sims and Alix-Garcia 2017).

**Regression 4:** (TWFE regression with aggregated unit of analysis) 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:

$$z_{j,t} = \beta_j \times D_j \mathbb{1}\{t \geq t_0\} + \lambda_t + \gamma_j + \epsilon_{j,t}$$

Regression 4 differs from Regression 3 in both the treatment variable,  $D_j$  and the outcome variable,  $z_{jt}$ . 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  $j$ . If treatment is assigned at the level of  $j$ ,  $D_j = 1$ .

The outcome variable,  $z_{jt}$ , denotes the event rate within unit  $j$  in period  $t$ . Arguably the most commonly used formula for  $z_{jt}$  in the deforestation literature is the deforestation rate calculated using the forested share of unit  $j$  in period  $t$  and its single-period lag  $t - 1$  (e.g. Carlson et al. 2018; Busch et al. 2015):

$$z_{jt} = \frac{F_{j,t-1} - F_{j,t}}{F_{j,t-1}} \quad (5)$$

, where  $F_{j,t}$  and  $F_{j,t-1}$  are area of forest cover within unit  $j$  at times  $t$  and  $t - 1$ , respectively. Although we use this definition of  $z_{jt}$  throughout the rest of the main text, we explore the relative merits and performance of alternative formulas for calculating deforestation rates in Section A.8.

### 2.2.5 Cox proportional hazards DID model

Survival analysis has emerged as a common approach to modelling the length of time until the occurrence of an irreversible event (Emmert-Streib and Dehmer 2019). Such survival models, including the Cox proportional hazards model, are frequently used to model events such as mortality (e.g., Puterman et al. 2020) and recidivism (e.g., Luallen, Edgerton, and Rabideau 2018), but can also be applied to deforestation contexts. 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 (e.g., Mastrobuoni and Pinotti 2015):

**Regression 5:** (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:

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

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

## 3 Methods

### 3.1 Analytical proofs

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). However, researchers have not adequately considered how the irreversible, binary structure of these data may affect the properties of some estimators. We use analytical proofs to demonstrate several concerns that arise with the use of specific model specifications in the context of an irreversible, binary outcome. All of our analytical results apply not only to forest conservation, but to any setting with an irreversible, binary outcome.

For our analytical results, we restrict ourselves to the case where  $t \in \{1, 2\}$  and  $t_0 = 2$ . We can then denote  $y_{iv1}^0$  and  $y_{iv2}^0$  as the outcomes for individual unit  $i$  in the first and second periods, respectively. In this setting,  $y_{iv1}^0 \in \{0, 1\}$  and  $y_{iv2}^0 \in \{0, 1, -\}$ , meaning that although only outcomes in the second period are dropped, there exists variation in deforestation across both periods. We present the core analytical results in the main text, and the detailed proofs can be found in the Appendix.



## 3.2 Simulation models

To validate our analytical results and explore the relative performance of different models frequently used in conservation impact evaluation, 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.

### 3.2.1 Landscape configuration

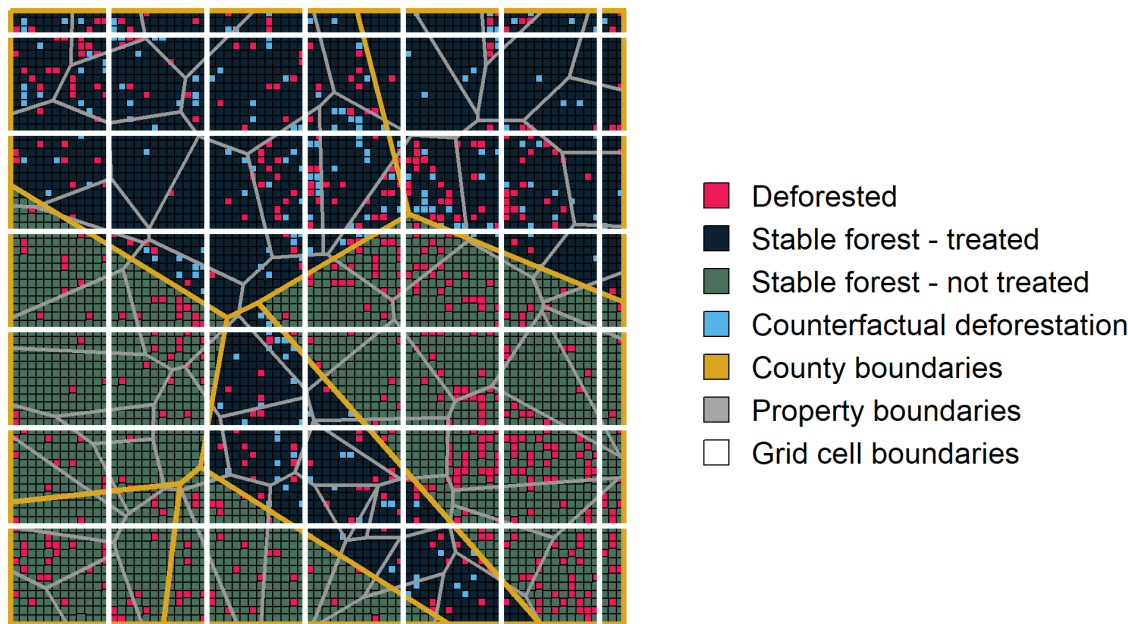


Figure 1: A map of a simulated landscape depicting patterns of deforestation under an effective conservation intervention, as well as counterfactual deforestation illustrating what would have happened in the absence of the intervention.

Figure 1 displays a simulated conservation intervention that reduced deforestation rates in treated areas. The landscape depicts both what is observed by the researcher at the end of the study period, as well as 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. We begin each Monte Carlo simulation by creating a synthetic landscape consisting of 150 rows and 150 columns of square pixels (22,500 total 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  $20.25 \text{ km}^2$ . 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

### 3.2.2 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, each of which consists of multiple years.

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.

The potential outcomes for the latent variable,  $y_{ivt}^*$ , are as follows:

$$y_{ivt}^*(0) = \beta_0 + \beta_1 D_i + \beta_{2,0}(1 - D_i)\mathbb{1}\{t \geq t_0\} + \beta_{2,1} D_i \mathbb{1}\{t \geq t_0\} + \alpha_i + u_{it} + \rho_v \quad (6)$$

and

$$y_{ivt}^*(1) = \beta_0 + \beta_1 + \beta_{2,1} + \beta_3 + \alpha_i + u_{it} + \rho_v \quad (7)$$

The *ATT* in our simulated setting is, therefore, defined as:

$$ATT = P(y_{ivt}^*(1) > 1 | D_i = 1, t \geq t_0) - P(y_{ivt}^*(0) > 1 | D_i = 1, t \geq t_0) \quad (8)$$

$$= P(\beta_0 + \beta_1 + \beta_{2,1} + \beta_3 + \alpha_i + u_{it} + \rho_v > 1) - P(\beta_0 + \beta_1 + \beta_{2,1} + \alpha_i + u_{it} + \rho_v > 1) \quad (9)$$

### 3.2.3 Assumed parameter values and evaluation criteria

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: 6 years in each of the pre-treatment and post-treatment periods ( $T = 12$ ,  $t_0 = 7$ ); 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 5.5. Note that Assumptions 1 and 2 are satisfied by construction. The derivations detailing the mapping from the landscape characteristics to the corresponding parameters in  $y_{iwt}^*$  can be found in Appendix A.4.

We compare econometric models using a combination of estimate bias, root mean squared error (RMSE), and coverage probability based on 500 replications of each set of parameters. 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. Factors such as bias and treatment of standard errors impact coverage.

## 4 Analytical results

### 4.1 TWFE with binary, irreversible outcomes yields biased estimate of $ATT$

Despite widespread use of pixel-level analyses of deforestation, the application of TWFE models to a binary, irreversible process yields a biased estimate of the  $ATT$ . Specifically, we show that the coefficient of interest from the unit-level TWFE model ( $\beta_{TWFE}$ ) estimates the post-treatment difference in outcomes (single difference), rather than the desired  $ATT$  (full proof contained in Appendix A.2.2).

**Result 1:** (Irreversible unit FE bias)

$$\beta_{TWFE} = ATT + \underbrace{E[y_{iwt}(0)|D_i = 1] - E[y_{iwt}(0)|D_i = 0]}_{\text{pre-treatment difference in deforestation rates}} \quad (10)$$

Regression 2 thus forgoes the benefits that panel methods provide, failing to properly control for time-invariant differences in treated and control units. If the treated area has a different baseline deforestation rate than the control, Regression 2 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.

Intuition for this result stems from the use of pixel-level, unit fixed effects in the regression specification. By including pixel-level 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 observations in the periods after the irreversible event is first realized, these fixed effects do not behave as the researcher expects. Observations that realize the event (i.e. are deforested) in years prior to treatment are, by definition, not observed for the entire panel. As a result, all pixels that contribute to the estimates of post-treatment effects must have a common pre-treatment outcome of 0. Interestingly,  $\beta_{TWFE}$  differs from  $\beta_{DID}$  in our context, even in the 2x2 case.

Section 5 corroborates Result 1 through simulations, and shows that the traditional DID does not suffer from similar concerns in the context of forest conservation. In addition, in Section A.3, we also show that

$\hat{\beta}_{TWFE}$  is equivalent to the coefficient from a TWFE regression on a dataset where any pixel deforested in years prior to treatment is completely excluded from the dataset (which would clearly be poor practice).

## 4.2 Alternative construction of $y_{ivt}^o$

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 challenges, especially when researchers are interested in interpreting the *ATT* in terms of a policy’s impact on deforestation rates. The *ATT* as defined in Section 2 represents an estimate of the impact of an intervention on the frequency of deforestation events (i.e. the decision to clear). Economists have typically focused their attention on this clearing decision, estimating the impact of a policy on the deforestation rate (flow) rather than on the stock of aggregate forest remaining. Under this interpretation, keeping deforested pixels 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 (Appendix A.2.4).

If researchers do keep cleared pixels in their sample after they are deforested, the correct interpretation of the *ATT* is that it is a measure of a policy’s impact on the aggregate stock of ever cleared forests. However, when reforestation timing is unaccounted for and deforestation is treated as an absorbing state, the interpretation of this measure becomes more challenging. Rather than measuring the area of land that is not in a forested state area at time  $t$ , this outcome variable measures the stock of ever-deforested area through the current time period. Therefore, if reforestation timing is not accounted for, researchers should be careful not to interpret these aggregate measures as deforested area (or forest cover if looking at  $1 - \text{deforested area}$ ), but rather as cumulative deforestation over the study period.

Importantly, researchers should carefully consider how the treatment of binary and irreversible data influences the interpretation of their coefficients of interest. Examples from outside of the deforestation context can highlight how alternate model structures can be used to generate estimates of the most policy-relevant or theoretically interesting framing of the *ATT*. For example, in the case of mortality, researchers may be most interested in communicating impacts on mortality rates, not aggregate mortality. Here, structuring the data to focus on mortality events is preferable. In contrast, a model of educational attainment might seek to estimate the stock of education within a population. In this case, the outcome could be adjusted to reflect a binary indicator of whether an individual has ever achieved a specific level of education.

## 4.3 Survival analysis

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

Multiple studies across a wide variety of settings have interpreted the exponentiated coefficient,  $\exp(\beta_{coxDID})$ , as a hazard ratio that describes the causal impact that treatment has had on the relative likelihood of survival. Specifically, 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, which we denote as the *HRTT*, can be considered a reframing of the traditional *ATT* as a ratio rather than a difference:

$$HRTT = \frac{E[y_{iv2}(1)|D_i = 1]}{E[y_{iv2}(0)|D_i = 1]}$$

Both in conservation and alternative settings, several studies using Regression 5 assess the plausibility of Assumption 1 (e.g., Li et al. 2022; Mastrobuoni and Pinotti 2015; Bueno and Sass 2018) in order to motivate

causal interpretation of  $\exp(\beta_{coxDID})$ . However, Appendix A.2.3 shows that  $\exp(\beta_{coxDID})$  only identifies the  $HRTT$  under an alternative assumption:

**Assumption 3:** (Proportional trends)

$$\frac{E[y_{iv2}(0)|D_i = 1]}{E[y_{iv1}(0)|D_i = 1]} = \frac{E[y_{iv2}(0)|D_i = 0]}{E[y_{iv1}(0)|D_i = 0]}$$

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 Assumption 1 and Assumption 3 cannot simultaneously hold (unless there is no trend at all). 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.

#### 4.3.2 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:

$$HR_1 = \frac{E[y_{iv2}(1)|D_i = 1]}{E[y_{iv1}(0)|D_i = 1]} \quad (11)$$

$$HR_2 = \frac{E[y_{iv2}(1)|D_i = 1]}{E[y_{iv2}(0)|D_i = 0]} \quad (12)$$

$$HR_3 = \frac{E[y_{iv2}(0)|D_i = 0]}{E[y_{iv1}(0)|D_i = 0]} \quad (13)$$

$$HRTT = \frac{E[y_{iv2}(1)|D_i = 1]}{E[y_{iv2}(0)|D_i = 1]} = \frac{1}{1/HR_1 + 1/HR_2 - 1/(HR_2 * HR_3)} \quad (14)$$

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  $h(t) = \lambda_0(t)\exp(\alpha 1\{t \geq t_0\})$ ;
- $HR_2 = \exp(\beta)$ , 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  $h(t) = \gamma_0(t)\exp(\beta 1\{D_i = 1\})$ ; and
- $HR_3 = \exp(\delta)$ , 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  $HRTT$ ,  $E[y_{iv2}(1)|D_i = 1]$ , 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  $HRTT$ :

$$\widehat{ATT}_{Cox} = \widehat{defor}_{D_i:1, t \geq t_0} - \frac{\widehat{defor}_{D_i:1, t \geq t_0}}{\widehat{HRTT}} \quad (15)$$

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 the traditional common trends assumption (Assumption 1). 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.

#### 4.4 Non-random sample selection can generate bias in irreversible settings

Irreversibility in observed deforestation 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.

$$\beta_{DID} = ATT + \underbrace{E[y_{iv2}^o | D_i = 1] - E[y_{iv2} | D_i = 1] - (E[y_{iv2}^o | D_i = 0] - E[y_{iv2} | D_i = 0])}_{\text{bias emerging from non-random sample selection}} \quad (16)$$

**Proof:** Appendix A.2.5 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.

## 5 Simulation results

### 5.1 TWFE bias

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, we have shown that TWFE models with pixel fixed effects are not a viable approach to estimate the  $ATT$ . Figure 2 shows the bias associated with Regression 2, a pixel-level 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.

Figure 2 further shows that the TWFE and DID estimators are not numerically equivalent, even in our 2x2 context. We further examine this finding in section A.3 and show that the coefficient of interest from Regression 2 is numerically equivalent to that from the same regression on a dataset where all pixels deforested in the first period are excluded from the dataset completely.

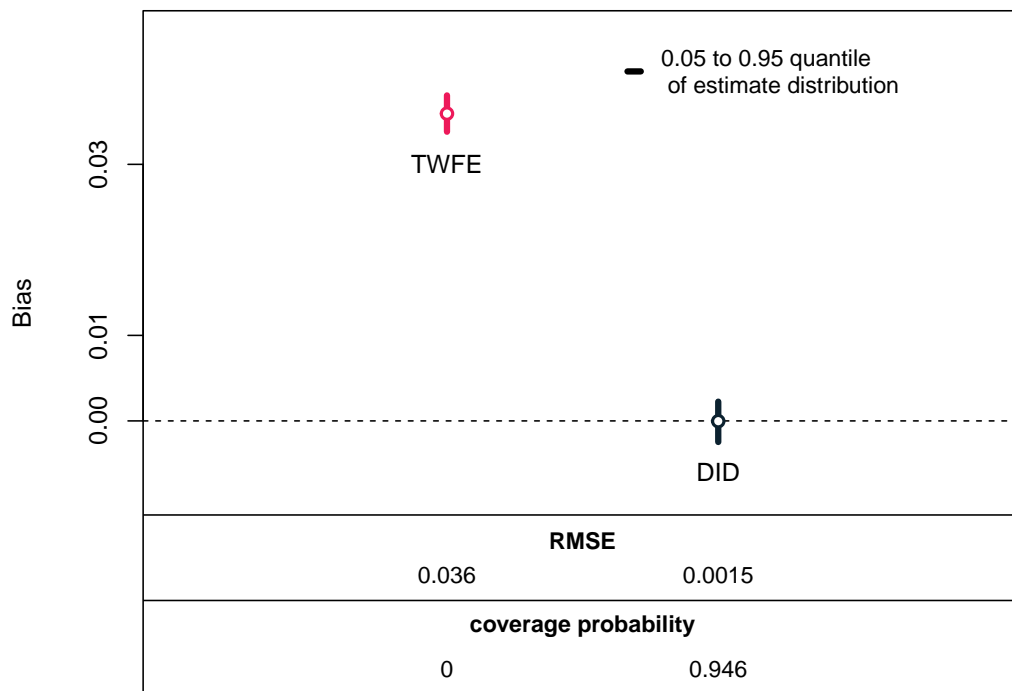


Figure 2: Comparison of model performance under pixel-level DID and TWFE with pixel unit fixed effects. Point illustrates mean bias while the confidence interval illustrates the 0.05 to 0.95 quantile range of this bias from the Monte Carlo simulations. Standard errors are clustered at the pixel.

## 5.2 Unbiased estimates from models with spatial aggregation

In the two-group, two-period case the traditional DID (Regression 1) is an unbiased estimator of the  $ATT$ , as shown in Figure 2 and column 1 of Figure 3 (The traditional 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.

Alternate forms of aggregation may provide a useful way to avoid the issues associated with integrating TWFE models with binary outcomes when the outcome is irreversible. Columns 2-4 of Figure 3 show that in pixel level TWFE regressions in the form of Regression 3,  $\beta_{FE,j}$  is an unbiased estimator of the  $ATT$ . These regressions use grid, county, or property fixed effects 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.

Similarly, columns 5-7 show that  $\beta_j$  from TWFE regressions in the form of Regression 4 is an unbiased estimator of the  $ATT$ . These TWFE regressions use an aggregated unit of analysis, where the deforestation rate for unit  $j$  is explicitly calculated in each time  $t$ . The following results are based on Equation (5). Neither the bias nor RMSE of the estimates vary dramatically across different levels of aggregation.

## 5.3 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 5), however, is not a viable solution (Section 4.3.1) under the typical common trends assumption (Assumption 1). Column 8 of Figure 3 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 9 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 changes in pixels after 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 the data generating process underpinning deforestation is influenced by unobservable characteristics of more aggregated spatial units such as the preferences of a property-owner. While it is straightforward to account for this in OLS, factors such as incidental parameters problems and computational difficulty in non-linear models may complicate easy implementation in survival analysis (Lancaster 2000; Fernández-Val and Weidner 2016)



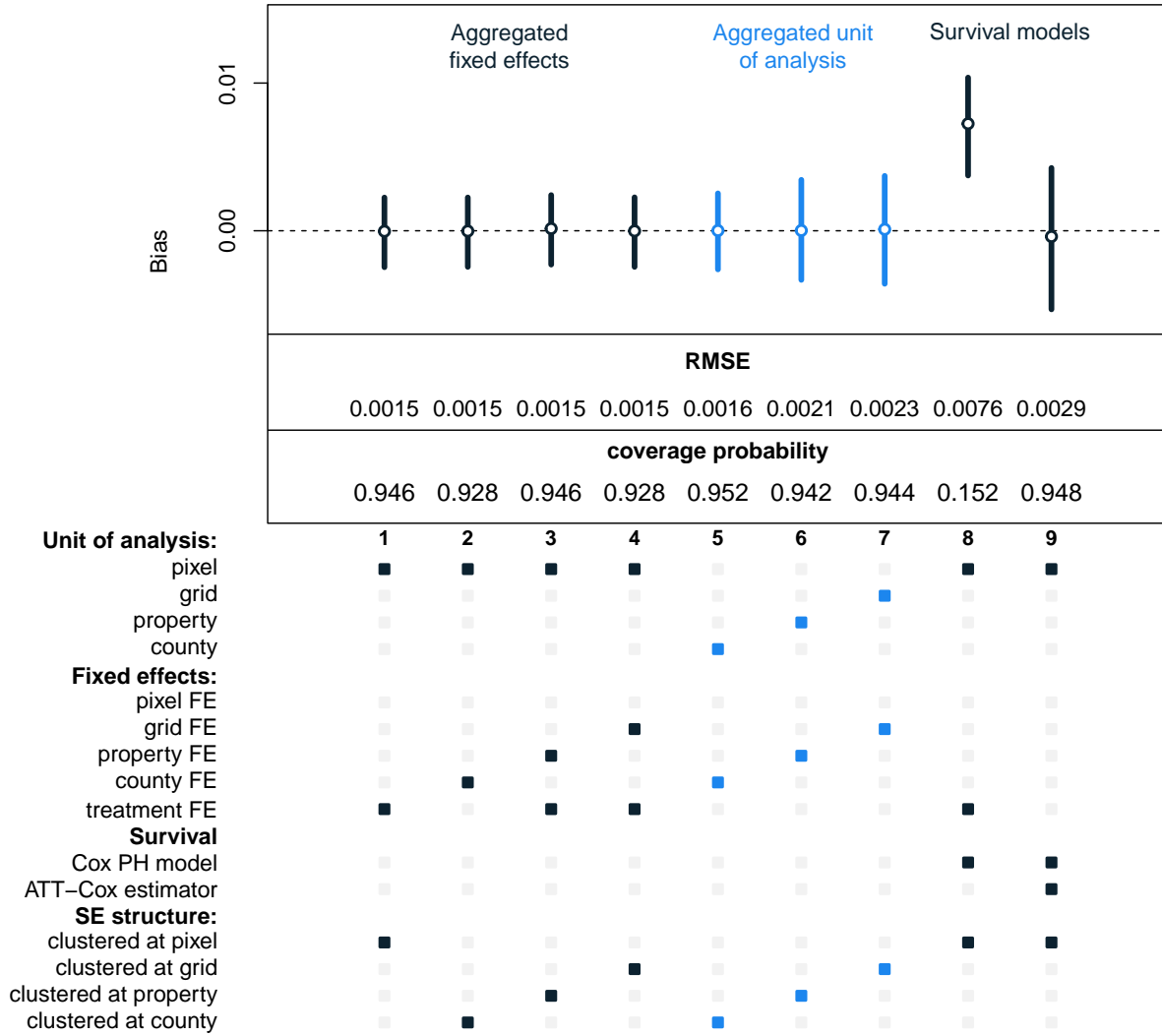


Figure 3: Comparison of model performance, where candidate models are separated by whether they incorporate aggregated fixed effects in pixel-level specifications, aggregated units of analysis, or survival analysis. Point illustrates mean bias while the confidence interval illustrates the 0.05 to 0.95 quantile range of this bias.

## 5.4 Non-random sample selection

We now explore how the non-random sample selection described in section 4.4 may influence estimates in our simulated landscapes. Non-random sample selection did not bias our initial simulations as presented in Figure 3 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  $y_{iwt}^o$ s were i.i.d.. However, once time-invariant disturbances enter the DGP, the distribution of the  $y_{iwt}^o$ s is potentially different in each year of the panel. This is likely to be the case in reality, since each plot of land will have time-invariant characteristics that impact its expected returns to clearing such as market access or agricultural suitability.

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 4 shows that non-random selection introduces a slight downward bias across every specification.

In practice, researchers cannot recover the second and fourth terms of the bias term in Equation 16, meaning that the magnitude of this bias is unknown to the researcher. However, this bias is likely to be of a smaller magnitude than in our simulated setting if deforestation rates are lower or more similar across treated and control groups.

## 5.5 Selecting the appropriate spatial structure

### 5.5.1 Model structures that match the spatial process of 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 5, 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.

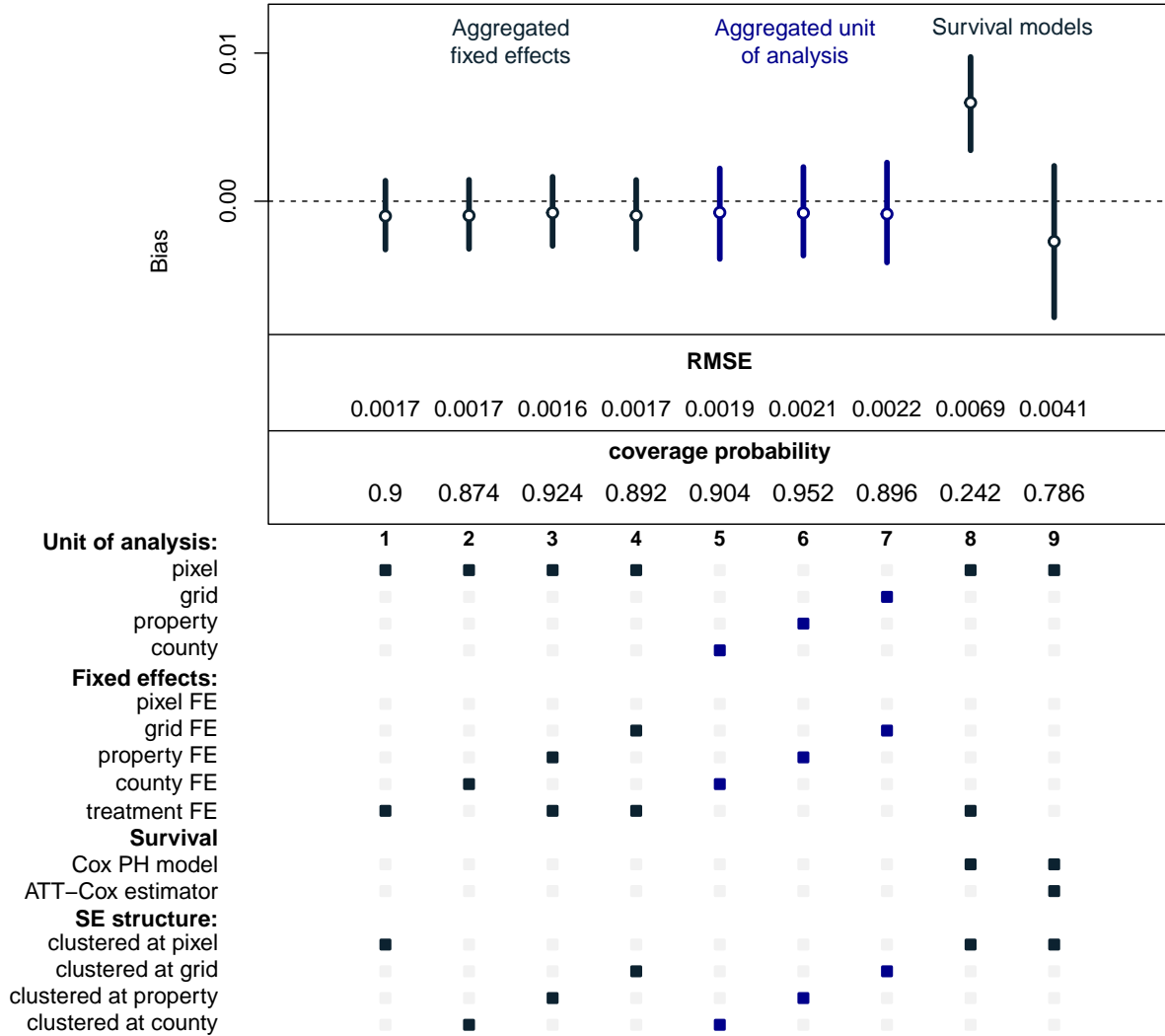


Figure 4: Comparison of model performance in the presence of non-random sample selection. Candidate models are separated by whether they incorporate aggregated fixed effects in pixel-level specifications, aggregated units of analysis, or survival analysis. Point illustrates mean bias while the confidence interval illustrates the 0.05 to 0.95 quantile range of this bias.

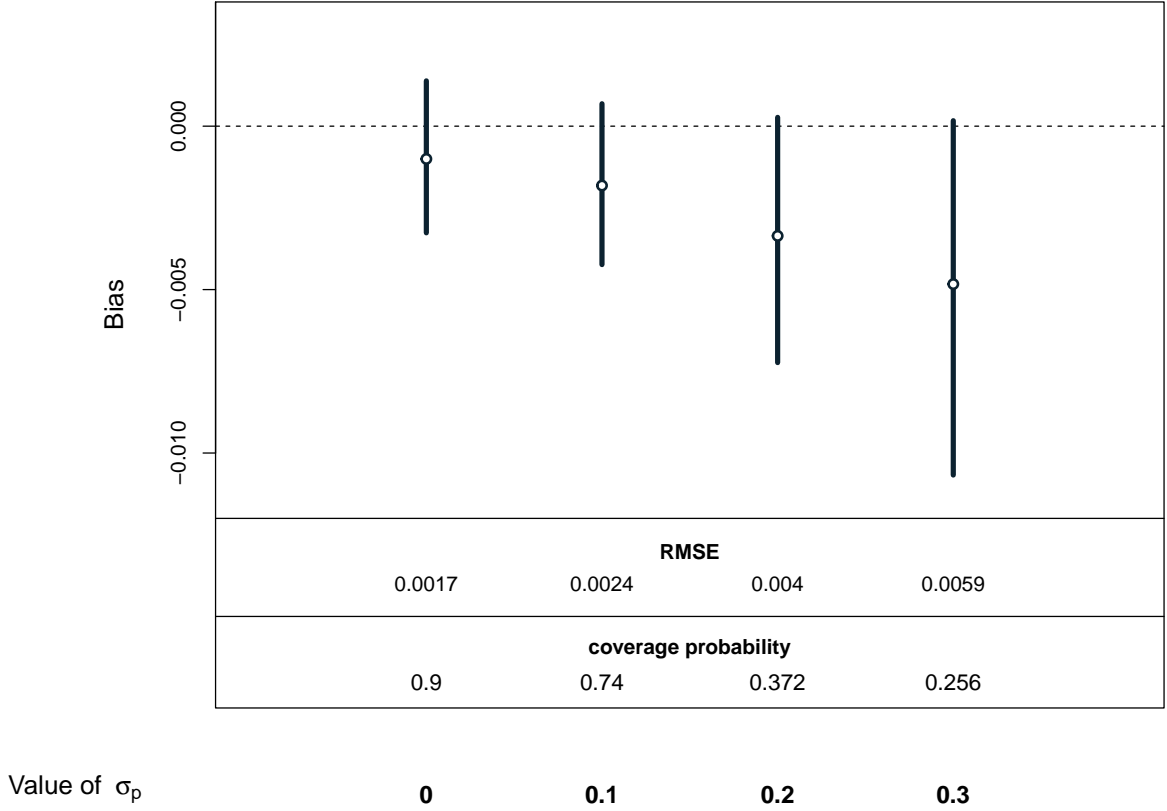


Figure 5: Comparison of pixel-level DID performance as the relative scale of property level disturbances ( $\sigma_p$ ) increases. Point illustrates mean bias while the confidence interval illustrates the 0.05 to 0.95 quantile range of this bias from the Monte Carlo simulations. Standard errors are clustered at the pixel.

In Figure 6 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 3; left panel) or use a spatially aggregated unit of analysis (Regression 4; right panel). Further, we see that the scale of spatial aggregation plays a role. Models incorporating property-level fixed effects suffered from relatively less bias, had lower RMSE, and yielded coverage closest to the expected 0.95 relative to models using larger or smaller scales. In Appendix A.6, we show that these findings generally hold under alternate parameterizations of the baseline deforestation rates and true treatment effect sizes.

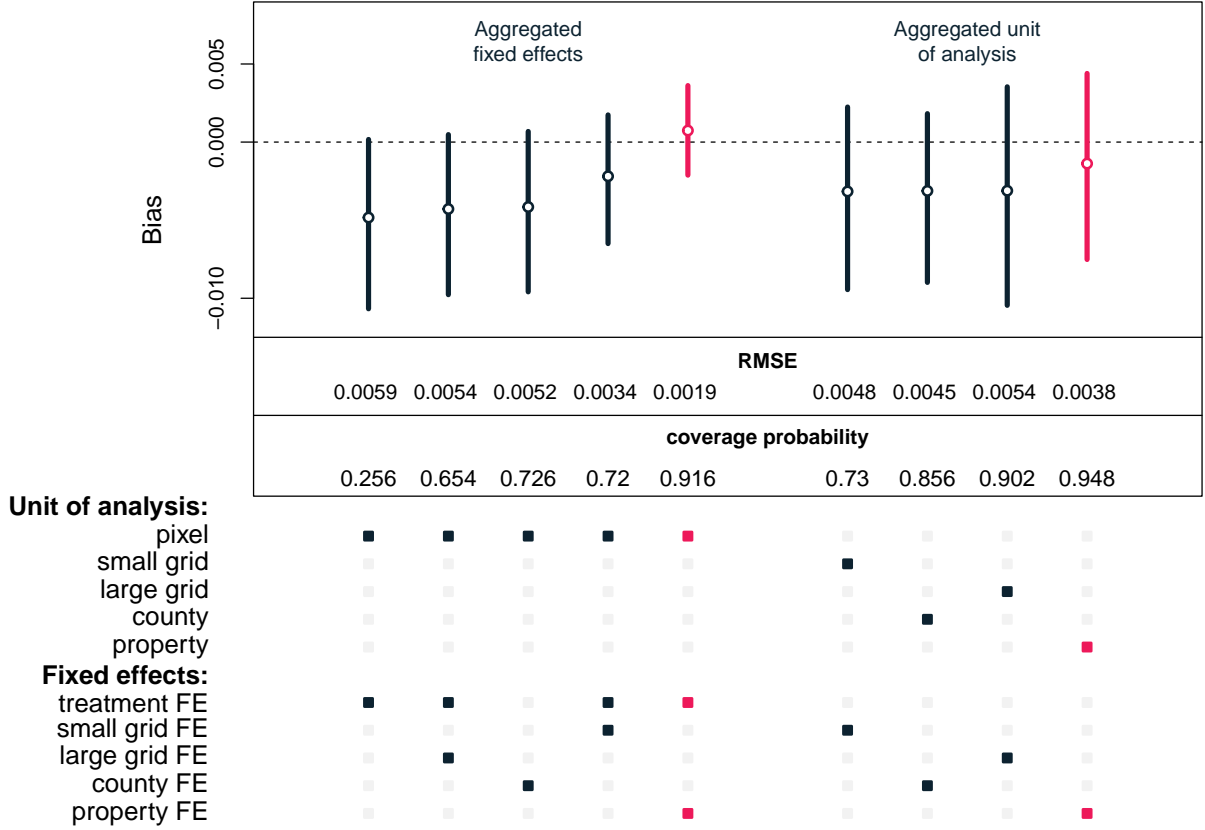


Figure 6: Comparison of model performance under specifications with aggregated unit fixed effects (left panel) and specifications with aggregated units of analysis (right panel) when  $\sigma_p = 0.3$ . Models are ordered by absolute value of bias for ease of comparison in this plot. Point illustrates mean bias while the confidence interval illustrates the 0.05 to 0.95 quantile range of this bias from the Monte Carlo simulations. Standar errors are clustered at the level of unit fixed effects.

Although spatial aggregation can improve the performance of OLS-based model specifications, there is no clear analog for survival models. Figure 13 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. Survival analysis-based estimators may perform better when the preferred unit of analysis is actually the individual (i.e., mortality or recidivism of individual people).

### 5.5.2 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 A.4.1.

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 calculated directly, because  $y(0)_{iv2}$  is not observable 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 7 shows the sample  $ATT$ s for both the property and landscape. In our simulation,  $ATT_{property} = -0.0136$ , and  $ATT_{ls} = -0.0092$ . 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.

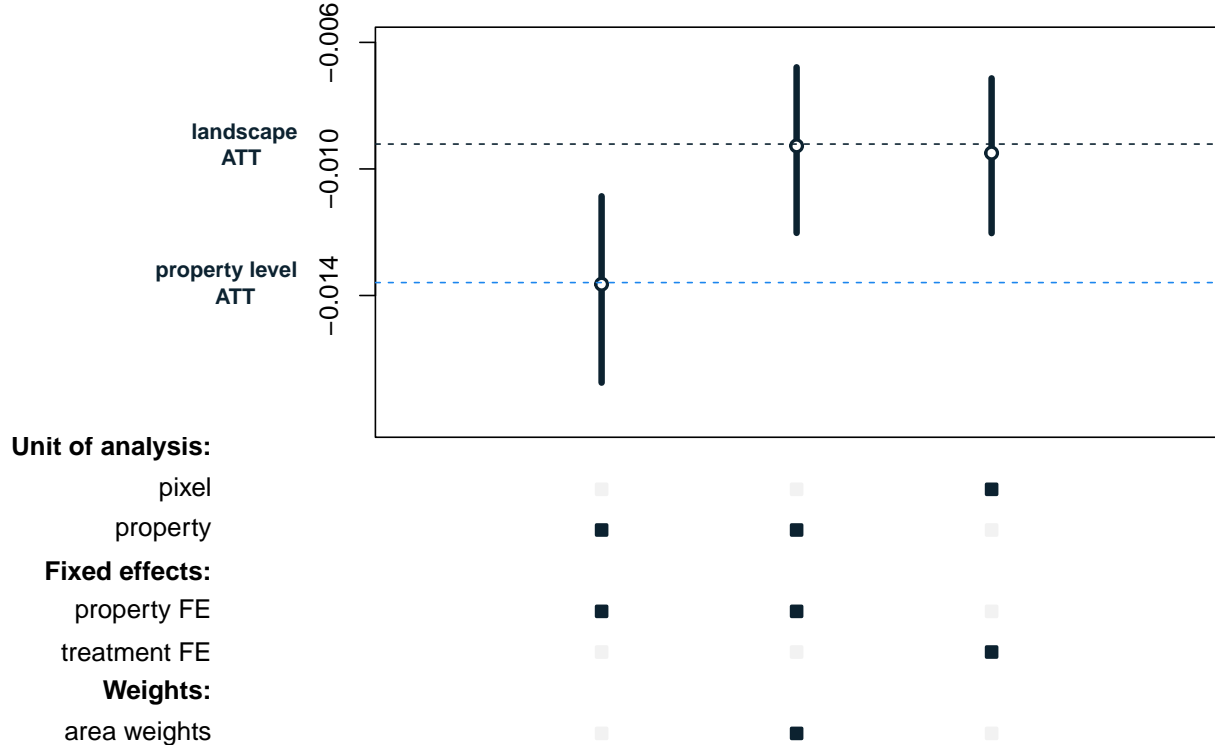


Figure 7: Comparison of model coefficients when treatment effect varies across properties. Point illustrates mean bias while the confidence interval illustrates the 0.05 to 0.95 quantile range of this bias from the Monte Carlo simulations. Standard errors are clustered at the unit of analysis.

## 5.6 Estimating the $ATT$ under staggered treatment

### 5.6.1 Staggered setup

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’Haultfoeuille 2020). Newly developed DID estimators seek to 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 5.1.

### 5.6.2 New DID estimators 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 A.7). 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 8 shows the observed deforestation rates ( $E[y_{it}^o]$ ) from one iteration of our simulation in this setting.

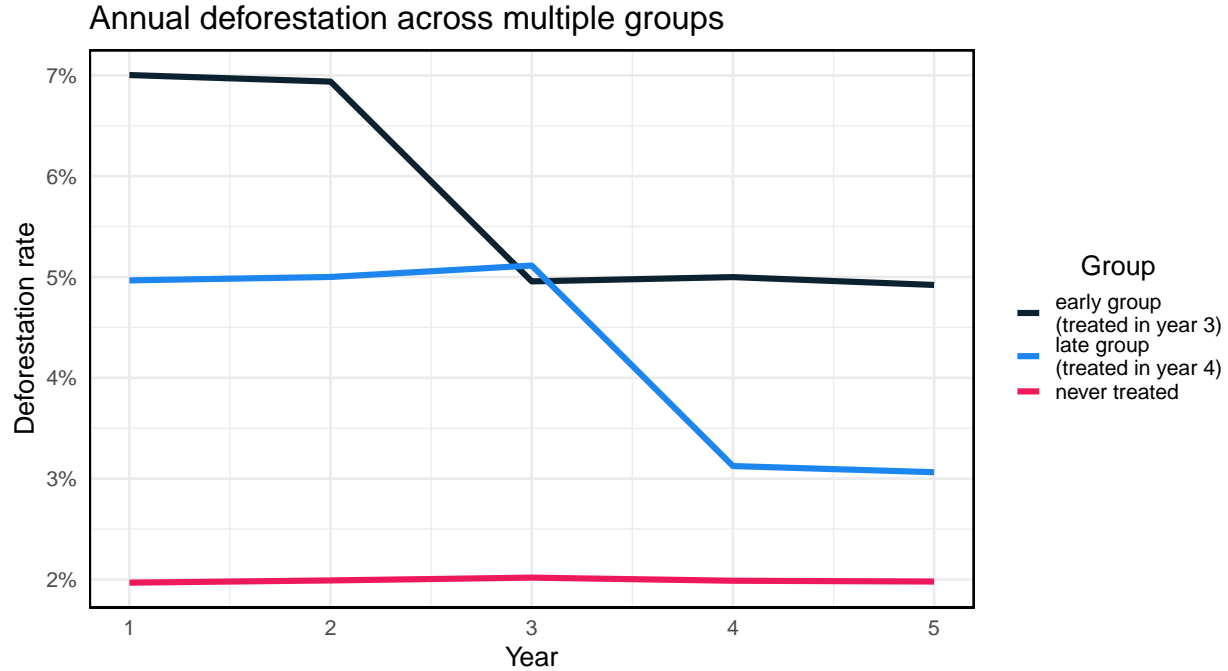


Figure 8: Observed deforestation in simulated landscape with multiple groups and variation in treatment timing.

The left panel of Figure 9 shows that the estimators developed in Sun and Abraham (2021), Gardner (2021), and Roth and Sant’Anna (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. Interestingly, the Callaway and Sant’Anna (2020) estimator did recover the true treatment effect. When allowing for an unbalanced panel, the estimator computes each separately available 2x2 group-time DID, before aggregating them into a summary measure of the *ATT* by event time.

The right panel of Figure 9 shows that the bias associated with TWFE and most newer estimators 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 at this time.



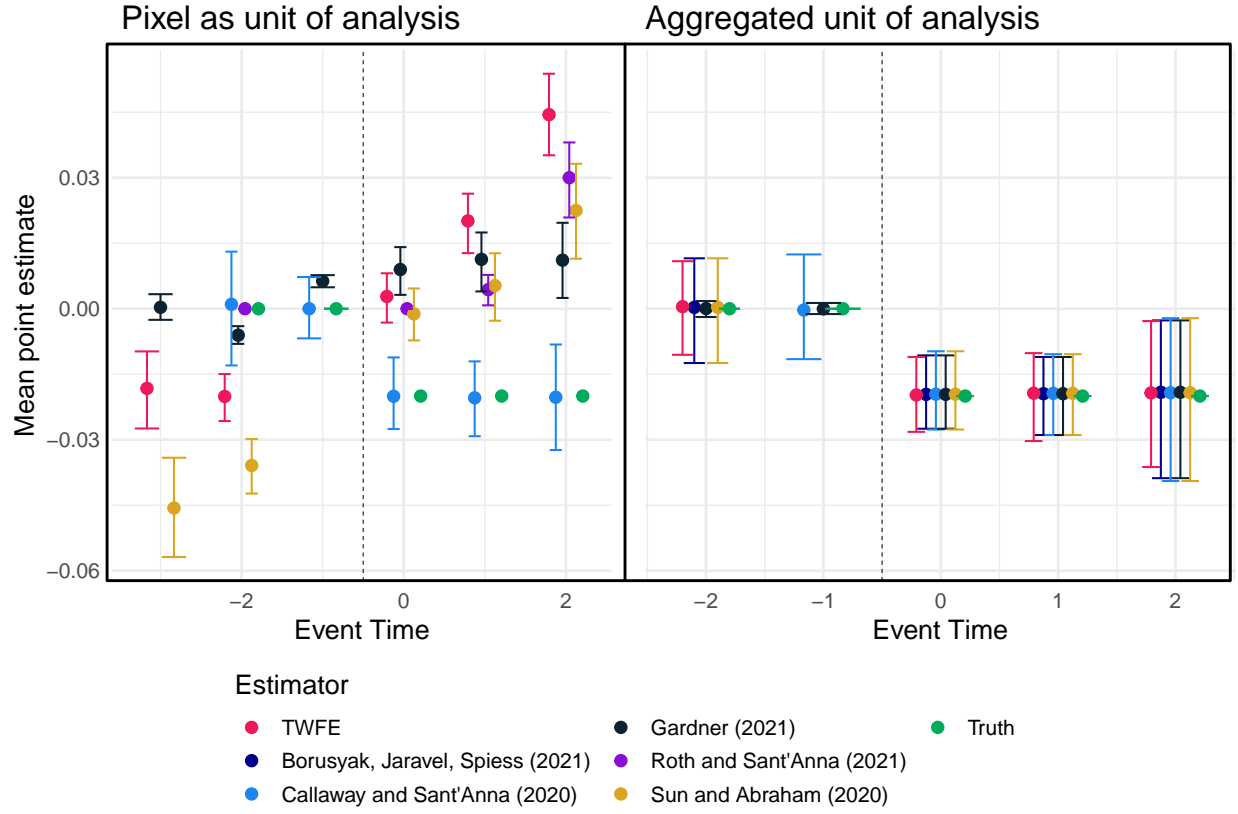


Figure 9: Comparison of event time coefficient estimates by estimator under pixel unit of analysis (left panel) and spatially aggregated (county) unit of analysis (right panel). Point illustrates mean coefficient estimate while the confidence interval illustrates the 0.05 to 0.95 quantile range of estimates. Standard errors clustered at unit of analysis.

### 5.6.3 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 A.7.1. Figure 10 shows deforestation rates in each of the three groups through time.

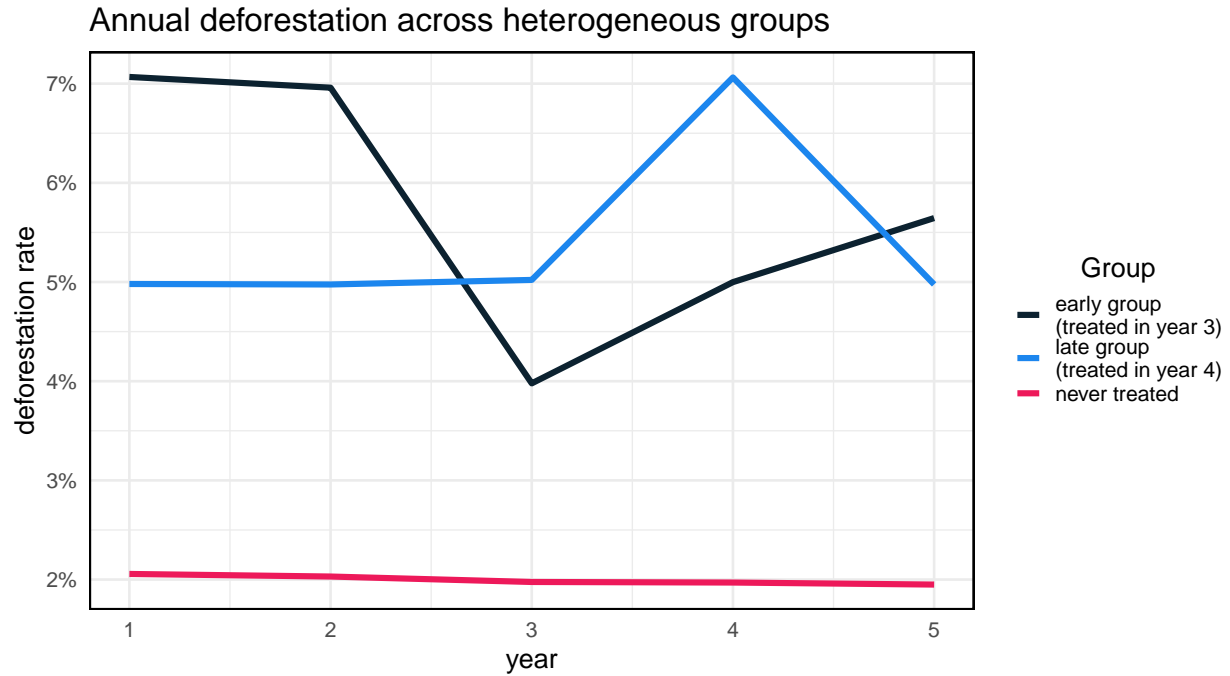


Figure 10: Observed deforestation in simulated landscape when treatment effects vary across groups and through time.

Figure 11 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.

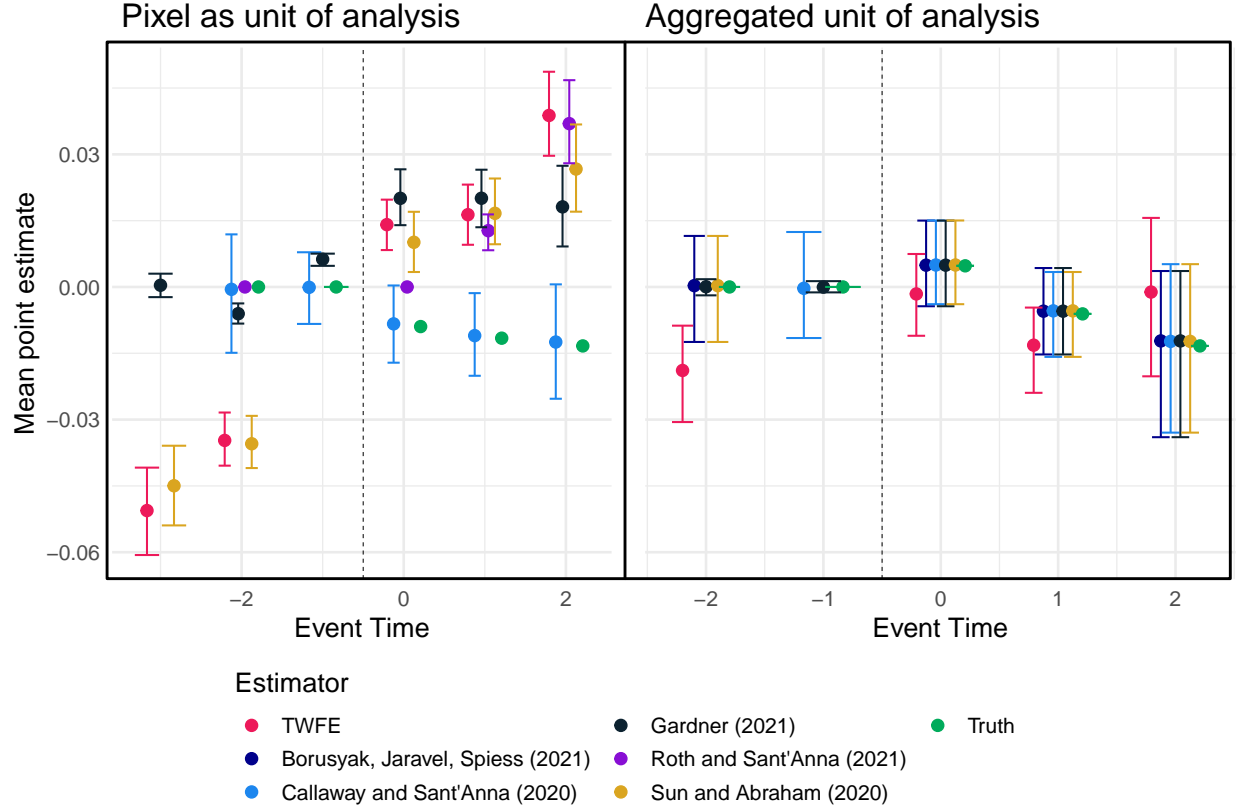


Figure 11: Comparison of event time coefficient estimates by estimator under pixel unit of analysis (left panel) and spatially aggregated (county) unit of analysis (right panel). Point illustrates mean coefficient estimate while the confidence interval illustrates the 0.05 to 0.95 quantile range of estimates. Standard errors clustered at unit of analysis.

## 6 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.

This paper complements 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 2022; Proctor, Carleton, and Sum 2023). 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.

This paper focuses upon efforts to identify the impacts of conservation policies on deforestation. However, the lessons we highlight are relevant to a wider audience, and many of our key findings apply to diverse settings in which the outcome of interest represents an irreversible, binary event. For example, estimates of the impacts of policies on unidirectional technology adoption, recidivism, or mortality may suffer from the same biases that 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.

## 7 Code availability

This entire paper, including the underlying data, results, figures, and tables can be reproduced using code available at this link.

## 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 and TWEEDS. 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.

## 8 References

- Agan, Amanda Y., and Michael D. Makowsky. 2018. “The Minimum Wage, EITC, and Criminal Recidivism.” *NBER WORKING PAPER SERIES* Working Paper 25116.
- 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 Millimet. 2022. “Remotely Incorrect? Accounting for Nonclassical Measurement Error in Satellite Data on Deforestation.” *Journal of the Association of Environmental and Resource Economists*, December, 723723. <https://doi.org/10.1086/723723>.
- 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>.
- 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>.

- Araujo, Claudio, Catherine Araujo Bonjean, Jean-Louis Combes, Pascale Combes Motel, and Eustaquio J. Reis. 2009. "Property Rights and Deforestation in the Brazilian Amazon." *Ecological Economics* 68 (8-9): 2461–68. <https://doi.org/10.1016/j.ecolecon.2008.12.015>.
- Arriagada, R. A., P. J. Ferraro, E. O. Sills, S. K. Pattanayak, and S. Cordero-Sancho. 2012. "Do Payments for Environmental Services Affect Forest Cover? A Farm-Level Evaluation from Costa Rica." *Land Economics* 88 (2): 382–99. <https://doi.org/10.3368/le.88.2.382>.
- Avelino, Andre Fernandes Tomon, Kathy Baylis, and Jordi Honey-Rosés. 2016. "Goldilocks and the Raster Grid: Selecting Scale When Evaluating Conservation Programs." *PLOS ONE* 11 (12): e0167945. <https://doi.org/10.1371/journal.pone.0167945>.
- Baehr, Christian, Ariel BenYishay, and Bradley Parks. 2021. "Linking Local Infrastructure Development and Deforestation: Evidence from Satellite and Administrative Data." *Journal of the Association of Environmental and Resource Economists* 8 (2): 375–409. <https://doi.org/10.1086/712800>.
- Baylis, Kathy, Jordi Honey, and M Isabel Ramírez. 2012. "Conserving Forests: Mandates, Management or Money?" *Selected Paper Prepared for Presentation at the Agricultural & Applied Economics Association's 2012 AAEA Annual Meeting*, August, 17.
- 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>.
- BenYishay, Ariel, Silke Heuser, Daniel Runfola, and Rachel Trichler. 2017. "Indigenous Land Rights and Deforestation: Evidence from the Brazilian Amazon." *Journal of Environmental Economics and Management* 86 (November): 29–47. <https://doi.org/10.1016/j.jeem.2017.07.008>.
- 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>.
- . 2015. "Strict Versus Mixed-Use Protected Areas: Guatemala's Maya Biosphere Reserve." *Ecological Economics* 112 (April): 14–24. <https://doi.org/10.1016/j.ecolecon.2015.01.009>.
- Blackman, Allen, Leonardo Corral, Eirivelthon Santos Lima, and Gregory P. Asner. 2017. "Titling Indigenous Communities Protects Forests in the Peruvian Amazon." *Proceedings of the National Academy of Sciences* 114 (16): 4123–28. <https://doi.org/10.1073/pnas.1603290114>.
- 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, A. Justin Kirkpatrick, and Steven Sexton. 2022. "Visibility and Peer Influence in Durable Good Adoption." *Marketing Science* 41 (3): 453–76. <https://doi.org/10.1287/mksc.2021.1306>.
- Börner, Jan, Dario Schulz, Sven Wunder, and Alexander Pfaff. 2020. "The Effectiveness of Forest Conservation Policies and Programs." *Annual Review of Resource Economics* 12 (1): 45–64. <https://doi.org/10.1146/annurev-resource-110119-025703>.
- 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>.
- Bueno, Carycruz, and Tim Sass. 2018. "The Effects of Differential Pay on Teacher Recruitment and Retention." *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3296427>.
- 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>.
- Butsic, Van, Catalina Munteanu, Patrick Griffiths, Jan Knorn, Volker C. Radeloff, Juraj Lieskovský, Daniel Mueller, and Tobias Kuemmerle. 2017. "The Effect of Protected Areas on Forest Disturbance in the Carpathian Mountains 1985-2010: Carpathian Protected Areas." *Conservation Biology* 31 (3): 570–80. <https://doi.org/10.1111/cobi.12835>.
- 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>.
- Dolan, Carrie B., Ariel BenYishay, Karen A. Grépin, Jeffery C. Tanner, April D. Kimmel, David C. Wheeler, and Gordon C. McCord. 2019. "The Impact of an Insecticide Treated Bednet Campaign on All-Cause Child Mortality: A Geospatial Impact Evaluation from the Democratic Republic of Congo." Edited by Luzia Helena Carvalho. *PLOS ONE* 14 (2): e0212890. <https://doi.org/10.1371/journal.pone.0212890>.
- Donaldson, Dave, and Adam Storeygard. 2016. "The View from Above: Applications of Satellite Data in Economics." *Journal of Economic Perspectives* 30 (4): 171–98.
- 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.
- Emmert-Streib, Frank, and Matthias Dehmer. 2019. "Introduction to Survival Analysis in Practice." *Machine Learning and Knowledge Extraction* 1 (3): 1013–38. <https://doi.org/10.3390/make1030058>.
- Feng, Li, and Tim R. Sass. 2018. "The Impact of Incentives to Recruit and Retain Teachers in 'Hard-to-Staff' Subjects: Incentives to Recruit and Retain Teachers." *Journal of Policy Analysis and Management* 37 (1): 112–35. <https://doi.org/10.1002/pam.22037>.
- Fernández-Val, Iván, and Martin Weidner. 2016. "Individual and Time Effects in Nonlinear Panel Models with Large N , T." *Journal of Econometrics* 192 (1): 291–312. <https://doi.org/10.1016/j.jeconom.2015.12.014>.
- Ferraro, P. J., and M. M. Hanauer. 2014. "Quantifying Causal Mechanisms to Determine How Protected Areas Affect Poverty Through Changes in Ecosystem Services and Infrastructure." *Proceedings of the National Academy of Sciences* 111 (11): 4332–37. <https://doi.org/10.1073/pnas.1307712111>.
- 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): 5311–18. <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.
- Gibson, John, Susan Olivia, Geua Boe-Gibson, and Chao Li. 2021. "Which Night Lights Data Should We Use in Economics, and Where?" *Journal of Development Economics* 149 (March): 102602. <https://doi.org/10.1016/j.jdeveco.2020.102602>.
- 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, 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, and Eric F. Lambin. 2016. "Impacts of Nonstate, Market-Driven Governance on Chilean Forests." *Proceedings of the National Academy of Sciences* 113 (11): 2910–15. <https://doi.org/10.1073/pnas.1600394113>.
- 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>.
- Herrera, Diego, Alexander Pfaff, and Juan Robalino. 2019. "Impacts of Protected Areas Vary with the Level of Government: Comparing Avoided Deforestation Across Agencies in the Brazilian Amazon." *Proceedings of the National Academy of Sciences* 116 (30): 14916–25. <https://doi.org/10.1073/pnas.1802877116>.
- Holland, Margaret B., Kelly W. Jones, Lisa Naughton-Treves, José-Luis Freire, Manuel Morales, and Luis

- Suárez. 2017. “Titling Land to Conserve Forests: The Case of Cuyabeno Reserve in Ecuador.” *Global Environmental Change* 44 (May): 27–38. <https://doi.org/10.1016/j.gloenvcha.2017.02.004>.
- Holland, Paul W. 1986. “Statistics and Causal Inference.” *Journal of the American Statistical Association* 81 (396): 945–60. <https://doi.org/10.1080/01621459.1986.10478354>.
- Imai, Kosuke, and In Song Kim. 2021. “On the Use of Two-Way Fixed Effects Regression Models for Causal Inference with Panel Data.” *Political Analysis* 29 (3): 405–15. <https://doi.org/10.1017/pan.2020.33>.
- 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.
- Lancaster, Tony. 2000. “The Incidental Parameter Problem Since 1948.” *Journal of Econometrics* 95 (2): 391–413. [https://doi.org/10.1016/S0304-4076\(99\)00044-5](https://doi.org/10.1016/S0304-4076(99)00044-5).
- 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>.
- Li, Longxiang, Francesca Dominici, Annelise J. Blomberg, Falco J. Bargagli-Stoffi, Joel D. Schwartz, Brent A. Coull, John D. Spengler, Yaguang Wei, Joy Lawrence, and Petros Koutrakis. 2022. “Exposure to Unconventional Oil and Gas Development and All-Cause Mortality in Medicare Beneficiaries.” *Nature Energy* 7 (2): 177–85. <https://doi.org/10.1038/s41560-021-00970-y>.
- Luallen, Jeremy, Jared Edgerton, and Deirdre Rabideau. 2018. “A Quasi-Experimental Evaluation of the Impact of Public Assistance on Prisoner Recidivism.” *Journal of Quantitative Criminology* 34 (3): 741–73. <https://doi.org/10.1007/s10940-017-9353-x>.
- 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>.
- Miteva, Daniela A., Subhrendu K. Pattanayak, and Paul J. Ferraro. 2012. “Evaluation of Biodiversity Policy Instruments: What Works and What Doesn’t?” *Oxford Review of Economic Policy* 28 (1): 69–92. <https://doi.org/10.1093/oxrep/grs009>.
- Nolte, Christoph, Beatriz Gobbi, Van Butsic, and Eric F Lambin. 2017. “Decentralized Land Use Zoning Reduces Large-scale Deforestation in a Major Agricultural Frontier.” *Ecological Economics*, 7.
- Panlasigui, Stephanie. 2018. “Impacts of Certification, Uncertified Concessions, and Protected Areas on Forest Loss in Cameroon, 2000 to 2013.” *Biological Conservation*, 7.
- 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>.
- Proctor, Jonathan, Tamma Carleton, and Sandy Sum. 2023. “Parameter Recovery Using Remotely Sensed Variables.” *NBER WORKING PAPER SERIES* Working Paper 30861.
- Puterman, Eli, Jordan Weiss, Benjamin A. Hives, Alison Gemmill, Deborah Karasek, Wendy Berry Mendes, and David H. Rehkopf. 2020. “Predicting Mortality from 57 Economic, Behavioral, Social, and Psychological Factors.” *Proceedings of the National Academy of Sciences* 117 (28): 16273–82. <https://doi.org/10.1073/pnas.1918455117>.

- 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).
- Rico-Straffon, Jimena, Zhenhua Wang, Stephanie Panlasigui, Colby J. Loucks, Jennifer Swenson, and Alexander Pfaff. 2023. "Forest Concessions and Eco-Certifications in the Peruvian Amazon: Deforestation Impacts of Logging Rights and Logging Restrictions." *Journal of Environmental Economics and Management* 118 (March): 102780. <https://doi.org/10.1016/j.jeem.2022.102780>.
- 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>.
- Roth, Jonathan. 2022. "Pretest with Caution: Event-Study Estimates After Testing for Parallel Trends." *American Economic Review: Insights* 4 (3): 305–22. <https://doi.org/10.1257/aeri.20210236>.
- Roth, Jonathan, and Pedro H. C. Sant'Anna. 2021. "Efficient Estimation for Staggered Rollout Designs." *arXiv:2102.01291 [Econ, Math, Stat]*, June. <http://arxiv.org/abs/2102.01291>.
- 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 Brazil's 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>.
- Shah, Payal, and Kathy Baylis. 2015. "Evaluating Heterogeneous Conservation Effects of Forest Protection in Indonesia." Edited by Dafeng Hui. *PLOS ONE* 10 (6): e0124872. <https://doi.org/10.1371/journal.pone.0124872>.
- Sims, Katharine R. E., and Jennifer M. Alix-Garcia. 2017. "Parks Versus PES: Evaluating Direct and Incentive-Based Land Conservation in Mexico." *Journal of Environmental Economics and Management* 86 (November): 8–28. <https://doi.org/10.1016/j.jeem.2016.11.010>.
- Sun, Liyang, and Sarah Abraham. 2021. "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects." *Journal of Econometrics* 225 (2): 175–99. <https://doi.org/10.1016/j.jeconom.2020.09.006>.
- Tabor, Karyn, Kelly W. Jones, Jennifer Hewson, Andriambolantsoa Rasolohery, Andoniaina Rambelison, Tokihenintsoa Andrianjohaninarivo, and Celia A. Harvey. 2017. "Evaluating the Effectiveness of Conservation and Development Investments in Reducing Deforestation and Fires in Ankeniheny-Zahemena Corridor, Madagascar." Edited by Mauro Villarini. *PLOS ONE* 12 (12): e0190119. <https://doi.org/10.1371/journal.pone.0190119>.
- Torchiana, Adrian L, Ted Rosenbaum, Paul T Scott, and Eduardo Souza-Rodrigues. 2022. "Improving Estimates of Transitions from Satellite Data: A Hidden Markov Model Approach\*."
- Wendland, K. J., M. Baumann, D. J. Lewis, A. Sieber, and V. C. Radeloff. 2015. "Protected Area Effectiveness in European Russia: A Postmatching Panel Data Analysis." *Land Economics* 91 (1): 149–68. <https://doi.org/10.3368/le.91.1.149>.
- Williams, David R., Andrew Balmford, and David S. Wilcove. 2020. "The Past and Future Role of Conservation Science in Saving Biodiversity." *Conservation Letters* n/a (n/a): e12720. <https://doi.org/10.1111/conl.12720>.
- Wooldridge, Jeffrey M. 2010. "Econometric Analysis of Cross Section and Panel Data, Second Edition." The MIT Press.

## Appendix

### A.1 Illustrative pixel-level panel with irreversible data structure

Below, we present a simple example of panel data with the structure described in section 2. Suppose we observe a pixelated map of nine pixels with structure similar to Hansen et al. (2013) and are interested in deforestation over four years, where treatment of some pixels begins in year 3. Following the above guidance yields the panel dataset. Note that panel datasets with this structure still maintain variation pre- and post-treatment.



Pixel 1: never cleared	Pixel 2: cleared Year 2	Pixel 3: cleared Year 3
Pixel 4: cleared Year 1	Pixel 5: cleared Year 4	Pixel 6: cleared Year 2
Pixel 7: cleared Year 4	Pixel 8: never cleared	Pixel 9: cleared Year 3

Figure 12: Example pixelated deforestation data, indicating year of deforestation.

Period	Year	Pixel 1	Pixel 2	Pixel 3	Pixel 4	Pixel 5	Pixel 6	Pixel 7	Pixel 8	Pixel 9
Pre-treatment	1	0	0	0	1	0	0	0	0	0
	2	0	1	0	-	0	1	0	0	0
Post-treatment	3	0	-	1	-	0	-	0	0	1
	4	0	-	-	-	1	-	1	0	-

## A.2 Analytical results

### A.2.1 Setup

Let  $y_{it}$  be the binary outcome of interest for individual unit  $i$  at time  $t$ . We assume that researchers have access to outcome data pre-treatment ( $t = 1$ ) and post-treatment ( $t = 2$ ). Some units ( $D_i = 1$ ) are exposed to a policy treatment in the second time period ( $t_0 = 2$  denotes the time of first treatment for treated points). Let  $W_{it} = 1$  if unit  $i$  is treated before time  $t$  and  $W_{it} = 0$  otherwise. Using the potential outcome notation, denote  $y_{it}(0)$  the outcome of unit  $i$  at time  $t$  if it does not receive treatment by time  $t$  and  $y_{it}(1)$  the outcome for the same unit if it receives treatment.

Thus, the realized outcome for unit  $i$  at time  $t$  is

$$y_{it} = W_{it}y_{it}(1) + (1 - W_{it})y_{it}(0)$$

.

The parameter of interest, the *ATT* is defined:

$$ATT = E[y_{it}(1) - y_{it}(0)|D_i = 1]$$

Make the following common trends assumption, under which we evaluate these methods:

$$E[y_{i2}(0) - y_{i1}(0)|D_i = 1] = E[y_{i2}(0) - y_{i1}(0)|D_i = 0]$$

Lastly, define  $C_i$  as the first year in which an irreversible event of interest (e.g., deforestation) is realized for individual unit  $i$  and suppose  $y_{it}$  is not observable when  $t > C_i$ .

### A.2.2 TWFE regression models with point fixed effects do not identify $ATT$

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

We define the observed outcome  $y_{it}^o$ :

$$y_{it}^o = \begin{cases} 1 & t = C_i \\ 0 & t < C_i \\ - & t > C_i \end{cases} \quad (17)$$

, where  $y_{it}^o = -$  indicates that the outcome for pixel  $i$  has been dropped from the panel in time  $t$ .

Lastly, define the traditional individual unit level TWFE regression:

$$y_{it}^o = \alpha + \beta_{TWFE} \times W_{it} + \gamma_i + \lambda_t + u_{it}$$

, where

- $\gamma_i$  indicate point fixed effects
- $\lambda_t$  indicate year fixed effects \

In the 2x2 case, we can write

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

and

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

, where  $\eta_{t=2}$ , an indicator for the post-treatment period, subsumes  $\lambda_t$ . Note that we substituted  $W_{it}$  for  $D_i$ , since the two are equivalent post-treatment.

In the 2x2 case, the TWFE estimator is equivalent to the first differences estimator, and yields:

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

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

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

The general expression can be restated as:

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

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:

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

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

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

Thus far, we have shown that  $\hat{\beta}_{TWFE}$  is equal to the ex-post difference in means between treatment and control units.

We now examine what this means for estimating the parameter of interest, the *ATT*.

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

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

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

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

Taking the expectation gives:

$$E[\hat{\beta}_{TWFE}] = ATT + E[y_{i2}(0)|D_i = 1] - E[y_{i2}(0)|D_i = 0]$$

, where the expectation of the  $y_{i2}^o$ s is equal to that of the  $y_{i2}$ s if they are i.i.d..

$$\beta_{TWFE} = ATT + E[y_{i2}(0)|D_i = 1] - E[y_{i2}(0)|D_i = 0]$$

Now, imposing the common trends assumption, substituting for  $E[y_{i2}(0)|D_i = 1]$ , and simplifying:

$$\beta_{TWFE} = ATT + E[y_{i1}(0)|D_i = 1] - E[y_{i1}(0)|D_i = 0]$$

### A.2.3 Cox PH DID identifies *HRTT* when proportional trends assumption holds

Consider the cox proportional hazards model of the censored  $y_{it}$  regressed on the treatment dummy,  $D_i$ , the post dummy,  $\mathbb{1}\{t \geq t_0\}$ , and their interaction:

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

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

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:

$$\exp(\beta_{coxDID}) = \frac{E[y_{i2}|D_i = 1]/E[y_{i1}|D_i = 1]}{E[y_{i2}|D_i = 0]/E[y_{i1}|D_i = 0]} \quad (18)$$

Introducing potential outcomes and simplifying:

$$\exp(\beta_{coxDID}) = \frac{E[y_{i2}(1)|D_i = 1]E[y_{i1}(0)|D_i = 0]}{E[y_{i2}(0)|D_i = 0]E[y_{i1}(0)|D_i = 1]} \quad (19)$$

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

$$\exp(\beta_{coxDID}) = \frac{E[y_{i2}(1)|D_i = 1]E[y_{i1}(0)|D_i = 1]}{E[y_{i2}(0)|D_i = 1]E[y_{i1}(0)|D_i = 1]} \quad (20)$$

, showing that under proportional trends,  $\exp(\beta_{coxDID}) = HRTT$

### A.2.4 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*.

Define an alternative observed outcome  $y_{it}^{alt}$ , where rather than dropping units from the panel, the outcome is imputed as a 1 when  $t > C_0$ .

$$y_{it}^{alt} = \begin{cases} 0 & t < C_0 \\ 1 & t \geq C_0 \end{cases} \quad (21)$$

Now, consider the difference-in-differences estimand with respect to outcome  $y_{it}^{alt}$ :

$$\psi = E[y_{i2}^{alt}|D_i = 1] - E[y_{i1}^{alt}|D_i = 1] - (E[y_{i2}^{alt}|D_i = 0] - E[y_{i1}^{alt}|D_i = 0])$$

Because  $y_{it}^{alt}$  is binary,  $\psi$  can be re-expressed using probabilities:

$$\psi = P(y_{i2}^{alt} = 1|D_i = 1) - P(y_{i1}^{alt} = 1|D_i = 1) - (P(y_{i2}^{alt} = 1|D_i = 0) - P(y_{i1}^{alt} = 1|D_i = 0))$$

Because we observe  $y_{it}^{alt} = 1$  anytime after  $C_0$ , we can express the probability that  $y_{it}^{alt} = 1$  as a function of the probability that  $C_0$  occurred prior to time  $t$ . Then, the terms with  $y_{it}^{alt}$ s are re-expressed.

$$\begin{aligned} \psi = & P(y_{i2} = 1|D_i = 1) \cup P(y_{i1} = 1|D_i = 1) - P(y_{i1} = 1|D_i = 1) - \\ & (P(y_{i2} = 1|D_i = 0) \cup (P(y_{i1} = 1|D_i = 0) - P(y_{i1} = 1|D_i = 0))) \end{aligned}$$

Applying the potential outcomes notation:

$$\begin{aligned} \psi = & P(y_{i2}(1) = 1|D_i = 1) \cup P(y_{i1}(0) = 1|D_i = 1) - P(y_{i1}(0) = 1|D_i = 1) - \\ & (P(y_{i2}(0) = 1|D_i = 0) \cup (P(y_{i1}(0) = 1|D_i = 0) - P(y_{i1}(0) = 1|D_i = 0))) \end{aligned}$$

The bias that results from using  $y_{it}^{alt}$  is expressed as the difference between the  $ATT$  and  $\psi$ . After re-expressing the  $ATT$  in terms of probability:

$$\begin{aligned} \psi - ATT = & P(y_{i2}(1) = 1|D_i = 1) \cup P(y_{i1}(0) = 1|D_i = 1) - P(y_{i1}(0) = 1|D_i = 1) - \\ & (P(y_{i2}(0) = 1|D_i = 0) \cup (P(y_{i1}(0) = 1|D_i = 0) - P(y_{i1}(0) = 1|D_i = 0))) \\ & - [P(y_{i2}(1) = 1|D_i = 1) - P(y_{i2}(0) = 1|D_i = 1)] \end{aligned}$$

Under the common trends assumption and substituting for the unobserved term,  $P(y_{i2}(0) = 1|D_i = 1)$ , yields

$$\begin{aligned} \psi - ATT = & P(y_{i2}(1) = 1|D_i = 1) \cup P(y_{i1}(0) = 1|D_i = 1) - P(y_{i1}(0) = 1|D_i = 1) - \\ & (P(y_{i2}(0) = 1|D_i = 0) \cup (P(y_{i1}(0) = 1|D_i = 0) - P(y_{i1}(0) = 1|D_i = 0))) \\ & - [P(y_{i2}(1) = 1|D_i = 1) - (P(y_{i2}(0) = 1|D_i = 0) - P(y_{i1}(0) = 1|D_i = 0) + P(y_{i1}(0) = 1|D_i = 1))] \\ = & P(y_{i2}(1) = 1|D_i = 1) \cap P(y_{i1}(0) = 1|D_i = 1) + P(y_{i2}(1) = 1|D_i = 1) - \\ & (P(y_{i2}(0) = 1|D_i = 0) \cap (P(y_{i1}(0) = 1|D_i = 0) + P(y_{i2}(0) = 1|D_i = 0))) \\ & - [P(y_{i2}(1) = 1|D_i = 1) - (P(y_{i2}(0) = 1|D_i = 0) - P(y_{i1}(0) = 1|D_i = 0) + P(y_{i1}(0) = 1|D_i = 1))] \end{aligned}$$

, where the second equality uses the definition of union operator.

Simplifying now yields:

$$\begin{aligned} \psi - ATT = & P(y_{i2}(1) = 1|D_i = 1) \cap P(y_{i1}(0) = 1|D_i = 1) - \\ & (P(y_{i2}(0) = 1|D_i = 0) \cap (P(y_{i1}(0) = 1|D_i = 0) \\ & - [P(y_{i1}(0) = 1|D_i = 1) - P(y_{i1}(0) = 1|D_i = 0)]) \end{aligned}$$

### A.2.5 Analytical expression of non-random sample selection bias in two-period two-group setting

Consider again, the observed outcome,  $y_{it}^o$ . We begin with the *DID* estimand in the two-group, two-period case:

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

Now the bias generated due to non-random sample selection can be represented as the difference between this estimand and the *ATT*:

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

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

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

Applying potential outcomes:

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

Applying our common trends assumption:

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

Finally, simplifying:

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

### A.3 Further Monte Carlo evidence that individual unit-level TWFE is equivalent to coefficient from same regression on dataset without pixels deforested pre-treatment

Table 3 shows coefficient estimates from the Monte Carlo setup described in the main text on altered datasets. It demonstrates that the coefficient of interest from Regression 2 is numerically equivalent to that from the same regression on a dataset where all pixels deforested in the first period are excluded from the dataset completely. The estimated coefficient is not numerically equivalent to the ex-post difference in means, although this is true in the 2x2 case. This exercise provides further evidence that this commonly used TWFE regression does not use the pre-treatment variation in deforestation at all, which is necessary to recover the *ATT* in this setting.

Table 3: TWFE with pixel fixed effects is numerically equivalent to TWFE on dataset with all pixels deforested pre-treatment 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

#### A.4 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:

$$\begin{aligned}
ATT &= E[y_{it}(1) - y_{it}(0)|t \geq t_0, D_i = 1] \\
&= E[y_{it}(1)|t \geq t_0, D_i = 1] - E[y_{it}(0)|t \geq t_0, D_i = 1] \\
&= P(y_{it}(1) = 1|t \geq t_0, D_i = 1) - P(y_{it}(0) = 1|t \geq t_0, D_i = 1) \\
&= P(y_{it}^*(1) > 0|t \geq t_0, D_i = 1) - P(y_{it}^*(0) > 0|t \geq t_0, D_i = 1) \\
&= P(\beta_0 + \beta_1 + \beta_{2,1} + \beta_3 + \alpha_i + u_{it} > 0) - P(\beta_0 + \beta_1 + \beta_{2,1} + \alpha_i + u_{it} > 0) \\
&= P(-\alpha_i - u_{it} < \beta_0 + \beta_1 + \beta_{2,1} + \beta_3) - P(-\alpha_i - u_{it} < \beta_0 + \beta_1 + \beta_{2,1}) \\
&= F(\beta_0 + \beta_1 + \beta_{2,1} + \beta_3) - F(\beta_0 + \beta_1 + \beta_{2,1}) \\
\\
trend_0 &= E[y_{it}(0)|t \geq t_0, D_i = 0] - E[y_{it}(0)|t < t_0, D_i = 0] \\
&= P(y_{it}(0) = 1|t \geq t_0, D_i = 0) - P(y_{it}(0) = 1|t < t_0, D_i = 0) \\
&= P(y_{it}^*(0) > 0|t \geq t_0, D_i = 0) - P(y_{it}^*(0) > 0|t < t_0, D_i = 0) \\
&= \frac{(1 - P(y_{it}^*(0) > 0|t < t_0, D_i = 0))P(y_{it}^*(0) > 0|t \geq t_0, D_i = 0)}{(1 - P(y_{it}^*(0) > 0|t < t_0, D_i = 0))} - P(y_{it}^*(0) > 0|t < t_0, D_i = 0) \\
&= P(-\alpha_i - u_{it} < \beta_0 + \beta_{2,0}) - P(-\alpha_i - u_{it} < \beta_0) \\
&= F(\beta_0 + \beta_{2,0}) - F(\beta_0)
\end{aligned}$$

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

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

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

, 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  $\beta_3$

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

#### A.4.1 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

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

#### A.5 Full summary figure from all specifications and values of $\sigma_p$

Using Figure 13 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.

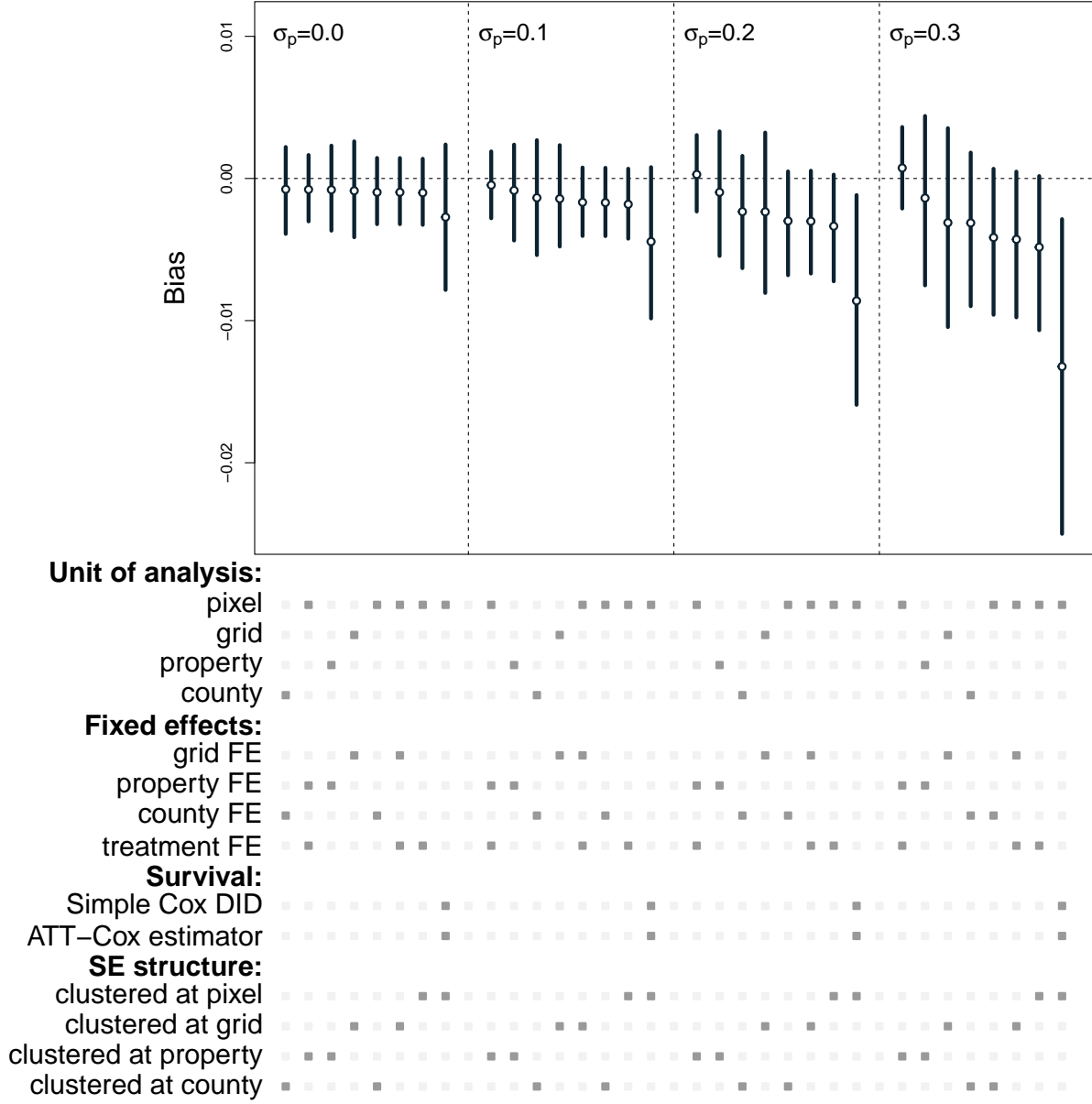


Figure 13: Comparison of model performance, separated by value of  $\sigma_p$ . Point illustrates mean bias while the confidence interval illustrates the 0.05 to 0.95 quantile range of this bias.

## A.6 Property-level models still suffer from least bias in alternative parameterizations

Figure 14 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 with a positive  $ATT$  (i.e.  $ATT = 0.01$  rather than  $-0.01$ ).

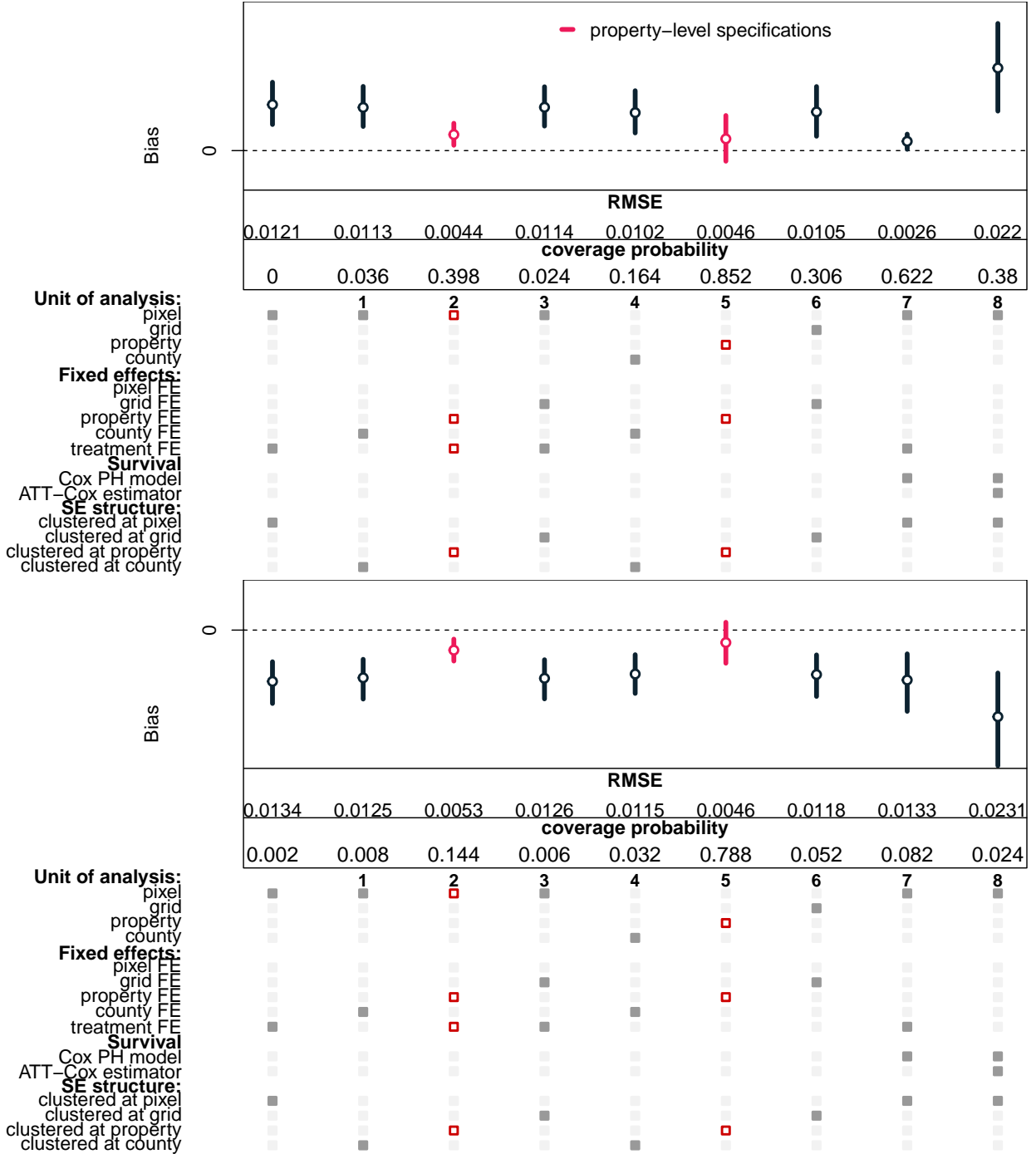


Figure 14: Comparison of model performance in alternative landscape parameterizations. Candidate models are separated by whether they incorporate aggregated fixed effects in pixel-level specifications, aggregated units of analysis, or survival analysis. Point illustrates mean bias while the confidence interval illustrates the 0.05 to 0.95 quantile range of this bias.

## A.7 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:

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

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

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

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

### A.7.1 parameterization for heterogeneous treatment effects example

The following parameters and their definitions inform the simulation parameterizations.

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

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  $N(0, \sigma_a^2 + \sigma_u^2)$

## A.8 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.

The formula used in the main text is:

$$\text{Outcome 1:} = \frac{F_{i,t-1} - F_{it}}{F_{i,t-1}} \quad (22)$$

, 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:

$$\text{Outcome 2:} = \frac{F_{i0} - F_{it}}{F_{i0}} \quad (23)$$

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:

$$\text{Outcome 3:} = \ln(F_{i,t-1}/F_{it}) \quad (24)$$

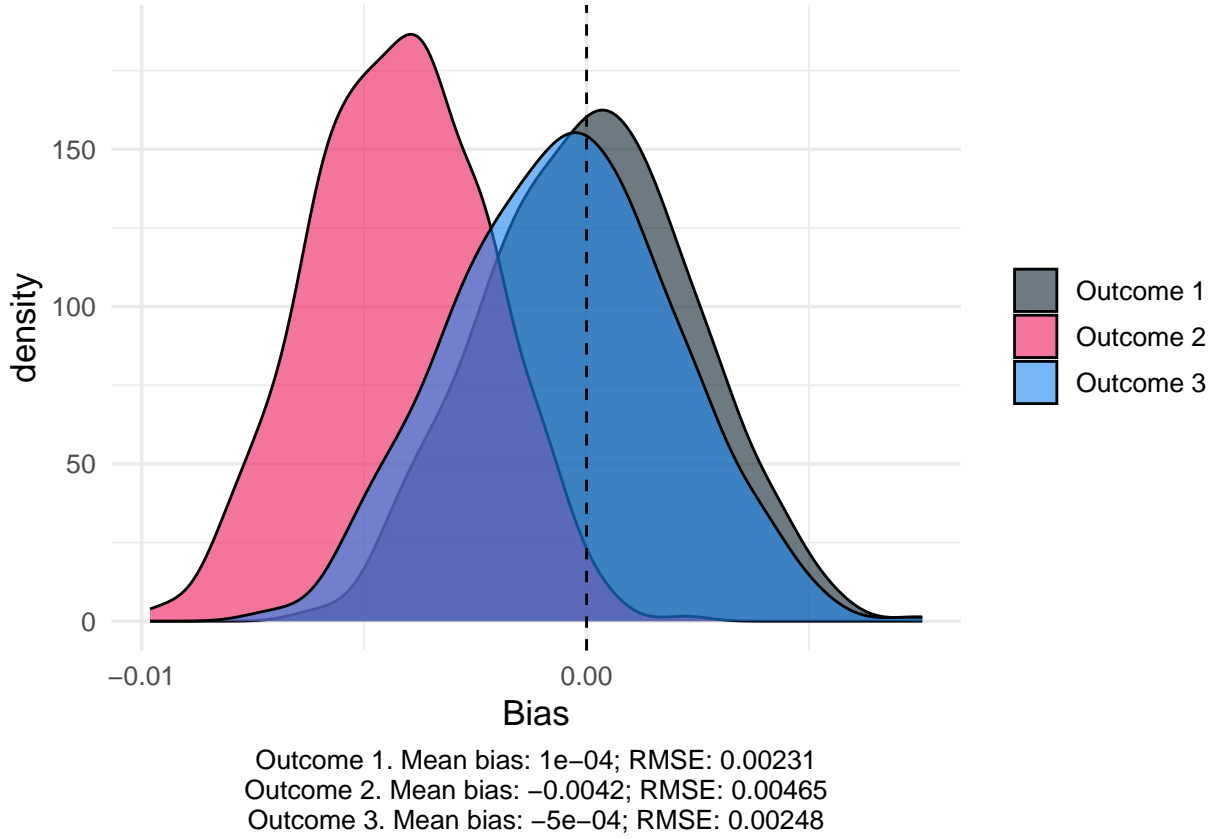


Figure 15: Distribution of estimates produced by different outcome variable formulae

Figure 15 demonstrates that outcome 1 results in the least bias and lowest RMSE in our guiding example. The other outcomes result in somewhat greater bias, although the differences between outcomes 1 and 3 is minimal in our setting. We do raise a concern about the use of outcome 2, especially for studies spanning relatively long time periods. In later time periods, the outcome variable will monotonically increase, even if the true rate of deforestation has not changed through 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 the formula used should be explicitly stated in a 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. Throughout our paper, all specifications using aggregated data use outcome 1. In our guiding example, it resulted in the least bias and lowest RMSE, and our understanding is that it is currently the most common deforestation rate calculation used in the literature.