What Are the Effective Policy Interventions That Have Reduced Deforestation in Cocoa-Producing Countries?

by

Jiangyu (Jayden) Peng
Bachelor of Arts (Honours), University of Victoria, 2020

An Extended Essay Submitted in Partial Fulfillment
of the Requirements for the Degree of

MASTER OF ARTS
in the Department of Economics

We accept this extended essay as conforming
to the required standard

Dr. Felix Pretis, Supervisor (Department of Economics)

Dr. Sophia Carodenuto, Co-supervisor (Department of Geography)

Abstract

In this paper, I employ a data-driven approach to identify potential effective policy interventions in cocoa-producing countries that have reduced deforestation by exploring a reverse causal question: "What are the effective policy interventions in these countries?" This approach, originally introduced by Pretis (2022) in a panel model to assess the impacts of the introduction of carbon pricing on Canadian CO_2 emissions, utilizes machine learning to select variables over a full set of individual-time interaction variables. Specifically, it aims to detect structural breaks in a balanced panel setting. These structural breaks can be interpreted as step-shift changes in the fixed effects of individual units at specific points in time. These shifts may signify noteworthy policy interventions or past events. By employing this innovative machine-learning model-selection method, I can identify treatment effects without a priori knowledge of their existence. This paper offers fresh insights into the policies that have effectively reduced deforestation in cocoa-producing countries. My model detects 15 structural breaks, nine of which show negative signs, indicating significant reductions in tree cover loss in the corresponding countries and years. Among these nine breaks, with effect sizes ranging from 45% to 70%, policy attribution suggests that two are linked to economic incentives, while five can be attributed to improvements in forest law and policy. These remarkably high effect sizes require discussion within the context of the significant year-on-year variations in the tree cover loss observed in the sampled countries. I assess the robustness of the estimation results by comparing point estimates and their standard errors across models with varying target significance levels γ_c and a model that controls for the area of cocoa harvested; $\gamma_c$is the primary calibration parameter in the break detection procedure.

Keywords: TWFE panel model, fixed effects, model selection, break detection, machine-learning, tree cover loss, cocoa-producing countries.

Table of Contents

1. **Introduction**

According to a recent study by the World Resources Institute (2023), globally, 4.1 million hectares of primary rainforest were lost in 2022, representing a 10% rise from 2021, and resulting in 2.7 gigatonnes of carbon dioxide emissions. The study suggests that forests burned in temperate and boreal regions are mostly caused by wildfires, which are considered "a natural and important part of the ecology." On the other hand, the majority of deforestation in the tropics happens to primary forests through human-caused events. The study also suggests that more than 96% of total deforestation occurs in tropical primary forests, and the upward trend (Figure 1) poses a growing challenge. Agriculture expansion and logging remain the top human activities that cause deforestation in the tropics. Brazil and the Democratic Republic of the Congo (DRC), the top two countries, in terms of primary forest loss area in several years, contributed significantly to both clear-cut deforestation and small-scale agriculture-related deforestation. The disadvantaged living conditions and technological development in some tropical countries, like Haiti (Saint, 2022), DRC, and Liberia (Alfaro & Jones, 2018), often lead to massive charcoal production, which acts as their main source of domestic energy as well as a medium of greenhouse gas emissions. As for the deforestation occurring in the tropics, Brazil and Indonesia accounted for nearly half of it, while three quarters are driven by agroindustries such as beef production and oilseeds (Ritchie, 2021 & Pendrill, 2019). The studies suggest that palm oil and soybean[1] dominate the cropland expansion for oilseeds, which also include sunflower, rapeseed, and sesame. While cattle meat is primarily consumed domestically, Pendril et al. argue that oilseeds are largely produced as export-oriented commodity crops. With the help of international trade, 62-76% of global oilseeds production embodied deforestation emissions is transferred from the countries of production, mostly Asia-Pacific and Latin America, to Europe, China, and the Middle East. The same goes for other high-value crops such as cocoa, coffee, tea, and spices.

Cocoa production, the foundation of the chocolate industry, has also been acting as one of the major sources driving deforestation in the humid tropics (Igini, 2023). Cocoa comes from the fruit of theobroma cacao, a tree native to the Amazon forest (Budiansky & Freire, 2018). Because theobroma cacao trees are small plants that only grow in tropical forest conditions, a massive amount of primary forests has been cleared by humans for cocoa production, especially in the four western African countries that together produce the amount of cocoa more than half the global demand: Côte d'Ivoire, Ghana, Nigeria, and Cameroon (Igini, 2023). To worsen the problem, slash-and-burn is a widely adopted method to clear forests for cocoa production (Johnson, 2023). The land is often abandoned afterward when it becomes infertile from monoculture farming (The Chocolate Journalist, 2023). Unsurprisingly, cocoa-producing countries largely coincide with the main contributors to deforestation. All 23 countries selected for the analysis in the present paper are both cocoa-producing countries as well as major deforestation contributors.

Despite the upward trend of the deforested area in the tropics and high deforestation rate in countries like Brazil, DRC, and Ghana, some tropical countries achieved certain levels of success in reducing deforestation. For example, the data shows the average primary forest loss has been going downward since 2016 in

[1] Feed of livestock and biofuels are the main consumers of global soybean production, of which only 6% is used for human food such as tofu or soy milk (Ritchie, 2021 & Pendrill et al., 2019).

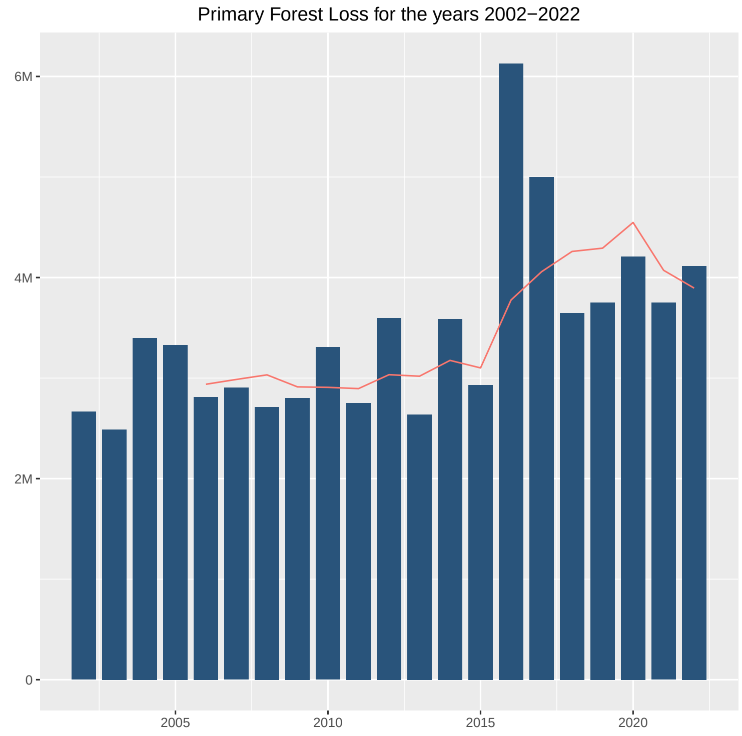

Figure 1: This bar chart shows the area, in hectares, of primary forest lost globally in the years 2002 to 2022. The red line represents 5-year moving average.

both Indonesia and Malaysia[2] (World Resources Institute, 2023). The differences in the deforestation rates and trends among the tropical countries pose questions like - What are the factors driving deforestation downward in countries like Indonesia and Malaysia and why are they successful? On a national level, government intervention is likely one of the biggest forces that can reduce deforestation at a large scale. Many countries have policies and laws to fight deforestation, especially those that joined the REDD+[3] program. In this paper, I locate interventions that potentially made an impact on reducing deforestation nationally from 2002 through 2021 by using a break-detection approach powered by machine learning.

As discussed by Pretis (2022), Pretis and Schwarz (2023), and Koch et al. (2022), break detection has been a well-established tool used in time series analysis. However, it falls short of enabling researchers to draw causal interpretations in time series applications due to the absence of control groups. Nevertheless, the novel approach of model selection using break detection, first introduced by Pretis (2022) and adapted in this study, allows for the inclusion of a control group in the final retained model. This inclusion, in turn, facilitates the drawing of causal interpretations. Through this approach, researchers can potentially address reverse causal questions that were challenging to tackle before this method. Reverse causal questions, as defined by Gelman and Imbens (2013), involve seeking the "causes of effects," in contrast to forward causal questions that estimate the "effects of causes." Pretis and Schwarz (2023) demonstrate that this approach can identify significant variables without a priori knowledge of their existence by saturating a complete set of individual-time interaction terms in a balanced panel model. This step saturation detects step-shift

[2]The primary forest loss is calculated with 30% minimum tree cover canopy density. The data is from the University of Maryland and available on the World Resources Institute's Global Forest Watch platform

[3]REDD+ stands for Reducing Emissions from Deforestation and forest Degradation in developing countries and additional forest-related activities that fight climate change. Established as part of the Paris Agreement, REDD+'s framework motivates developing countries to cut deforestation by result-based payments (United Nations).

changes in the treated individuals' intercepts as structural breaks, which they argue are equivalent to the binary variables included in the conventional difference-in-differences setting as known treatment effects.

The question being explored in the present paper is a reverse causal question I attempt to answer and is split into two steps by the break-detection procedure: first using machine learning to detect notable changes (step-shifts) in tree cover loss in each sample country and subsequent attributing these breaks to policy interventions or other past events. The group of countries in which no breaks are detected in tree cover loss becomes the control group that is used to compare with the treated countries that have breaks detected. This approach also allows me to check the robustness of my detected breaks by selecting different target levels of significance for the break detection. When breaks are detected in models of various target significance levels, these breaks are considered robust. Stable magnitudes of the breaks' estimated effects also imply robustness. The same applies when breaks and their magnitudes remain stable when adding new potentially relevant variables, such as when I add the area harvested for cocoa beans as a variable controlling for the cocoa effect in these cocoa-producing countries (last column in Table 1).

Using a model controlling the natural logarithms of GDP and population as well as their quadratic terms while pre-setting the target significance level at $\gamma_c = 0.01$ for the break detection procedure, I detect 15 structural breaks, and 9 of them have a negative sign, indicating a step-shift reduction in tree cover loss in the corresponding country and year. The size of these nine effects ranges from 45.7% to 72.5% of reduction in tree loss, compared to their counterfactual cases in the absence of the breaks (i.e. potential treatments). One of the nine breaks is likely led by a data fluke, while the policy attribution of the rest can be categorized into three types: forest law and policy improvement and enforcement (5), economic incentives (2), and other (1). The effect achieved by the law and policy improvement approach varies the most but sometimes does come larger than the effect realized by economic incentives, which on the other hand is medium-sized but more stable. The economic incentives related treatments are the carbon credit market established in Costa Rica in 2013 and results-based payments from REDD+ sought by Ecuador in 2014, both of which are associated with more than 50% reduction in tree cover loss, again, compared to their counterfactual scenarios.

The rest of this paper is structured in the following order: first, in section 2, I talk about the data sets for tree cover loss, population, and GDP, including their sources and the existing data issues, especially the inconsistency issue of the tree cover loss data. Then I introduce the break-detection methodology I adapt in the paper to detect notable treatments for each country throughout the years 2001-2021, as well as the final retained model after the break-detection procedure. In section 4, I analyze the estimation results and attribute the breaks to potential policies or past events. Last but not least, section 5 provides a further discussion of the analysis and concludes the paper.

2. **Data**

Tree Cover Loss Data

Tree cover loss is the key variable used in my econometric models of this paper. I obtain tree cover loss data from the "Tree Cover Loss by Driver" dataset collected by the University of Maryland, through the Global Forest Watch (2022) organization. Their dataset on deforestation rates offers an overview of tree cover loss, associated with five dominant drivers, for each country, spanning 21 years from 2001 to 2021. The five dominant drivers are commodity-driven deforestation, shifting agriculture, forestry, wildfire, and urbanization. To run my models, I use deforestation data driven by shifting agriculture. The data source suggests a 93% global user accuracy for the class of shifting agriculture, which is the highest accuracy amongst that of the five dominant drivers, making it comparatively safe to use as the dependent variable for the models. Another reason for choosing the shifting agriculture class over commodity-driven is that deforestation driven by cocoa production is demonstrably different from agriculture systems for most commodity crops. For example, with the current remote sensing techniques, perennial commodity crops are mostly easily detectable either in their mature or immature stages, owing to their monoculture practices, but cocoa is an exception (Ordway et al., 2017). Unlike crops such as palm oil, rubber, and banana that thrive in full sun conditions, Ordway et al. argue cocoa production in sub-Saharan African countries is largely carried out by smallholder agriculture and occurs in "shade grown systems under secondary forest canopies or inter-cropped with food crops, rendering it indistinguishable from forest [and] mixed crop," making it undetectable using Landsat images. Therefore, shifting agriculture has a higher chance of (partially) capturing cocoa-driven deforestation which enables my models to explore the potential of identifying related cocoa policies. In each country, tree cover loss associated with shifting agriculture is overall volatile throughout the sample years (see Figure 2), which might act as an amplifying factor for the effect size of breaks detected.

Although this tree cover loss data is the current best available spatial data on deforestation trends (Weisse and Potapov, 2021), it's not without caveats. Because data collection methods and technologies evolve, inconsistency emerges in the data as newer data is often more accurate than the older. In general, when a change in relevant methods takes place, the tree cover loss data of the same year or the following year would have a sudden change and thus potentially create a spurious structural break in the trend, but this break is not one related to any meaningful treatment effect as it's not caused by a real-world political or economic intervention.

Weisse and Potapov (2021), in their study discussing how to use this dataset to assess trends in tree cover loss, suggest that over the available time span of the data, inconsistencies have been caused by approximately three factors: algorithm adjustments, improved satellite data, and variations in satellite image availability. First, the algorithm improvement that took place in 2013 resulted in enhanced detection of tree cover loss, especially from 2015 onward. Also, since the algorithm uses all available images from the Landsats for tree cover loss detection, the retirement of an old Landsat and the launch of a new one would interrupt data consistency by changing the number of available Landsat images. All Landsat images from 2001 to 2012 were provided by the two Landsats, 5 and 7 (NASA, TIMELINE); in 2013, the replace-

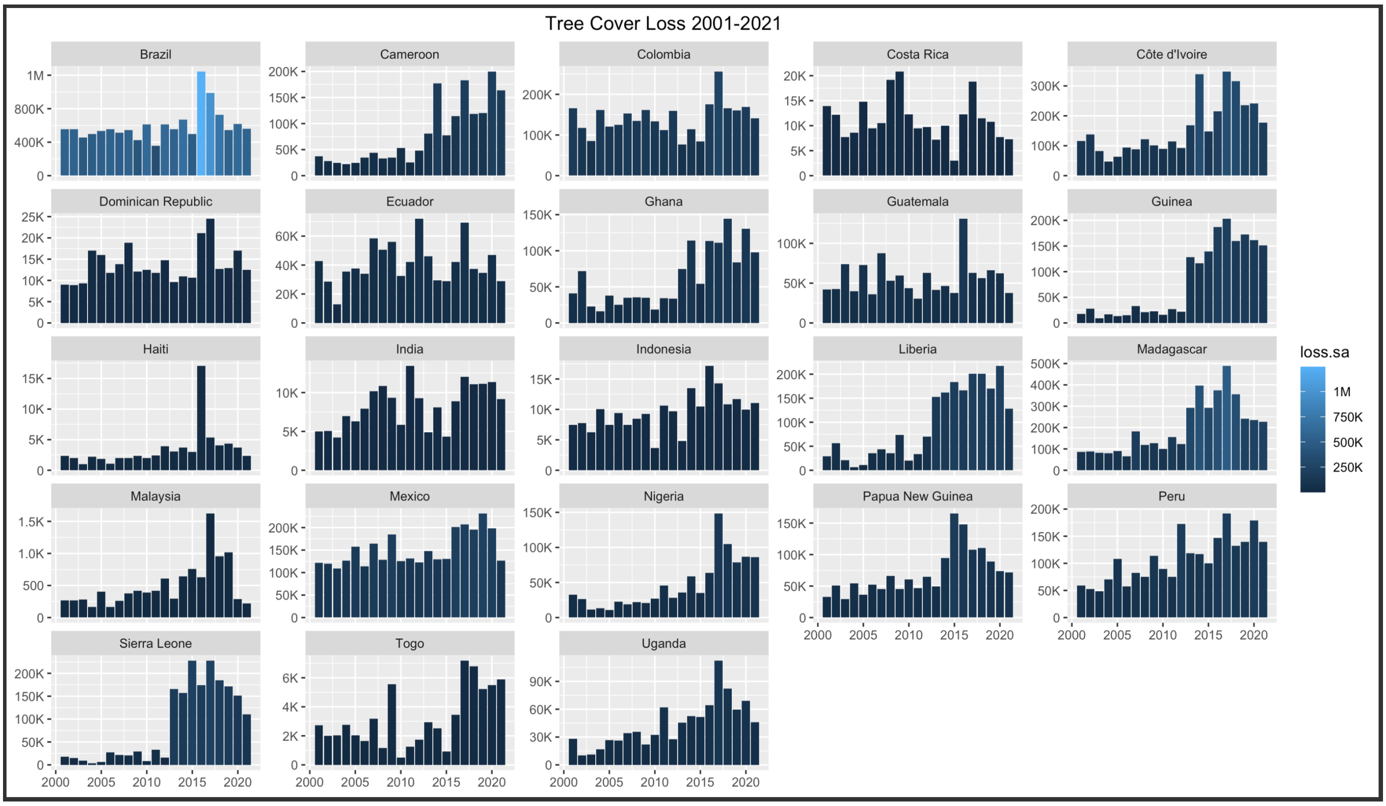

Figure 2: This figure presents the time series of tree cover loss caused by shifting agriculture (loss.sa) from 2001 to 2021 for each country separately. The y-axis varies across countries since they, have varying levels of tree cover loss. The legend uses a gradient colour scheme that place lighter colour in bars with higher numbers. For example, Brazil has an overall higher tree cover loss level compared to the others, so the bars in Brazil are generally in a lighter blue.

ment of Landsat 5 with Landsat 8, which has the same resolution but "improved sensor that can better resolve features on the ground," enabled the dataset to start capturing "tree cover loss smaller than an individual pixel [such as] selective logging" (Weisse and Potapov, 2021). Third, in general, the available Landsat images have been increasing over time, while increasing data accuracy, it technically also creates data inconsistency. Weisse and Potapov suggest that 2012 was a year when available Landsat images were particularly low due to the gap between the retirement of Landsat 5 and the launch of Landsat 8 in 2013. Although decommissioned in January 2013, Landsat 5 stopped acquiring pictures in November 2011 due to a "rapidly degrading electronic component" (NASA, Landsat 5). In addition, Weisee and Potapov also suggest that generally, the increased sensitivity in detection due to the first two factors (algorithm adjustment and improved satellite data) cast a larger effect in the data for regions with smaller-scale forest clearing such as West and Central Africa and Southeast Asia. Theoretically, it remarkably increases the chance of detecting structural breaks in countries of these regions from 2013 onward.

Figure 3 shows that amongst the five dominant drivers, commodity crops and shifting agriculture contribute the most to deforestation, followed by forestry, in my sampled cocoa-producing countries. About 32% of the total deforestation is due to shifting agriculture while 60% is driven by commodity crops, nearly double the amount of the former. Commodity-driven deforestation is mostly contributed by just three countries: Brazil, Indonesia, and Malaysia. From the bar chart, it's easy to spot that Brazil acts as an outlier in this dataset because it not only drives more than half of commodity-related deforestation but also contributes the most tree cover loss among all the cocoa-producing countries. By choosing shifting agriculture data over commodity-driven, the model works with data that has more variation across countries. The difference in the data variation is made visible in Figure 4.

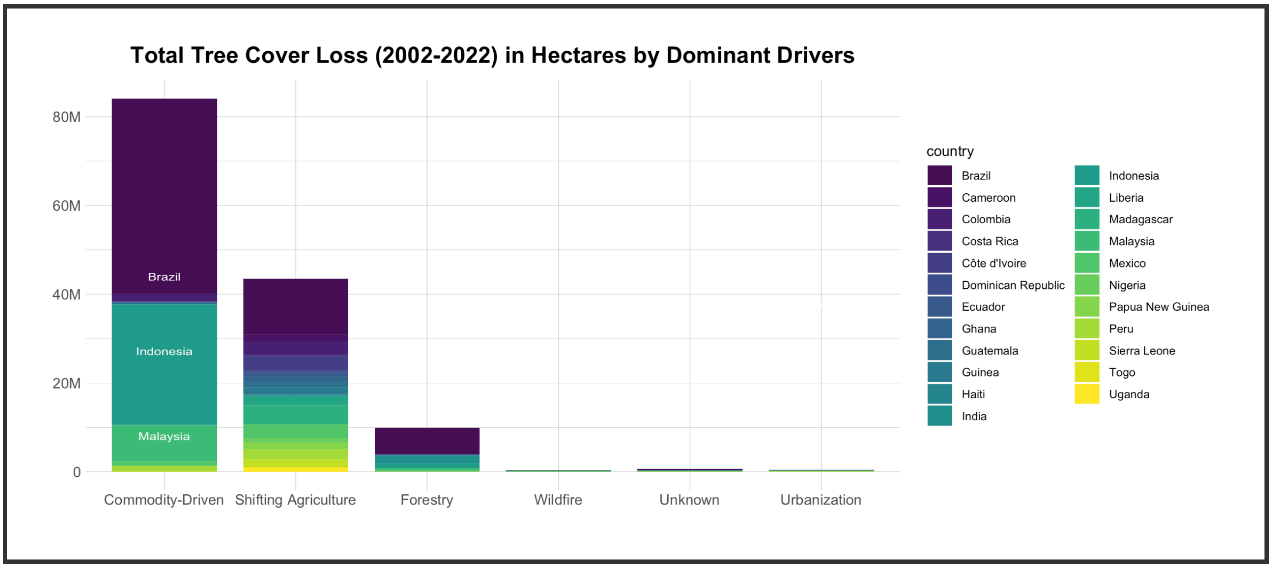

Figure 3: This stacked bar chart shows the area, in hectares, of tree cover loss broken down by the five dominant drivers and the uncategorized loss (unknown), stacked by countries. The bars from left to right are ordered by the amount of total tree cover loss led by each driver.

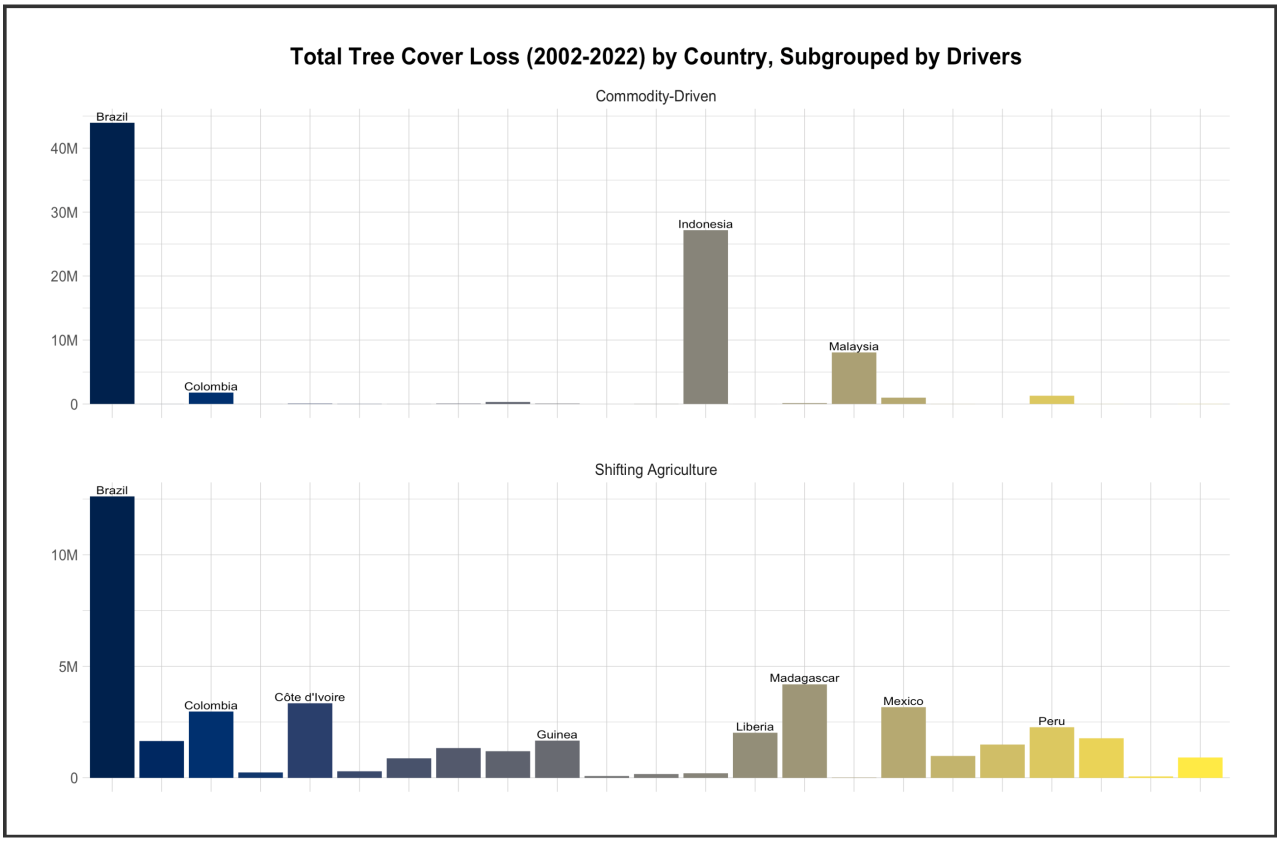

Figure 4: These bar charts show the area, in hectares, of tree cover loss broken down by cocoa-producing countries, in commodity-driven and shifting agriculture category, respectively.

GDP & Population Data

I retrieved the data for population and GDP (constant 2015 US$) from the World Bank (2022). They are included in the model as quadratic terms to account for the effect population and GDP, as well as their growths, have on deforestation. In the 1950s, Simon Kuznets hypothesized an inverted U-shaped relationship between GDP per capita and income inequality, namely the Kuznetz Curve where income inequality first rises with GDP per capita and then falls as the population gets richer on average (Bo, 2011). Grossman and Krueger later found the same type of relationship between economic growth and environmental quality, which Panayotou named the Environmental Kuznetz Curve (EKC) in 1993(Stern, 2004 & Bo, 2011). Generally speaking, a country with a small population (relative to its land/forest size) need not worry about large-scale deforestation because 1) either it's a rich developed country whose economy doesn't predominantly depend on agriculture or forestry 2) or smallholder agriculture expansion is popular in the country but there isn't enough labour to cause a large amount of deforestation. As the population grows, the rich country in scenario one might consider forestry for its economic growth; and the small country would have more labour to support activities that drive deforestation. The Kuznets curve concept applied to the GDP variables can be analyzed analogously. During early development, a country often has more workers in the rural area who will shift to cities while the country undergoes industrialization. When the country becomes developed enough so its residents don't have to worry about food and shelter, they often start to care more about the environment and tend to advocate for policies that target fixing environmental issues such as large-scale deforestation. This income elasticity of demand for environmental quality is the driving force for the shape of EKC (McConnell, 2001).

Cocoa Area Harvested Data

The data for the area of cocoa harvested is obtained from the Food and Agriculture Organization of the United Nations (2022). It is included as the independent variable to capture the effect of cocoa production on tree cover loss in the model that checks the estimation robustness, while also in the hope of discovering the cocoa-related treatments.

In general, cocoa production is carried out by smallholders in shade-relying farming systems, often under secondary forest canopies or other taller food crops that create shade (Ordway et al. 2017), and it induces small-scale deforestation events unlike those induced by commodities such as palm oil plantation. Shifting-agriculture-driven deforestation is defined as "Temporary loss or permanent deforestation due to small-to-medium-scale agriculture" (Global Forest Watch, 2022) and is the closest to cocoa-farming-driven deforestation. Since the data used for my models is shifting-agriculture-driven deforestation, controlling for the cocoa area harvested is a plausible way of checking the robustness of the break estimates.

3. **Methodology**

Reverse causal questions

Gelman and Imbens (2013) categorize the causal questions into two broad classes - forward causal questions as "the estimation of 'effects of causes'"; and reverse causal questions as "the search for 'causes of effects'" - policy evaluation in existing literature of econometrics largely focuses on single, known policy interventions with the forward causal question approach, such as the conventional difference-in-differences approach. Koch et al. (2022) suggest that uncertainties of policy effectiveness often remain in such analyses because of their nature of focusing on single, known interventions. A common cause of inaccuracy led by said policy evaluations is the omission of important attributes (variables) in the model, being "effective but unknown interventions or the falsely deemed ineffective and therefore untested ones" (Koch et al., 2022). Gelman and Imbens (2013) demonstrate this problem using the potential outcome framework (Neyman, 1923 & Rubin, 1974) and argue that reverse causal questions could be asked in such scenarios and answered within the traditional statistical frameworks where forward causal question analyses predominate. Specifically, they suggest that the search for causes could be reflected by model checking and hypothesis generation, the result of which becomes improved econometric models with more accurate variable sets enabling the reproduction of the actual patterns we observe in data. That is, reverse causal questions (the "why" questions) act as a more natural way of thinking when anomalies arise in data and cannot be explained by pre-existing models, and the search for the causal explanation can be achieved by adding new variables to improve pre-existing models. Inspired by this idea, the first step to answering the reverse causal question of interest in the present paper is the search for an appropriate econometric model. However, such model checking procedure can be difficult because it involves searching for variables that researchers do not yet know exist, but it was made possible owing to today's statistical software advancement.

A novel approach - break detection

A novel approach of general-to-specific model selection via indicator saturation powered by machine learning was recently introduced to answer reverse causal questions (Pretis and Schwarz 2023, Koch et al 2022, Pretis 2022). Pretis and Schwarz (2023) show that this approach can reveal meaningful variables without a priori knowledge of their existence, by selecting new variables over a full set of individual-time interactions in a balanced panel without putting restrictions on the number or effect size of selected variables. They argue that in the conventional setting of two-way fixed effects (TWFE) panel estimators, heterogeneous treatment effects can be detected as structural breaks conditional on the treatment effects being non-zero. In the standard difference-in-differences (DID) setting, known treatment effects are included as dummy variables that indicate the interactions of treated units and the post-treatment periods. They suggest that having these treatment dummy variables are equivalent to allowing for step-shift breaks in the treated units' intercepts (i.e. fixed effects) and that these structural breaks are detectable in the units' individual fixed effects using the machine-learning powered break-detection method provided by the "getspanel" R package (used together with "gets"[4]).

[4]"gets" stands for general to specific approach. Both "gets" and its update "getspanel" are introduced by Pretis and Schwarz (2023)

The pair of break detection and subsequent policy attribution has become common practice in time series analysis, but it's difficult to make causal interpretation in most of those analyses due to the lack of control groups (Pretis and Schwarz 2023, Koch et al 2022, Pretis 2022). On the other hand, this novel approach adopted by the present paper uses the group of units that do not have breaks detected in them as the control group, revealing possible causal links. The main calibration parameter of this approach is the target level of significance that controls the false positive rate of detection γ_c, which can be scaled to target either "a stable false-positive *rate* of classifying the treatment group... [or] a stable *number* of false-positive treated units." In other words, the false-positive rate can be used to determine the rate of a single unit being falsely classified as treated or the total number of units being falsely classified as treated. Pretis (2022) demonstrates this by adjusting the target significance level and subsequently successfully controlling the false-positive rate in a Monte Carlo simulation, meanwhile proving unknown treatments can be detected at a rate close to "the optimal case" in a DID setting with known treatments.

While Koch et al. (2022) argue that the search for "causes of effects" is rarely sought after by researchers perhaps because the method is less obvious compared to the various tools tackling forward causal questions, they use this novel approach to tackle the reverse causal question - "What reduced the CO_2 emissions?" in the road transport sector among the EU member states - within the framework of break detection in a panel DID setting. They first use machine learning to detect notable changes (structural breaks) in emissions in each EU member relative to a control group and then attribute the breaks to a single or mix of policy interventions. Since the break detection and the subsequent policy attribution are separate steps, this approach frees up some common requirements including "any a-priori knowledge of changes in emissions" or the causes of the changes, and the number of the breaks. This means the approach makes it possible and simple to search for any unknown single or mixes of policies in any context, regardless of the number of unknown polices, while being able to control the effect size of the policies retained in the model. Similarly, Pretis (2023) uses this approach to answer the question - does BC carbon tax reduce CO_2 emissions? - and obtains a final model without the BC Carbon tax detected as a break, implying its insignificance. Meanwhile, he finds a significant reduction in one branch of BC CO_2 emissions, namely BC transportation emissions, brought by the introduction of the same tax.

In addition, this break detection procedure can function as a tool that checks coefficient robustness, owing to its feature of identifying previously unknown but statistically significant treatments that are subsequently retained in the final model, a sparse model that drops all the insignificant units through the procedure given the target level of significance. Therefore, when assessing the effectiveness of policies on reducing CO_2 emissions, Pretis (2022) and Koch et al. (2022) use this procedure to check the robustness of coefficients on known treatments in DID settings. For example, Pretis (2022) uses the break detection approach to agnostically search for notable changes in CO_2 emissions, and the fact that the retained breaks (i.e. treatments) do not include the BC Carbon Tax policy, which is the variable of interest in the context, indicates that the effect of the policy being not robust. This implies either too small an effect size of the policy or statistical insignificance of the effect. In the meantime, the break detection procedure also helps to uncover the existence of notable effective policies previously unknown. Similarly, Koch et al. evaluate the robustness of their result by altering the target significance level among models, *i.e.* a break is more robust when detected by more models with different target significance levels. They suggest the stability in the magnitude of the estimated effects indicates robustness as well. Another way of robustness check for DID analysis is forcing the known treatment into the model of break detection that includes new retained breaks (interventions), stable coefficient size for the known treatment indicates it is robust and vice versa (Pretis, 2022); the same holds for the robustness of the detected breaks. One of my models controls for the area of cocoa harvested, compared which with the original model without this variable checks the robustness of the estimates of the detected breaks in the original model.

Adapting the break detection approach

Suppose an existing model is assessing whether a certain policy intervention (policy A) in a cocoa-producing country is successful in reducing deforestation, and because the nature of the model is to focus on a single and known intervention, uncertainties of the policy effectiveness remain. In the scenario where the model suggests that policy A has a statistically significant effect on deforestation reduction in the country, we wouldn't be entirely sure whether the effect is caused by policy A or another policy that the analyst didn't think was related and thus didn't include in the model. This is when we want to ask the "why" question and work on model checking and hypothesis generation. Theoretically, we could improve the model by adding every single potential variable (policy interventions or other attribute/events of the country, as well as the time effect) to the model and observe whether the estimation results and conclusion change, in the hope for unchanged and robust estimations. However, doing so manually would be tedious heavy-loaded work and an impossible job in some cases. In the present paper, I rely on the machine learning model-selection approach first introduced by Pretis (2022) in a paper assessing the effect of the BC carbon tax, using the "gets" R package with the "getspanel" update, both introduced by Pretis and Schwarz (2023).

To explore the "why" question regarding policy interventions that led to effective reductions in deforestation among cocoa-producing countries, I'm answering the reverse causal question in the following order: 1) Have there been changes in deforestation rates in any of the sample countries over the time span of my data? 2) If there exist such changes that are large and statistically significant, can they be attributed to possible interventions (i.e. why these changes)? These two steps correspond to the steps in the adapted break detection approach: 1) detecting breaks and 2) attributing the detected breaks to potential policy interventions. Both steps together are almost equivalent to asking - what interventions have been effective in reducing deforestation in cocoa-producing countries?

Given the equivalence between the treatment dummy variables in a DID setting and the step shifts in treated units' fixed effects, this break-detection method enables me to estimate a TWFE panel model and search for potential structural breaks (step shifts) at the same time. With the identified treatment breaks, my final model includes interaction terms that denote unit-heterogeneous piece-wise constant treatment effects (step shifts). In brief, my approach first sets a basic TWFE panel model and then uses the machine-learning powered variable selection method to detect notable step-shift structural breaks that can be subsequently interpreted as potential intervention effects attributed to potential causes, being events or policies that took place. This approach can also identify "unit-heterogeneous fully-time-varying" treatment effects, as Pretis and Schwarz discuss in more detail in their paper. Regardless of the types of treatment effects being identified, the units without a break together act as a control group against those with identified breaks.

I now set up a model (1) that allows for potential treatment effects for any individual at any point in time - in the present context, they would be potential interventions for any of the 23 cocoa-producing countries in any year spanning from 2001 to 2021 - using a full set of step-function dummies denoting individual-time (country-year) interactions (NT = 21×23 = 483) that are being selected over from a balanced panel. However, since breaks won't be detected in the first year (2001) because it would coincide with the fixed effects (i.e. intercepts for each unit), the time of step-functions starts from 2002 in the model. With more variables than observations, this model goes through the process of machine-learning variable selection, embedded in the "getspanel" R package, that detects the step shifts. This process drops statistically insignificant variables, keeping only the ones that matter. The retained interaction terms are identified potential treatments (*i.e.* detected structural breaks in country-specific fixed effects c_i) that can subsequently be attributed to past events or policies conditioning on the effect being non-zero as well as big enough to be detected given the preset target significance level γ_c. Figure 5 presents the detection result at $\gamma_c = 0.01$, and the coefficient values are displayed in the second column of Table 1. While enforcing two-way fixed effects on country and year, this break detection procedure uses fixed effects step indicator saturation (FESIS) from "getspanel", and the retained fixed effects step-shifts are represented by red vertical lines in Figure 4.

$$\begin{aligned} log(Loss_{i,t}) = c_i + g_t + \Sigma_{j=1}^{23}\Sigma_{s=2002}^{2021}\tau_{j,s} \cdot 1_{\{i=j,t\geq s\}} + \beta_1 log(Pop)_{i,t} + \beta_2 log(Pop)_{i,t}^2 \\ +\beta_3 log(GDP)_{i,t} + \beta_4 log(GDP)_{i,t}^2 + u_{i,t} \end{aligned} \tag{1}$$

$$\begin{aligned} Loss_{i,t} &= \text{tree cover loss caused by shifting agriculture in country i at year t} \\ c_i &= \text{country fixed effect for country i} \\ g_t &= \text{time fixed effect for year t} \\ \Sigma_{j=1}^{23}\Sigma_{s=2002}^{2021}\tau_{j,s} \cdot 1_{\{i=j,t\geq s\}} &= \text{step-functions denoting potential treatment in country i and time t} \\ Pop_{i,t} &= \text{population of country i at year t} \\ GDP_{i,t} &= \text{GDP of country i at year t} \end{aligned}$$

The Environmental Kuznetz Curve is a U-shaped relationship between economic growth and environmental quality (Grossman and Krueger, 1991 & Panayotou, 1993). To reproduce the U-shape, economic growth appears as quadratic terms on the right-hand side of the equation while the environment variable appears on the left as the dependent variable. In my model, the natural logarithm of tree cover loss acts as the environmental quality variable; the natural logarithms of GDP and population appear as quadratic terms capturing economic growth.

Figure 5 displays the breaks detected in each country included in this study. Among the 23 chosen cocoa-producing countries, breaks (step shifts) are detected in 10 countries for the years between 2002 and 2021 (breaks cannot be detected in 2001 as it's the first year of the panel dataset), denoting piece-wise constant treatment effects. Countries with identified effects are considered to have had potentially effective interventions that impacted the amount of tree cover loss driven by shifting agriculture activities, either positively or negatively.

4. Results and Analysis

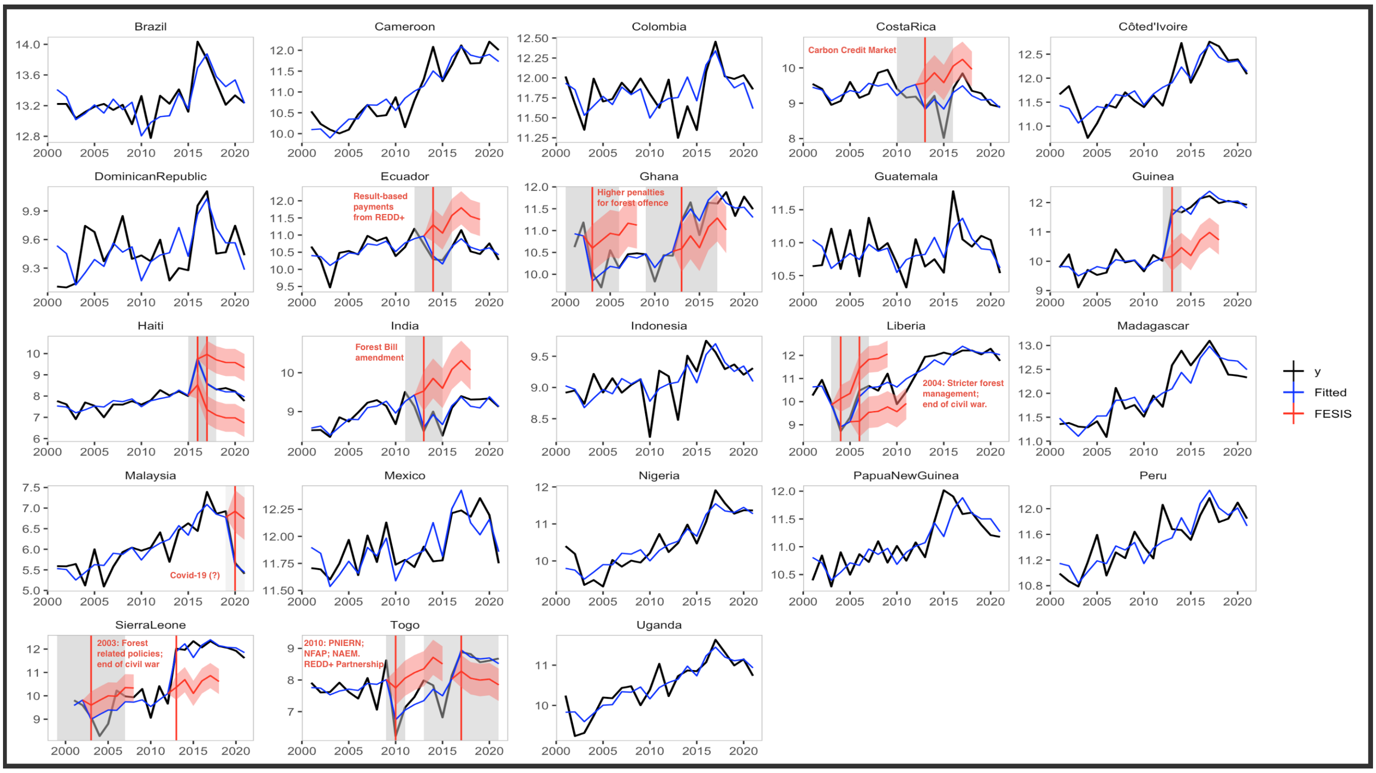

Figure 5: In this TWFE panel of tree cover loss, constant treatment effects are detected using fixed effect step indicator saturation ("FESIS") that saturates the full set of country-year interactions (NT = 483). The retained step shifts are represented by the red vertical lines. The red curves are the estimated counterfactual states of each retained break for the period of five years after the break; and the red shadings represent the range of confidence intervals for the counterfactual states. The saturation algorithms is available in "getspanel" R package and is run with a preset target significance level $\gamma_c = 0.01$ for this figure.

Using the target significance level $\gamma_c = 0.01$, fifteen structural breaks are detected across the sample countries (Figure 5), Nine of which are negative with magnitudes ranging from 0.61 to 1.38 (84.0% - 297.5% deforestation reduction), indicating potential sizable treatment effects that could be attributed to past events or policy interventions that successfully helped reduce tree cover loss. Those are potential interventions that had negative-signed impacts on tree cover loss (but a positive impact regarding the deforestation issue) conditional on the effects being non-zero and not attributable to the control variables in the model. On the other hand, the six positive-signed breaks represent potential past events that increased the tree cover loss in the corresponding countries. Half of these positive-signed step-shift breaks are detected in the year 2013 (see Table 1), which are likely due to the aforementioned existing data inconsistency incorporated in my panel data of tree cover loss by the University of Maryland (2022) given that no plausible interventions could be identified to which the structural breaks can be attributed. These positively signed breaks of the year 2013 are detected in Ghana, Guinea, and Sierra Leone. Figure 6 shows the jump in mean loss in 2013 in those three countries. What's worth noting is that the government of Sierra Leone did issue the Agenda for Prosperity - Road to Middle Income Status (2013), a five-year (2013-2018) poverty reduction strategy that promoted the expansion of agriculture, among other sectors. Being the backbone of the country, agriculture contributed 40-50% of GDP and employed as many as two-thirds of the country's population (The Government of Sierra Leone, 2013). This Agenda for Prosperity aimed to promote medium-to-large scales of agriculture activities and value-added activities regarding agriculture products, making large investments in those sectors of the economy. Therefore, the promotion of agriculture expansion might have worsened the deforestation situation in Sierra Leone, partially accounting for the drastic increase in tree cover loss in the country in 2013. This makes sense because the data for the three countries (Ghana, Guinea, and Sierra Leone) all suffer the same issue, but the magnitude of the structural break detected in Sierra Leone seems to be larger compared with the other two (See Table 1). On the other hand, the structural break in Ghana in 2013 is only detected with $\gamma_c = 0.01$ and not under the other two more conservative target significance levels, aligning with its smaller effect size and perhaps the weaker statistic significance.

Because of mapping algorithm adjustments for the tree cover loss data, comparison before and after 2011, or before and after 2015 requires caution (Weisse and Potapov, 2013). For this very reason, the large percentage increase (approximately 238.7%) in Haiti's deforestation in 2016 (see Figure 7) is likely caused by a data fluke. There is a chance that the identified Haiti's negatively signed break in the flowing year is caused by the spurious jump in 2016 and hence by the same data fluke, but I would note that Haiti signed the Paris Agreement in 2016-Apr-22, which became effective on 2016-Nov-04 and hence could have potentially led to a success in forest protection in 2017. In addition, the United States Agency for International Development (USAID) launched a reforestation project in 2017 to help plant trees and preserve forests in the North and North-east of Haiti (USAID, 2020). Although a reforestation project would help increase tree cover in Haiti, it wouldn't affect the tree cover loss data I use here because the loss and the gain are absolute values measured separately. In other words, I do not use *net* tree cover loss in this paper. However, this reforestation program helped diversify the income of Haitians to discourage them from chopping down trees for charcoal (Lo, 2022). Given that charcoal represents the main source (80%) of Haiti's domestic energy (Saint, 2022), USAID's deforestation program might also contributed to the decrease in Haiti's deforestation in 2017. It remains unclear whether the mapping algorithm adjustments, the Paris Agreement signage, or USAID's deforestation program played the main roles in Haiti's deforestation reduction in 2017.

As discussed in the data section, the increased sensitivity in tree cover loss detection due to algorithm adjustment and improved satellite images cast a larger effect on West and Central Africa and Southeast Asia. This could be the reason Figure 6 (West African countries) and Figure 7 (a Latin American country) illustrate different patterns of impact caused by improved algorithms and satellite images. Furthermore, Haiti suffered the infamous Hurricane Matthew in late 2016, which likely destroyed a large amount of trees, causing a temporary jump in the data.

Table 1: Detecting Fully Time Varying Treatment: Panel Model

Dependent Variable: Log(Loss)				
Model:	$\gamma_c = 0.01$	$\gamma_c = 0.005$	$\gamma_c = 0.001$	TWFE
Variable	*Coefficient (standard error in parentheses)*			
Log(Pop)	-15.33	-13.00	-25.73	-1.40
	(22.13)	(22.46)	(19.51)	(22.66)
Log(Pop_sq)	0.41	0.35	0.76	0.02
	(0.66)	(0.67)	(0.60)	(0.67)
Log(GDP)	12.23 ***	12.05 ***	14.95 ***	10.43 ***
	(3.42)	(3.48)	(3.35)	(3.47)
Log(GDP_sq)	-0.25 ***	-0.24 ***	-0.29 ***	-0.21 ***
	(0.07)	(0.07)	(0.07)	(0.07)
Log(Cocoa)				-0.14 **
				(0.01)
- – — Breaks — – -				
τ: i=Costa Rica, t $\geq$ 2013	-0.75 ***			-0.74 ***
	(0.26)			(0.25)
τ: i=Ecuador, t $\geq$ 2014	-0.90 ***	-0.89 ***		-0.87 ***
	(0.25)	(0.25)		(0.25)
τ: i=Ghana, t $\geq$ 2003	-0.75 ***	-0.99 ***		-0.72 ***
	(0.27)	(0.26)		(0.27)
τ: i=Ghana, t $\geq$ 2013	0.62 **			0.64 **
	(0.27)			(0.27)
τ: i=Guinea, t $\geq$ 2013	1.40 ***	1.41 ***	1.42 ***	1.28 ***
	(0.26)	(0.26)	(0.28)	(0.26)
τ: i=Haiti, t $\geq$ 2016	1.22 ***	1.23 ***		1.22 ***
	(0.34)	(0.34)		(0.33)
τ: i=Haiti, t $\geq$ 2017	-1.38 ***	-1.38 ***		-1.65 ***
	(0.33)	(0.33)		(0.34)
τ: i=India, t $\geq$ 2013	-0.93 ***	-0.93 ***		-0.95 ***
	(0.26)	(0.26)		(0.26)
τ: i=Liberia, t $\geq$ 2004	-1.22 ***	-1.22 ***		-1.28 ***
	(0.29)	(0.30)		(0.29)
τ: i=Liberia, t $\geq$ 2006	1.06 ***	1.02 ***		0.88 **
	(0.35)	(0.35)		(0.35)
τ: i=Malaysia, t $\geq$ 2020	-1.29 ***	-1.29 ***	-1.27 ***	-1.32 ***
	(0.26)	(0.26)	(0.29)	(0.26)
τ: i=Sierra Leone, t $\geq$ 2003	-0.61 **	-0.63 **		-0.66 **
	(0.29)	(0.30)		(0.29)
τ: i=Sierra Leone, t $\geq$ 2013	1.53 ***	1.54 ***	1.79 ***	1.58 ***
	(0.27)	(0.28)	(0.28)	(0.27)
τ: i=Togo, t $\geq$ 2010	-1.01 ***	-1.00 ***	-1.37 ***	-1.12 ***
	(0.28)	(0.29)	(0.28)	(0.29)
τ: i=Togo, t $\geq$ 2017	0.67 **	0.67 **		0.51 *
	(0.26)	(0.27)		(0.27)
Fixed Effects				
Country	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
Fit Statistics				
Observations	NT = 483	NT = 483	NT = 483	NT = 483
Within R^2	0.66	0.65	0.58	0.67

*Signif. codes: (***:1%); (**:5%); (*:10%).*

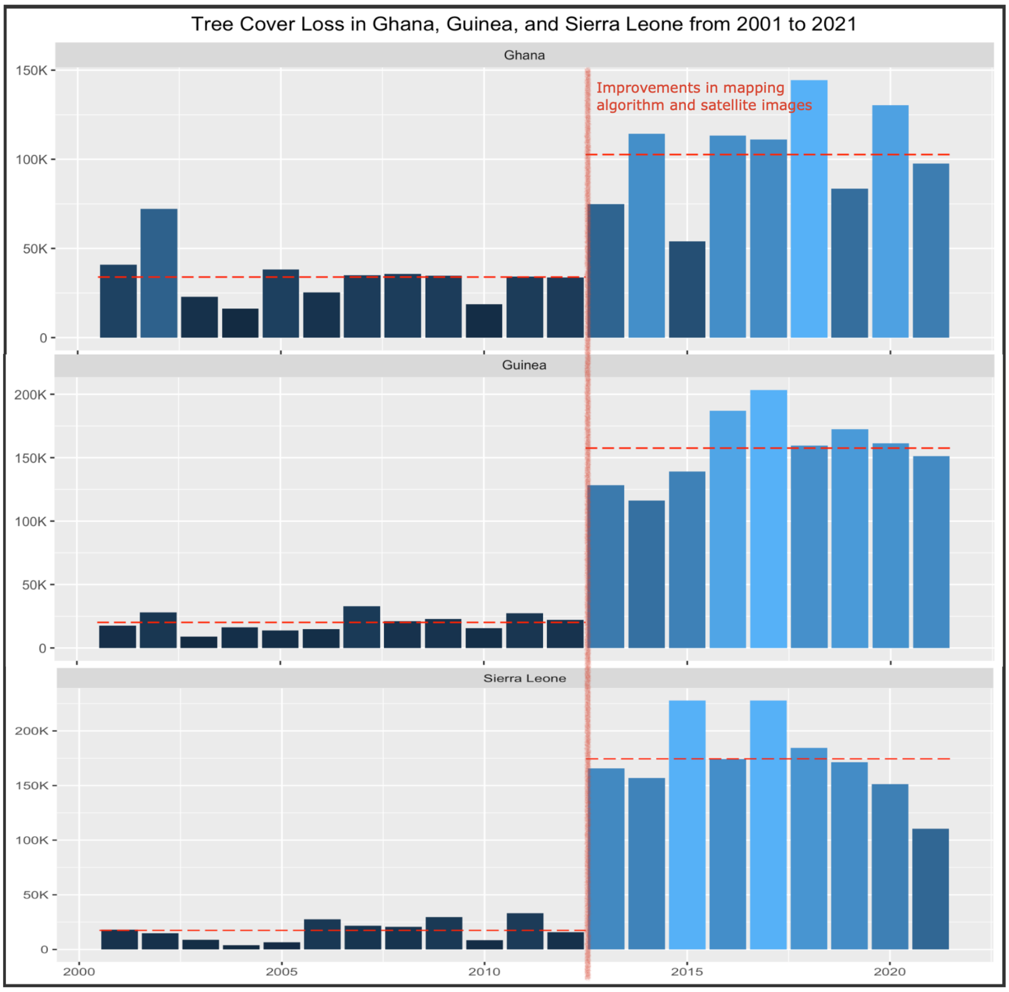

Figure 6: This graph gathers three bar charts visualizing the tree cover loss from 2001 to 2021 in the three countries with detected structural breaks that are likely caused by a combination of mapping algorithm improvements and improved satellite data. The three countries are Ghana, Guinea, and Sierra Leone. In each chart chart, the red dashed horizontal lines represent the mean of tree cover loss during 2001-2012 and 2013-2021, respectively.

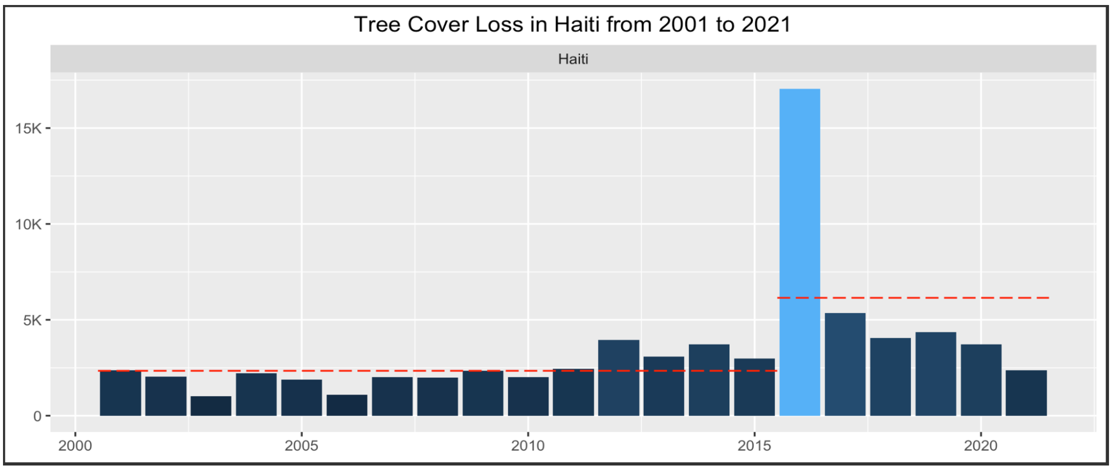

Figure 7: This bar chart shows the tree cover loss from 2001 to 2021 in Haiti. The red dashed horizontal lines represent the mean of tree cover loss during 2001-2015 and 2016-2021, respectively. The jump in 2016 is likely a data fluke caused by mapping algorithm adjustment.

The other 2 positively signed breaks are detected in Liberia in 2006 and Togo in 2017, separately. Bar graphs of the tree cover loss over time for these two countries (Figure xx) can be found in the Appendix. In September of 2006, The government of Liberia approved a National Forestry Reform Law of 2006, whose section 5.4 - Timer Sale Contracts may award to allow "Forest Land to be cleared for agriculture or for the establishment of plantations" (FAO, 2006). This would inevitably lead to more deforestation in Liberia, if not all of the 188.6% increase in tree cover loss in the country in 2006. On the other hand, Togo did not approve any new law or policy in 2017 that would've otherwise contributed to its 95.4% of deforestation that year. Instead, 2017 was the year when Togo's five-year Strategy for Accelerated Growth and Employment Promotion ended, through which Togo was committed to "a process of long-term development based on low-carbon emissions" (FCPF, 2013). Hence, Togo's increase in deforestation in 2017 could be just a rebound after a restrictive law ended.

Potential Policy Interventions at $\gamma_c = 0.01$

While over half of the detected positive-signed structural breaks can be attributed to the pre-existing data inconsistencies, the negative-signed breaks show greater promise, excluding Haiti's 2017 break. Concerning the effect size of the latter, the top three are observed in Malaysia, Liberia, and Togo (see Table 2), ranging from 63.6% to 72.5%. While these estimated effect sizes appear large, they are well within the year-on-year variability of observed deforestation. It is not unusual to see changes of $+100\%$ or -100% from one year to the next.

The structural break in **Malaysia** is detected in 2020, revealing a 72.5% reduction in deforestation. This is likely attributed to the onset of the COVID-19 pandemic, as no other discernible political or economic interventions coincide with this reduction. The global shift to remote work during COVID-19 significantly curtailed mobility, potentially impacting deforestation-related activities and resulting in fewer tree cover losses.

While COVID-19 had a worldwide impact as a pandemic, affecting every country in this study, no other breaks are detected for the year 2020. This absence is explained by the year fixed effects, capturing the universal effects of COVID-19 as the pandemic unfolded globally. Conversely, the 2020 break in Malaysia suggests a disproportionate response to COVID-19 compared to other sampled countries. There may exist specific Malaysian policies directly or indirectly influencing their deforestation rate during the pandemic, though such detail is not found in this study. Although the pandemic extended into 2021, no breaks would be detected for that year as it is the last year of the dataset.

Liberia's negative-signed break is detected in 2004 with an effect size of 70.5% decrease in tree cover loss. This could have been led by a few factors, the main one being the approved amendment of Liberia's New National Forest Law On October 10, 2003 (FAO, 2003). Specifically, this amendment established the Division of National Parks and Wildlife of the Forest Development Authority, whose responsibility includes protecting tropical lowland rainforests threatened by factors such as shifting agriculture. It also emphasized the

The eight negative-signed breaks and their Policy Attribution					
Country	Break Year	Effect Size	Avoided Deforestation (Ha)	% of Avoided Deforestation	Policy/Event Attribution
Malaysia	2020	-1.29	762.8	72.5%	(O) COVID-19.
Liberia	2004	-1.22	14,704.2	70.5%	(LP) Stricter forest-management.
Togo	2010	-1.01	884.8	63.6%	(LP) PNIERN; NFAP; NAEM; REDD+ Partnership.
India	2013	-0.93	7,513.2	60.5%	(LP) Forest Bill Amendment.
Ecuador	2014	-0.90	43,056.3	59.3%	(E) Results-based payments from REDD+.
Costa Rica	2013	-0.75	7,997.5	52.8%	(E) Carbon credit market.
Ghana	2003	-0.75	25,652.4	52.8%	(LP) Higher penalties for forest offence; improved work environment.
Sierra Leone	2003	-0.61	7,451.8	45.7%	(LP) Forest related policies.

Table 2: This table summaries the eight detected negative-signed structural breaks, excluding the 2017 break in Haiti, given the target level of significance equals 1%. The breaks are listed in the order of the effect size of the corresponding interventions. *i.e.* the top one has the largest effect size. Avoided deforestation is calculated using the formula $Avoided_{i,t} = exp(log(Loss_{i,t}) - \tau_{i,t}) - Loss_{i,t}$, where $\tau_{i,t}$ denotes the break detected in country i and year t and the $exp()$ part denotes the estimated counterfactual values of tree cover loss for country i and year t in the absence of the attributed intervention. The percentage of avoided deforestation is the ratio of avoided deforestation (in hectares) and the counterfactual deforestation. Other numeric values used in the calculation can be found in Table 3 in the appendix. In the column of policy attribution, entries are categorized into three types: economic incentives (E), Law and Policy (LP), and other (O).

importance of setting boundaries. Within one year of the establishment of any category of protected forest area, such as a National Forest, the authority is required to clearly mark the boundary between the protected area and its surrounding lands. In addition, section 9.9 of the document requires the authority to identify and establish smaller areas such as Controlled Hunting Areas and Communal Forests, other than the more general category such as National Forests and Nature Reserves, to be reserved for particular activities. These smaller reserved areas make up "Conservation Corridors to facilitate sustainable protected forest management and biodiversity protection." Last but not least, the following section clearly states that, among other activities, farming is prohibited in all types of protected forest areas. With the four-year Second Liberian Civil War ending in 2003, a more stable government and law, enforcement accompanied by newly established policies, likely was the largest force that brought down deforestation in the country in the following year.

Togo's negative-signed break is detected in 2010 with an effect size of 63.6% decrease in tree cover loss. During the Oslo Climate and Forest Conference in May 2010, Togo joined the REDD+ Partnership (Forest Carbon Partnership Facility, 2013), in the hope of mitigating the forest crisis in the country. Since then, REDD+ has been integrated into Togo's development programs. For example, Togo validated its PNIERN[5] in October 2010 and held regional workshops of PNIERN across its economic regions before the validation, as well as advocacy sessions outside the country during international meetings attended by foreign technical and financial partners. in the meantime, the forestry sub-sector was one of the focus points in the process of formulating the National Forestry Action Plan (NFAP) framework founded by FAO[6] between 2010 and 2011. The NFAP further refined some urgent actions in the PNIERN concerning sustainable forest management. Together the NFAP and PNIERN potentially provided a significant reduction effect on

[5]National Investment Programme for Environment and Natural Resources [Programme National d'Investissements pour l'Environnement et les Ressources Naturelles PNIERN).
[6]Food and Agriculture Organization of the United Nations.

Togo's deforestation. What's worth noting is that Togo also passed its National Agency for Environmental Management in October 2009 (Convention on Biological Diversity. n.d.), not long before it participated in the REDD+ Partnership.

India's break denotes a 60.5% decline in tree cover loss in 2013. Both data algorithm improvement in 2013 and the lack of available Landsat images (*i.e.* low tree cover loss detection) in the previous year strongly encourages India to experience a jump in tree cover loss data in 2013, which is the opposite of this detected break. Therefore, the 2013 reduction in India indicates a particularly large drop in the tree loss statistics. One potential cause for the reduction is The Indian Forest (Amendment) Bill, 2012, which was presented on the 26th of November 2012, but it seems implausible because the biggest policy change made by the amendment was simply the increased transparency of law enforcement on forest offence. The principal act being amended by this bill is the Indian Forest Act, 1927, which consolidated laws associated with forests and forest produce. This Act prohibits certain activities including making fresh clearance in forests and setting fires to reserved forests, which could lead to imprisonment or fines (The Indian Forest (Amendment) Bill, 2012). However, section 68 of the principal act empowers a forest officer to accept money or property that is worth the same value as compensation for most types of forest offence from a person when an offence is committed or simply when a suspicion exists. Upon the reception of the compensation, the forest officer is required to discharge the person. Therefore, the enforcement of this law could lead to fraudulent claims of forest offence by incompetent forest officers for corrupt reasons. The objective of the 2012 amendment bill was to make the law more relevant to the changed socio-economic scenario and to protect tribals and forest dwellers from harassment of petty forest offences. The amendment attempted to solve the issue by adding a new subsection 68(4) that requires a forest officer to "obtain and record the views of the Gram Sabha[7] concerned." The Gram Sabha is the head of a village government whose decision-making is only valid if the decision is endorsed by its Gram Sabha (Nambiar M, 2001). In conclusion, the 2012 amendment of The Indian Forest Bill increased the transparency in the process of forest offence accusations by forest officers, but it remains unclear whether there's any direct or indirect association between this amendment and the detected reduction in deforestation in India for the year 2013. One of the possible explanations is that the amendment helped achieve more public awareness of the existence of such forest laws, which subsequently contributed to its sharing in reducing tree cover loss.

Ecuador's break appeared in 2014, 6 years after the country implemented its conservation program "Socio Bosque" to fight deforestation and increase native forests, among other goals (United Nations Development Program, 2018). According to the Program, The Green Climate Fund promised to issue funding through results-based payments based on an Environmental and Social Assessment Report. The payment was sought for the period from 2009 to 2014, starting soon after the implementation of the Socio Bosque program in November 2008. One reasonable assumption of this detected 59.3% reduction in deforestation is that the government of Ecuador invested more effort in reducing tree cover loss in 2014 specifically, the

[7] Gram Sabha is the Indian term for People's Forum. It is "an institution direct democracy" commonly exists in Indian village communities. A Gram Sabha is said to be a people's body that "[gives] approvals for development plans, [prioritizes, identifies] beneficiaries, [and promotes] active participation of people in implementing developing programmes" (Nambiar M, 2001).

last year of the six-year period for which the results-based payments are sought, to achieve better results that would in turn lead to higher payments from the Green Climate Fund.

Using the target significance level of $\gamma_c = 0.01$, statistically significant breaks of the same magnitude (-0.75) with similar standard errors are detected in **Costa Rica** for 2013 and in Ghana for 2003, representing a decrease in deforestation by 52.8%. Although similar breaks, they were likely realized through different approaches. One of the most effective policies in reducing deforestation in Costa Rica was the Payments for Ecosystem Services (PES). Nevertheless, the country was still facing a deforestation struggle since the launch of PES in 1997. For example, it lost approximately 45 thousand hectares of forest vegetation between 2008 and 2013 from the creation of pastures (Climate Chance, 2020), representing nearly 0.9% of the country's land size (5.1M Ha) (The World Bank, n.d.). According to the case study of Costa Rica's land use by Climate Chance in 2020, this loss of tree cover accounts for 68% of the total deforestation throughout the 6 years. This loss happened largely in private lands, in which the main deforestation driver was the conversion of woodlands to agriculture or livestock farming. In October 2013, Costa Rica established an environmental bank named BanCO2 that created a voluntary market for carbon dioxide credit, aiming to mitigate climate change and develop a low-carbon business model (Salazar, n.d. & The Tico Times, 2013). This environmental bank would buy and sell carbon credits enabling individuals and companies to offset their GHG emissions. However, According to the World Bank Group et al (2020), this domestic carbon market never became fully operational. It remains a question whether this downward step-shift in deforestation was directly caused by the carbon credit market or was indirectly associated with it, or even unrelated. Whichever path the impact comes into effect, it is potentially larger than the point estimate due to the algorithm adjustment and Landsat 5 retirement that were supposed to increase tree cover loss data, which took place in the same year.

On the other hand, concerning the step-shift of the same direction and magnitude in **Ghana**, things are a little more complex. Among other factors, the one that seems to have the most direct connection with the deforestation drop in 2003 is the assent of Act 624 of The Forest Protection (Amendment) Act, 2002. This act was assented to in May 2002 and raised the penalties for forest offences, such as fraudulent taking, marking, and destruction of any tree or timber (World Resources Institute, 2013). In particular, the act forbade farm creation and cultivation as well as erecting any buildings, which prevented any farming activities from being conducted within the boundaries of reserved forests, especially cocoa farms whose products took up more than half of Ghana's export value. While Act 624 took its time to be enforced in each of Ghana's regions in late 2002 and early 2003, the country approved another law - the Act 651 of Labour Act 2003, which established Ghana's National Labour Commission to promote a conducive industrial environment that provided maximum working hours and overtime work (Client Earth, 2022). Act 651 also prohibited "the employment of young people in hazardous work" of all sorts. Under the influence of Act 651, a healthier and more sustainable work environment possibly helped reduce reforestation from two indirect paths. One is through the higher employment rate in the industrialized jobs, switching workers from farming to alternative employments that are less associated with felling trees; the other is banning young people from working in the forests, which could be considered hazardous.

Last but not least, our approach detects a shift in **Sierra Leone**'s deforestation with the approximate effect size of -45.7% in 2003, one year after the Sierra Leone Civil War ended. In that same year, the government of this country started the formulation of a new forest policy statement, but the top priorities during the post-war reconstruction were given to increasing economic and social goals as well as food self-sufficiency (Grainger and Konteh, 2007). Formulated in 2003, the Sierra Leone Biodiversity Strategy and Action Plan has several sections dedicated to forest biodiversity, proposing preservation actions such as better forest management and discouragement of forest-related agriculture expansion, especially slash-and-burn activities (FAO, 2003). Specifically, the preservation actions include, amongst many others, "[developing and implementing] alternatives to slash-and-burn agriculture in the forest zones;" encouraging sustainable exploitation and management of forest resources by providing incentives through a reward system; subsiding cleaner energy source development to replace firewood and charcoal; adopting stringent licensing systems; and planning allowable forest cuts in piror to exploitation. All of these clauses had the potential to reduce the tree cover loss in Sierra Leone conditional on they were formally adopted and strictly enforced. Meanwhile, another statement of policy was drafted in the country, namely the Forestry and Wildlife Sector Policy. However, revealed by the documentation of Sierra Leone's Forest Policy 2010, for years, this draft remained the most complete statement of policy but was never officially adopted by the government (FAO, 2010). More than one policy statement was formulated in 2003 to protect Sierra Leone's forests, but one cannot be entirely sure all the policies became formal across the country's land. It is reasonable to assume the effort Sierra Leone made to preserve forests right after a decade-long civil war ended was little, if not none, compared to what they would do to regain economic and social stability and food self-sufficiency. All the promising and efficient-sounding policies formulated, if not adopted, would be nothing but words living in documentation.

All the above policies that are potentially effective in reducing deforestation in the eight countries in which negative-signed structural breaks are detected can be framed in three broad categories: economic incentives, related law and policy improvement and enforcement, and other. Table 2 summarizes them with a list of policy/event attribution along with the magnitude of the detected breaks. As listed, among these eight breaks, two of them are largely driven by economic incentives, five are likely caused by improvement in related law and policy enforcement, and the one left is potentially led by other factors such as a global pandemic (COVID-19) during which quarantine was a social norm. Among these eight policy attributions, stricter forest law enforcement is the most popular, perhaps owing to its feature of being the most straightforward and likely the fastest to see improved results, which are desired in countries suffering severe and rapid tree cover loss. Interestingly, the same type of policies seem to offer different effect sizes in different treated countries. For example, clearer forest classification and stricter forest management had a more significant effect in Liberia compared to similar forest law improvements made in other countries. Both the negative-signed breaks in Liberia (2004) and Sierra Leone (2003) are detected for the period after their civil wars, separately, but they sit on the opposite ends of the effect size range when considering breaks attributed to forest laws alone (-45.7% to -70.5%). This could imply that Liberia had a more stable post-war government, among other potential factors, at the time of enforcing the improved forest laws.

Togo and Ghana's efforts on forest law improvement went in slightly different directions and achieved differently sized impacts. Since joining the REDD+ Partnership, Togo integrated REDD+ into its national development programs and made great efforts to advocate environmental issues as well as related policies designed to solve such issues. It is likely that the advocating led to greater public awareness, which encouraged a more significant effect in Togo, compared to the scenario in Ghana that focused more on increasing the penalties for forest offences.

What's interesting to note is the method of economic incentives used in Ecuador and Costa Rica achieved effects of very similar effect sizes (-59.3% and -52.8%). A marginally higher effect realized by Ecuador might be due to its already well-established system of economic incentives. The Socio Bosque program that started in 2008 aimed to conserve native forests by letting property holders voluntarily enter agreements with the Ministry of the Environment using incentive payments paid bi-annually to those who place all or part of their land into conservation areas (United Nations Development Program, 2018). Another potential reason would point to the data issue where the algorithm adjustment and Landsat 5 retirement could have repressed the tree cover loss data for Costa Rica in 2013 and subsequently resulted in a smaller change when compared to the data of the previous year. Aside from this factor, the similarity in effect sizes for the detected breaks in Costa Rica and Ecuador implies that using economic incentives to help reduce tree cover loss would achieve a more stable statistically significant effect, though not necessarily the most remarkable. Therefore, it potentially indicates a trade-off, between a moderate-sized but stable effect and a more volatile effect in a range of effect sizes with a higher upper limit, when policymakers choose between forest law improvement and adapting economic incentives that would reduce tree cover loss.

On the other hand, these two breaks attributed to economic incentives might not be the most robust estimates. Costa Rica's 2013 break is only detected when the target significance level equals 0.01 (1%) while Ecuador's 2014 break is detected both at 0.01 (1%) and 0.005 (0.5%). The breaks are considered more robust when they are detected in more different models. Nevertheless, the fact that these two breaks are detected only with bigger false-positive rates could be attributed to their smaller effect sizes. For example, the four breaks retained in the model with $\gamma_c = 0.001(0.1\%)$ have the largest effect size among all the fifteen breaks in the original model (see Table 1).

More robustness check is done via the model that includes the cocoa variable controlling for the natural logarithm of the area of cocoa harvested (last column in Table 1). Since cocoa cultivation is one of the major drivers of deforestation in the tropics and all 23 countries in the data sample are major cocoa-producing countries, controlling for cocoa area harvested in the model potentially reveals the breaks that are related to cocoa production. None of the breaks from the original model is dropped when cocoa area harvested is controlled, unfortunately, indicating that none of these breaks are significantly correlated with the cocoa industry. However, this provides further evidence of the robustness of the estimates in the model with a 1% target significance level, given that all the breaks are retained with stable effect sizes and standard errors while controlling for extra variables. The largest size variation in the estimates is contributed by the break in Haiti (2017) from -1.38 to -1.65, a -0.27 change, which is within an acceptable range. The variations for the other breaks are on average very small.

5. Conclusion and Discussion

The exploration of reverse causal questions is rarely sought after by researchers in the econometrics literature, perhaps because the method of tackling them is less obvious. However, the break detection approach that relies on machine learning can change that. The backbone of this approach is the model selection procedure based on a complete variable selection done by saturating a full set of individual-time interaction terms and subsequently keeping only the significant ones in the retained final model. By using this approach, first introduced by Pretis (2022), I identify potential effective policy interventions that cast an impact on deforestation trends in cocoa-producing countries. Although break detection has become a well-established tool in time series analysis, such applications often lack control groups that would otherwise enable researchers to draw causal interpretations. However, the novel approach adopted by the present paper allows for a control group, consisting of the countries where structural breaks are not detected. This inclusion, in turn, allows me to draw causal links between the identified interventions and the reduction in tree cover loss. Among the twenty-three cocoa-producing countries included in the dataset, ten countries have detected structural breaks in the natural logarithm of their tree cover loss, meaning the other thirteen countries together act as the control group. Of the detected breaks, eight can be attributed to potential policy interventions that effectively reduced tree cover loss in the corresponding countries. Five of these interventions are improvement and enforcement of new laws or policies while two are economic incentives. On one hand, the approach of Law and policy enforcement achieved effects that vary in a large size range, perhaps due to the varying enforcement methods and level of strictness adopted by different governments. On the other hand, using economic incentives as a tool to drive down deforestation seems to help achieve more stable but comparatively medium-sized effects. This potentially would help the decision-makers to forecast a more accurate effect size and market response time.

As for the positive-signed breaks that are detected by the model with 1% target significance level, although more than half of them are likely due to either the pre-existing data inconsistency or other data-related issues, the rest implies increases in tree cover loss caused by improper forest laws or inferior method of policy enforcement. For example, in Liberia, the approval of the forestry law that would award forest clearing for agriculture expansion was followed by a remarkable increase in tree cover loss; when the policy in Togo that promoted low carbon emissions ended after its scheduled five-year period, the level of tree cover loss jump perhaps due to the lack of another similar policy in place, indicating that inconsistent policy enforcement would create interruptions in the result variable.

Some caveats persist in my estimation results, primarily due to issues within my tree cover loss dataset.

Assessing trends using data that include years before 2015 may lead to questionable conclusions due to data inconsistency. While employing 3-year moving averages on a simple line graph can help smooth inconsistencies between years, it proves more challenging in the variable selection procedure. Caution is warranted when comparing periods before and after 2011 or before and after 2015, as it involves a comparison across model changes due to mapping algorithm adjustments in the dataset.

Additionally, the suggested percentages of avoided deforestation (45-70%) resulting from the attributed policy and event interventions appear surprisingly high, raising doubts about the estimation. Therefore, it's crucial to discuss the effect sizes within the context of large year-on-year variations in the tree cover loss data in the sampled countries. Moreover, cocoa production is undetectable using Landsat images due to its shade-grown systems, often under secondary forest canopies or mixed with food crops. Hence, the data of tree cover loss likely has less desired accuracy in capturing cocoa production, rendering a less accurate robustness check while controlling cocoa harvested area. Future analyses are encouraged, either when more data after 2015 is available from this dataset or when other spatial data that better capture deforestation trends become available.

Planting new trees is a great way to achieve carbon sequestration, but it is fairly less efficient compared to protecting longstanding forests, which are far more valuable carbon sinks (Moseman, 2022). While new growth takes its time to become efficient media for carbon sequestration, the cutting down of old growth contributes far more greenhouse gas emissions, let alone the emissions coming from the logging machinery and log transportation (Weir, 2022). Nevertheless, all these reasons don't seem to have slowed the global deforestation rate as it remains a crucial world-class problem. It is worth noting that numerous developed nations are expanding their forested areas, while rampant deforestation is taking place in the tropical and subtropical regions, where less developed countries are situated (Pendrill et al., 2019). As per the study findings, between 2005 and 2013, 26% of global deforestation can be traced back to the international demand for commodity crops, and strikingly, 87% of this demand was exported to countries where deforestation rates were decreasing or forest cover was on the rise. Since a large amount of commodity-driven deforestation is attributed to crops produced primarily for exports, maybe the consumer countries could collect specific taxes on such crops and dedicate them to mitigating deforestation in the locations of production.

6. **Reference**

Abbott, P. Distortions to Agricultural Incentives in Cote d'Ivoire. https://documents1.worldbank.org/curated/ar/546141468150872673/pdf/560300NWP0CI0v1Cote1dIvoire10708rev.pdf

Alfaro, J.F. and Jones, B. (2018). Social and environmental impacts of charcoal production in Liberia: Evidence from the field. Energy for Sustainable Development, 47, 124-132. https://doi.org/10.1016/j.esd.2018.09.004

Bo, S. (2011). A Literature Survey on Environmental Kuznets Curve. *Energy Procedia 5(2011)*, 1322-1325. https://doi.org/10.1016/j.egypro.2011.03.229

Budiansky, E. and Freire, R.M. (2018). Forest Positive Cocoa in the Amazon. *World Cocoa Foundation.* https://www.worldcocoafoundation.org/blog/cocoa-agroforestry-breaking-the-cycle-of-deforestation-in-the-amazon/

Client Earth and Taylor Crabbe. (2022). Cocoa Research – Briefing 1: Legal & institutional framework for cocoa production & trade in Ghana. https://www.clientearth.org/media/bqnajojt/cocoa-research-briefing-1-legal-institutional-framework-for-cocoa-production-trade-in-ghana.pdf

Climate Chance. (2020). *Costa Rica Land Use.* https://www.climate-chance.org/wp-content/uploads/2020/09/costarica_landuse_climatechance_engl.pdf

Convention on Biological Diversity. (n.d.). *Togo – Main Details.* https://www.cbd.int/countries/profile/?country=tg

FAO. (2003). Act for the Establishment of a Protected Forest Area Network and Amending Chapter 1 and 9 of the new National Forestry Law, Part II of Title 23 of the Liberian Code of Laws Revised. *FAOLEX Database.* https://www.fao.org/faolex/results/details/en/c/LEX-FAOC061868

FAO. (2003). Sierra Leone – Sierra Leone Biodiversity Strategy and Action Plan. *FAOLEX Database.* https://www.fao.org/faolex/results/details/en/c/LEX-FAOC175795/

FAO. (2006). National Forestry Reform Law of 2006 - An Act Adopting the National Forestry Reform Law of 2006. *FAOLEX Database.* https://www.fao.org/faolex/results/details/en/c/LEX-FAOC067626. Date of Access: September 6, 2023.

FAO. (2010). Sierra Leone – Forest Policy 2010. *FAOLEX Database.* https://www.fao.org/faolex/results/details/en/c/LEX-FAOC143754/

Food and Agriculture Organization of the United Nations. (2022). https://www.fao.org/faostat/en/#data/QCL

Forest Carbon Partnership Facility (FCPF). (2013). READINESS PREPARATION PROPOSAL (R-PP) – COUNTRY: TOGO. https://www.forestcarbonpartnership.org/system/files/documents/Togo_110813_RPP_English.pdf

Gelman A. & Imbens G. (2013) Why ask why? Forward causal inference and reverse causal questions.

Grainger. A, and Konteh, W. (2007). Autonomy, ambiguity, and symbolism in African politics: The development of forest policy in Sierra Leone. *Land Use Policy. (24)*1, 42-61. https://doi-org.ezproxy.library.uvic.ca/10.1016/j.landusepol.2005.09.001

Grossman G.M. and Krueger. A.B. (1991). Environmental Impacts of a North American Free Trade Agreement. *National Bureau of Economic Research. Working Paper.* DOI 10.3386/w3914

Igini, M. (2023). How Does Cocoa Farming Cause Deforestation? *Earth.Org.* https://earth.org/how-does-cocoa-farming-cause-deforestation/

Johnson, A. (2023). From Slash-and-Burn to Sustainable: How Biochar is Changing Cocoa Farming in Côte d'Ivoire. *Beyond Beans.* https://beyondbeans.org/2023/03/27/from-slash-and-burn-to-sustainable-how-biochar-is-changing-cocoa-farming-in-cote-divoire/

Koch N. et al. (2022) What reduces road CO_2 emission? Policy attribution using break detection.

Lo, J. (2022). US-funded trees 'not likely to survive' in Haiti when project ends. *CLIMATE HOME NEWS.* https://www.climatechangenews.com/2022/08/05/us-funded-trees-not-likely-to-survive-in-haiti-when-project-ends/

Marcelli, T. (2020). USAID/HAITI REFORESTATION & AGROFORESTRY – FACT SHEET – January 2020. *United States Agency International Development.* https://2017-2020.usaid.gov/sites/default/files/documents/1862/USAID_Haiti_Reforestation_Fact_Sheet_-_January_2020.pdf

McConnell, K.E. (1997). Income and the demand for environmental quality. *Environment and Development Economics, 2*(4), 383-399. https://doi.org/10.1017/S1355770X9700020X

Moseman, A. and Harvey, C. (2022). How many new trees would we need to offset our carbon emissions? *MIT Climate Portal.* https://climate.mit.edu/ask-mit/how-many-new-trees-would-we-need-offset-our-carbon-emissions

Nambiar, M. (2001). Making the Gram Sabha Work. *Economic and Political Weekly, 36*(33), 3114-3117. https://www.jstor.org/stable/4410988

NASA Earth Observatory (2023a). *July 2023 Was the Hottest Month on Record.* https://earthobservatory.nasa.gov/images/151699

NASA Earth Observatory (2023b). *World of Change: Global Temperatures.* https://earthobservatory.nasa.gov/world-of-change/global-temperatures

National Aeronautics and Space Administration. (n.d.). LANDSAT 5. Retrieved September 9, 2023, from https://landsat.gsfc.nasa.gov/satellites/landsat-5/

National Aeronautics and Space Administration. (n.d.). TIMELINE. Retrieved September 9, 2023, from https://landsat.gsfc.nasa.gov/satellites/landsat-5/

Neyman, J. (1923). On the application of probability theory to agricultural experiments. Essay on principles. Section 9. Translated and edited by D. M. Dabrowska and T. P. Speed. *Statistical Science* 5, 463–480 (1990).

NOAA Climate.gov (2023). *Climate Change: Global Temperature.* https://www.climate.gov/news-features/understanding-climate/climate-change-global-temperature#:
:text=The%202022%20surface%20temperature%20was,period%20(1880%2D1900).

Pahlevi A. de Souza, Lunian F. Moreira, Diógenes H.A. Sarmento, Franciscleudo B. da Costa (2018). Exotic Fruits. American Press. https://doi.org/10.1016/B978-0-12-803138-4.00010-1.

Pendrill, F., Persson, U.M., Godar, J., Kastner, T., Moran, D., Schmidt, S., Wood, R. (2019). Agricultural and forestry trade drives large share of tropical deforestation emissions. *Global Environmental Change.* 56, 1-10. DOI: 10.1016/j.gloenvcha.2019.03.002.

Pretis, F. (2022). Does a Carbon Tax Reduce CO_2 Emissions? Evidence from British Columbia.

Pretis, F. & Schwarz, M. (2022). Discovering What Mattered: Answering Reverse Causal Questions by Detecting Unknown Treatment Assignment and Timing as Breaks in Panel Model.

Ritchie, H. (2021). Cutting down forests: what are the drivers of deforestation? *Our World in Data.* https://ourworldindata.org/what-are-drivers-deforestation

Rubin, D. B. (1974b). Estimating causal effects of treatments in randomized and nonran- domized studies. Journal of Educational Psychology 66, 688–701.

Saint, E. (2022). The Struggle to Conserve Threatened Forests in Haiti. Earth. Org. https://earth.org/conserve-forests-in-haiti/

Salazar, M. (n.d.). Costa Rica Aims For Carbon Neutrality With Payments For Ecosystem Services. *Ecosystem Marketplace – A Forest Trends Initiative.* https://www.ecosystemmarketplace.com/articles/costa-rica-aims-for-carbon-neutrality-with-payments-for-ecosystem-services/

Stern, D.I. (2004). Environmental Kuznets Curve. *Encyclopedia of Energy.* https://doi.org/10.1016/B0-12-176480-X/00454-X

The Chocolate Journalist. (2022). Here Is Exactly How Cocoa Farming Is Causing Deforestation. https://www.thechocolatejournalist.com/blog/cocoa-deforestation

The Forest Protection (Amendment) Act, 2002. Act 624. (2002). Act 624 THE SIX HUNDRED AND TWENTY-FOURTH ACT OF THE PARLIAMENT OF THE REPUBLIC OF GHANA. https://www.clientearth.org/media/sm1jfwic/forest-protection-amendment-act-2002-ext-en.pdf

The Government of Sierra Leone. (2013). THE AGENDA FOR PROSPERITY - ROAD TO MIDDLE INCOME STAT - Third Generation Poverty Reduction Strategy Paper (2013 – 2018). *Food and Agriculture Organization of the United Nations. FAOLEX Database.* https://www.fao.org/faolex/results/details/en/c/LEX-FAOC149110/

The Indian Forest (Amendment) Bill, 2012. https://prsindia.org/billtrack/the-indian-forest-amendment-bill-2012#:~:text=The%20Act%20consolidates%20the%20law,fire%20to%20a%20reserved%20forest

The Sustainability Consortium, World Resources Institute, and University of Maryland. "Tree Cover Loss by Driver." Accessed through Global Forest Watch on 13/10/2022. www.globalforestwatch.org.

The Tico Times. (2013). *Costa Rica's carbon market is open for business.* https://ticotimes.net/2013/10/18/costa-rica-s-carbon-market-is-open-for-business

The World Bank. (2022). https://data.worldbank.org/indicator/SP.POP.TOTL https://data.worldbank.org/indicator/NY.GDP.MKTP.KD?most_recent_value_desc=true

The World Bank. (n.d.). *Land area (sq. km) - Costa Rica.* https://data.worldbank.org/indicator /AG.LND.TOTL.K2?end=2020&locations=CR&most_recent_value_desc=true&start=1961&view=chart

United Nations Development Program. (2018). Environmental and Social Assessment of Ecuador's Socio Bosque Programme. https://www.ambiente.gob.ec/wp-content/uploads/downloads/2018/09/ESA-Socio-Bosque-Ecuador-Report-11-09-18-f.pdf

United Nations. What is REDD+? *Climate Change.* https://unfccc.int/topics/land-use/workstreams/redd/what-is-redd

Weir, K. (2022). Forest Carbon 101. *The Nature Conservancy.* https://www.nature.org/en-us/magazine/magazine-articles/forest-carbon-101

Weisse and Peter Potapov. (2021). Assessing Trends in Tree Cover Loss Over 20 Years of Data. *Global Forest Watch.* https://www.globalforestwatch.org/blog/data-and-research/tree-cover-loss-satellite-data-trend-analysis/

World Bank Group, Partnership for Market Readiness Costa Rica, and Direccion de Cambio Calimatico. (2020). *Partnership for Market Readiness - Costa Rica Program.* https://documents1.worldbank.org/curated/en/469421617604271877/pdf/Partnership-for-Market-Readiness-Costa-Rica-Program-The-Costa-Rican-Offset-Mechanism-MCCR.pdf

World Resources Institute. (2013). Ghana. Forest Legality Initiative. https://forestlegality.org/risk-tool/country/ghana#:~:text=The%20Forest%20Protection%20Decree%2C%201974,of%20the%20competent%20forest%20authority.

World Resources Institute. (2023). Forest Pulse: The Latest on the World's Forests. Global Forest Review. https://research.wri.org/gfr/latest-analysis-deforestation-trends

7. **Appendix**

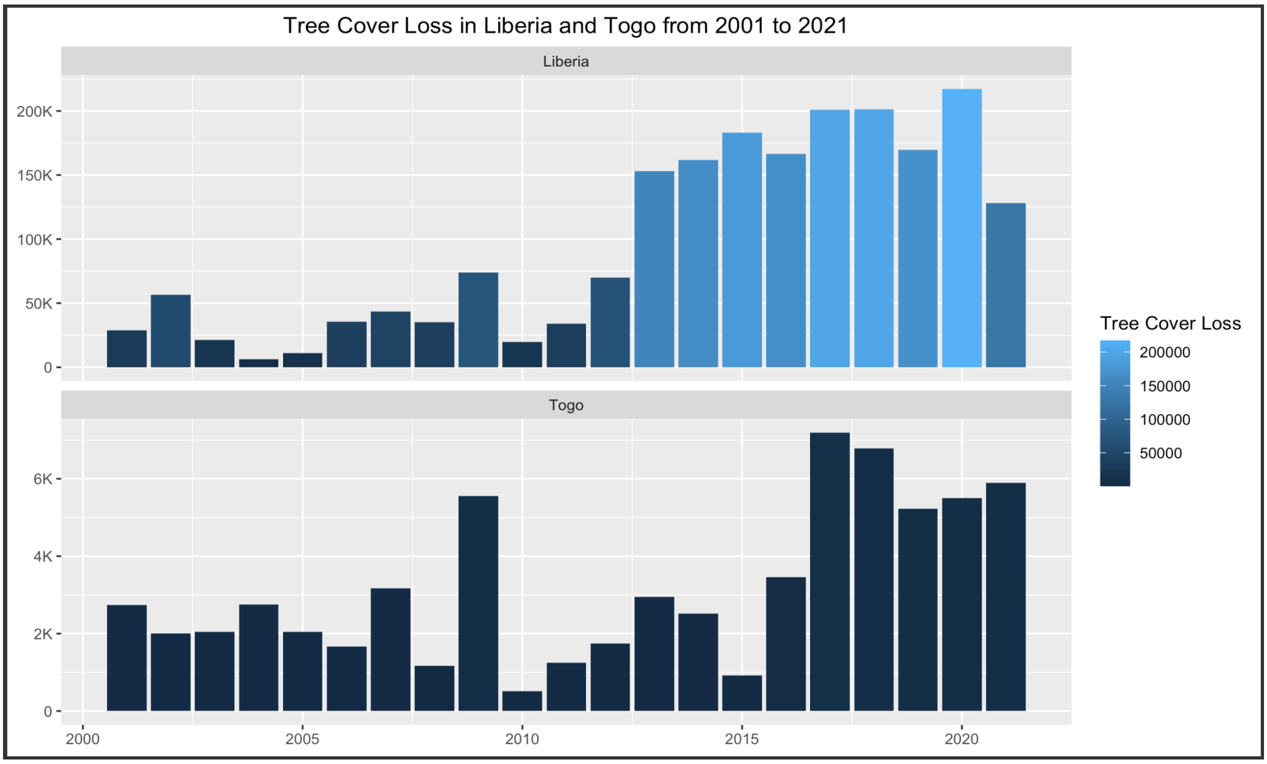

Figure 8: This bar chart shows the tree cover loss from 2001 to 2021 in Haiti. The red dashed horizontal lines represent the mean of tree cover loss during 2001-2012 and 2013-2021, respectively.

The eight negative signed breaks and their avoided deforestation							
Country	**Break Year**	**Effect Size**	**Log(Loss)**	**Counterfactual Loss (Ha)**	**Actual Loss (Ha)**	**Avoided Deforestation (Ha)**	**% of Avoided Deforestation**
Malaysia	2020	-1.29	5.668892	1,052.47	289.7	762.8	-72.5%
Liberia	2004	-1.22	8.725769	20,863.76	6159.6	14,704.2	-70.5%
Togo	2010	-1.01	6.228202	1,391.59	506.8	884.8	-63.6%
India	2013	-0.93	8.496203	12,409.32	4896.1	7,513.2	-60.5%
Ecuador	2014	-0.90	10.2921	72,554.99	29498.7	43,056.3	-59.3%
Costa Rica	2013	-0.75	8.876229	15,157.17	7159.7	7,997.5	-52.8%
Ghana	2003	-0.75	10.04175	48,618.04	22965.6	25,652.4	-52.8%
Sierra Leone	2003	-0.61	9.090051	16,318.44	8866.6	7,451.8	-45.7%

Table 3: This table is an extension of Table 2. It shows the numeric values for the natural logarithm of the tree cover loss, the counterfactual tree cover loss in the absence of the attributed treatment, and the actual tree cover loss. Avoided deforestation is computed as Counterfactual Loss - Actual Loss and the percentage of avoided deforestation $\frac{\text{Avoided Deforestation}}{\text{Counterfactual Loss}}$.