

Practical guidance for conservation impact evaluation using remotely sensed data

Alberto Garcia and Robert Heilmayr

April 22, 2020

Abstract

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

Introduction

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

Increasingly, conservation science has turned to econometric methods of impact evaluation to disentangle causal relationships (Butsic et al. 2017; Baylis et al. 2016). While randomized experiments are the gold standard for scientific discovery in both the natural and social sciences, conservation often poses questions that are impossible or unethical to pursue through experimentation. In such settings, a growing portfolio of statistical techniques enable researchers to draw causal conclusions using observational data. When implemented carefully, these methods can yield conclusions that are comparable to what a researcher would discover if they were able to run a randomized experiment (Ferraro and Miranda 2017). (Ferraro, Sanchirico, and Smith 2019)

Econometric methods generate accurate estimates of an intervention's impact when they effectively control for the non-random assignment of the intervention. A low rate of deforestation within a remote protected area may reflect the protected area's effectiveness, or it may be indicative of the remote location's poor suitability for agricultural development (Andam et al. 2008; Pfaff et al. 2009). Econometric methods control for both observed and unobserved... Panel data settings, in which units are observed repeatedly through time, greatly facilitate the measurement of treatment effects and program impacts (Blackman 2013). XX - ELABORATE ON SUITE OF METHODS - panel, matching... XX importance of counterfactual thinking (Ferraro 2009; meyfroidt2016; Ribas et al. 2020)

Econometric methods of causal inference have become particularly useful for conservation impact evaluation as a result of the proliferation of remotely sensed data (Blackman 2013; Jones and Lewis 2015). For example, NASA’s landsat missions provide detailed and consistent information on land use spanning the entirety of the world since the 1970s. As a result, a scientist hoping to quantify the impacts of a land use policy adopted decades ago can assemble longitudinal data for treated and control units that span both pre- and post-implementation periods. (Jain 2020)

However, many of the remotely sensed metrics used in these applications have structural differences from the data used in more traditional applications of causal inference. These differences include... Here we demonstrate that, as a result of these differences, many of the econometric models used in this growing literature are likely biased - significance, magnitude and even direction of estimated effects might be incorrect. These biases arise even when researchers follow common guidance to adopt “rigorous” research designs with valid counterfactuals [Blackman (2013); jones2015]. Based on a review of the existing literature, we identify key model design decisions that researchers need to make. Using Monte Carlo simulations and analytical proofs, we show how these design decisions affect the validity of , we illustrate ways in which scientists can tailor these design decisions to minimize the bias in their impact evaluations.

Key considerations for impact evaluations of deforestation

We consider the case in which a researcher would like to quantify the impact that an intervention has had on deforestation rates. We assume that the intervention has clearly defined boundaries (e.g. a protected area, certified concession, or indigenous territory), and that the researcher has access to spatially explicit observations of forest cover and forest loss spanning the periods before and after the intervention was adopted. The researcher’s goal is to measure the Average Treatment Effect on the Treated (*ATT*), which quantifies the change in deforestation occurring inside treated units when compared to an unobservable counterfactual world in which no conservation intervention was adopted (Figure 1). The fundamental problem of causal inference is that, for every treated unit, we fail to observe the value that the outcome would have taken in the absence of treatment ((???)). Figure 1 displays this problem in the context of our simulated conservation intervention. Assuming that the intervention reduced deforestation, we do not observe the counterfactual deforestation that would have occurred in the intervention area in the absence of the intervention. This general setting describes a broad array of research studies that apply panel methods to remotely sensed data (Table XX).

- need to add Holland, Paul W. 1986. “Statistics and causal inference.” Journal of the American statistical Association 81 (396): 945–960.

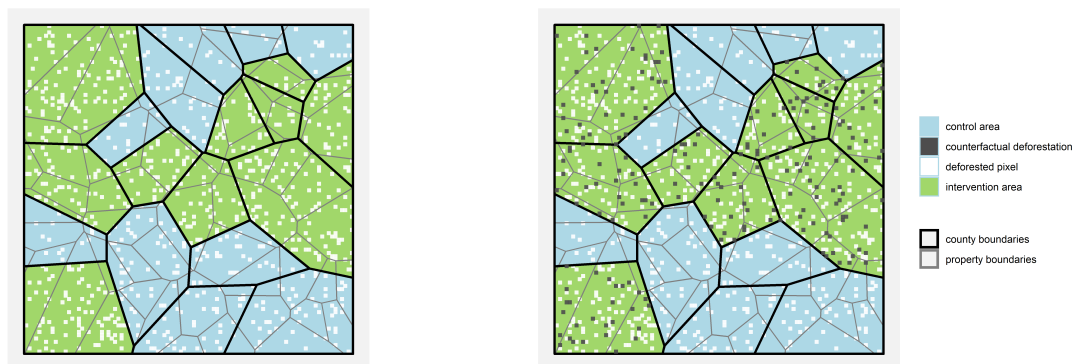


Figure 1: Left: deforestation observed in the period after the policy has been implemented in the intervention area; Right: deforestation that would have occurred in the absence of the conservation intervention. Deforestation in the control area is the same in both cases, since no intervention occurred.

In panel data settings, two methods in particular are often used to measure the impact of conservation interventions: Difference-in-Differences (DID) and Two-way Fixed Effects (FE) regression models (Blackman 2013; Jones and Lewis 2015).

The typical DID regression model includes a dummy equal to one for units in the treatment group, a dummy equal to one for observations in the period after the intervention, and their interaction. Conceptually, the DiD estimator calculates the treatment effect as the difference between the differences of the treated and non-treated observations before and after treatment (Butsic et al. 2017). In our example, this is the difference between post and pre-treatment deforestation rates in observations with the policy, minus the difference between post and pre-treatment deforestation rates in the observations without the policy.

FE regression models are often used to apply DID methods to multiple groups or treatment periods. This amounts to estimating a regression that controls for unit and time fixed effects. Intuitively, this can be thought of as including a dummy variable for each unit of analysis and each time period. The fixed effects account for any unobservable confounding variables that may vary across units or through time. In the case of two groups and two time periods, the FE regression should give an estimate equivalent to that of the DID model (Wooldridge, n.d.). Because FE regression models are often used to generalize the DID method, they are used in a wider variety of settings. Settings in which units undergo treatment in more than two distinct time periods may be amenable to an FE regression, but not the standard DID method. For example, a researcher may use an FE 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, which enrolls properties in annual cohorts.

The DID and FE models identify the ATT in the two-period two-group case under one main assumption. This assumption is known as the common trends assumption, which amounts to assuming that both units in the intervention area and untreated units would have experienced the same average change in the outcome in the absence of treatment.

Unit of analysis

- unit of analysis and aggregation
- Alix-Garcia and Gibbs 2017: We avoided two challenges by using the point as the analysis unit rather than an entire property or a grid cell of arbitrary size. The first was that results could be interpreted directly as the average effect across the landscape rather than the effect for a property with average characteristics, without re-weighting by property size. This was useful for assessing the full avoided deforestation impacts rather than property-level responses to policy. The second advantage came in the assignment of property boundaries to the points – each point had an easily attributable first year of application to the CAR. This is not the case with properties, which can potentially have shifts or overlaps in boundaries over time that make it difficult to assign a unique date of first application to a property.
- Carlson et al. 2018: We conducted our analysis at the plantation level to match the scale at which many plantation management decisions occur
- Alix-Garcia and Wolff 2014: Indirectly addressing slippage, Arriagada et al. (2012) conduct a whole-farm analysis of Costa Rica’s PES program. By including the entire area owned by an individual farmer, rather than just the area enrolled in the program, the authors can implicitly measure the program impact net of leakage.
- Blackman et al. 2018: Our FMU-level analysis has advantages and disadvantages. The disadvantage is that we are not able to measure effects that occur only inside harvest zones. The advantage is that we control for spatial spillover effects that occur when forest management certification on one part of an FMU spurs deforestation on other parts
- We had property boundaries from a cadastral survey of all titled households in our study area; thus, our unit of analysis represented the household decision-making unit
- Jones and Lewis 2015: Since PES is a household-level program, the ideal unit of analysis is the household parcel. spatial boundaries of household parcels are available for this study area from a cadastral survey conducted by the Government of Ecuador. If this information were not available then it would be difficult to rigorously evaluate the impact without collecting property boundaries using GPS, even

though some evaluations of PES have used pixels, or grid cells, for evaluation [36], these do not represent the decision-making unit

- JL 15: a plot is removed from the dataset once change occurs so that forest disturbed on this plot—the outcome variable—is not double counted

calculating deforestation rates

area weights

- papers weight the regression or outcome to account for units of analysis that are heterogeneous in size
- ex:

Functional form

- Busch 15 pnas: Poisson is theoretically consistent with forest cover loss within a 3 km x 3 km grid cell being the count of many independent, discrete binary observations of forest cover loss or maintenance at the level of 30m x 30 m remote sensing data. A Poisson QMLE model tolerates zero values, and generates a distribution of predicted values that fits the distribution of observed data, which is concentrated nearest to zero deforestation and diminishes toward greater levels of deforestation
- Carlson 18 Since our four outcome variables were all nonnegative, include zero, and not overdispersed, we used a Poisson model to assess the impact of certification on deforestation and fire activity
- Mohebalian and Aguilar 2018: We chose logit models, over probit regression, because the former provides similar results but can be more easily interpreted as odd ratios and can be more robust at either bounds of a probabilistic distribution curve
- Many papers, ex: Blackman 2015, do not justify use of non-linear functional form. Most default to OLS or simply check that their estimates are robust to different functional forms.
- Wendland et al. : We choose a linear probability model over non-linear Probit or Logit models for two reasons. First, plot fixed effects cannot be easily included in non-linear discrete-choice models in a flexible manner. Fixed plot effects cannot be included in a Probit model due to the incidental parameters problem, and fixed effects Logit models do not allow for calculation of marginal effects since marginal effects are non-linear functions of the un-estimated fixed effects (see Wooldridge 2010, Ch. 15).
standard error structure JL15: For cross-sectional and fixed effects regressions we use cluster robust standard errors to control for spatial autocorrelation. Clustering standard errors relaxes the assumption of no correlation across observations within the spatial unit used for clustering.

Methods

Monte Carlo simulations

To study how choices pertaining to model selection influence estimates of conservation program effectiveness, we employ a set of Monte Carlo simulations to generate and analyze data. To begin, consider a setting in which a policy is implemented to reduce deforestation at a discrete point in time. The researcher observes treatment and control units in two time periods, before and after the implementation of the policy. Forest cover outcome data is obtained at the pixel resolution, where pixels are determined to either be forested or deforested for a given time period.

The data generating process underlying our Monte Carlo simulations begins with the assignment of four parameters: The pre-treatment period deforestation rate for untreated units, $baseline_0$; the pre-treatment period deforestation rate for treated units, $baseline_1$; the difference in the deforestation rate between the period in which the treatment has been implemented and the pre-treatment period for the untreated units, $trend$; and lastly, the average treatment effect of the policy on the untreated units, ATT . The ATT is the primary parameter the researcher is interested in uncovering.

We then define the latent variable,

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

The β coefficients are derived from the above four parameters assigned by the researcher (see appendix 1). The treatment variable, D_i , is equal to 1 when pixel i is treated. The period in which the treatment is implemented is denoted t_0 . The pixel specific parameter is generated according to $\alpha_i \sim N(0, .1^2)$ and the error term is generated according to $u_{it} \sim N(0, .25^2)$. This latent variable encapsulates the underlying spatial processes determining the policy’s impact on deforestation within the area represented by a given pixel. This is unobservable to the researcher in its current form.

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

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

Here, the observed outcome variable, y_{it} , is equal to 1 if pixel i is observed as deforested in time t and 0 otherwise. The observed variable is the binary outcome visible to the researcher. It tells the researcher whether the pixel is classified as deforested or forested. The area represented by a pixel may not be entirely deforested but may still be classified as deforested using the binary outcome metric.

Set up and evaluation measures

Based on baseline parameters that appeared to be common in the literature, we’ve selected a guiding example to explore for the remainder of the paper. We’ve set $baseline_0 = 0.13$, $baseline_1 = 0.17$, $trend = 0.02$, and lastly, the true $ATT = -0.08$. The primary criteria we use to evaluate different methods are bias and coverage probability. Using our Monte Carlo simulations, we determine bias by computing the difference between the mean of the coefficient estimates and the ATT parameter. Coverage probability is defined as the proportion of simulations in which the true ATT lies within the simulation’s 95% confidence interval (CI). As such, we would expect the ATT to lie within this CI 95% of the time, however, factors such as the bias of the estimates, their distribution, and treatment of standard errors may impact coverage. If the estimator is biased, for example, it is ex-ante less likely to contain the true parameter within the CI.

Results

Bias inherent to binary deforestation DGP

Because of the data generating process inherent to remotely sensed metrics of deforestation as well as our framework, there is often bias. In the case of a binary pixel as in our DGP, the researcher does not observe the underlying spatial process of deforestation, but instead, a pixelated version according to a forest cover share cutoff. The area represented by a pixel may experience a decline in forest cover, but this deforestation will only be observable if it pushes the pixel’s total forested share below the cutoff (expressed by equation (2) in our DGP). This is representative of most analyses using binary satellite measures of deforestation. For example, in the widely used Hansen et al. (2013) global forest change dataset, forest loss is defined as a stand-replacement disturbance and disaggregated by reference percent tree cover stratum (e.g. >50% crown cover to 0% crown cover) and by year. Note that any time the latent DGP has error generated according to a distribution with a nonlinear cumulative distribution function (e.g. normal), this type of bias will arise from the DGP. The magnitude and direction of the resulting bias depend largely on the $baseline_0$, $baseline_1$, $trend$, and ATT parameters. Understanding at a minimum the direction of the bias this imposes on a researcher’s estimates can aid in proper inference.

In Figure 1, we can see this bias in the context of our guiding example. Allowing the outcome to vary between 0 and 1 across time periods allows us to see the bias due only to the disparity between the underlying spatial data generating process and the observed binary outcome.

- figure allowing outcome to vary between 0 and 1 showing bias

This bias is separate from that which may arise from satellite sensor characteristics, satellite angle, or atmospheric conditions (Jain (2020)). We do not explore the interaction of these two potential sources of bias in our paper, however, it is likely an important consideration. For the remainder of the paper, we net out the bias due to the discussed underlying spatial processes in order to focus on bias that arises from various model selection decisions.

Unit of analysis

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

In order to account for these dynamics in the context of deforestation, it has been advised to drop deforested pixels in the periods after they first become deforested (Jones and Lewis 2015; Alix-Garcia and Gibbs 2017). The logic for doing so is as follows. A forested pixel switches from its assigned value of 0 to a value of 1 following a discrete deforestation event. Keeping the deforested pixel in the panel beyond the first period in which it was observed as deforested may imply that it has actively been deforested in each subsequent time period. In fact, no new deforestation event has occurred, but it simply remains deforested from the prior event. Indeed, we see in our Monte Carlo simulations that regression models failing to drop deforested pixels in subsequent periods incur severe bias, while dropping the pixels lessens this bias.

Figure 2 demonstrates the magnitude and direction of the bias incurred from keeping deforested pixels in the panel after they are first realized as deforested in the context of our guiding example. We see that the direction of the bias is positive, which aligns with our intuition from above. The positive bias seems to stem from deforested observations in years subsequent to the actual deforestation event contributing to the deforestation rate. As a result, pixels that were deforested prior to the implementation of the policy continued to contribute to the deforestation rate in the post period in both the treatment and control groups. * it might be possible to get an expression for this bias

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

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

We now bring attention to the distinction between specifications (1) and (2). In both specifications, observations are dropped from the panel in the periods after which they are first realized as deforested. As discussed, this is preferable in terms of bias when using the pixel as the unit of analysis. The figure shows that the FE model returns a biased measure of the ATT, and in fact, estimates an ex-post single difference. In our guiding example, the ex-post single difference is -0.04 , representing a decrease of 4% in the deforestation rate, while the assigned *ATT* is equal to -0.08 . Thus, the ex-post single difference as well as

the FE estimate, is biased positively by a rate of 4%, as evidenced both in Figure 3 and the analytical proof found in the appendix. This shows that FE models using the pixel as the unit of analysis is not a viable method to estimate the ATT in deforestation impact evaluations. In contrast, the DID...

Aggregated Outcomes Since pixel level analyses are not feasible in the context of FE regressions, researchers should be aware of the tradeoffs using aggregated units of analysis. The following results apply to both DID and FE regression models, as both are equivalent in the two-period, two-group example. For simplicity, we assume the researcher can choose between three levels at which to aggregate the data: grid cell, county, and property. Grid cells are uniform grids layered over the study area and may have a treatment value between 0 and 1 following aggregation of pixels. Counties are heterogeneous administrative units at which we now assign the treatment. Lastly, properties are smaller administrative units within a county.

We find little evidence that any one level of aggregation is consistently preferable in terms of bias, meaning that coverage probability will play a larger role as an evaluation metric. We explore the bias and coverage probabilities corresponding to different levels of aggregation in the context of our guiding example. Figure 4 shows that the bias of the estimates is not critically different depending on level of aggregation, however the distributions are slightly different. Notice that the distribution of estimates when aggregating to the property level is wider than the others. This is likely due to the heterogeneity in property size. The distribution of the estimates as well as the standard errors will affect the coverage probabilities. Looking at Table 3

Property level unobservables Property level unobservables may impact both treatment effect estimates and coverage probabilities. This is likely to be a factor when land use decisions are made at the property level by the landowner or the policy intervention seeks to alter landowner incentives underlying certain land use activities. We introduce an additional error term to the initial DGP that varies at the property level in order to account for these unobservables.

As property level unobservables play a larger role, the treatment of standard errors also becomes more important.

*plot of p_err vs. coverage by agg method and error structure

Functional form

- options: ols, probit, logit, poisson
- context dependent

Understanding the observed data as well as the underlying DGP is therefore particularly important when it comes to choices of functional form. Factors that may affect a researchers decision are:...

Calculating deforestation rates

- outcome calculations and weighting

Upon aggregating data, the researcher must determine how to calculate deforestation rates in the outcome. With no clear guidance on how this deforestation rate should be computed, there have been a variety of techniques used. We outline some of these methods in Table 2.

	Outcome	Papers
1	fractional deforestation based on lag	Shah and Baylis 2015, Busch et al. 2015, Carlson et al. 2018
2	fractional deforestation based on baseline forest cover	Pfaff 1999

Figure showing outcomes.

As seen in Figure... , outcome 1 results in the least bias in our guiding example, and the distribution is nearly identical to the binary outcome when pixels are dropped in the DID case. The other outcomes result in relatively greater bias.

It is also common practice to weight the outcome by the unit’s area, particularly when the units of analysis are heterogeneous in size. In our simulations, we find that weighting the regression does not impact the bias of the estimators significantly (Figure 5), however, the distribution of estimates is drastically widened. This may have implications for inference. We find that in cases where the bias is relatively low, it is preferable not to weight the regression, as coverage probability is negatively impacted. In Table 3, we show that coverage probability is negatively impacted when the regression is weighted by area.

- Figure 5 showing that weighting doesn’t significantly affect bias, but widens distribution

Level	Weighted	Unweighted
grid		
property		
county		

Discussion

Acknowledgements and data

References

- Alix-Garcia, Jennifer, and Holly K. Gibbs. 2017. “Forest Conservation Effects of Brazil’s Zero Deforestation Cattle Agreements Undermined by Leakage.” *Global Environmental Change* 47 (November): 201–17. <https://doi.org/10.1016/j.gloenvcha.2017.08.009>.
- Andam, K. S, P. J Ferraro, A. Pfaff, G. A Sanchez-Azofeifa, and J. A Robalino. 2008. “Measuring the Effectiveness of Protected Area Networks in Reducing Deforestation.” *Proceedings of the National Academy of Sciences* 105 (42): 16089.
- Baylis, Kathy, Jordi Honey-Rosés, Jan Börner, Esteve Corbera, Driss Ezzine-de-Blas, Paul J. Ferraro, Renaud Lapeyre, U. Martin Persson, Alex Pfaff, and Sven Wunder. 2016. “Mainstreaming Impact Evaluation in Nature Conservation.” *Conservation Letters* 9 (1): 58–64. <https://doi.org/10.1111/conl.12180>.
- Blackman, Allen. 2013. “Evaluating Forest Conservation Policies in Developing Countries Using Remote Sensing Data: An Introduction and Practical Guide.” *Forest Policy and Economics* 34 (September): 1–16. <https://doi.org/10.1016/j.forpol.2013.04.006>.
- 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>.
- Ferraro, Paul J. 2009. “Counterfactual Thinking and Impact Evaluation in Environmental Policy.” *New Directions for Evaluation* 2009 (122): 75–84. <https://doi.org/10.1002/ev.297>.
- Ferraro, Paul J., and Juan José Miranda. 2017. “Panel Data Designs and Estimators as Substitutes for Randomized Controlled Trials in the Evaluation of Public Programs.” *Journal of the Association of Environmental and Resource Economists* 4 (1): 281–317. <https://doi.org/10.1086/689868>.
- Ferraro, Paul J., James N. Sanchirico, and Martin D. Smith. 2019. “Causal Inference in Coupled Human and Natural Systems.” *Proceedings of the National Academy of Sciences* 116 (12): 5311–8. <https://doi.org/10.1073/pnas.1805563115>.
- 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>.
- 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>.

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

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

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

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

Wooldridge, Jeffrey M. n.d. “Clusterextended.Pdf.”

Appendix

Initial parameter to β coefficient mapping

The researcher sets the following four parameters:

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

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:

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

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

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

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

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

Now solving for the β coefficients:

solving for β_0

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

solving for β_1

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

solving for β_2

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

solving for β_3

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

Bias inherent to binary deforestation DGP

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

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

Two-way fixed effects proof