UC San Diego
UC San Diego Electronic Theses and Dissertations

Title
Essays on the political economy of the environment; methods and applications

Permalink
https://escholarship.org/uc/item/5zf1c11j

Author
Sanford, Luke Coyne

Publication Date
2021

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA SAN DIEGO

Essays on the political economy of the environment; methods and applications

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Political Science & International Affairs

by

Luke Coyne Sanford

Committee in charge:

Professor Jennifer Burney, Co-Chair
Professor Margaret Roberts, Co-Chair
Professor John Ahlquist
Professor Prashant Bharadwaj
Professor Megumi Naoi

2021

Copyright
Luke Coyne Sanford, 2021
All rights reserved.

The dissertation of Luke Coyne Sanford is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2021

DEDICATION

To Xiaomin: You hold me together and make me whole.

To Cassia for the perspective that you un-knowingly lent

To Izzi Woof.

EPIGRAPH

What you make from a tree should be at least as miraculous as what you cut down.

—Richard Powers, The Overstory

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I want to start by acknowledging that I have spent the better part of the last six years living and studying on the unceded homelands Kumeyaay people. I have much to learn as a person and as a scholar from indigenous peoples. I acknowledge your contributions to the region and thank you for your stewardship.

Surviving a PhD takes a village, and I am so thankful for mine. This dissertation is the product of years and years of support from faculty, classmates, colleagues, friends and family. I have too many people to thank and I undoubtedly left at least a few people who are deserving of thanks out. Xiaomin, wherever you are is where my home is. Thank you for your endless support. I love you. Cassia, your future inspires me to ask questions about how to make the world a better place for you to live. Izzi, I wouldn't have made it without you. You deserve many more days at dog beach than I could provide.

I was advised by two of the best people I know: Jen Burney and Molly Roberts. Jen, you have been an amazing role model since I first took your class in the first days of 2014. I hope that I can be a fraction of the mentor that you have been for me one day. Molly, your work and guidance has been inspirational for me. I have learned so much about the discipline of political science from you. You both have given me opportunities which I did not deserve but strive to live up to. I will strive to make both of you proud.

Megumi Naoi, you guided me through my two most difficult years in graduate school, helped me craft the paper which forms the core of my research agenda, and helped me to pick out an amazing committee. Most of what I know about acting strategically in academia has come from you. I am sure that I would not have had the same success on the job market without your guidance.

John Ahlquist, you consistently provide the feedback I need to continue improving. Thanks for always pushing me to think critically and not take shortcuts.

Prashant Bharadwaj, your enthusiasm and perspective have helped me to discover some of the most important aspects of my own work.

Barbara Walters and Emilie Hafner-Burton provided invaluable guidance and support when I decided to move from the professional masters program to the PhD program. Your support changed my life.

I would like to thank the Institute on Global Conflict and Cooperation and the Big Pixel Initiative for their generous funding of my research and study.

My extended family, particularly my mom Sara Molden and sister Kate Sanford always believed in me, even when I didn't believe in myself. My aunt and uncle Dan and Marty Molden provided the comfort and support of a home and family in San Diego. Dan has also been my best friend and taught me how to surf, one of the greatest joys of living in San Diego. My aunt Karen Conniff and uncle David Molden hosted me in their home for nine months and showed me what a good research-life balance looked like while providing an opportunity to be a part of my first real research projects. My parents in law Jianguo Hu and Guilan Guan helped us survive our first year as parents during a pandemic. You are better in-laws than most people can imagine.

My friends, classmates, and mentors have been indispensable. Most of my best ideas were probably their ideas and everything I have written or presented was made better by their input. Rachel Schoner, Thomas Flaherty, and Mariana Carvalho-Barbosa were my inner circle. Surviving the econometrics sequence together forged an unbreakable bond between us which will continue across continents. You supported me and helped take care of my family when I couldn't. Jazmin Flaherty-Barajas, and Ricardo Dahis, you became some of my best friends over the last several years and provided important outside perspectives. Wesley, Lucy and Sandy, you helped to tire Izzi out when I couldn't.

Joe Riley, Kyle Hemes, Arman Khachiyan, Eric Geier, Thomas Flaherty, and Uncle Dan–with your help surfing became a passion which kept my body healthy and my mind free. See you out there.

So many friends provided support, comfort, advice and an open heart. Thank you to Simeng Zeng, Chris Lee and Bryn Taylor, Jackson Somers, and my teammates on California Burrito, especially Peter Osseward and Sarah Zhang. San Diego would not have been the same without you.

A number of older students played a role in my development as a scholar. Abby Vaughn and Brandon Merrell, Kristy Pathakis, Taylor Carleson and Mike Seese all showed me what good peer mentorship looks like and guided me through many of the small difficulties of academia.

Many friends made an effort to stay in touch even when I was buried in research. Matt Cameron (awwooo) and Willie Kunkel provided great friendship from afar, and it has been fun to grow into adulthood

with you two. Nate Cohn, Jake Ginsbach and Anjali Vats were always there for me when I needed them and have provided some of my favorite memories over the last few years. Mark Buttweiler and Colin Doberstein were the best housemates I could have asked for and were somehow just as nerdy as I was in exactly the right way. Nic Beaudoin has been a touchstone, great friend, and one of my earliest surfing partners.

A number of faculty have been hybrid friend-mentors over the years and made me feel accepted as myself in academia. Thanks to Gordon McCord, Teevrat Garg, Sean Ingham, David Weins, and Morgan Levy for befriending a graduate student and sharing your leisure time activities and advice with me.

Special thanks to all of the administrators who tirelessly worked to make sure that I had funding, was enrolled in classes, had health insurance, and was supported at UCSD. Special thanks to Julie Choi, Ariane Parks, Vanessa Pool, Stephanie Boomhower and Jerry Pang.

To the 2015 cohort—I love you all, you truly are the best cohort ever, thank you for everything.

Chapter 2, in full, has been accepted for publication and is presented as it may appear in "Sanford, L. *Democratization, Elections, and Public Goods: The Evidence from Deforestation* AJPS."

Chapter 4 is coauthored with Sanford, L. Roberts, Margaret and Li, Kevin. The dissertation author was a lead author of this chapter.

VITA

2009	Bachelors of Arts in Politics Whitman College
2011	Research Assistant International Water Management Institute
2016	Master of Political Science University of California San Diego
2015-2021	Graduate Student Researcher and Teaching Assitant University of California San Diego
2021	Doctor of Philosophy in Political Science & International Affairs University of California San Diego

PUBLICATIONS

Sanford, L. & Burney, J. *Cookstoves Illustrate the Need for a Comprehensive Carbon Market* Environmental Research Letters 2015, 10 (8), 084026. https://doi.org/10.1088/1748-9326/10/8/084026

Lautze, J.; de Silva, S.; Giordano, M. & **Sanford, L.** *Putting the Cart Before the Horse: Water Governance and IWRM* Natural Resources Forum 2011, 35 (1), 1-8. https://doi.org/10.1111/j.1477-8947.2010.01339.x

Molden, D.; Lautze, J.; Shah, T.; Bin, D.; Giordano, M. & **Sanford, L.** *Governing to grow enough food without enough water–Second Best Solutions Show the Way* International Journal of Water Resources Development 2010, 26 (2), 249-263. https://doi/10.1080/07900621003655643

Sanford, L. *Democratization, Elections, and Public Goods: The Evidence from Deforestation Accepted at the American Journal of Political Science*

ABSTRACT OF THE DISSERTATION

Essays on the political economy of the environment; methods and applications

by

Luke Coyne Sanford

Doctor of Philosophy in Political Science & International Affairs

University of California San Diego, 2021

Professor Jennifer Burney, Co-Chair
Professor Margaret Roberts, Co-Chair

Trying to study environmental politics is like trying to juggle flaming bowling pins while riding a unicycle–it requires the acrobat to be good enough at both juggling (political science) and unicycle riding (environmental science) that they do not immediately crash and burn. My attempt to do so is below. In my dissertation I take three different approaches to answering questions which can contribute to our understanding of the interface between the political forces and the environment. In the first I explore electoral deforestation cycles, where deforestation rates are higher surrounding elections in young democracies.These cycles are most pronounced when the elections are highly competitive, occur in young or weakly institutionalized democracies, and are held in majoritarian systems where politicians can effectively target voters with geographic policies. Here, a set of rules designed to expand political

power have the unintended consequence of increased environmental destruction. In the second paper we develop a method for discovering and testing influential concepts and phrases in text. We adapt a neural network with recurrent and convolutional layers designed to make the network's decisions more interpretable to a different task—to identify phrases and concepts which are highly persuasive to a reader. We then apply this to climate change communication to try to uncover some of the most persuasive concepts both for and against climate change mitigation. In it I evaluate the effects of receiving a formal land title on the behavior of plot owners—usually smallholder farmers—using satellite imagery and machine learning. This case speaks to a substantive question that drives millions of dollars in aid annually—how can we reduce barriers to increasing productivity for the world's poorest and least food secure regions. It also demonstrates new methods which use existing data to more effectively evaluate future (or past!) interventions. Specifically, I evaluate whether having the boundaries of one's plot officially demarcated and recorded and the ability to obtain a legal title increase the probability that part of the plot will be converted from annual to perennial crops, whether there will be cropland expansion in the plot, and whether the built-up area in these plots increases. I also test whether land titling results in land-sparing in surrounding areas by increasing productivity on the intensive margin. To do this I develop a set of methods for using satellite imagery to measure changes in land cover based on sub-annual variations in surface reflectance of different wavelengths of light. This allows me to observe outcomes at an annual scale and detect the proportion of plots under different types of landcover. Evidence from a pilot area of analysis shows increased conversion from natural forest to cropland as a result of land formalization.

Chapter 1

Introduction: The Political Economy of Natural Resources and the Environment

I study the political economy of environmental stewardship: how do citizens and politicians value the preservation or exploitation of natural resources, and how are those values converted into actions and policies? My research combines deep understanding of environmental processes and governance with methods from computer vision and natural language processing that facilitate use of novel imagery and text for causal inference around these questions. I firmly believe that environmental politics research is interdisciplinary by nature and that understanding environmental processes is imperative to understanding the politics of those processes. So far my research has been focused on three main projects: 1) Do elections change the value that incumbent politicians place on forest preservation? 2) Do formal property rights change how landowners value different land uses? 3) What concepts or phrases are most persuasive when discussing enacting climate policy? I also develop methodological tools to answer these questions. I develop methods for integrating data sources like satellite imagery and text into a causal inference framework and I show how we can we make use of learned representations from machine learning algorithms to understand causal processes of environmental politics.

In the first paper in my dissertation Chapter 2 titled “Democratization, Elections and Public Goods: the Evidence from Deforestation” I find evidence of “electoral deforestation cycles” where on average deforestation rates in election years are higher than those in non-election years. These cycles are most pronounced when the elections are highly competitive, occur in young or weakly institutionalized democracies, and are held in majoritarian systems where politicians can effectively target voters with geographic policies. I use satellite measured global gridded forest cover data from 1982 to 2015, results from all national-level elections held during that time period, and control variables including economic and demographic variables. The research builds upon case studies of this dynamic during particular elections in Kenya, Brazil and Indonesia to develop a theoretical framework for when we should expect elections to be associated with higher rates of deforestation. I argue that in non-election years politicians have incentives to preserve forests because of the flow of goods that they provide through ecosystem services and because they will preserve forests to be "used" when they have the highest political value. During an election year when the politician could plausibly lose, the future value of the flow of ecosystem services is greatly reduced. Meanwhile the politician can maximize the political credit they get from allocating access to forested land to people who are influential in the election—farmers in important districts or firms with deep pockets, among others. This means politicians have much greater incentives to allow the exploitation of forests during competitive election years than during non-election years. Furthermore, these effects are likely to be most pronounced in settings where the politician is unlikely to be held accountable for their actions (weakly institutionalized democracies) and where they can identify and target key voters (majoritarian systems).

The second paper in my dissertation Chapter 3 is titled "The Effects of Land Tenure Security on Agricultural and Environmental Outcomes in Benin: New Evidence from Satellite Data" which received the best poster award at Political Methodology annual conference and Outstanding Student Presentation Award at American Geophysical Union annual conference. In it I evaluate the effects of receiving a formal land title on the behavior of plot owners—usually smallholder farmers—using satellite imagery and machine learning. This case speaks to a substantive question which drives millions of dollars in aid annually—how can we reduce barriers to increasing productivity for the worlds poorest and least food secure regions. It also demonstrates new methods which use existing data to more effectively evaluate future (or past!) interventions. Specifically, I evaluate whether having the boundaries of one's plot officially demarcated and recorded and the ability to obtain a legal title increase the probability that part of the plot will be converted from annual to perennial crops, whether there will be cropland expansion in the plot, and whether the built up area in these plots increases. I also test whether land titling results in land-sparing in surrounding areas by increasing productivity on the intensive margin. To do this I develop a set of methods for using satellite imagery to measure changes in land cover based on sub-annual variations in surface reflectance of different wavelengths of light. This allows us to observe outcomes at an annual scale and detect the proportion of plots under different types of landcover.

The statistical contribution comes from the use of a double machine learning estimator to use pre-treatment satellite imagery to adjust for differences between titled and untitled land which are only visible through characteristics of land cover. I compare results from two machine learning methods that I developed specifically for this task: a random forest classifier combined with statistical time-series methods and a deep convolutional neural network imported from natural language processing techniques. At the moment the results are limited to one department of Benin but they show that plots which were demarcated for titling showed had a higher conversion of forest to cropland and also a higher rate of conversion to perennial crops. The paper's new methods pave the way for many retrospective evaluations of similar interventions, especially in places where treatment was not randomized or failed to include important physical or environmental factors. I am in the process of developing methods which can incorporate high-resolution imagery to improve the model's ability to learn different characteristics of built up areas.

In a third coauthored Chapter 4 paper titled "Discovery of Influential Text in Experiments Using Deep Neural Networks" we develop a method for discovering and testing influential concepts and phrases in text. We adapt a neural network with recurrent and convolutional layers designed to make the network's decisions more interpretable to a different task—to identify phrases and concepts which are highly persuasive to a reader. We developed methods to extract phrases from the intermediate layers to facilitate the concepts in the text which were most influential. We validate the method on a conjoint experiment where respondents were asked about the desirability of immigrants with different characteristics, showing that if one didn't know what the treatments were in the conjoint experiment we can still extract the most influential treatments and their approximate effects. We then apply the model to climate change communication.

In summary, the work in this dissertation advances our understanding of environmental politics by building new theories of the political economy of natural resource use and developing new methodological tools to help test those theories. I hope that by taking a multi-faceted approach informed by fields outside of political science I can edge the field a step closer to answering some of the difficult but important questions which lurk on the horizon. I also hope that this research can be used to improve the lives of the millions who depend on ecosystem services for their well-being or who are vulnerable to climate change and environmental destruction.

Chapter 2

Democratization, Elections, and Public Goods: The Evidence from Deforestation

2.1 Abstract

This paper shows that over the last three decades competitive elections were associated with increased deforestation. Protection of forested areas provides long-term, public goods while their destruction provides short-term, private goods for particular voters. Politicians facing a competitive election offer voters access to forested areas mainly for small-scale farming or commercial use of timber in exchange for electoral support. I test this theory of political deforestation using satellite generated global forest cover data and the results of over 1000 national-level elections between 1982 and 2016. I find that countries which undergo a democratic transition lose an additional .8 percentage points of their forest cover each year, that years with close elections have over 1 percentage point per year higher forest cover loss compared to non-election years, and that as the margin of victory in an election decreases by 10 points the amount of deforestation increases by .7 percentage points per year. These increases are on the order of five to ten times the average rate of forest loss globally. This suggests democratization is associated with under-provision of environmental public goods and contested elections are partially responsible for this under-provision.

Replication Materials: The data, code, and any additional materials required to replicate all analyses in this article are available on the American Journal of Political Science Dataverse within the Harvard Dataverse Network, at: https://doi.org/10.7910/DVN/EF7R0Z.

2.2 Introduction

"If an election were held every year, there would be no forest left."

–High level Kenyan official, December 1998[1]

Leading up to Kenya's first competitive election in 1992, President Moi signed a series of excisions granting key voters access to protected forested areas [3]. The World Resource Institute noted, "Recent forest loss has resulted from government approved, politically motivated, and dubiously legal excisions of forest land from protected areas, reserves, and plantations" [1]. Decades earlier, [4] wrote "securing the backing of the Mourides became more urgent with the advent of self-government in Senegal... the government of Senegal curried favor with the Marabouts by giving them privileged access to publicly subsidized inputs: fertilizers, mechanical equipment, *land carved out from forest reserves*..." (emphasis added). Deforestation rates increase 8-10% in mayoral election years in Brazil [5, 6]. How widespread is political deforestation, and what are the common mechanisms that underlie these electoral deforestation cycles?

Deforestation is one of the most important environmental issues of our time. When forests are cleared most of the carbon in their biomass is released into the atmosphere, accounting for over one third of all greenhouse gas emissions [7]. Preventing deforestation is one of the most cost-effective climate change mitigation measures [8]. Deforestation is also the leading cause of habitat loss and species extinction and is associated with higher aridity, increased soil erosion, and lower water quality. Globally, only 6.2 million square kilometers of the preindustrial 16 million square kilometers of forest remain [9], nearly 90% of which is on publicly owned land. Recent estimates point to a slowing rate of deforestation but continued tropical deforestation remains an important problem [10]. Because nearly 90% of remaining forests are publicly owned, deforestation is a political problem.

Some argue that democratic governments tend to provide more public goods than autocratic

[1]The original quote appears in [1] and later appears in [2].

governments, the provision of which improves the lives of those who democracies enfranchise [11, 12]. Others suggest that competitive elections in weakly institutionalized democracies incentivize politicians to forgo public goods provision and instead target electorally influential people with private goods (for example, [13, 14]). This paper adds nuance to the research on democratic governance and environmental protection [15, 16] by considering the effects of competitive elections on The provision of environmental public goods through forest protection.

I demonstrate that democratic transitions and closely contested elections in weakly institutionalized democracies result in deforestation. I start with a trade-off politicians face: provide short-term, private goods offered by cutting down forests or provide long-term, public goods offered by forest preservation. When a politician faces a more competitive election the short-term electoral advantage they gain from giving key voters access to forested land outweighs the long-term support a politician gains by preserving forests. This results in increased rates of deforestation during competitive elections–as observed in Kenya [3] or Brazil [6].

I test this theory globally using satellite derived data on deforestation from 1982 to 2016 [17]. I combine this with national level electoral data, and economic and demographic covariates. Across all countries with any forest from 1982-2016, forests in countries which undergo a democratic transition have higher rates of forest loss after the transition than before, controlling for changes in population and economic growth. Election years have higher rates of forest loss than non-election years in weakly institutionalized democracies. Elections with smaller margins of victory are associated with a higher rate of forest cover loss compared to elections with larger margins of victory in these countries. These tests eliminate many alternative mechanisms including economic growth, population changes, time-invariant characteristics of a location (such as topography or agricultural suitability), and year-to-year changes that affect all cells similarly (such as global commodity prices).

In generalizing across such a wide array of countries over such a long time this paper is limited in its ability to isolate specific causal mechanisms. Increased deforestation in an election year can occur as a result of agricultural expansion, logging, mining, infrastructure expansion, or others. However, the decision to allow destruction of forests so that other land uses can occur is one which is most likely to occur in weak democracies in an election year–something which is both important and merits further investigation.

This paper provides the first cross-national, longitudinal study of the link between elections and deforestation using data generated from satellite imagery. I show that forests are a resource politicians can use for political gain, and that electoral competition may lead politicians to prioritize short-term gain at the expense of longer-term environmental public goods provision. This means that competitive elections themselves, a foundational component of democracy, provide perverse incentives to cause long-term environmental damage. This runs counter to the common conception of democracy as protective of natural capital.

2.3 Democracy and deforestation

It is a stylized fact that democratic governments improve public goods provision resulting in welfare gains for the majority of the population [11, 12], and that the mechanism that drives this relationship is competitive elections [18]. This line of reasoning extends into the research on democracy and the environment. Developed democracies tend to have higher levels of environmental protection and lower levels of pollution than non-democracies [19] because of their tendencies to provide public goods [16, 20, 21] and reduce corruption [8, 22].

Other research clarifies when democracy reduces public goods provision and environmental protection. Often in young or developing democracies an introduction or increase in political competition can decrease the availability of public goods [14, 23], increase patronage politics [13, 24], and hurt the poor [25]. Others have found that democracies are associated with worse environmental outcomes [26] and specifically deforestation [20, 27] though these findings lack theoretical motivation.

Several case studies examine the link between electoral incentives and deforestation. [2] argues that the destruction of several forest reserves in Kenya can be attributed to increased demand for patronage in pre-election periods. Also set in Kenya, [3] argues that the introduction of multiparty elections in 1991 led to targeted excisions of protected forested land in areas pivotal for the election and deforestation rates increased thereafter. [6] finds that in Brazilian municipalities where mayors run for re-election deforestation rates are 8-10% higher than in non-election years. [28] discover “political logging cycles” in Indonesia, where deforestation rates increase during competitive elections. These studies join brief observations by other authors that in competitive elections, politicians use protected forested areas as a bargaining chip to win the support of key voters [4, 29, 30].

Natural sciences research identifies the mechanisms by which most deforestation occurs. The most common land use transition over the last fifty years has been from forest to agricultural land and pasture. Over the last thirty years, more than 80% of new agricultural land was previously forest [8]. Economic growth causes deforestation through logging as well as infrastructure expansion and mining [31, 32].

Drawing on theories of electoral competition and public goods provision, case studies, and the natural sciences literature on causes of deforestation, this paper systematically develops and tests a theory of the link between deforestation and elections across countries. It contributes a general theoretical model for when and where electoral deforestation cycles should be strongest. Methodologically and empirically, this study adds higher quality data and panel methods to the debate on democracy and environment, and performs the first cross-national tests of electoral deforestation cycles.

2.4 Trading the forest for the trees?

Kenya: A motivating case

Figure 1 shows the Mau Forest Reserve, an area of government-owned protected forest, more than half of which has been converted into smallholder farms. The map on the left shows that the formally protected (darker) area falls into three counties: Nakuru, Narok, and Bomet. The satellite image on the right shows (light colored) cropland areas and (dark colored) forested areas. Nakuru county is an electorally competitive county with a population of over one million divided among the major Kenyan tribes. Narok county is a primarily Maasai county that consistently voted for the opposition to the incumbent Kenya African National Union (KANU) party by a large margin in the 1990s and early 2000s. Bomet county has consistently voted for KANU by a wide margin. Most of the forest preserve in Nakuru county has been converted from forest to cropland (with the largest losses occurring around elections) while the majority of the preserve in Narok and Bomet county remains standing. In the Mau Forest Reserve the important difference was political: Nakuru was pivotal for control of the national legislature while Narok and Bomet's representatives were all but guaranteed to represent the opposition and incumbent parties respectively.

President Moi and the KANU party distributed patronage in the form of explicit and de jure land grants to voters in pivotal counties to maintain political control. [2] describes two possible benefits the ruling party obtained through these land-grants: sell the timber to finance re-election campaigns or distribute the land to potential supporters in exchange for their electoral support.

This and other anecdotes describe how deforestation can result from electoral competition at the district level. If this is generally true then years with competitive elections should have higher rates of deforestation at the national level, even if that deforestation is concentrated in competitive districts. I generalize the theoretical mechanism for how and when politicians exchange trees for votes and then consider the implications on a *cross-national level* using the overall competitiveness of national level elections rather than district-level competitiveness. This allows observation of global patterns across the many countries and years for which sub-national electoral returns are not available.

The value of forests

Protected[2] forests are valuable to voters for several reasons. First, when left undisturbed, forests provide ecosystem services to surrounding areas [33]. They host pollinators that are essential to seed production and predators that control pest populations. Additionally, forests reduce air pollution, decreasing respiratory and cardiac illnesses. They act as natural filters that purify water and help to recharge groundwater basins that are important for agriculture. Forests mitigate floods and droughts by preventing large fluctuations in the flow of rivers while preventing erosion and sediment loading that can make water more difficult to consume and shortens the lifespan of dams [34]. Finally, they attract tourists and bring foreign spending. Most of these benefits accrue to populations beyond those that are adjacent to the forest, and fall somewhere on the spectrum of positive externalities (sediment reduction) to pure public goods (CO2 emissions reduction) [35]. These benefits accrue slowly, for example flood mitigation would not be apparent except in high-runoff events, and the effects of air quality on health can be latent for tens of years. However there is growing evidence that negative environmental effects on voters' livelihoods can reduce support for an incumbent politician [36]. As a result, politicians who protect forests may receive some additional electoral support from those who benefit from the public goods that forests provide.

[2]Protected means requiring government authorization for use.

Forested land is also valuable through the sale of timber or the potential to use the land for crops or other commercial purposes. The timber itself has value for firms which benefit from decreased protections for forests [5, 28, 37]. More commonly or in conjunction with the above is commodity driven deforestation where the value of removing forests comes from what replaces the trees [3]. Forested land is high in nutrients like nitrogen and phosphorous and is extremely productive when converted to agriculture [38]. Rather than providing value over time, the value associated with cutting down forests is immediately realized and clearly attributable to the politician who provided it. Furthermore, the stored value of forested land accrues directly to the firm who is able to log the region or to the people who gain access to agricultural land. This choice over the distribution of value from forests mirrors the choice faced by politicians in electoral business cycles where politicians are more likely to increase spending in competitive districts and on projects for which politicians can easily claim credit [39–41].

The mechanism of granting access to forests generally takes one of two forms: use permits or property rights. Both of these mechanisms vary in their formality–they range from being transparent and formal to hidden and informal. Use permits grant firms the right to log or mine an area of land, as in Brazil and Indonesia [5, 28]. The politician can target a firm, which can provide jobs or economic growth to a particular area, or can contribute additional money to the politician either through higher tax revenue or political donations. Politicians can achieve a similar outcome by reducing protections or even reducing enforcement of protections for forests; this strategy has lower target-ability but is less visible to the general public. In either situation, continued use is often contingent on the re-election of the politician who provided permits or reduced protection [42]. Property rights grant farmers the ability to clear forests and plant crops or graze livestock with the understanding that those who benefit will vote for the politician who provided those rights. The Kenya case is instructive for how this transaction occurs. Politicians can either target a particular forested with reduced protection, benefiting the people who live nearby and can expand onto the de-protected land or by granting property rights to a group of people whose votes the politician wants to secure. Property rights may be reversible should the incumbent who provided the property lose the subsequent election.

Forests differ from other classes of goods governments provide such as roads, clinics, and schools [25, 43]. Forests take decades to regenerate and are thus a one-off opportunity in a political lifetime.

Additionally the allocation of forested areas does not require government spending that trades off with other projects. The exploitation of forests in the present only trades off with either their future exploitation or the future public goods that they could provide. Even if the government absolutely discounted the diffuse public goods that forested areas provide, it might choose to preserve some forested areas for future use and smooth its consumption of forested areas [44]. The implication is that even if officials place little or no value on the public goods forests provide, they should tend to preserve forested areas until the present need for the goods exploitation of forests provides is greater than the expected future need for those goods. In other words, politicians should only grant access when they need to provide short-term benefits to an important group of constituents or when they are afraid they might lose the ability to grant access.

Political incentives

Seeking to stay in power, politicians possess two strategies with respect to forested areas: one is to allocate some access to publicly owned forests to the constituents on whose support they rely. The other is to protect forests and rely on the public goods protected forests provide to convince constituents that they will be better off if the politician stays in power. A politician must distribute benefits in such a way that they generate enough support to stay in power.

How will a politician use limited forest resources to maximize their chances of staying in power? Most research focuses on how politicians supply goods to potential supporters. In an autocracy where a politician must please a small winning coalition, providing private goods tends to be more efficient than providing public goods [12], and we expect politicians to allocate more access to public forests [15, 16, 22]. In a democracy where the winning coalition is large, providing public goods is more efficient at generating support, and politicians can be expected to preserve forest at a higher rate [11, 45, 46].

However, there are two demand-side reasons deforestation rates increase when a country transitions from autocracy to democracy: the relative political empowerment of those who demand land and the shortened political time horizons that come with regular elections. In a new democracy the selectorate includes recently enfranchised small-holder farmers (or others who may benefit from deforestation) for whom forested land is an extremely valuable resource [3, 29]. In an autocracy, the selectorate tends to consist of a small wealthy group of industrialists who do not have incentives to quickly deforest[3] [47–50]. Note that this can occur in countries where agriculture plays a small role in the economy–all that is required is the combination of newly enfranchised forest consumers. When a democratic transition occurs, the political value of removing protections for forested land increases.

Weakly institutionalized democracies are more clientelistic because parties are weaker and their promises are less credible in the mind of voters [51] and lower public goods provision [14]. The introduction of additional electoral competition can exacerbate clientelism [13] by increasing the stakes of clientelist relationships or by increasing the demand for clientelist goods [24]. Thus, after democratic transitions and during close elections targeted forest allocation should be more likely, especially in weakly institutionalized democracies.

I focus first on countries which experience a democratic transition. This allows me to isolate the relationship between political incentives and deforestation rates in the two different systems while holding other conditions relatively constant. Based on the empowerment of farmers who demand cropland and the introduction of elections which emphasize short-term political gains, I hypothesize that:

Hypothesis 1. Countries which transition regime type have higher rates of deforestation under democratic government.

There are further observable implications for election and non-election years. Politicians have a shortened time horizon ahead of a competitive election because they may not be reelected. Short horizons reduce the value of the long-run goods forests provide and make the short-run benefits of granting access to the land more appealing. Should the politician lose re-election, the long-term goods protected forests provide are worthless to them, rendering the immediate increase in political support from immediate allocation even more valuable in comparison. Additionally, if the politician is able to identify pivotal voters, the efficiency of granting access to those voters likely exceeds the efficiency of providing public goods. However, this comes at the expense of the increase in support generated by protecting forested lands until they are allocated at some point in the future and the benefit the politician might get in a future election by allocating those goods. Because of this, a politician should generally only choose to allocate

[3]This is for two reasons. The value for industrialists tends to come from selling timber products rather than planting crops meaning that they would smooth their consumption of forest over time. Second, as the number of plausible consumers of forest increases, the incentive to deforest now becomes stronger [44]

forested land when they feel truly threatened, and only if institutions are weak enough that the politician will face minimal backlash for these allocations. Given that autocracies rarely, if ever, have competitive and meaningful elections and institutionalized democracies have mechanisms to punish politicians for clientelist behavior, the following two hypotheses apply primarily to weakly institutionalized democracies.

Hypothesis 2. Election years will have higher rates of deforestation than non-election years.

Hypothesis 3. Years with competitive elections will have higher rates of deforestation than years with non-competitive elections.

Finally, all of the hypotheses should be strongest in places where politicians can observe and target key constituents. In a single-district proportional representation system policies which by their nature target certain geographic areas are less likely to be useful than in majoritarian systems with low district magnitude (like Kenya) where politicians can identify pivotal districts [52, 53].

Hypothesis 4. Majoritarian systems will amplify the effects of democratic transitions and elections on deforestation compared to non-majoritarian systems.

2.5 Empirical Strategies

Data

The dependent variable for this study is the percentage point change in primary forest cover in a $.05^o \times .05^o$ cell of land in one year. The total area of a cell is $30.25km^2$ near the equator, but as small as $8.90km^2$ near the poles. Forest is characterized by the presence of vegetation with a canopy over 5 meters tall. The data used to construct this measure is from [17] which uses data from Advanced Very High Resolution Radiometer instruments to measure vegetation cover over the globe on an annual basis. This type of data is remarkable for a few reasons: the coverage is global, the method is accurate, and the data are not susceptible to interference from parties that seek to conceal or misrepresent information. I use data from 1982 to 2016, the full extent of the available data [10].

I extract the percent land-cover of forest of each cell in each year, resulting in 34 years of global forest cover data, and calculate the year-on-year differences for the dependent variable. The dependent variable exhibits a unit root in levels, which suggests taking a first difference will produce more consistent results than including a lagged dependent variable. Alternately, the dependent variable of interest is the rate of forest cover change rather than the level of forest cover. A value of -1 for a cell indicates a one percentage point loss of forest.

I merge national boundaries with this data, so each observation contains a unit level measure of forest cover and a set of national-level independent variables. I exclude all observations that never have forest cover from 1982-2016 because such places are never eligible to lose forest. Many areas gain forest, particularly in China, Russia, and Canada where large scale tree planting or climatic changes have resulted in more forests. Forest increases are included in the data but are difficult to link to a political event because the different growth rates of various species of trees mean that new trees may take many years to appear in the data. However, a slower than normal rate of gain in a particular year could indicate forest loss in some parts of the cell.

Right-hand side variables come from several sources and are merged with forest cover data by country-year. This analysis uses a dichotomous indicator of democracy from [54] for the democratic transition test. They define a minimum threshold for both contestation and participation to determine whether a country is a democracy or not in a given year. Data on election years and votes come from the Database of Political Institutions (DPI) [55] and the V-Dem project [56]. The variable **election year** takes a value of 1 if a national-level legislative election occurred in that country in a given year and 0 otherwise.[4]

[57] note that ideally researchers should use past vote swings and seats-votes elasticities to calculate electoral risk, but even this is complicated by implicit assumptions about how effort maps to votes [58]. This is further complicated by two features of this study: seat-vote elasticities are not available for most of the countries considered, and using previous vote swings requires 6 previous elections with relatively stable parties. Because the focus of this paper is young democracies, previous elections both are not available and do not provide good information on which to base an expectation of the competitiveness of the current election. Instead, in the spirit of [57] I use the percent seat difference between the two largest parties. This most clearly captures the margin by which the largest party holds the prime minister position in a parliamentary system or the presidency in a two-party democracy as measured by [56]. Alternately, I use

[4] I use the V-Dem measure in the paper but the results are robust to DPI

two measures derived from [55]: the difference between the incumbent coalition's vote proportion and 0.5; and the difference between the incumbent coalition's seat proportion and 0.5. For interpretability I transform these variables so a value of 100 represents a tie election and a value of 0 represents one party garnering 100% of the vote or seats[5]. I use the Polity IV data to divide countries into "autocracies" (polityIV < -5), "anocracies" (polityIV $\leq$ 5 and $\geq$ 5) and "democracies" (polityIV > 5) [59]. Anocracies are the weakly institutionalized democracies for which I expect electoral deforestation cycles to be most pronounced.

I create a variable which is 1 if proportional representation is used in national legislative elections and 0 otherwise [55]. This allows a test of Hypothesis 4 by isolating majoritarian systems where geographic targeting is more feasible.

In each regression I include the following controls from the World Bank World Development Indicators: per capita GDP (thousands of US Dollars), change in per capita GDP (% change), and change in population (% change) [60]. Each is lagged by one year to prevent the inclusion of post-treatment controls. This means a variable such as per capita GDP is included from time $t-1$ and change in per capita GDP is included as the change from time $t-2$ to time $t-1$. I also include a control for the amount of forest remaining in a cell at the start of the year because I expect deforestation rates might be higher in places that are partially forested than places that have 100% forest cover. Supporting Information Tables A.1 and A.3 and Figure A.1 (pages 3-8) present specifications which include percent of the population employed in agriculture and agriculture as a percent of the GDP.

I include unit and year fixed effects. The unit fixed effect absorbs any time-invariant characteristics at the unit level, including location, country, elevation, average climate, soil type, etc. It also de-means the forest cover loss variable, considering only deviations from the average forest cover loss in each cell. Year fixed-effects absorb global-level changes specific to a single year, like food, lumber or fuel prices. The remaining variation is composed of deviations from each observation's average forest cover loss that are also deviations from the global average forest cover loss in that year. Because election shocks should appear only in cell-years that experience an election, this specification should control for most

[5]Formally: margin $= 100 - |\%\ \text{votes}_i - \%\ \text{votes}_j|$ where i and j are the two parties with the most seats, or margin $= |50 - \%\ \text{votes or seats}_i| \times 2$ where i is the incumbent coalition.

Table 2.1: Observations in different levels of aggregation

Number of forested cell-years	157,586,802
Number of forested cells	4,397,228
Number of countries with forested area	162
Number of country-Years	5,665
Number of elections	1,244

variables that are associated with both election years and deforestation. It should also control for most of the non-political drivers of deforestation including economic and population growth. Simply, the variation I explain is: changes in forest cover that are not associated with development, economic growth, population growth, size of the agricultural sector, growth in the agricultural sector, and changes idiosyncratic to a particular location or year.

I cluster standard errors at the country and year level to account for correlation in residuals between cells in the same country, possibly over many years, and to account for correlation in residuals between distant cells in the same year. This reduces the effective number of observations to a number much closer to the number of country-years (thousands) instead of cell-years (tens of millions) (Table 2.1). A second set of regressions reported below aggregates forested cells to the country level, generating a dependent variable which is the average change in forest cover among forested cells in a country in a year. Supporting Information Tables B.1 to B.3 (pages 8-10) aggregate to cells which are 100 times larger than those described above, and level 1 and 2 sub-national administrative units. The country-level specification targets the question "for a country experiencing a democratic transition or election, what is the national expected rate of forest loss for forested areas" while the cell-level specification targets the question "for a patch of forest, what is the expected rate of forest loss during a transition or election." The cell level has a fixed geographic area as the unit of analysis, so it tends to upweight countries with lots of forest. The country level tends to upweight the effects on forests in countries without much forest.

Summary Statistics

There are approximately 4.4 million cells which ever have forest, 136 countries which have some forest and appear in them sample, and 1,146 elections. Those elections predominantly occur in strong

Table 2.2: Elections by Regime and Electoral type

Regime Type	System	Elections	Margin <20	Margin <10
Anocracy	PR	92	24	5
Anocracy	Majoritarian	105	15	17
Autocracy	PR	28	2	2
Autocracy	Majoritarian	67	12	5
Democracy	PR	144	64	36
Democracy	Majoritarian	544	340	220

democracies, but 197 of them are in anocracies and 95 in autocracies. Only 22 elections in anocracies have a margin of victory of less than 10 percentage points.

Test 1: Democratic Transitions

First, I test whether cells in democratic countries that experience regime type transitions transitions have higher rates of deforestation than cells in non-democratic countries which experience such a transition. The main independent variable is whether a country is a democracy, where democracies are coded 1 and non-democracies coded 0. The dependent variable is percentage point change in forest cover for a cell in a year. The main specification uses unit and year fixed effects which project out time-invariant characteristics of each cell (and thus country):

$$\text{ForestChange}_{i,c,t} = \alpha_i + \gamma_t + \beta_1 * \text{Democracy}_{c,t} + \lambda * X_{c,t} + \delta * X_{i,t} + u_{i,c,t} \quad (2.1)$$

β_1 represents the within-country difference between years when a country was a democracy (according to [54]) and years when that country was not. α and γ are cell and year fixed effects, $X_{c,t}$ is a vector of country-level controls, $X_{i,t}$ is a vector of cell-level controls, and $u_{i,c,t}$ is the unexplained variation, clustered at the country and year level.

This regression targets whether a patch of forest was more likely to lose forest under a democratic or non-democratic regime. It gives each equal-sized area the same weight, focusing on the total amount of deforestation. I focus on these cell-level regressions, however, I also present the country-level results in Figure 2.1 and in Supporting Information Figures A.1 - A.3 (pages 3-7).

Table 2.3: Regressions of forest change on democracy. 61 countries experienced a regime-type transition. Columns 2 and 4 replicate the non-fixed effects specifications used in previous work

	Cell	Cell	National	National
Democracy	-1.10^{*}	-0.35	-0.25	0.04
	(0.44)	(0.39)	(0.29)	(0.10)
Forest	-0.77^{***}	-0.05^{***}	-0.56^{***}	-0.01
	(0.03)	(0.01)	(0.06)	(0.01)
PCGDP	0.08	0.00	0.11^{**}	0.00
	(0.05)	(0.01)	(0.03)	(0.01)
Δ PCGDP	-7.86	10.26	-9.33	26.70
	(20.33)	(28.39)	(10.40)	(16.96)
Pop Growth	-0.13	-0.01	-0.08	-0.06
	(0.27)	(0.12)	(0.08)	(0.06)
Constant		1.78^{***}		0.46
		(0.31)		(0.27)
FE	cell + year	none	country + year	none
Num. obs.	136743524	136743524	4375	4375
Adj. R^2 (full model)	0.38	0.02	0.34	0.01
Adj. R^2 (proj model)	0.37	0.02	0.26	0.01

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Hypothesis 1 is countries which transition regime type have higher rates of deforestation under democratic government.

Table 2.3 shows that among countries that experience at least some years as a democracy and a non-democracy, forested areas have about 1 percentage point greater forest cover loss during democratic years compared to the nondemocratic years. Columns 2 and 4 show that without including fixed effects this relationship is not detectable because it compares deforestation rates *between* fundamentally different countries rather than within a single country. This change in forest cover is estimated to be negative at a $\alpha = 0.05$ confidence level in the cell-level specification. The rate of forest cover loss associated with democracy is higher than the global average has been in any year since 1982. Democratization is responsible for approximately nine million square kilometers of forest loss, or an area roughly the size of

Brazil. This represents approximately 9% of the total forest cover as measured in 1982 lost in countries which have experienced years as a democracy and non-democracy. Furthermore, this estimate is *after* the main structural economic and demographic drivers of forest cover loss have been taken into account.

Hypothesis 4 posits that the effect of democratization will be stronger in majoritarian systems. To test this, I run the following regression:

$$\text{ForestChange}_{i,c,t} = \alpha_i + \gamma_t + \beta_1 * \text{Democracy}_{c,t} + \beta_2 * PR_{c,t} + \\ \beta_3 * (\text{Democracy}_{c,t} * \text{PR}_{c,t}) + \lambda * X_{c,t} + \delta * X_{i,t} + u_{i,c,t} \quad (2.2)$$

The β_1 coefficient can be interpreted as the the relationship between a democratic transition and forest cover change for majoritarian systems. Figure 2.1 shows the democracy coefficient from the original specification and the coefficient from this specification. When a country transitions to a majoritarian system the forest cover loss associated with this transition is consistently large, negative and significant across aggregation levels.

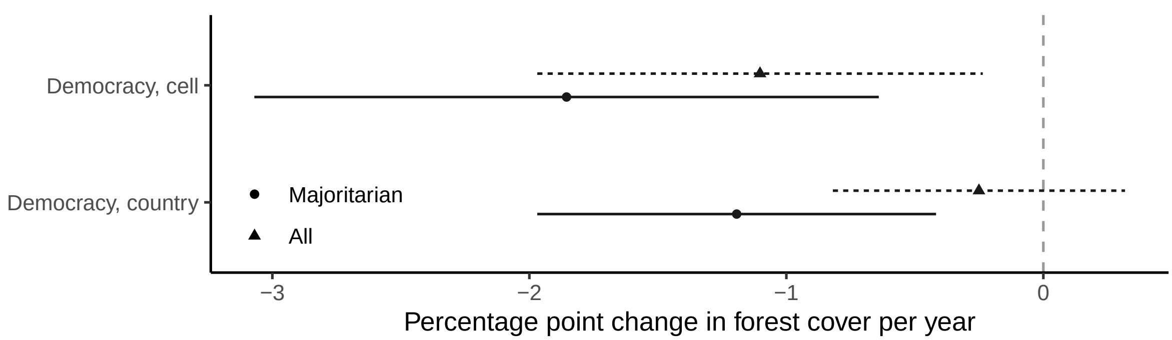


Figure 2.1: Forest cover change across electoral systems

Coefficients for the regression in table 2.3. Dashed triangles correspond to columns 1 and 3, black circles are the same regressions for transitions to democracies with majoritarian systems.

When the data are aggregated to a national level the effect size is smaller and not statistically distinguishable from 0. This points to the possibility that the effect is more pronounced in countries with more forested cells or that the effects are localized to hot-spots which carry little weight in the national-level regressions. To further investigate I run geographically weighted regressions and regressions at different levels of spatial aggregation, including aggregating cells to 55km $\times$ 55km, and aggregating to first and second level sub-national units. As [61] point out, identifying the 'correct' unit of analysis can be difficult and lead to different results, so results from various levels of aggregation can be found in Tables B.1 - B.3 (pages 8-10) of the Supporting Information. When focusing on majoritarian systems the effect of a regime type transition is larger and significant for all levels of aggregation, suggesting that geographic targeting is for the political and ecological versions of this question.

Where previous results attempt to answer the question of whether democracies deforest more than non-democracies, this paper poses a more tractable question: how do rates of forest cover loss change when a country switches between being a non-democracy to being a democracy? This refrains from comparisons between vastly different countries and instead relies on variation within countries over time. Second, by including data at the cell area it allows me to ask a modified version of the question–for an area of forested land, what is the likely result of having the country in which it is located change regime types. This focuses on an effect which is substantively important: in the large countries where most of the remaining forest in the world resides, what were the consequences of democratic transitions? The evidence shown here is that democratic transitions are associated with higher rates of deforestation, especially when the transition is to an electoral system with geographic targeting.

Test 2: Election years

Competitive elections create a unique set of incentives for politicians to allocate more forested land to voters than they do in non-election years. I expect this to be strongest in weakly institutionalized democracies and weak in both autocracies (where leaders' positions of power are not contingent on elections) and strong democracies (where institutions can prevent clientelism). A blunt test of Hypothesis 2 considers forest cover loss in all national-level election years across the government-type trichotomy and compares it to forest cover loss in non-election years. Because business cycles are known to be connected with elections and could drive deforestation, I control for change in per capita GDP from $t-2$ to $t-1$. While this estimation strategy cannot rule out the possibility of some unobserved confounder, such a confounder would have to cause elections and increase forest loss in many countries over the course of multiple elections. Unit and year fixed effects prevent unit, country, or year-specific characteristics from confounding the estimates:

$$\text{ForestChange}_{i,c,t} = \alpha_i + \gamma_t + \beta_1 * \text{Election}_{c,t} + \beta_2 * \text{GovType}_{c,t} + \\ \beta_3(\text{Election}_{c,t} * \text{GovType}_{c,t}) + \lambda * X_{c,t} + \delta * X_{i,t} + u_{i,c,t} \quad (2.3)$$

β_1 represents the within-country difference between election years and non-election years for anocracies. β_2 represents the within-country difference between regime types. β_3 represents the difference between election-year effects for anocracies versus democracies or autocracies. The second and third panels only include elections within the specified margin of victory. α and γ are unit and year fixed effects, $X_{c,t}$ is a vector of country-level controls, $X_{i,t}$ is a vector of cell-level controls, and $u_{i,c,t}$ is the unexplained variation, clustered at the country and year level. The goal is to isolate deviations from each cell or country's average rate of deforestation that cannot be explained by economic or demographic characteristics, and test whether those deviations align with election years and close elections.

In an average election there is no more deforestation than usual. However, as the competitiveness of the election increases so does the expected rate of deforestation, culminating in a two percentage point increase in the rate of deforestation (Over ten times the average rate of forest loss in the Brazilian Amazon). Columns 1-3 of Table 2.4 show results at the cell level. Election years themselves are not significantly associated with forest change, but elections with less than a 20 point and less than a 10 point margin of victory are increasingly associated with deforestation. Columns 4-6 show that when aggregated to a national level these results at not significant. At the cell level the rows 4-6 show that in democracies the effect of elections are counteracted, but row 7 supports the above conclusion that being in a democracy is a net negative for forest cover.

Figure 2.2 shows this relationship across levels of aggregation and electoral system types. Refer to Table 2.1 for the number of elections in each category. The top of Figure 2.2 shows the coefficients from the first three rows of Table 2.4. The bottom half of the figure shows the results from the electoral system regressions–at the cell level the point estimates are smaller but still significant. At the national level focusing on majoritarian systems increases the precision of the estimates where competitive and close

Table 2.4: Regressions of forest change on election year. Election years are subset by how competitive they were: columns 1 is all election years, 2 is years with a margin of victory less than 20 points, 3 is years with a margin of victory less than 10 points. Election year is interacted with regime type with anocracy as the base case.

	Cell	Cell	Cell	National	National	National
Election Year	−0.10			−0.12		
	(0.38)			(0.23)		
Margin < 20		−1.38			−0.39	
		(0.69)			(0.38)	
Margin < 10			-2.17^{**}			−0.37
			(0.70)			(0.35)
Election:Democracy	0.06			0.07		
	(0.37)			(0.22)		
Margin<20:Democracy		1.22			0.19	
		(0.76)			(0.42)	
Margin<10:Democracy			2.08^{**}			0.16
			(0.65)			(0.41)
Democracy	-0.91^{*}	-0.87^{*}	-0.96^{*}	−0.33	−0.38	−0.35
	(0.37)	(0.35)	(0.37)	(0.23)	(0.23)	(0.25)
Autocracy	−0.50	−0.54	−0.56	0.11	0.21	0.21
	(0.60)	(0.63)	(0.63)	(0.38)	(0.37)	(0.37)
Forest	-0.77^{***}	-0.77^{***}	-0.77^{***}	-0.55^{***}	-0.56^{***}	-0.56^{***}
	(0.03)	(0.03)	(0.03)	(0.06)	(0.06)	(0.06)
PCGDP	0.07	0.08	0.09	0.11^{**}	0.11^{**}	0.11^{**}
	(0.06)	(0.05)	(0.05)	(0.03)	(0.03)	(0.03)
Δ PCGDP	35.20	7.91	13.68	−1.48	1.93	3.14
	(37.62)	(29.47)	(31.86)	(10.02)	(10.10)	(10.21)
Pop Growth	−0.14	−0.23	−0.18	−0.06	−0.08	−0.07
	(0.27)	(0.26)	(0.27)	(0.07)	(0.06)	(0.06)
Country + Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Num. obs.	132801614	118463788	110285677	4081	3701	3517
Adj. R^2	0.38	0.38	0.38	0.34	0.35	0.35
Num. groups: Country	136	136	136	136	136	136
Num. groups: year	33	33	33	33	33	33

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

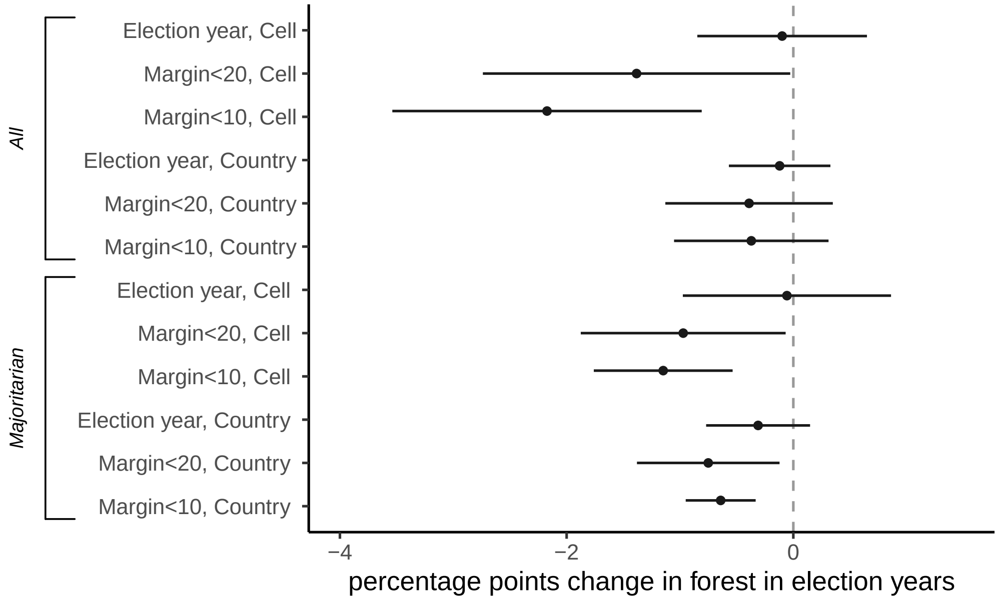


Figure 2.2: Forest cover change across levels of electoral competitiveness, regime type and electoral system

elections both have higher rates of deforestation than non-election years. When aggregated over elections in anocracies I find that close elections are responsible for around *additional* 500,000 square kilometers of deforestation, or larger than the size of the state of California.

In many countries we should expect sub-national variation in effect sizes due to variation in importance or competitiveness across districts. This or specific locations of key constituencies should lead us to expect only some parts of countries to exhibit higher rates of deforestation in election years. This should bias against finding results because it averages across areas where no change in forest cover should be expected.

Supporting Information Figures C.1 - C.4 (pages 11-12) show results of geographically weighted regressions. Notably, the sub-equatorial region which has experienced the most intense deforestation in the last 40 years also has the strongest relationship between elections and forest cover loss, stretching across Brazil, Argentina, Uruguay, DRC, Congo, Angola, Indonesia and up into South Asia. This analysis also shows heterogeneity within countries–something which merits analysis in future work.

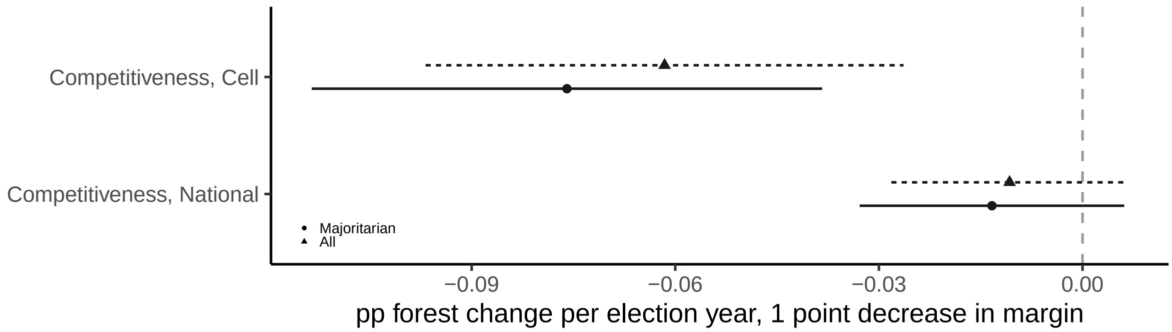


Figure 2.3: Forest cover change across electoral competitiveness, regime type, and electoral system

Test 3: Competitiveness

In this section the sample is restricted to years in which elections occurred, and close elections (in which the margin of victory is low) are compared to elections where one party got a preponderance of the votes. The independent variable measures the competitiveness of an election where 100 corresponds to a tie vote between the two largest parties, 0 corresponds to an election in which one party got 100% of the votes cast. This simplifies the interpretation of the coefficient–as elections get more competitive rates of forest cover loss increase. Once again the main independent variable is interacted with the trichotomized polity variable to isolate the effects in weakly institutionalized democracies. The main test includes unit fixed effects[6], and the same controls:

$$\text{ForestChange}_{i,c,t} = \alpha_i + \beta_1 * \text{Competition}_{c,t} + \beta_2 * \text{GovType}_{c,t} + \beta_3(\text{Competition}_{c,t} * \text{GovType}_{c,t}) + \lambda * X_{c,t} + \delta * X_{i,t} + u_{i,c,t} \quad (2.4)$$

β_1 represents the within-country difference between election years and non-election years for anocracies. β_2 represents the differences between rates of deforestation in election years across regime type estimated at a competitiveness of 0. β_3 represents the difference between margin-of-victory effects for anocracies versus democracies or autocracies. α is unit fixed effects, $X_{c,t}$ is a vector of country-level controls, $X_{i,t}$ is a vector of cell-level controls, and $u_{i,c,t}$ is the unexplained variation, clustered at the country and year level.

[6] but not year FE because demeaning forest cover change only in cells which have an election by year doesn't make sense and loses several years in which there was only one election

Table 2.5: Regressions of forest change on electoral competitiveness. Competitiveness is interacted with government type with Anocracy as the base case.

	Cell	National
Competitiveness	-0.06^{**}	-0.01
	(0.02)	(0.01)
Comp:Democracy	0.06	0.00
	(0.03)	(0.01)
Democracy	-4.53	0.16
	(2.52)	(0.96)
Forest	-0.71^{***}	-0.43^{***}
	(0.03)	(0.07)
PCGDP	0.07	0.14^{**}
	(0.09)	(0.05)
Δ PCGDP	43.02	12.29
	(28.60)	(35.22)
Pop Growth	-0.03	-0.34
	(0.47)	(0.26)
FE	cell	country
Num. obs.	35346878	864
Adj. R^2	0.35	0.22
Num. groups: Country	128	128

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 2.5 shows that at the cell level a one percentage point increase in competitiveness is associated with a 0.06 percentage point decrease in forest cover among election years in anocracies. While the previous set of regressions compared election years to non-election years this regression compares competitive to uncompetitive elections. However, the results are consistent with a ten percentage point increase in competitiveness linked to a .6 percentage point decrease in forest cover, or the difference between a 20 point margin and a ten point margin. While the coefficient on the interaction between democracy and competitiveness is not significant, it exactly cancels out the size of the anocracy competitiveness coefficient, suggesting that the relationship between electoral competitiveness and deforestation is mitigated there. At the national level the effect is not distinguishable from zero, likely for the same reasons cited above–that elections in countries with large forested areas are driving the effect.

Figure 2.3 shows this relationship across cell and national levels of aggregation, and across electoral systems. Here like above subsetting to majoritarian countries only marginally changes the estimate of the effect size, and the two estimates are not significantly different. This test demonstrates that the degree of competitiveness can have a large effect on the deforestation rate. Brazil's average rate of deforestation between 1982 and 2016 was -0.13 percentage points per year which means going from a tied election to a 55-45 split would be expected to increase the deforestation rate by 50%.

A variety of other mechanism and robustness tests can be found in the Supporting Information: Tests of heterogeneous treatment effects by forest type and agricultural influence in (A), different levels of spatial aggregation results (B), geographically weighted regressions for the other tests (C), different measures of electoral competitiveness (D), including neighboring and lagged forest as controls (E), and examining the timing of deforestation with respect to elections (F).

2.6 Implications and Limitations

Implications

A number of influential papers argue that democratic governments are more likely to provide public goods than non-democratic governments. [11] and [62] argue that more contestable political markets decrease the monopoly rents the state can extract from its provision of public goods, implying that the more competitive the election, the more public goods politicians are likely to provide. This paper demonstrates that in young or weakly institutionalized democracies this relationship does not hold; rather than politicians choosing between state rents and public goods, politicians choose between strategies that maximize their chances of victory, sacrificing long-term provision of public goods for short-term transfers of private goods. As a result we should not expect political competiton to increase state production of *environmental public goods*. Instead, political competiton may fuel exploitation of natural resources in a way that is more consistent with [63]'s description of common pool resources. [51] and [23] argue that young democracies are more prone to clientelism and corruption, which reinforces the effect observed in this paper.

[46] and [12] argue that because democratic politicians rely on the support of a larger subset of the population to stay in power, providing public goods is a more efficient way to generate public support than providing private goods. My findings run contrary to Deacon's and Mesquita's theses. First, they do not consider differences in demand for different types of goods across different selectorates. When the newly enfranchised population is largely agrarian, politicians may choose to distribute private goods with a higher marginal utility to that population rather than providing public goods. Second, selectorate models do not incorporate changes in the marginal utility of public or private goods leading up to an election. As the time of an election grows nearer or if geographic targeting is easy politicians can exercise a price-discriminating strategy where they distribute just enough goods to secure pivotal districts. When a politician can do this, the efficiency of providing public goods decreases (because it essentially offers a single price for the vote of a selectorate member) and politicians will choose to offer private goods (forest access) to low-price members of the winning coalition even if doing so reduces the well-being of other constituents. This effect may be amplified in places where the distribution of private goods is highly attributable but the utility provided by ecosystem services is not easily attributable. [2] notes that as the attributability of environmental destruction increased in Kenya, forest for votes exchanges became less common.

In addition to the question of democratic provision of public goods, the findings have implications for how we categorize the goods natural systems like forests provide when they are preserved. The default framework for natural resources in political science work is Common Pool Resources (CPRs) as in [63]. These resources are notoriously hard to preserve because consumers face an N-player prisoners dilemma game where defection from preservation is a strictly dominant strategy for each player [44, 64]. This paper characterizes forests differently: rather than only considering the value forests provide when they are cut down or "consumed," it evaluates the value these forests provide when they are preserved. The ecosystem services outlined above are public goods (non-rival, non-excludable), which changes the way we might think about their preservation. In non-election years, government control of the resources produces an efficient outcome (contrary to what one might expect with a CPR). However, in election years CPR problems begin to crop up. Perhaps forested areas are a class of goods that are best described as "public goods with common pool resource problems."

With this categorization, the CPR literature can offer some insight into why election years have such an effect on forest change. [63] argues that rapid changes in the value of a common pool resource can reduce the ability of any governance system to prevent overuse, but does not consider when political systems themselves might induce this change. As elections approach the value protected forests provide politicians (through the political support they help to generate) undergoes rapid changes. The value to a politician of removing protections and granting access increases relative to the value of preserving that resource, triggering a situation where the governance system (democratic governance) fails.

Limitations

A few limitations exist for these findings. These include the vast heterogeneity among countries and years in the sample, and potential measurement issues for independent variables across such a heterogeneous sample. These limitations generate possibilities for future work including testing the hypotheses here with higher resolution data, exploiting surprise elections, and examining the factors that might mediate the effect including political institutions, the demand for forested land, and the type of forest. The results for democracy only apply to countries which switch regime type, not stable democracies. The other effects are often estimated based on relatively few elections and as a result are more likely to be idiosyncratic to the sample, but this is a fundamental limitation of the data and our political history.

Second, because the vote totals are an outcome of the level of deforestation there may be some reverse causality. However,the rate of deforestation likely only explains a very small part of the variation in electoral competitiveness (compared to vote-buying [30], constructing roads and clinics [43], and agricultural taxation and subsidation [50]), limiting the size of the potential reverse causality bias. Future work could use pre-election polling results to directly measure the effect that deforestation has on voting behavior ([3] uses a similar strategy).

This paper is limited in how it can address specific mechanisms. Using national election returns means that I cannot assess whether politicians target core or swing voters, or whether the effect is driven by particular sectors. It also means that deforestation as a result of electoral business cycles is observationally equivalent to targeted allocation of protected forested land. I also cannot rule out the possibility that firms try to extract timber faster when there is political uncertainty. While I expect these effects to be relatively small compared to agriculturally and pastorally driven electoral deforestation cycles, they are potentially important mechanisms.

2.7 Conclusion

To summarize, I argue that democratic transitions are associated with higher rates of deforestation, competitive election years have higher rates of deforestation than non-election years, and the more competitive an election the higher the rate of forest loss. Politicians choose to allow, induce, or even subsidize deforestation to garner political support when they fear they might not be re-elected. Doing this is costly for the politicians–they give up both the additional support the public goods provided by forests might provide them as well as the ability to allocate that land in the future.

These findings are a first step towards demonstrating that natural resources might not fit neatly into the democracy and public goods provision literature. This is in part because natural resources differ from the "normal" public or private goods politicians offer their constituents in exchange for political support. However, this is also because an electoral mechanism leads to changes in demand for particular types of goods, leading politicians to take actions that do not seem efficient if one only considers the supply of public and private goods. Finally, it shows natural resources that provide environmental services might not fit neatly into a CPR framework, opening possibilities for new lines of research into environmental preservation.

The policy implications of this work are twofold. First, international institutions should note that democratic transitions and especially closely contested elections during a transition pose a threat to forests.

Preventing forest cover loss is one of the most cost-effective methods to combat global warming, and politically motivated deforestation is something a process that international environmental institutions might be uniquely suited to address. Second, this research illuminates behavior by politicians that is inefficient in the long-term as a contributor to deforestation. Recognizing the situations in which democratic elections do not promote public goods provision but rather the provision of goods to a small politically important subset of the population is an important first step towards understanding when democracy fails to live up to its promise.

Chapter 2, in full, has been accepted for publication and is presented as it may appear in "Sanford, L. *Democratization, Elections, and Public Goods: The Evidence from Deforestation* AJPS."

2.8 Heterogeneous Treatment Effects: agriculture and election type

In this section I investigate heterogeneous treatment effects across forest type. In addition, for the tests of democratization I investigate whether the effect is more pronounced for countries with a larger share of agricultural workers or agriculture as a percent of GDP.

in Figure A.1 I find that the effects uncovered across forest types are not significantly different from the effects in either tropical and subtropical forests nor from the effects in temperate and boreal forests, as classified by [65]. Unconditional on level of aggregation or forest type the transition to a majoritarian democratic system is always significantly associated with higher rates of deforestation–though I should note that the samples for some subgroups are quite small.

Given the theory in the main text it would be easy to infer that countries with a larger share of agricultural employment or economic activity would experience higher rates of deforestation after a regime type change to democracy. The reasoning is that either of these conditions would translate to more political power for that group and thus more incentive for politicians to allow deforestation. However, previous research on the political power of the agricultural sector shows that there is not a monotonic relationship between size of the sector and political power [53]. [66] point out that as the size of the agricultural sector grows farmers face a collective action problem as a lobby group, so the power of farmers decreases with the size of the agricultural sector. The tests presented in Figure A.1 and Table A.1 are consistent with

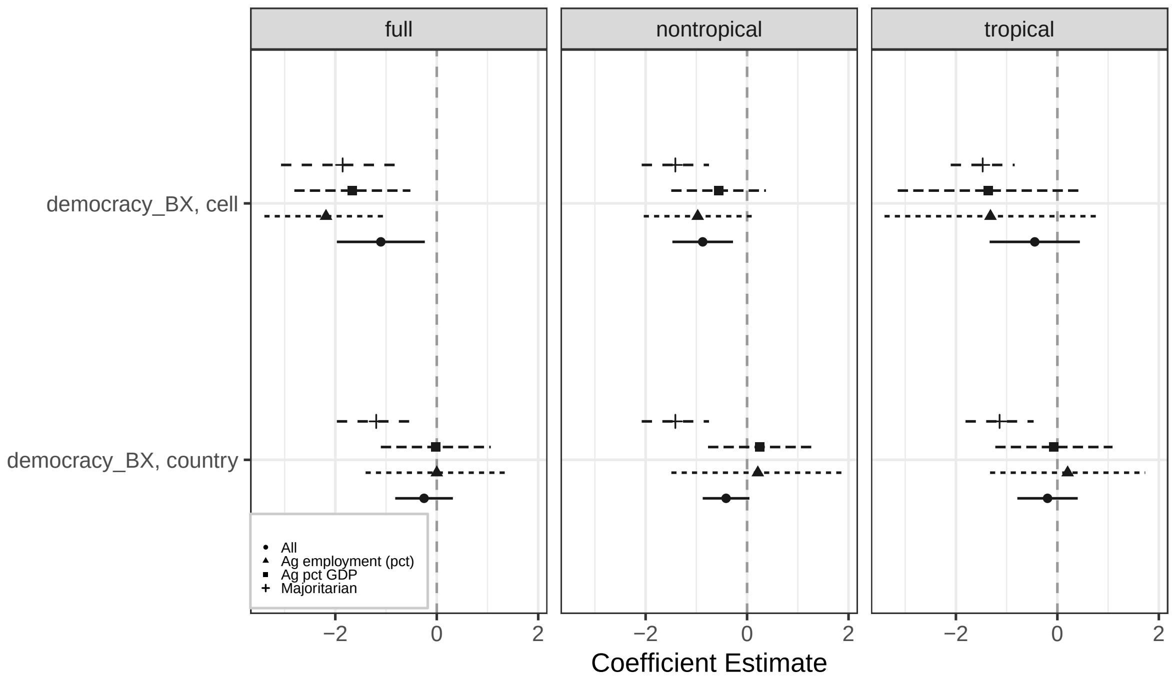


Figure A.1: Relationship between democracy regime type and forest cover change across forest types and three moderators: electoral system, agricultural percent of GDP, and agricultural percent of the population. The coefficients shown are for the coefficient on democracy and represents the estimate of the effect with agricultural employment or agricultural percent of GDP set to 0, or in purely majoritarian systems (PR system set to 0).

Table A.1: Regressions of forest change on democracy (Model 1) interacted with percent of workforce employed in agriculture (Model 2), Agriculture as a percent of GDP (Model 3), and Electoral system (Model 4). Note that it is possible though rare for nondemocracies to have a "voting system." Control variables are omitted for length.

	Model 1	Model 2	Model 3	Model 4
Democracy	-1.10^{*}	-2.18^{**}	-1.66^{**}	-1.86^{**}
	(0.44)	(0.62)	(0.58)	(0.62)
Ag employment		-0.05		
		(0.03)		
Ag emp:Democracy		0.05^{***}		
		(0.01)		
Ag pct GDP			-0.02	
			(0.04)	
Ag pct GDP:Democracy			0.05^{*}	
			(0.02)	
PR				-0.21
				(0.72)
PR:Democracy				1.16
				(0.67)
Num. obs.	136743524	108628844	112662730	117037436
Adj. R^2 (full model)	0.38	0.38	0.38	0.38
Adj. R^2 (proj model)	0.37	0.38	0.37	0.37

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

this constraint on agricultural power–the size of the agricultural sector does not strongly moderate the relationship between regime type transition and deforestation rate.

Table A.1 shows heterogeneous treatment effects by agricultural employment, agriculture share of GDP, and electoral system. In all cases the main effect is negative and significant, but in both of the agricultural interactions the interaction effect is positive and significant, indicating that the effect is strongest in countries which do not have a large agricultural sector. The dynamic here merits further research in future work.

Figure A.2 shows the results from Figure 4 in the main text broken out across forest types. Several patterns emerge from this analysis. First, in the cell-level tests elections with a margin of victory of less than 10 points are always significantly associated with forest cover loss. In national-level regressions these close elections are almost always associated with forest loss-except for in all electoral systems and all forest cover or tropical and subtropical forests. The mechanisms behind this finding deserve analysis in

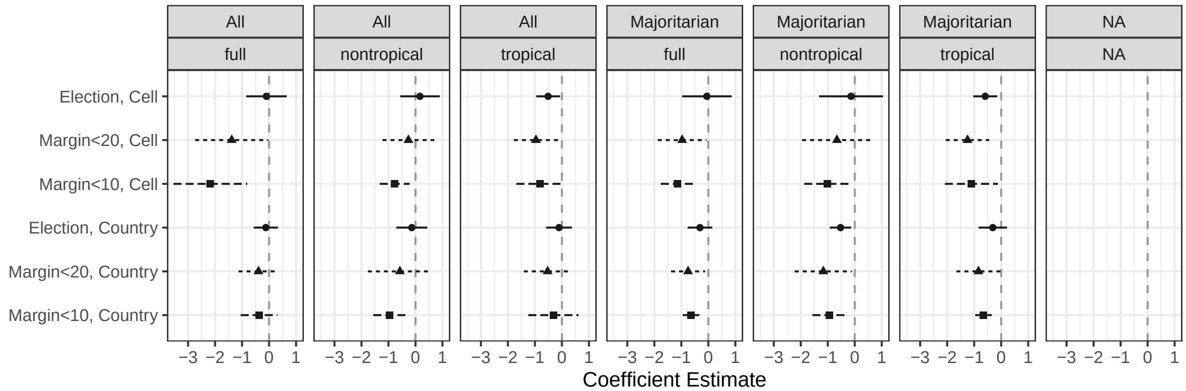


Figure A.2: Coefficient plots for forest type–elections

Relationship between elections with varying margins of victory and forest cover change across forest type and electoral system.

future work. Second, in Majoritarian electoral systems the relationship between competitive (margin<20) and close (margin<10) elections and forest cover change is always significant and negative with the sole exception in temperate and boreal forests in competitive elections. Despite the fact that each of these categories has relatively few elections there is suprising homogeneity in the results.

Figure A.3 shows results from Figure 5 in the main text broken out along forest types. What clearly stands out is that the relationship is driven by temperate and boreal forests at the cell level and null in tropical and subtropical forests. Future investigations of this effect will need to focus on the deforestation mechanisms associated with different forest types and the years and countries which are driving this effect.

Table A.2 shows the results from the main text in Models 1-3 and shows that if I interact electoral system with election year the results are present though smaller (and notably not significant for elections with the smallest margin of victory). This is evidence that regime type is the more important dimension. The results from Figure 4 in the main text further demonstrate that electoral system does not seem to matter too much, at least in the cell-level regressions.

Table A.3 shows regressions where election years are interacted with agricultural share of employment (Models 1-3) and agricultural share of GDP (figures 4-6). The table shows all null effects–that there are no detectable effects of agricultural strength on whether election years have higher rates of deforestation (without accounting for regime type).

Table A.2: Models 1-3 show the results across regime type and level of electoral competition from the main paper, Models 4-6 show interactions between electoral system and level of electoral competition (instead of regime type). Control variables are omitted for length.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Election Year	-0.10			-0.20^{***}		
	(0.38)			(0.05)		
Margin $<$ 20		-1.38			-0.34^{**}	
		(0.69)			(0.10)	
Margin $<$ 10			-2.17^{**}			0.16
			(0.70)			(0.11)
Election:Autocracy	0.14					
	(0.54)					
Margin$<$20:Autocracy		2.56^{*}				
		(1.22)				
Margin$<$10:Autocracy			3.95^{**}			
			(1.36)			
Election:Democracy	0.06					
	(0.37)					
Margin$<$20:Democracy		1.22				
		(0.76)				
Margin$<$10:Democracy			2.08^{**}			
			(0.65)			
Election:PR				0.35		
				(0.31)		
Margin$<$20:PR					0.20	
					(0.30)	
Margin$<$10:PR						-0.49^{*}
						(0.19)
PR				0.14	0.44	0.38
				(0.81)	(0.86)	(0.88)
Democracy	-0.91^{*}	-0.87^{*}	-0.96^{*}			
	(0.37)	(0.35)	(0.37)			
Autocracy	-0.50	-0.54	-0.56			
	(0.60)	(0.63)	(0.63)			
Num. obs.	132801614	118463788	110285677	117028441	103133799	95218053
Adj. R^2 (full model)	0.38	0.38	0.38	0.38	0.38	0.39
Adj. R^2 (proj model)	0.37	0.37	0.37	0.37	0.37	0.37

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table A.3: Models 1-3 show interactions between election years of different levels of electoral competition and agricultural share of employment, Models 4-6 show interactions between election years of different levels of electoral competition and agricultural share of GDP. Control variables are omitted for length.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Election Year	0.09			0.06		
	(0.20)			(0.25)		
Margin $<$ 20		-0.42			-0.44	
		(0.29)			(0.36)	
Margin $<$ 10			-0.48			-0.74
			(0.48)			(0.55)
Ag employment	-0.02	-0.02	-0.03			
	(0.03)	(0.03)	(0.03)			
Election:Ag employment	-0.01					
	(0.00)					
Margin$<$20:Ag employment		0.00				
		(0.01)				
Margin$<$10:Ag employment			0.00			
			(0.01)			
Ag pct of economy				-0.02	-0.03	-0.03
				(0.05)	(0.05)	(0.05)
Election:Ag pct of economy				0.00		
				(0.01)		
Margin$<$20 : Ag pct of economy					0.01	
					(0.02)	
Margin$<$10 : Ag pct of economy						0.03
						(0.03)
Num. obs.	107944319	96575335	90622563	112138852	100941895	95368927
Adj. R^2 (full model)	0.38	0.39	0.39	0.38	0.38	0.38
Adj. R^2 (proj model)	0.38	0.37	0.38	0.37	0.36	0.36

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

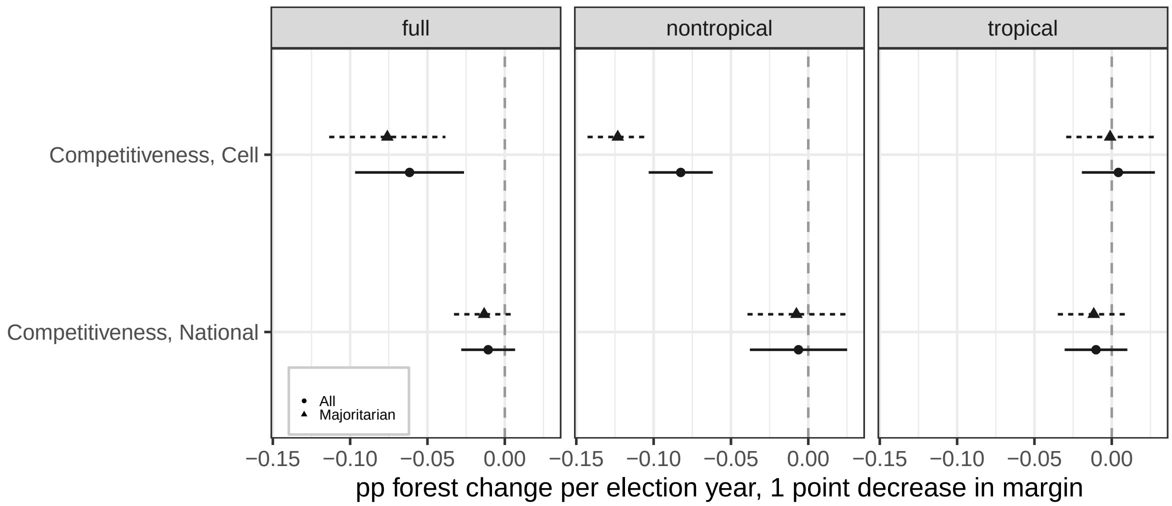


Figure A.3: Coefficient plots for forest type–margin

Relationship between margin of victory and forest cover change across forest cover types and electoral systems.

2.9 Levels of aggregation

This section presents Tables 3-5 in the main text across different levels of spatial aggregation. The extremes here are Cell and National, with intermediate levels of cells aggregated to one hundred times the native resolution (about 55×55 km near the equator), second level adminstrative units, and first level administrative units. The first two columns are implicitly weighted by forest area–countries which have more forest will be weighted more highly but each cell is weighted equally. The third to the fifth tests are weighted by adminstrative unit–for example each state or province receives equal weight regardless of the amount of forest. This implicitly downweights the result of cells which are in adminstrative units with many other forested cells. For example, the relationship uncovered in Liechtenstein is weighted equally to the relationship uncovered in Brazil. I include smaller administrative units because I think of those as possibly the “true” level of treatment–as the Kenya example shows, in many cases it is the electoral competitiveness at the district or county level which determines whether deforestation occurs or not.

Table B.1: Democracy results across levels of aggregation. 0.05 decimal degree is the cell level discussed in the paper, 0.5 dd is aggregated to cells 100 times as large, L2 is level two administrative units, L1 is level 1 administrative units, National is national borders as discussed in the paper.

	0.05dd	0.5dd	L2	L1	National
Democracy	-1.10^{*}	-1.05^{*}	-0.85	-0.35	-0.25
	(0.44)	(0.40)	(0.58)	(0.35)	(0.29)
Forest	-0.77^{***}	-0.65^{***}	-0.63^{***}	-0.59^{***}	-0.56^{***}
	(0.03)	(0.04)	(0.04)	(0.05)	(0.06)
PCGDP	0.08	0.06	0.09	0.12^{**}	0.11^{**}
	(0.05)	(0.04)	(0.05)	(0.04)	(0.03)
Δ PCGDP	-7.86	-4.64	-18.86	-10.07	-9.33
	(20.33)	(15.49)	(17.62)	(9.02)	(10.40)
Pop Growth	-0.13	-0.13	0.05	-0.16	-0.08
	(0.27)	(0.19)	(0.23)	(0.09)	(0.08)
Num. obs.	136743524	1545318	1300578	79914	4375
Adj. R^2 (full model)	0.38	0.34	0.34	0.33	0.34
Adj. R^2 (proj model)	0.37	0.31	0.30	0.27	0.26

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

2.10 Geographically weighted regressions

Geographically weighted regressions are a way of examining spatial heterogeneity in the results of this analysis. In each GWR below every point corresponds to the coefficient of interest from a regression which includes only cells within two decimal degrees from that point. The goal is to visually investigate heterogeneities in the relationship tested by the regression.

2.11 Measures of competitiveness

This section explores the results of Tables 3 and 4 in the main text but with two other measures of electoral competition. The first is the difference between the incumbent coalition's vote share and 50 which measures how close that coalition was to winning/losing an election. This measure does not capture the seat-vote elasticity discussed in [58] and only roughly captures the notion of incumbent victory or loss, see for example the US election in 2016. It is pre electoral system so a small margin of victory in the popular vote might not mean a close election in terms of seats or control of the executive branch. It is

Table B.2: Regressions of forest change on election year. 0.05 decimal degree is the cell level discussed in the paper, 0.5 dd is aggregated to cells 100 times as large, L2 is level two administrative units, L1 is level 1 administrative units, National is national borders as discussed in the paper.

	Cell	Cell	Cell	0.5dd	0.5dd	0.5dd	L2	L2	L2	L1	L1	L1	National	National	National
Election Year	-0.91^{*}	-0.87^{*}	-0.96^{*}	-0.78^{*}	-0.74^{*}	-0.82^{*}	-0.55	-0.79	-0.82	-0.11	-0.26	-0.24	-0.33	-0.38	-0.35
	(0.37)	(0.35)	(0.37)	(0.32)	(0.31)	(0.33)	(0.51)	(0.56)	(0.61)	(0.24)	(0.26)	(0.27)	(0.23)	(0.23)	(0.25)
Margin $<$ 20		-1.38			-1.23			-0.80			-0.38			-0.39	
		(0.69)			(0.61)			(0.59)			(0.40)			(0.38)	
Margin $<$ 10			-2.17^{**}			-2.04^{**}			-1.67			-0.43			-0.37
			(0.70)			(0.61)			(1.00)			(0.46)			(0.35)
Election:Autocracy	0.14			0.14			0.44			0.19			0.43		
	(0.54)			(0.42)			(0.39)			(0.27)			(0.36)		
Margin<20:Autocracy		2.56^{*}			1.93^{*}			1.28			0.53			0.72	
		(1.22)			(0.86)			(1.18)			(0.87)			(1.07)	
Margin<10:Autocracy			3.95^{**}			3.32^{**}			3.34^{*}			1.78			2.08
			(1.36)			(1.11)			(1.52)			(1.34)			(1.34)
Election:Democracy	0.06			0.05			0.14			0.04			0.07		
	(0.37)			(0.31)			(0.30)			(0.19)			(0.22)		
Margin<20:Democracy		1.22			1.07			0.35			0.16			0.19	
		(0.76)			(0.67)			(0.59)			(0.49)			(0.42)	
Margin<10:Democracy			2.08^{**}			1.94^{**}			1.06			0.13			0.16
			(0.65)			(0.58)			(0.90)			(0.53)			(0.41)
Democracy	-0.50	-0.54	-0.56	-0.39	-0.42	-0.42	0.06	0.10	-0.03	0.47	0.50	0.47	0.11	0.21	0.21
	(0.60)	(0.63)	(0.63)	(0.45)	(0.47)	(0.47)	(0.80)	(0.78)	(0.77)	(0.40)	(0.43)	(0.42)	(0.38)	(0.37)	(0.37)
Autocracy	-0.10			-0.12			-0.62			-0.22			-0.12		
	(0.38)			(0.33)			(0.33)			(0.16)			(0.23)		
Forest	-0.77^{***}	-0.77^{***}	-0.77^{***}	-0.65^{***}	-0.65^{***}	-0.66^{***}	-0.63^{***}	-0.63^{***}	-0.63^{***}	-0.58^{***}	-0.58^{***}	-0.59^{***}	-0.55^{***}	-0.56^{***}	-0.56^{***}
	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)
PCGDP	0.07	0.08	0.09	0.05	0.06	0.06	0.08	0.08	0.11	0.10^{*}	0.10^{*}	0.12^{**}	0.11^{**}	0.11^{**}	0.11^{**}
	(0.06)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)	(0.06)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)
Δ PCGDP	35.20	7.91	13.68	31.53	8.35	13.21	-6.89	-8.78	-12.73	-4.57	-4.33	-4.57	-1.48	1.93	3.14
	(37.62)	(29.47)	(31.86)	(32.20)	(23.84)	(25.70)	(21.50)	(18.53)	(19.31)	(9.86)	(8.92)	(9.54)	(10.02)	(10.10)	(10.21)
Pop Growth	-0.14	-0.23	-0.18	-0.13	-0.20	-0.15	0.08	0.11	0.05	-0.13	-0.12	-0.14	-0.06	-0.08	-0.07
	(0.27)	(0.26)	(0.27)	(0.19)	(0.19)	(0.19)	(0.24)	(0.23)	(0.21)	(0.10)	(0.08)	(0.09)	(0.07)	(0.06)	(0.06)
Num. obs.	132801614	118463788	110285677	1502050	1337802	1243855	1259194	1136341	1055911	75545	67751	63947	4081	3701	3517
Adj. R^2 (full model)	0.38	0.38	0.38	0.33	0.34	0.35	0.33	0.34	0.34	0.32	0.33	0.33	0.34	0.35	0.35
Adj. R^2 (proj model)	0.37	0.37	0.37	0.31	0.31	0.32	0.30	0.30	0.30	0.27	0.27	0.27	0.26	0.26	0.26

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table B.3: Regressions of forest change on electoral competitiveness. Competitiveness is interacted with government type with Anocracy as the base case. 0.05 decimal degree is the cell level discussed in the paper, 0.5 dd is aggregated to cells 100 times as large, L2 is level two administrative units, L1 is level 1 administrative units, National is national borders as discussed in the paper.

	0.05dd	0.5dd	L2	L1	National
Competitiveness	-0.06^{**}	-0.05^{**}	-0.01	-0.01	-0.01
	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
Comp:Autocracy	0.09^{***}	0.08^{***}	0.04^{*}	0.04^{*}	0.02
	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)
Comp:Democracy	0.06^{*}	0.05^{*}	-0.00	0.00	0.00
	(0.03)	(0.02)	(0.03)	(0.02)	(0.01)
Autocracy	-5.99^{***}	-5.00^{***}	-1.14	-1.50	-0.95
	(1.18)	(1.10)	(1.49)	(0.77)	(1.02)
Democracy	-4.53	-3.87	0.78	0.42	0.16
	(2.36)	(2.05)	(2.10)	(1.30)	(0.89)
Forest	-0.71^{***}	-0.58^{***}	-0.58^{***}	-0.48^{***}	-0.43^{***}
	(0.03)	(0.04)	(0.04)	(0.04)	(0.06)
PCGDP	0.07	0.05	0.18^{***}	0.19^{***}	0.14^{**}
	(0.08)	(0.07)	(0.03)	(0.05)	(0.05)
Δ PCGDP	43.02	43.99	94.11^{*}	52.08	12.29
	(26.81)	(22.68)	(39.27)	(30.80)	(32.50)
Pop Growth	-0.03	-0.13	-1.14	-0.46	-0.34
	(0.44)	(0.35)	(0.56)	(0.25)	(0.24)
Num. obs.	35346878	397758	341318	17227	864
Adj. R^2 (full model)	0.35	0.29	0.32	0.25	0.22
Adj. R^2 (proj model)	0.27	0.20	0.23	0.12	0.07

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

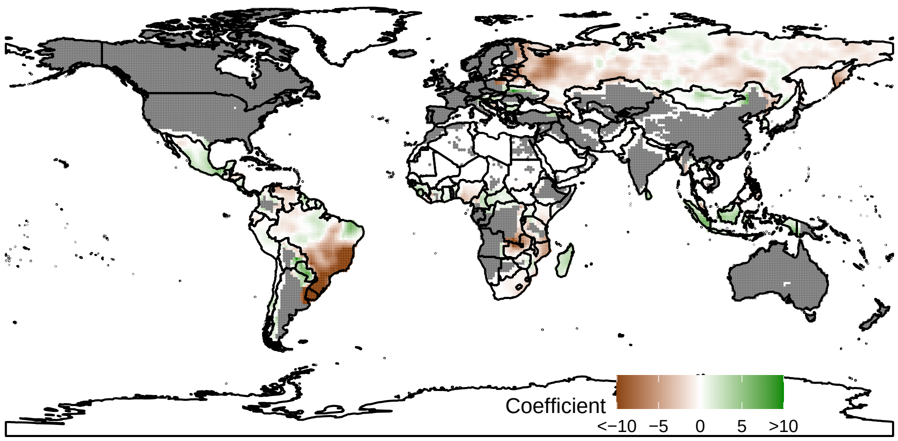


Figure C.1: Geographically Weighted Regression: Coefficient on Democracy

The color of a cell corresponds to the relationship between democracy and forest cover change for the cells within two decimal degrees of that cell. White cells are areas where there was no change, including places in which there was never forest present. Grey cells indicate areas for which the coefficient of interest is not estimable because there were no cells which had variation in regime type.

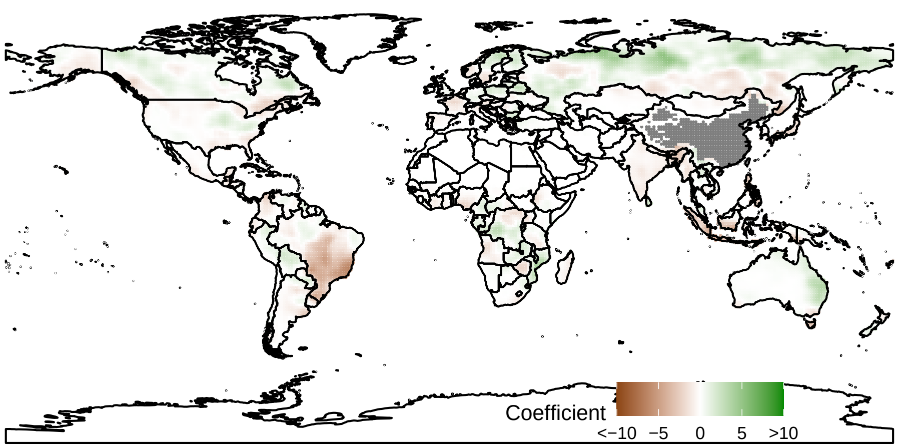


Figure C.2: Geographically Weighted Regression: Coefficient on < 20% Margin

The color of a cell corresponds to the relationship between election years and forest cover change for the cells within two decimal degrees of that cell. cells included in the regression are weighted by their inverse distance to the target cell. White cells are areas where there was no change, including places in which there was never forest present. Grey cells indicate areas for which the regression failed, mainly due to no variation in whether there were any elections.

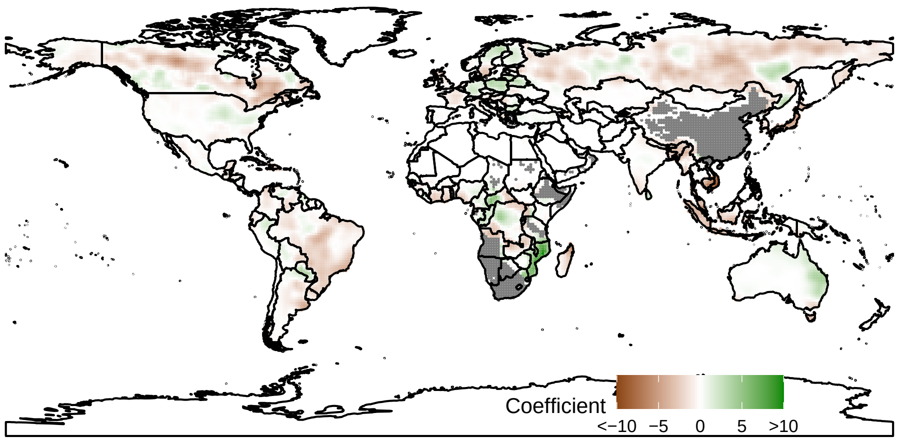


Figure C.3: Geographically Weighted Regression: Coefficient on < 20% Margin

The color of a cell corresponds to the relationship between election years with a margin of victory of less than 20% and forest cover change for the cells within two decimal degrees of that cell. White cells are areas where there was no change, including places in which there was never forest present. Grey cells indicate areas for which the regression failed, mainly due to no variation in whether there were any elections.

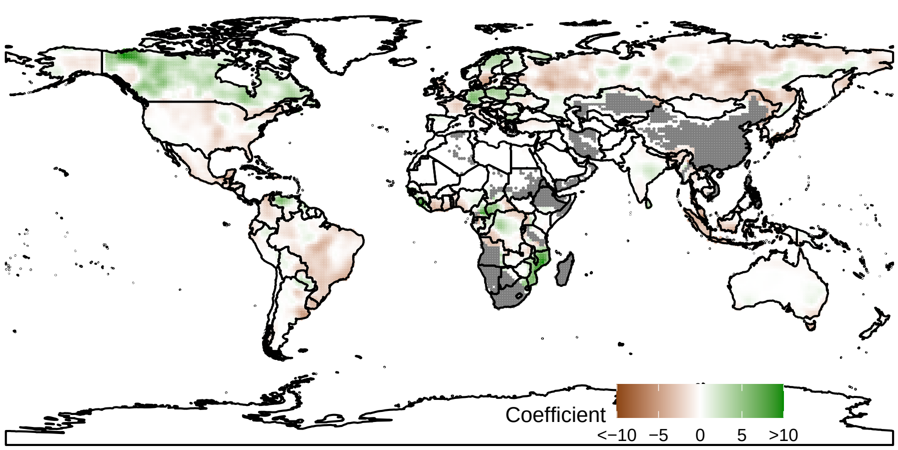


Figure C.4: Geographically Weighted Regression: Coefficient on < 10% Margin

The color of a cell corresponds to the relationship between election years with a margin of victory of less than 20% and forest cover change for the cells within two decimal degrees of that cell. White cells are areas where there was no change, including places in which there was never forest present. Grey cells indicate areas for which the regression failed, mainly due to no variation in whether there were any elections.

useful for single district PR systems, but those are places where geographic targeting already makes the expected effect small. Tables 3 and 4 are replicated in Tables D.1 and D.2.

The second is the difference between the incumbent coalition's seat share and 50 which implicitly includes seat-vote elasticities, but fails to capture what [57] say is a better measure of competitiveness, which is which party holds the prime minister position or the presidency. Tables 3 and 4 are replicated in Tables D.3 and D.4.

Both measures have the expected sign and magnitude at the National level of aggregation but null or opposite results at the cell level. Future work should investigate how electoral systems interact with measures of competitiveness to help resolve this issue. Subnational measures of competitiveness would also provide additional clarity.

2.12 Neighboring forest and lagged forest

In this section I explore what happens when a spatial lag of forest cover is included along with previous forest cover, and when their functional form is modeled by a set of interacted fixed effects. In theory the forest cover of neighboring cells should predict whether a cell will lose forest cover. Cells on the edge of a forest are more likely to be deforested than those surrounded by forest, while cells which have begun to be deforested but are not too close to 0 forest remaining should see the highest rates of deforestation. To explore the degree to which this matters for this analysis I create a set of fixed effects for each decile of forest cover and neighboring forest cover: 0-10 percent forest cover through 90-100 percent forest cover for both previous forest cover and avrage neighbor forest cover. Then I interact these sets of fixed effects and include them in regressions to achieve a closer approximation of the true functional form of these variables. Below the paper tables are replicated with this set of fixed effects included.

2.13 Timing of Deforestation

The final test presented here explores the timing of the forest loss with respect to an election. I expect deforestation rates to be highest in the time surrounding an election. The rate at which politicians

Table D.1: Regressions of forest change on election year. Election years are subset by how competitive they were: columns 1 is all election years, 2 is years with a margin of victory less than 20 points, 3 is years with a margin of victory less than 10 points. Election year is interacted with government type with Anocracy as the base case. Margin of victory is calculated as the absolute value of percent of votes for the incumbent coalition minus the percent of votes of the opposition coalition.

	Cell	Cell	Cell	National	National	National
Election Year	-0.10			-0.12		
	(0.38)			(0.23)		
Margin < 20		-0.51			-0.77^{*}	
		(0.74)			(0.33)	
Margin < 10			0.83^{*}			-1.42^{*}
			(0.39)			(0.58)
Election:Democracy	0.06			0.07		
	(0.37)			(0.22)		
Margin$<$20:Democracy		0.60			0.61	
		(0.69)			(0.37)	
Margin$<$10:Democracy			-0.51			1.42^{*}
			(0.33)			(0.62)
Democracy	-0.91^{*}	-0.90^{*}	-0.84^{*}	-0.33	-0.32	-0.34
	(0.37)	(0.39)	(0.38)	(0.23)	(0.22)	(0.23)
Autocracy	-0.50	-0.54	-0.48	0.11	0.17	0.17
	(0.60)	(0.60)	(0.59)	(0.38)	(0.36)	(0.36)
Num. obs.	132801614	124713884	119285168	4081	3816	3697
Adj. R^2 (full model)	0.38	0.38	0.38	0.34	0.35	0.35
Adj. R^2 (proj model)	0.37	0.37	0.37	0.26	0.26	0.26

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table D.2: Regressions of forest change on electoral competitiveness. Competitiveness is interacted with government type with Anocracy as the base case. Margin of victory is calculated as the absolute value of percent of votes for the incumbent coalition minus the percent of votes of the opposition coalition.

	Cell	National
Competitiveness	-0.03	-0.00
	(0.02)	(0.01)
Comp:Autocracy	0.00	0.01
	(0.03)	(0.02)
Comp:Democracy	0.03	-0.00
	(0.02)	(0.02)
Autocracy	-1.20	0.91
	(1.93)	(1.41)
Democracy	-4.19	0.66
	(2.12)	(1.66)
Forest	-0.70^{***}	-0.47^{***}
	(0.03)	(0.07)
PCGDP	0.07	0.12^{**}
	(0.05)	(0.04)
Δ PCGDP	168.92^{*}	31.61
	(67.50)	(32.33)
Pop Growth	0.34	-0.33
	(0.49)	(0.36)
Num. obs.	27696215	600
Adj. R^2 (full model)	0.33	0.21
Adj. R^2 (proj model)	0.21	0.03

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table D.3: Regressions of forest change on election year. Election years are subset by how competitive they were: column 1 is all election years, 2 is years with a margin of victory less than 20 points, 3 is years with a margin of victory less than 10 points. Election year is interacted with government type with Anocracy as the base case. Margin of victory is calculated as the percentage of seats won by the victorious coalition minus 50.

	Cell	Cell	Cell	National	National	National
Election Year	-0.10			-0.12		
	(0.38)			(0.23)		
Margin < 20		0.17			-0.94^{**}	
		(0.66)			(0.31)	
Margin < 10			0.71			-0.31
			(0.52)			(0.27)
Election:Democracy	0.06			0.07		
	(0.37)			(0.22)		
Margin$<$20:Democracy		-0.22			0.79^{*}	
		(0.65)			(0.34)	
Margin$<$10:Democracy			-0.59			0.13
			(0.50)			(0.39)
Democracy	-0.91^{*}	-0.84^{*}	-0.87^{*}	-0.33	-0.37	-0.37
	(0.37)	(0.40)	(0.40)	(0.23)	(0.23)	(0.24)
Num. obs.	132801614	118136134	110261053	4081	3572	3410
Adj. R^2 (full model)	0.38	0.38	0.38	0.34	0.35	0.35
Adj. R^2 (proj model)	0.37	0.37	0.37	0.26	0.27	0.27

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table D.4: Regressions of forest change on electoral competitiveness. Competitiveness is interacted with government type with Anocracy as the base case. Margin of victory is calculated as the percentage of seats won by the victorious coalition minus 50.

	Cell	National
Competitiveness	0.02	-0.01
	(0.01)	(0.01)
Comp:Autocracy	-0.02	0.01
	(0.02)	(0.01)
Comp:Democracy	-0.02	0.01
	(0.01)	(0.01)
Autocracy	1.05	0.13
	(1.15)	(0.78)
Democracy	-0.04	-0.37
	(0.91)	(0.78)
Forest	-0.69^{***}	-0.45^{***}
	(0.03)	(0.06)
PCGDP	0.09	0.12^{**}
	(0.06)	(0.04)
Δ PCGDP	125.60	21.30
	(65.20)	(25.69)
Pop Growth	-0.02	-0.35
	(0.40)	(0.19)
Num. obs.	34499480	907
Adj. R^2 (full model)	0.34	0.21
Adj. R^2 (proj model)	0.24	0.09

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table E.1: Regressions of forest change on democracy. Fixed effects for each ten percent of lagged forest cover and each ten percent of average neighbor forest cover and their interaction.

	Cell	Cell	National	National
Democracy	-1.07^{*}	-0.42	-0.18	-0.04
	(0.45)	(0.33)	(0.20)	(0.09)
PCGDP	0.09^{*}	0.01	0.04	0.00
	(0.04)	(0.01)	(0.03)	(0.01)
Δ PCGDP	-6.67	22.88	12.81	27.14
	(18.71)	(25.31)	(10.38)	(17.18)
Pop Growth	-0.10	0.00	0.03	-0.04
	(0.29)	(0.12)	(0.06)	(0.06)
Constant				0.12
				(0.20)
Num. obs.	122497614	122497614	4375	4375
Adj. R^2 (full model)	0.35	0.13	0.07	0.00
Adj. R^2 (proj model)	-0.04	0.00	-0.04	0.00

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table E.2: Regressions of forest change on election year. Election years are subset by how competitive they were: columns 1 is all election years, 2 is years with a margin of victory less than 20 points, 3 is years with a margin of victory less than 10 points. Election year is interacted with government type with Anocracy as the base case. Fixed effects for each ten percent of lagged forest cover and each ten percent of average neighbor forest cover and their interaction.

	Cell	Cell	Cell	National	National	National
Election Year	-0.17			-0.19		
	(0.38)			(0.28)		
Margin $<$ 20		-1.18			-0.59	
		(0.77)			(0.50)	
Margin $<$ 10			-1.99^{*}			-0.42
			(0.85)			(0.48)
Election:Democracy	0.10			0.05		
	(0.36)			(0.30)		
Margin$<$20:Democracy		0.91			0.20	
		(0.83)			(0.53)	
Margin$<$10:Democracy			1.80^{*}			0.10
			(0.79)			(0.54)
Democracy	-0.81^{*}	-0.74	-0.85^{*}	-0.17	-0.23^{*}	-0.20
	(0.39)	(0.38)	(0.40)	(0.12)	(0.11)	(0.12)
Num. obs.	118807786	106171502	99157113	4081	3701	3517
Adj. R^2 (full model)	0.34	0.35	0.35	0.08	0.08	0.08
Adj. R^2 (proj model)	-0.04	-0.04	-0.04	-0.04	-0.04	-0.05

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table E.3: Regressions of forest change on electoral competitiveness. Competitiveness is interacted with government type with Anocracy as the base case. Fixed effects for each ten percent of lagged forest cover and each ten percent of average neighbor forest cover and their interaction.

	Cell	National
Competitiveness	-0.03^{*}	-0.02
	(0.01)	(0.01)
Comp:Autocracy	0.05^{*}	0.02
	(0.02)	(0.02)
Comp:Democracy	0.03	0.00
	(0.02)	(0.01)
Autocracy	-2.84^{*}	-1.51
	(1.28)	(1.11)
Democracy	-1.98	0.48
	(1.31)	(0.82)
Num. obs.	35346878	864
Adj. R^2 (full model)	-0.00	0.01
Adj. R^2 (proj model)	-0.13	-0.17

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

will choose to allocate forested land will peak just before the election takes place for two reasons: the ability of a politician to efficiently allocate resources increases as the election approaches, and voters exhibit recency bias. First, as an election approaches a politician's expected probability of winning that election becomes more precise, so they can choose how much land to allocate to ensure victory without wasting too much of the resource. Second, voters tend to exhibit some myopia and give more weight to recent events than less-recent events. Simply, a politician wants the benefit a voter received from that politician to be in the front of his mind when he goes to the ballot box [23, 67].

While one might expect deforestation to be a slow process and thus expect higher rates of deforestation for several years after an election, I expect deforestation associated with elections to happen quickly for two reasons: opportunity costs of waiting and political uncertainty. Consider the two mechanisms highlighted above: smallholder farmers converting forest to cropland and logging firms extracting timber. Smallholder farmers have an incentive to clear forests quickly so they can plant crops during the next growing season. Failure to do so would be to sacrifice a year's worth of additional income. Furthermore, if farmers have to relocate to obtain this additional land like they did in the case of the Mau forest reserve in Kenya, their main priority is to clear the land and start growing crops. Alternatively, logging firms have different incentives to exploit forested resources quickly: their access might be contingent on the incumbent winning the upcoming election. Should a challenger win it would make sense for the challenger to revoke access to a firm that supported their opponent. Knowing this, logging firms should extract as much as possible quickly.

I test this by creating a lead and a lag of the **competitive election** variable ($t-1$ to $t+1$) to identify the timing of deforestation. Table F.1 shows that the year of an election with less than a 10% margin of victory in an anocracy has significantly higher rates of deforestation than other years while the years before and after an election are not impacted to a degree which is statistically significant. This relationship is strong across levels of aggregation.

Table F.1: Regressions of forest change on close elections, with election leads and lags

	Cell	Cell, lead lag	National	National lead lag
Close Election	-2.17^{**}	-1.52^{**}	-0.37	-0.52
	(0.70)	(0.48)	(0.35)	(0.49)
Close Lag		-0.24		-0.38
		(0.30)		(0.59)
Close Lead		-0.43		-0.94
		(0.38)		(0.75)
Close:Democracy	2.08^{**}	1.47^{**}	0.16	0.14
	(0.65)	(0.48)	(0.41)	(0.49)
Close Lag:Democracy		0.29		0.10
		(0.36)		(0.60)
Close Lead:Democracy		0.34		0.96
		(0.37)		(0.74)
Forest	-0.77^{***}	-0.77^{***}	-0.56^{***}	-0.59^{***}
	(0.03)	(0.04)	(0.06)	(0.06)
PCGDP	0.09	0.09	0.11^{**}	0.11^{*}
	(0.05)	(0.06)	(0.03)	(0.04)
Δ PCGDP	13.68	13.92	3.14	12.12
	(31.86)	(32.00)	(10.21)	(12.68)
Pop Growth	-0.18	-0.18	-0.07	-0.01
	(0.27)	(0.28)	(0.06)	(0.08)
Democracy	-0.96^{*}	-0.96^{*}	-0.35	-0.50
	(0.37)	(0.37)	(0.25)	(0.32)
Autocracy	-0.56	-0.55	0.21	0.14
	(0.63)	(0.64)	(0.37)	(0.40)
Num. obs.	110285677	109946876	3517	2366
Adj. R^2 (full model)	0.38	0.38	0.35	0.36
Adj. R^2 (proj model)	0.37	0.37	0.26	0.26

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

References

(1) *The Right Conditions: The World Bank, Structural Adjustment, and Forest Policy Reform*; Seymour, F. J., Hutter, C., Eds.; World Resources Institute: Washington, D.C, 2000.

(2) Klopp, J. M. *Journal of Eastern African Studies* **2012**, *6*, 351–370.

(3) Morjaria, A. In *Paper for the World Banks 2012 Annual Bank Conference on Development Economics*, 2012.

(4) Bates, R. H., *Markets and States in Tropical Africa*; U. of California Press: 1979.

(5) Pailler, S. Local Politics and Deforestation in the Brazilian Amazon., Ph.D. Thesis, CLARK UNIVERSITY, 2016.

(6) Pailler, S. *Journal of Environmental Economics and Management* **2018**, *88*, 345–365.

(7) Alkama, R.; Cescatti, A. *Science* **2016**, *351*, 600–604.

(8) Gibbs, H. K.; Ruesch, A. S.; Achard, F.; Clayton, M. K.; Holmgren, P.; Ramankutty, N.; Foley, J. A. *Proceedings of the National Academy of Sciences* **2010**, *107*, 16732–16737.

(9) Malhi, Y.; Roberts, J. T.; Betts, R. A.; Killeen, T. J.; Li, W.; Nobre, C. A. *science* **2008**, *319*, 169–172.

(10) Song, X.-P.; Hansen, M. C.; Stehman, S. V.; Potapov, P. V.; Tyukavina, A.; Vermote, E. F.; Townshend, J. R. *Nature* **2018**, *560*, 639.

(11) Lake, D. A.; Baum, M. A. *Comparative Political Studies* **2001**, *34*, 587–621.

(12) De Mesquita, B. B., *The Logic of Political Survival*; MIT Press: 2005.

(13) Pierskalla, J. H.; Sacks, A. *British Journal of Political Science* **2019**, 1–23.

(14) Gottlieb, J.; Kosec, K. *American Political Science Review* **2019**, *113*, 88–107.

(15) Li, Q.; Reuveny, R. *International Studies Quarterly* **2006**, *50*, 935–956.

(16) Buitenzorgy, M.; P. J. Mol, A. *Environmental and Resource Economics* **2011**, *48*, 59–70.

(17) Hansen, M. MEaSUREs Vegetation Continuous Fields (VCF) Yearly Global 0.05 Deg., 2018.

(18) Besley, T., *Principled Agents?: The Political Economy of Good Government*; Oxford University Press: 2007.

(19) Dasgupta, S.; De Cian, E. *Energy Research & Social Science* **2018**, DOI: 10.1016/j.erss.2018.05.023.

(20) Ehrhardt-Martinez, K.; Crenshaw, E. M.; Jenkins, J. C. *Social Science Quarterly* **2002**, *83*, 226–243.

(21) Bernauer, T.; Koubi, V. *Ecological Economics* **2009**, *68*, 1355–1365.

(22) Didia, D. O. *Global Environmental Change* **1997**, *7*, 63–76.

(23) Keefer, P. *American Journal of Political Science* **2007**, *51*, 804–821.

(24) Driscoll, B. *Studies in Comparative International Development* **2018**, *53*, 404–427.

(25) Harding, R.; Stasavage, D. *The Journal of Politics* **2014**, *76*, 229–245.

(26) Midlarsky, M. I. *Journal of Peace Research* **1998**, *35*, 341–361.

(27) Marquart-Pyatt, S. *International Journal of Sociology* **2004**, *34*, 33–51.

(28) Burgess, R.; Hansen, M.; Olken, B. A.; Potapov, P.; Sieber, S. *The Political Economy of Deforestation in the Tropics*; Working Paper 17417; National Bureau of Economic Research, 2011.

(29) Boone, C., *Political Topographies of the African State: Territorial Authority and Institutional Choice*; Cambridge Studies in Comparative Politics; Cambridge University Press: Cambridge, UK ; New York, 2003.

(30) Boone, C., *Property and Political Order in Africa: Land Rights and the Structure of Politics*; Cambridge Studies in Comparative Politics; Cambridge University Press: New York, 2014.

(31) Chupezi, T. J. *In Search of Common Ground* **2009**, 411.

(32) Mertens, B.; Lambin, E. F. *Applied Geography* **1997**, *17*, 143–162.

(33) Newell, P., *The Globalization and Environment Reader*; John Wiley & Sons: 2016.

(34) Boelee, E., *Ecosystems for Water and Food Security. [Background Paper]*; United Nations Environment Programme (UNEP); International Water Management Institute (IWMI).: 2011.

(35) Chazdon, R. L. *Science* **2008**, *320*, 1458–1460.

(36) Obradovich, N. *Climatic Change* **2017**, *140*, 135–147.

(37) Ross, M. L., *Timber Booms and Institutional Breakdown in Southeast Asia*, 1 edition; Cambridge University Press: Cambridge, UK ; New York, 2001.

(38) Rudel, T. K. *Phil. Trans. R. Soc. B* **2013**, *368*, 20120405.

(39) Rogoff, K. *Equilibrium Political Budget Cycles*; Working Paper 2428; National Bureau of Economic Research, 1987.

(40) Drazen, A. *NBER Macroeconomics Annual* **2000**, *15*, 75–117.

(41) Mayhew, D. R., *Congress: The Electoral Connection, Second Edition*, 2 edition; Yale University Press: New Haven, 2004.

(42) Albertus, M. *Comparative Political Studies* **2013**, *46*, 1082–1111.

(43) Harding, R. *One for the Road: Voting for Public Goods in Ghana*; tech. rep.; Working Paper, New York University, 2011.

(44) Ostrom, E.; Burger, J.; Field, C. B.; Norgaard, R. B.; Policansky, D. *Science* **1999**, *284*, 278–282.

(45) Olson, M., *The Logic of Collective Action*; Harvard University Press: 1965.

(46) Deacon, R. T. *Public Choice* **2009**, *139*, 241–262.

(47) Anderson, K., *The Political Economy of Agricultural Price Distortions*; Cambridge University Press: 2010.

(48) Swinnen, J. F. *The Political Economy of Agricultural Price Distortions* **2010**, 81–104.

(49) Bates, R. H.; Block, S. A. *The Journal of Politics* **2013**, *75*, 372–384.

(50) Kasara, K. *American Political Science Review* **2007**, *101*, 159–172.

(51) Keefer, P.; Vlaicu, R. *The Journal of Law, Economics, and Organization* **2008**, *24*, 371–406.

(52) Rogowski, R. *International Organization* **1987**, *41*, 203–223.

(53) Park, J. H.; Jensen, N. *American Journal of Political Science* **2007**, *51*, 314–329.

(54) Boix, C.; Miller, M.; Rosato, S. *Comparative Political Studies* **2013**, *46*, 1523–1554.

(55) Beck, T.; Clarke, G.; Groff, A.; Keefer, P.; Walsh, P. *The World Bank Economic Review* **2001**, *15*, 165–176.

(56) Coppedge, M.; Gerring, J.; Lindberg, S. I.; Teorell, J.; Altman, D.; Bernhard, M.; Fish, M. S.; Glynn, A.; Hicken, A.; Knutsen, C. H., et al. *Codebook. Version* **2020**.

(57) Kayser, M. A.; Lindstädt, R. *Political Analysis* **2015**, *23*, 242–253.

(58) Cox, G. W.; Fiva, J. H.; Smith, D. M. *Political Analysis* **2020**, *28*, 168–185.

(59) Marshall, M. G. **2019**, 86.

(60) WDI World Development Indicators | DataBank., http://databank.worldbank.org/data/, 2017.

(61) Lee, D. W.; Rogers, M. *Political Analysis* **2019**, *27*, 263–280.

(62) Olson, M. *The American Political Science Review* **1993**, *87*, 567–576.

(63) Ostrom, E., *Governing the Commons: The Evolution of Institutions for Collective Action*; Cambridge University Press: 1990.

(64) Hardin, G. *Science* **1968**, *162*, 1243–1248.

(65) FAO, I. Global Agro-Ecological Zones (GAEZ v3.0)., 2012.

(66) De Gorter, H.; Swinnen, J. In *Handbook of Agricultural Economics*; Agricultural and Food Policy, Vol. 2; Elsevier: 2002, pp 1893–1943.

(67) Keefer, P.; Khemani, S. *The World Bank Research Observer* **2005**, *20*, 1–27.

Chapter 3

Measurement and Inference using Satellite Data in Benin

3.1 Abstract

In this paper I show that satellite imagery can be used to improve estimates of the effects of programs or policies in settings where the location of the treatment is known and the outcome(s) of interest are observable in satellite imagery. We often only observe where and when an intervention occurs but lack an adequate control group with which to compare. I combine remote sensing methods with a "double machine learning" strategy to account for confounds that appear (and may only appear) in the satellite record, and which may contribute to both selection into treatment and outcomes of interest. In non-experimental settings this approach thus allows researchers to directly control for confounders that must otherwise be assumed away. I demonstrate this approach using both convolutional neural networks and random forests and show how both can be used to take advantage of multi-spectral, high frequency imagery. As a case study example, I apply this technique to estimate the effects of land tenure formalization on landowner investment behavior in Benin, West Africa and find that conventional techniques would over-estimate the effects of land titles on conversion from forest to cropland.

3.2 Introduction

A rapid expansion in both data availability and new analysis have demonstrated the utility of remotely-sensed data for capturing important on-the-ground outcomes in many settings. In addition, the recent ubiquity of satellite observations with rapid return times contains a second exciting promise for those who seek to estimate causal effects in settings where the location of the treatment is known and the outcome(s) of interest (particularly economic, agricultural, or environmental) are observable in satellite imagery. Here I demonstrate one set of methods that shows how time series land surface observations can improve impact evaluations by isolating the effect of the intervention from confounds that my contribute to both selection into treatment as well as outcomes. This method is particularly powerful because it accounts for confounds which are encoded into the satellite record but do not need to be explicitly parameterized or specified by the researcher.

This method is intended to be useful for anyone studying a geographically defined treatment with an outcome which can be measured using remote sensing. However, as a case study and demonstration, I here focus on land tenure formalization. Land tenure formalization has been a goal of international organizations and governments around the world for the last 50 years and was intended to spur investment by providing more secure property rights. However, these projects are almost never conducted at scale and evaluated experimentally, leaving researchers with the difficult task of analyzing if, when, where, and how these projects work. I begin with a brief introduction land tenure formalization for readers who may come from a remote sensing background.

I then review the fundamental problem of causal inference using the potential outcomes framework, how randomized control trials (RCTs) provide a solution to this problem, and what issues face the vast majority of (non-RCT) studies like most land-tenure formalization efforts. Briefly, these issues arise when those who receive a treatment (or policy) are different from those who don't such that we would not expect them to have the same outcome had they both received (or not received) the treatment. In this setting researchers cannot identify whether observed differences in the treated and control groups are because of the treatment or because of the initial differences. Remote sensing can resolve this issue by improving measurement of those initial differences so they can be accounted for in the study.

Briefly, the strategy is to use several years worth of satellite imagery from before the program was implemented to uncover systematic differences between the treated and untreated groups. Specifically, it seeks to uncover any pre-treatment differences which might affect whether the land will be eligible to receive a title and any differences which affect what the land will be used for in later years. It uses a procedure called "double machine learning" to adjust for those differences. Any confounding factors which appear in satellite imagery can be adjusted for in this way, including many physical or environmental factors which are otherwise difficult to measure. As long as these factors are encoded in satellite imagery, they can be adjusted for; they need not be explicitly measured.

Finally, I demonstrate the power of this methodology by estimating the effect of a land-titling reform in Benin on land use changes indicative of household investment, finding that places which received titles were more likely to convert natural forest to cropland than similar places which did not receive titles.

This method is of broad applicability for remote sensing-based impact evaluation. It offers researchers the ability to adjust for a range of potential confounding variables which are difficult or impossible to measure using traditional methods. It also offers the possibility for impact evaluations of large-scale, non-random policies, or more broadly the effect of political institutions on environmental outcomes.

3.2.1 Land Titling

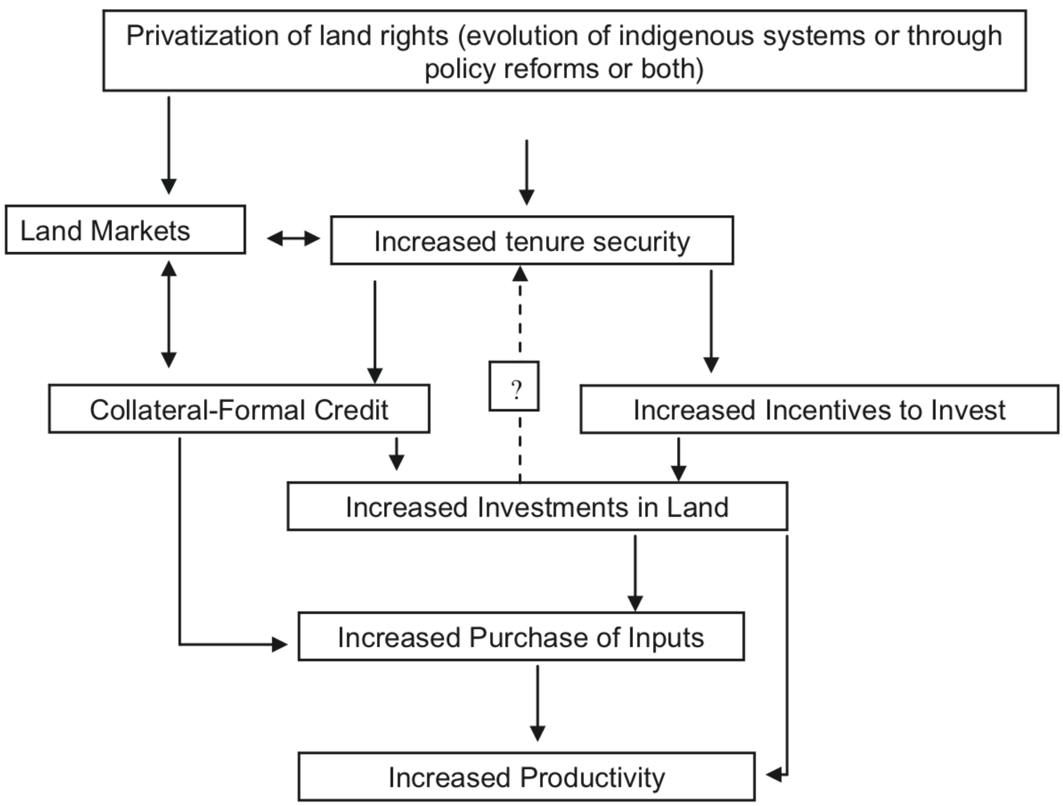
Land tenure formalization efforts begin from the foundational assumption that lack of formal property rights prevents efficient use of land. Over 2.5 billion dollars have been spent on land titling efforts over the last twenty years, and over 100 different studies have been conducted evaluating those efforts. Unfortunately, few strong conclusions emerge from this literature, likely in part due to the lack of rigorous measurement and reserach design associated with most interventions. [1] could find only four randomized control trials (RCTs) out of over one thousand considered publications, and of 117 studies which aimed to estimate the causal impact of land tenure security fewer than half were considered rigorous. Even among the most rigorous studies the methods the methods used remain susceptible to environmental confounds that may be associated with selection for titling in the first place.

Why is tenure formalization entrenched in development efforts? The potential pathways of impact are multiple: A land title could provide the owner with collateral and thus additional access to credit. Here, the land serves as an asset which owners can leverage to improve their access to credit. The downstream implications of additional access to credit are many, but include the ability to make additional investments in the land. Land tenure formalization could make a landowner feel more secure in areas where land has historically been seized or reallocated according to political whim or violence, and thus more willing to undertake productivity improving investments. In this case a landowner is more certain that her land will not be expropriated by the government or otherwise taken from her, making her more likely to invest in the long-term productivity of the land. This might include forgoing several harvests so that she can plant higher-value perennial crops. Finally tenure formalization could result in increased liquidity in land ownership by facilitating purchases and sales [2–4]. Then the landowner would be able to sell her land to a farmer for whom the land would be more productive. This too would result in higher productivity. These mechanisms are outlined in Figure 3.1.

Other effects of land titling programs that are outside of the mechanisms proposed by [2] include changes in labor-market participation and whether increased production occurs at the intensive or extensive margin. Some recent evidence shows that land titling can result in a decrease in unproductive labor which functions only to assert continued ownership of a parcel of land. This means that those formerly occupied with this unproductive labor can seek employment in productive jobs, improving the overall welfare of title-holding families through a mechanism wholly unconnected to agricultural productivity. Another line of research investigates whether land titles can reduce the amount of land use conversion associated with increased agricultural production. By incentivizing productivity-enhancing investments on land which is already used for agriculture these programs may be able to reduce the amount of land which needs to be converted to cropland to meet demand for agricultural products. This would generate positive externalities in the form of continued ecosystem services from un-converted natural habitat.

Unfortunately the effects of land titling seem to vary significantly across contexts and experimental designs[1, 3]. Agriculture in different agro-climatic zones has different characteristics and inputs, and traditional land tenure varies both across and within countries. Studies differ in the type of land tenure that they evaluate–sometimes offered by the government as part of a new policy, sometimes in conjunction with the World Bank or another NGO. Perhaps most importantly, these differences affect the degree of

Figure 3.1: Mechanisms by which formal land tenure can affect productivity. Source: [4]



tenure security under traditional land tenure systems *and* the degree of tenure security after a title has been assigned. If the difference is not large, or if agricultural behavior is not contingent on tenure security, we could expect (and some studies find) negligible effect sizes [**robinson_does_2014**, 3].

The main outcome of interest of most studies is a productivity-enhancing investment. This takes the form of soil-fallowing, use of fertilizer, improved seeds, mechanization, planting of non-crop trees, terracing, or planting of perennial or tree crops[3].[1] To date, no study that I am aware of has investigated the relationship between land titling and use of irrigation to improve productivity.

Most studies of land titling projects have taken place in African countries, including at least Rwanda [7], Madagascar[8], Ghana[2, 5, 6], Burkina Faso[9], Mali, Ivory Coast[6], Ethopia[10–12], Uganda[13], Benin[14–18], and Niger[19]. However, because of variations in study design and local context, there are few over-arching conclusions that can be drawn from the existing literature. The conclusion that seems to enjoy the greatest support is an increase in tree-planting, either as way to prevent soil erosion or in the form of perennial cash crops [1].

I focus on one empirical example: the effects of the Plan Foncier Rural (PFR) program conducted by the Milennium Challenge Corporation to assign legal land titles to around 70,000 landowners in Benin. To facilitate the land titling procedure, the borders of each titled plot were demarcated and recorded. However, plots which were eligible for but did not receive a title were not similarly demarcated. I estimate the effect that demarcation had on agricultural expansion, deforestation, expansion of built-up areas, and the introduction of tree-crops. Previous work analyzing this program has found an increase in tree-crops on treated parcels [14, 17] and a decrease in deforestation in areas adjacent to villages with treated parcels [18]. For now (for computation reasons) I focus on the Department of Alibori in which 3074 parcels of land from 33 villages received treatment.

[1] A difficulty for many previous studies is that many types of investment can serve to improve informal tenure security because some traditional systems allow de-facto ownership of land through improvement to that land [5, 6]. This makes the problem of identifying the effect of land titling on land improvement more difficult, especially outside the context of a randomized control trial.

3.2.2 Remote Sensing and Impact Evaluation

Satellite remote sensing is one of the fastest growing measurement methods in the social sciences. Researchers have using night-lights data to measure economic outcomes for years [20] but have recently begun to take advantage of multi-spectral daytime imagery to perform a broader array of measurement tasks [21]. These include measuring population [22], income, poverty[23], urban expansion [24], urban damage and rebuilding[25], infrastructure development [26], land cover and land use (especially deforestation) [27, 28], agricultural practices and crop characteristics [29], or crop yields are necessary for the study of many social science questions. With a public satellite record dating back to the 1970's and instruments with improved temporal, spectral, and spatial resolution coming online each year researchers have an ever-expanding set of measurement strategies available.

In the land titling literature several studies have begun to rely on remote sensing to measure outcomes of interest. [1] find 14 of their 117 studies use remote sensing. Remote sensing offers frequent and high spatial resolution measurement of outcomes, particularly changes in land use. This has enormous benefits. Researchers can cheaply extend their on the ground measurements, both across time (to evaluate longitudinal impacts) and across space (to evaluate impacts across a wider study area). It also allows researchers to avoid relying on respondents to accurately report the effects of a policy (when it might be in their interest to misrepresent). Third, it allows analysis at a level of geographic aggregation which matches the research question. Finally, and most importantly in this study, it allows researchers to adjust for factors which can be measured in satellite imagery but are extremely difficult to otherwise adjust for.

Because of the ubiquity of satellite imagery, it is simple to match on the ground observations from surveys to images which correspond to the timing and location of the observations. These ground truth observations can then be used as training data in a machine learning algorithm to scale up measurement of important outcomes across both space and time. Returning to the question of the impacts of land tenure, a commonly studied outcome is tree-planting (an activity which has long-term benefits to the landowner and thus should be more common among those who have higher land tenure security). A set of on the ground measurements of whether farmers planted trees in certain locations can easily be scaled up using those measurements as training data, leaving researchers with a measurement of the outcome variable at all locations rather than just those surveyed. Furthermore, with many measurements each year researchers can observe the timing of responses to allow a more complete picture of when impacts occur.

Researchers are also not tied to ground-truth measurements collected by researchers who are truly on the ground. Recent work in remote sensing uses very high resolution imagery to code outcomes which are observable by humans. Researchers can use the google basemap to code many outcomes, including construction of structures or infrastructure or changes in vegetation, and those observations can be scaled up using more ubiquitous and available data at lower spatial resolutions [30][2].

The majority of land tenure impact evaluations rely on the subjects of the study to self-report on the outcome of interest. This can be problematic if social desirability bias leads respondents to tell researchers that some program was effective. Worse, if respondents believe that future aid or interventions are contingent on the success of the study they may have incentives to misrepresent outcomes. Using satellite imagery which cannot be manipulated by the subjects of the study avoids this issue.

Relatedly, researchers questions tend to imply a unit of analysis which might not agree with the unit on which data is collected. For example, if the question is about the behavior of landowners it makes sense to collect data at the plot or household level. However, if the question is about changes to productivity or deforestation, data should be collected by spatial unit rather than household or plot so that the overall aggregate effect can be identified. Using remotely sensed data allows researchers to aggregate up to the appropriate level of analysis.

However, the use of satellite imagery comes with its own set of problems, many of which are effectively outlined in [31] and [32]. One issue stands out–errors in measurement are likely to be correlated with underlying environmental characteristics resulting in non-classical measurement error. Worryingly, if the things which result in measurement error are also associated with whether a parcel of land is assigned a land title then the measurement error can bias the results. For example: if an area has favorable growing conditions (soil nutrients, access to water, etc.) then natural vegetation might be more likely to be mistaken for cropland because of its increased productivity. Such an area might also induce a farmer to apply for a land title to improve her security for that parcel of land. If this pattern is persistent (favorable areas are both

[2] In my analysis of land titling in Benin I hand code land cover classes using the google basemap and then scale that classification up using landsat imagery.

more likely to be mis-classified and to receive a title) then any relationship between titling and agricultural area would be confounded by this measurement error.

Remote sensing also opens the possibility for researchers to include many more units in their study, but this is only useful if those units are similar enough to those with a different treatment status to be useful. In many land titling studies the treated units are known because they received a formal title, but the appropriate control group is not[3]. Including more control units is only helpful if they contain information about what would have happened to treated units had they been left untreated, and if that information is able to be extracted despite the differences between those units and the treated units. In the land titling example if researchers collect data on control units which either would never have been eligible for a title (located in a national park, for example) or which cannot experience the same landcover transition as the titled areas (because they are on extremely degraded land which is infrequently titled) then those units can bias the results of an analysis if they are not properly accounted for.

In this paper I demonstrate how a "double machine learning" (DML) approach [33] can help mitigate these issues with using satellite data and in doing so can also account for environmental and physical characteristics of land which are likely to bias traditional methods.

3.3 Methods

I show that remote sensing, when paired with a DML approach to inference, can improve the way that we conduct impact evaluation of land tenure formalization. The goal of this approach is to make sure that the treated and untreated groups are as similar as possible across two dimensions: how likely they were to receive a land title before titles were assigned, and what their eventual land cover would have been had they received the opposite treatment assignment[4]. I use machine learning methods designed for remote sensing applications to adjust both groups to ensure that a fair comparison is being made across these two dimensions.

In this section I review the goal and fundamental problem of causal inference, how randomization

[3] Implementors rarely want to spend the time and money to demarcate plots which do not receive a title, and doing so is unethical if land titles are expected to have a positive outcome.

[4] That is, we would like it to be the case that if no area had been treated, the average outcome of treated and control groups would have been identical *and* that the average outcome of both groups would have been the same had they all received titles.

helps, and the type of issues which arise without randomization. I contextualize this in the world of land tenure formalization studies and give a short overview of the toolkit most frequently employed by researchers. I show that pairing the use of satellite imagery with machine learning methods and a DML approach can improve inference by allowing us to measure and adjust for important differences between treated and control groups. Finally, I outline how this can resolve the two issues with using satellite imagery described above.

3.3.1 Potential Outcomes

To understand the effect of receiving a land title on whether a farmer makes a productivity-enhancing investment we must compare two different outcomes: what happened when the farmer received her title, and what would have happened had she not received a title (but with all other factors the same). The difference between those outcomes is the individual treatment effect of a land title on one subject. If we expand this to a number of study subjects, some of whom received a title and some of whom did not, we have the average treatment effect. The difficulty is in how we infer the outcome that we did not observe for each individual or group. This dilemma is known as the fundamental problem of causal inference [34]. The field of causal inference is built around how to best estimate the difference between what would have happened under a different treatment condition and what actually happened.

To clarify the process, I introduce some notation originally introduced by [34] and commonly used in econometrics today [35, 36]. Define Y_i^1 as the "potential outcome" of unit i had it received the treatment and Y_i^0 as the potential outcome of unit i had it not received the treatment. For any observation, both of these exist, but only one is ever observed. For example, Y_i^1 is what would occur on plot i had it received a title, and Y_i^0 is what would occur had it not received a title. Then $\delta_i = Y_i^1 - Y_i^0$ is the treatment effect for unit i, or how much of an impact a land title would have. The goal of most studies is to estimate an average treatment effect, or ATE:

$$\begin{aligned} ATE &= E[\delta_i] \\ &= E[Y^1] - E[Y^0] \\ &= E[Y^1 - Y^0] \end{aligned} \tag{3.1}$$

The ATE is the average difference between Y^1 (the potential outcome under treatment) and Y^0 (the potential outcome under control). $E[]$ is the expected value, usually operationalized as the sample average. We only ever observe one *potential* outcome: the one which matches the observed treatment assignment. In land tenure security studies each plot has two potential outcomes: what would have happened with a title, and what would have happened without one. For each plot, we only get to see one of those–the potential outcome which corresponds to the actual title status of the plot. We can only observe Y^1 for units which received treatment and Y^0 for units which did not. We can write the observed Y_i as:

$$Y_i = D_i Y_i^1 + (1 - D_i) Y_i^0 \tag{3.2}$$

Which simply gives the treated potential outcome if the unit is treated (leaving $1 * Y_i^1 + 0 * Y_i^0$) and the untreated potential outcome if the unit is untreated (leaving $0 * Y_i^1 + 1 * Y_i^0$. Because we cannot simultaneously observe Y^1 and Y^0, one common approach is to use the simple difference between the treated and untreated groups:

$$ATE = E\left[Y^1 \mid D = 1\right] - E\left[Y^0 \mid D = 0\right] \tag{3.3}$$

Here the ATE is the difference between the treated potential outcomes Y^1 of the treated group $D = 1$ and the untreated potential outcomes Y^0 of the control group. $D = 0$. For this strategy to uncover the true effect it must be true that the treated and control groups are similar across all the factors which influence the outcome. In our example, this means that the titled and un-titled plots should have looked exactly the same had they all been left with no title. The only setting where this is plausibly true is when titles are randomly assigned or not assigned to an eligible group. Formally, I follow [36] and decompose this estimator into two parts:

$$\begin{aligned} E\left[Y^1 \mid D = 1\right] - E\left[Y^0 \mid D = 0\right] = & \\ & E\left[Y^1\right] - E\left[Y^0\right] \\ & + E\left[Y^0 \mid D = 1\right] - E\left[Y^0 \mid D = 0\right] \end{aligned} \tag{3.4}$$

The first line is the simple difference in means estimator. The second line is the “true” average treatment effect which we would uncover if we knew all of the potential outcomes. The third line is a "selection bias" term which will make our estimate of the ATE differ from the true value if it is not equal to zero. It describes the difference in what would have happened absent treatment Y^0 across treated and control groups ($D = 1$ and $D = 0$).[5]

An example of selection bias in the context of land titling is when certain areas which are fertile are both more likely to receive a land title *and* more likely to be converted to tree crops. This can happen in a number of ways, for example if those implementing the project want it to be successful they may choose the most suitable areas to receive titles, or farmers on productive land may be more likely to apply for titles. In either case the land which received a title was different from the land which did not. Specifically, the titled land was more likely to be converted to tree crops even absent land titles. This means that the untreated potential outcomes for the treated group $E\left[Y^0 \mid D = 1\right]$ are not the same as for the control group $E\left[Y^0 \mid D = 0\right]$ resulting in an estimate of the ATE which is different from the true effect. Because conversion to tree crops was already more likely on the titled land we could expect to see a difference between titled and untitled outcomes even if the title truly had no effect. If there is a true effect of land titles, we will observe an effect size which is larger than the true effect because of this selection bias.

Formally, if the treatment status is the result of a set of factors X, then $D = g(X)$. If the potential outcomes Y^1 and Y^0 are also a result of both treatment and those factors, $Y^1 = m(X)$ and $Y^0 = m(X)$ then without accounting for X $E\left[Y^0 \mid D = 1\right] - E\left[Y^0 \mid D = 0\right] \neq 0$ and

$$E\left[Y^1 \mid D = 1\right] - E\left[Y^0 \mid D = 0\right] \neq E\left[Y^1\right] - E\left[Y^0\right]$$

This is also almost certain to happen when a researcher only knows the location of the title-receiving areas and seeks to use remote sensing to measure what occurred in untreated areas. Here, the researcher is almost guaranteed to select areas which have different potential outcomes than the treated areas. Ideally, the researcher would only collect data from areas which were also eligible to receive titles but were randomly selected not to, but in every case which I am aware of those areas were not demarcated

[5] I choose to focus on selection bias and rely on a homogeneous treatment effect assumption here for simplicity, as in [36]. For a discussion of heterogeneous treatment effect bias, see [35] and for a description of how this bias is also minimized by the estimation strategy employed here see [33].

(even if treatment was randomized). As a result, the researcher using remote sensing might gather data from areas which are covered by water or in protected areas–two areas for which the potential outcomes differ greatly from those in titled areas.

These examples illustrate that there must be balance across the treated and control groups potential outcomes, or at least balance in the set of factors which influence potential outcomes and treatment. There are a number of strategies available to researchers which can minimize selection bias and recover the ATE. I begin by describing the gold standard for estimating the ATE: the randomized control trial (RCT). A randomized control trial (RCT) is where a group of eligible units are selected and then each unit is randomly assigned to either treated or untreated status. This automatically breaks and relationship between X and D because D is generated exclusively as a result of some randomization process. This means that $(Y^1, Y^0, X) \perp\!\!\!\perp D$ because D is randomly assigned, meaning that $E\left[Y^1 \mid D=1\right] = E[Y^1]$ and $E\left[Y^0 \mid D=0\right] = E[Y^0]$. The selection bias term disappears and the ATE can be estimated using the simple difference in means (Equation 2). In the context of land titling, if a title is assigned or not according to a coin flip then there cannot be any factors which influence selection into treatment other than the coin flip and the groups are expected to be the same on average. In practice it is important to verify that there is balance across observed characteristics of the treated and control areas.

Unfortunately randomization of assignment to treatment is often impossible or not feasible. Government programs typically have objectives which are not compatible with randomization, and randomizing access to a beneficial program like access to formal land titles can anger constituents. Practitioners are also often resistant to RCTs because demarcating and evaluating areas is expensive, and if those areas are not titled then doing so can seem like a waste of money. As a result, very few RCTs exist in the land titling literature [1] and I am unable to find a study which included plot-level randomization. The closest example I could find was the Plan Foncier Ruraux program in Benin, where out of 576 villages eligible to receive land titles, 300 were randomly selected. In the treated villages, villagers were allowed to apply for the project and have their plots demarcated free of cost. In the untreated villages, no demarcation occurred. Randomization guarantees that the potential outcomes for the treated and control *villages* are equal, but does not provide any guarantee at the plot level.

Without knowledge of how treatment to the PFR was assigned inside the villages, we do not know which parts of the untreated villages would have gotten titles had the village be selected for treatment. The initial problem is that untreated plots are not even defined (because they were never demarcated) and thus outcomes cannot even be measured. If we instead use a pixel in a remote sensing image as the unit of analysis then untreated units are defined (pixels in untreated villages) and treatment status is 1 for all pixels in plots in treated villages and 0 for all pixels in untreated villages. Even setting aside the fact that a "village" does not have demarcated boundaries, the potential outcomes under treatment or control are not equivalent for treated and control units. This is because to be treated a pixel must both be in a (randomly assigned) treated village *and* in a non-randomly assigned titled plot. The process of deciding which land within a village is demarcated almost ensures that the treated areas have different potential outcomes than untreated areas.

For example, treated points do not occur in village centers where commercial property, communal areas, and most roads are located. These places are very likely to have their (treated and untreated) potential outcomes be "built up." In the untreated villages, and without any additional information, some pixels will be sampled from built up areas which would not have been eligible for a title. This is the most obvious of many possible situations which all but guarantee that $E\left[Y^1\right] \neq E\left[Y^0\right]$. Areas which are forested and far from a road are also both less likely to receive a title *and* more likely to remain forest, both with and without being titled. Note that these issues arise in one of the most carefully conducted RCTs evaluated by two of the only four studies which evaluate RCTs found by [1].

This suggests a strategy where we only sample control points from areas which are sufficiently similar across dimensions which might influence both assignment of treatment and post-treatment land cover. Using measured demographic or economic data to perform this selection is precisely the matching approach used by a number of land tenure formalization studies. More generally, this approach boils down to finding all of the factors X which might result in $E\left[Y^1\right] \neq E\left[Y^0\right]$ and adjust for those things by selecting control areas such that $p(D=1 \mid X) = p(D=0 \mid X)$.

In the Benin PFR setting, even though treatment was randomly assigned, we cannot use the RCT to evaluate impacts at the plot level using satellite data. Because control plots were never demarcated, the best we can do is to say that while suitable control plots exist, we do not know where they are located. A simple difference in means approach will be biased by which areas the researcher chooses to use as untreated areas unless those areas are similar enough to the treated plots. The matching strategy described above is one of the strategies used to evaluate programs which are not RCTs.

Without an RCT, the next best way to evaluate a program is to attempt to model the selection into treatment process sufficiently well that you only compare places which were identically likely to receive treatment ex-ante (treatment model, for example matching), or to model the outcome process sufficiently well that conditional on some other factors the potential outcomes of treated and control units are equivalent (outcome model, for example multivariate regression).

The data that a researcher would wish for in this setting includes all the information which went into determining which areas received a title and all the information which determines what the ultimate land cover classification will be. These include things like distance to village center, distance to roads, characteristics of the land like soil characteristics, access to water, rainfall, temperature, elevation, aspect, latitude, the skill of the farmers working there, access to other inputs, ecosystem service provision, and many others. Once in possession of this data, the researcher would attempt to adjust for this set of covariates X to recover the ATE.

The assumption that a specified set of covariates are sufficient to recover the ATE is often called "conditional ignorability," and written $(Y_i^1, Y_i^0) \perp\!\!\!\perp D \mid X_i = x$. This can be read "Treatment is as-if randomly assigned among units with the same values of X." Without going further into details, conditioning on X recovers:

$$\begin{aligned} E\left[Y^1 \mid D=1, X=x\right] \quad & -E\left[Y^0 \mid D=0, X=x\right] \quad = \\ E\left[Y^1 \mid X=x\right] \quad & -E\left[Y^0 \mid X=x\right] \end{aligned} \tag{3.5}$$

With the third line of equation (4) equal to zero via the conditional ignorability assumption [36]. [6] As long as X captures all of the information used to assign a title then once we adjust for these factors the treated and untreated areas will be equally likely to receive the treatment, like in an RCT. For example, if for each treated point we find a control point which shares the same X (all of the above factors: distance to roads, soil quality, etc.) then we are comparing points which were equally likely to be selected for treatment, like in an RCT.

[6] We also require the "common support" assumption, which is $0 < Pr(D_i = 1 \mid X = x) < 1$ or that for any values of X it is possible to have received either treatment status.

Alternatively, if X captures all of the information necessary to predict the untreated potential outcome Y^0 then once we condition on X the untreated potential outcomes for the two groups will be equal, meaning $E\left[Y^0 \mid D=1, X=x\right] - E\left[Y^0 \mid D=0, X=x\right] = 0$. For example, if we know that absent treatment, a place is more likely to be converted to tree crops if it has better access to water and higher quality of soil then if we only select control points which are similar to the treated points on those two characteristics we will be making a fair comparison and would expect to see no difference between the groups if the title did not have any effect on land use.

The many approaches–including matching, fixed effects, and DiD–described by [1] as "rigorous" studies are all attempts to condition on some observed variables. However, we are always limited by the data which we posses. Broadly, these studies do not control for environmental characteristics, including many of those mentioned above: soil characteristics, access to water, rainfall, temperature, elevation, aspect, latitude, the skill of the farmers working there, access to other inputs, ecosystem service provision, etc. The reason for this is that these can be extremely difficult to measure. However, remote sensing studies can either directly gather or use satellite imagery to measure many of these things!

This suggests two possible approaches. One is to use remote sensing data to generate estimates of all of the potential confounders that we can think of and then adjust for those via some estimator (regression, matching, etc.). The other is to find a way to directly use the imagery as X and let a model determine which features of the image are important. In this paper suggest the second. The approach can use a number of different estimators (and estimands) but here I focus on a simple case: a linear regression of an outcome Y on treatment status D and some covariates X. With the conditional ignorability assumption along with a constant treatment effect assumption ($E[Y_i^1 - Y_i^0 \mid X_i] = E[Y^1 - Y^0 \mid X_i] \forall i$) and a linearity assumption (Y is a linear, additive function of D and X) we can say that B_1 is an unbiased and consistent estimator of the ATE [36]:

$$Y = \beta_0 + \beta_1 D + \gamma X + \varepsilon \tag{3.6}$$

where β_0 is an intercept, β_1 is the coefficient on the vector of treatment statuses D, γ is a vector of coefficients corresponding to the control variable matrix X, and ε is a vector of residuals.

If we seek to use satellite images themselves as the controls, the linearity assumption is very unlikely to hold. That is, the land cover class is not likely to be a linear function of the reflectances of different wavelengths of light over time (and those of the pixel's neighbors). Instead, I imagine that there are a number of potentially important features encoded as a complicated, non-linear function of satellite imagery. In fact, when remote sensing researchers use a neural network or random forest or other machine learning classifier to predict characteristics of the land using satellite imagery, they are approximating one such complex, non-linear function. To include such a function we need a variation of regression called "partially linear regression" [37]:

$$\begin{gathered} Y = \beta_0 + \beta_1 D + m(X) + U, \quad E[U|X,D] = 0, \\ D = g(X) + V, \quad E[V|X] = 0 \end{gathered} \tag{3.7}$$

where Y is the outcome variable, D is the treatment, X is a vector of controls, $g()$ and $m()$ are some (non-linear) transformations, and U and V are vectors of residuals. In this type of model we do not have the linearity assumption for the covariates and can substitute any functions for $g()$ and $m()$.

The double machine learning approach [33, 38] estimates a partially linear model where $g()$ and $m()$ can be estimated using machine learning methods appropriate for very high dimensional covariates. [38] show that issues associated with regularization bias and overfitting in ML estimators can be overcome by employing Neyman-orthogonalization and sample splitting. This results in a $N^{-1/2}$ consistent estimate of the treatment effect which is not biased by variables which can be represented as functions of satellite images.

[38] suggest the following process:

Estimate: $D = \hat{g}(X) + \hat{U}$

Estimate: $Y = \hat{m}(X) + \hat{V}$

Estimate: $\hat{\beta}_1$ from $\hat{V} = \beta_0 + \beta_1 \hat{U} + \varepsilon$

where each of $\hat{U}$ and $\hat{V}$ are estimated out of sample through a sample-splitting process. Here the goal is to estimate effect (β_1) of some policy (D) on an outcome in a location in a given year (Y). X includes all of the images taken of that plot of land for some time before the policy began in addition to any other covariates that the researcher wishes to model. $\hat{g}()$ and $\hat{m}()$ are obtained through estimating suitable machine learning algorithm–ideally one which has shown some success in other satellite imagery applications.

More concretely, this process involves splitting the sample (say into 10 groups), estimating $\hat{g}()$ and $\hat{m}()$ using 9/10 of the data and generating $\hat{U}$ and $\hat{V}$ with the remaining 1/10 of the data, repeating until each group has been used to generate $\hat{U}$ and $\hat{V}$. Then, $\hat{U}$ and $\hat{V}$ have been generated through a sample-splitting process and orthogonalized and can be used to estimate β_1 in a linear regression[39], having had nuisance parameters partialled out [40, 41].

One useful property of this estimation strategy is that it is doubly-robust, meaning that if either $\hat{g}()$ or $\hat{m}()$ approximate the true functions $g()$ and $m()$ sufficiently well then the estimate will be unbiased[7]. Stated differently, if a confounder is captured in either the treatment assignment model or the outcome model it will not bias the estimate of β_1. A correct specification of the treatment model or the outcome model generates $(Y_i^1, Y_i^0) \perp\!\!\!\perp D \mid X_i = x$, removing the possibility for bias.

A variety of estimators can be used to approximate the $g()$ and $m()$ functions with $N^{-1/4}$ convergence (as this achieves $N^{-1/2}$ convergence of the overall estimator), and recent results indicate that deep neural networks can achieve this under many conditions [43]. While this result does not necessarily hold for deep, *convolutional* networks, it would be straightforward to pre-train the convolutional layers on a different set of points and then apply transfer learning so that the model used is simply a fixed transformation (by the convolutional layers) followed by a deep neural network. Doing so would explicitly satisfy the assumptions in [43] but has been left to future work. Random forests also achieve a $N^{-1/4}$ convergence rates, as demonstrated by [44].[8]

In practice, doubly-robust estimators can minimize bias by adjusting for both the major causes of treatment status and the major causes of the outcome, meaning that any remaining confounder is a relatively minor cause of each of the treatment and outcome. Another way to think of this is that any unit

[7] This also helps to take care of regularization bias as shown in [38, 42]

[8] Other assumptions necessary for the machine learner include that the distribution of the outcome does not have unbounded moments, the common support assumption, and that neither the outcome or the treatment can be perfectly predicted [38].

for which we can use pre-treatment information to perfectly predict treatment status is not useful to us because it does not contain any useful couterfactual information. If a treated unit is unique enough in its characteristics to be predicted perfectly, that uniqueness makes it difficult to generate a counterfactual for. Similarly, if a unit's eventual landcover can be perfectly predicted by satellite imagery before treatment was assigned, then it does not provide any useful information on the effect of a treatment. The more difficult it is for a good model to predict both the treatment status and the outcome using pre-treatment imagery, the more useful it is in estimating the causal effect of the treatment.

This corresponds to the use of $\hat{U}$ and $\hat{V}$. X is pre-treatment satellite imagery, and $g()$ is a model tuned to predict treatment. The easier it is for the model to correctly predict the treatment, the less useful that observation is, and the smaller the treatment residual. The more difficult it is, the more useful the observation is because of its similarity to units with the opposite treatment status. The model $m()$ is tuned to predict eventual landcover class from pre-treatment imagery, and like above the better able predict, the less useful the observation. The more different the predicted and true landcover classes, the larger V and the more likely it was that treatment had an effect on that outcome.

3.3.2 Inference with Imagery

Below I describe the specifics of how satellite imagery is used as an input to two machine learning models: a random forest paired with a harmonic regression (RF) and a deep, convolutional neural network (NN). Similar models have been shown to be extremely effective in measurement of landcover [45]. A more thorough description of the models and their estimation of land cover classification in Benin are in the Supporting Information.

3.3.3 Estimation strategy

I demonstrate how to implement a double machine learning estimator which estimates the effect of some geographically known treatment while using a series of satellite images to adjust for possible confounders. The researcher must know the locations of treated units–either as points or as polygons–and the date at which treatment occurred. They must also have access to satellite imagery[9] from before treatment was assigned, generally at least a year of bi-weekly images (as in Landsat or Sentinel). The researcher may also have access to other variables, including the geographic characteristics of treated areas (slope, aspect, elevation, distance to urban center, etc.). The untreated units may be known (as in the experimental setting) or unknown, in which they may comprise all of the untreated areas subject to constraints based on previously measured covariates (within the same administrative unit, or with a certain distance of treated units, for example).

With this information in hand, we proceed with four steps: measurement, the outcome model, the treatment model, and the inferential model. The measurement step is where I record the outcome(s) of interest for each unit in the study. These measurements may come from an off-the-shelf remotely sensed measure, for example forest cover [27], they may come from an on the ground survey, or they can be generated using remote sensing data (this is what I do in the Benin case).

In the outcome model I use satellite imagery from the years directly preceding treatment to predict the outcomes from the measurement model: $Y = \hat{m}(X) + \hat{V}$. This uses the information contained in that imagery to try to estimate the outcome, and then keeps only the part of the outcome which cannot be predicted using that data. If this $\hat{m}(X)$ model does a good job of approximating what the land would have looked like without the land title, then what is left over can be attributed to the title. $\hat{V}$ represents this left-over variation.

In the treatment model I use satellite imagery from the years directly preceding treatment to predict the treatment status of each unit: $D = \hat{g}(X) + \hat{U}$. This uses the imagery to estimate where treatment will occur, and then keeps only the portion of the treatment which cannot be predicted. This ensures that the remaining variation in treatment $\hat{U}$ is conditionally independent of any confounders which appear in those images. Alternately, this generates a treatment variable which is balanced across treatment groups after accounting for factors which appear in the satellite imagery. If this model accurately approximates the relationship between confounding variables and treatment, the estimates of the treatment effect will be unbiased.

[9] I assume this imagery shows surface reflectance and has had clouds and their shadows masked. Also that there are enough cloud-free images over the time period to obtain a good model.

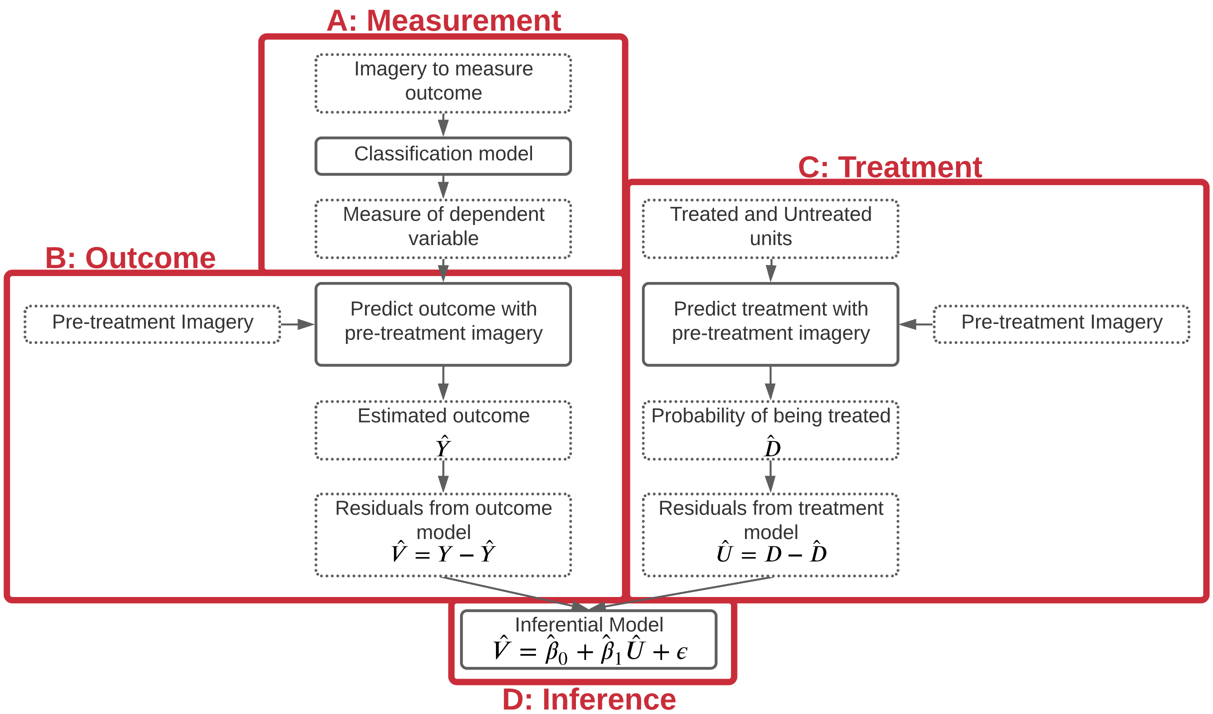


Figure 3.2: Components of the DML strategy for satellite imagery. A corresponds to the measurement model with hand-labeled landcover classes–however if the researcher is using an off the shelf measure of the outcome this does not require a model. B estimates the non-linear component of the outcome model and partials out functions of the satellite imagery and other variables which influence the dependent variable, producing $\hat{V}$. C estimates the non-linear component of the treatment model and partials out any functions of the satellite imagery and other independent variables which predict treatment producing $\hat{U}$. D is the inferential model which is linear in $\hat{\beta}_1$ with potential confounders "concentrated out" [33]

In the inferential model we compare the adjusted outcomes to the adjusted treatment using a linear regression to recover the ATE. If either $\hat{g}(X)$ or $\hat{m}(X)$ approximates either the land cover outcome or the selection into treatment process, the estimate of the ATE generated by $\hat{\beta}_1$ will be unbiased.

This is laid out in figure 3.2 where boxes with solid lines represent models and boxes with dotted lines represent data. The classification model (Panel A) is used to generate the dependent variable (if one does not already exist) for each unit in the study. The prediction models (Panels B and C) take satellite imagery and other data and cross-fit the models to estimate treatment status and outcome status. The inferential model (Panel D) estimates the treatment effect with possible confounders partialled out by the machine learning models.

3.4 Results

3.4.1 Benin Application

To demonstrate the methods explained above I consider an empirical example: the effects of the Plan Foncier Rural (PFR) program conducted by the Millennium Challenge Corporation to assign legal land titles to around 70,000 landowners in Benin. To facilitate the land titling procedure, the borders of each titled plot were demarcated and recorded. However, plots which were eligible for but did not receive a title were not similarly demarcated. I estimate the effect that demarcation had on agricultural expansion, deforestation, expansion of built-up areas, and the introduction of tree-crops. Previous work analyzing this program has found an increase in tree-crops on treated parcels [14, 17] and a decrease in deforestation in areas adjacent to villages with treated parcels [18]. For now (for computation reasons) I focus on the Department of Alibori in which 3074 parcels of land from 33 villages received treatment, but future iterations of this project will expand the study area to the whole of Benin.

3.4.2 Benin PFR

The *Plan Foncier Rural* or PFR is a policy experiment which has been implemented in various forms across several African countries. One objective of the program is to improve agricultural productivity by increasing land tenure security. While originally designed to protect natural resources from encroachment, the project was scaled up in Benin with the involvement of the Millennium Challenge Corporation (MCC) in 2006. The project seeks to formalize traditional land rights within customary practices through the following steps. 1) an informational campaign to make villagers aware of how the process will work. 2) A study to record all land claims in the village and resolve any differences. 3) A topographic survey to demarcate all agreed-upon land parcels with cornerstones and digitally record the bounds of those parcels. 4) Plots demarcated in 3) are associated with their owners based on the results of 2). At the time of the follow up survey in 2011 and published by [14] owners had not received formal titles, and later in a 2015 follow up only about 30% of titles had been claimed [17].

The implementation of the PFR program in Benin took place mainly from 2009-2011 included over 70,000 plots distributed throughout 283 villages spread across all 12 departments. NGOs went to each

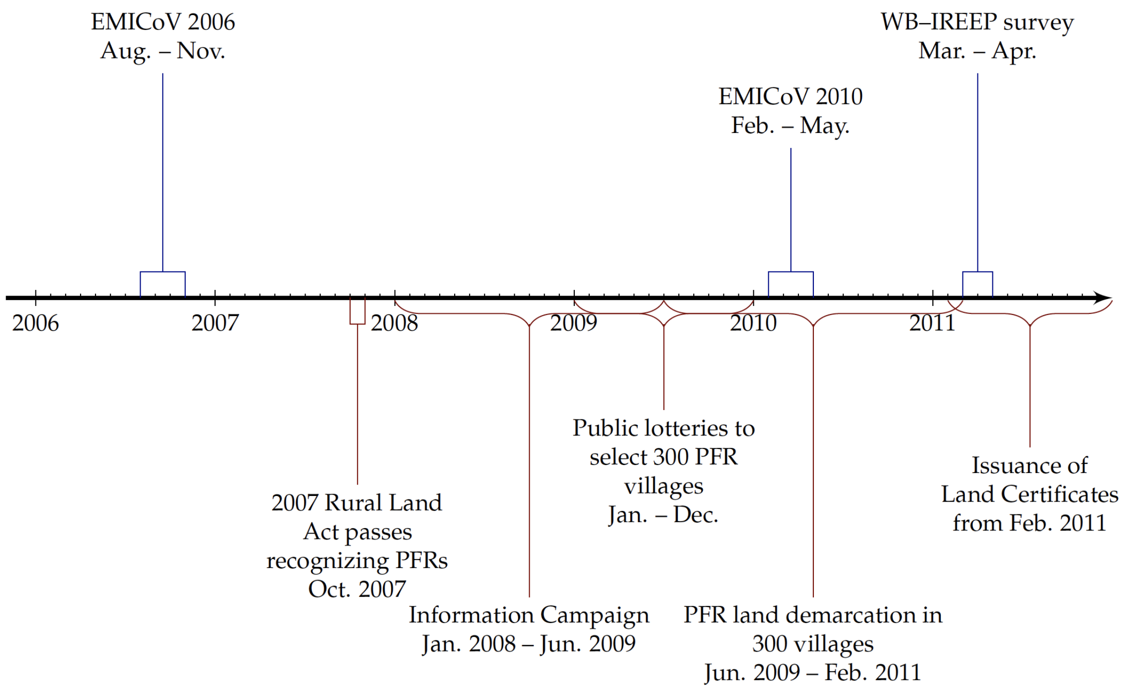


Figure 3.3: Timeline of project rollout. Source: [14]

treated village and set up meetings where landowners worked to resolve land disputes and then set out to demarcate the boundaries of each plot. This process included recording existing or placing new markers on the corners of plots of land and then recording those markers as well as the geographical positioning system coordinates of the boundaries of the plot. The NGOs also recorded individual attributes of the owners.

In 2011 a follow-up survey was conducted where a group of landowners from both treated and untreated villages were asked to answer questions about their beliefs and behavior. This report found that people whose plots were demarcated by the program were more likely to plant trees and to plant perennial cash crops (likely highly correlated with tree planting). They also found that women were more likely to fallow their land, possibly a sign of improved tenure security. On the other hand, yields for female land owners were lower than those for men on demarcated plots [14]. In the 2015 follow-up survey households in treated villages had a 1.7 % increase in tree planting and a 2.4% increase in perennial crops. It also found that women experienced the largest increases in perceived tenure security and this resulted in a decision to shift production from plots which had been demarcated to non-demarcated areas further from the village [17].

The goal of work on land tenure is to determine the effect of land tenure status on the behavior of those who received the intervention. The best practice for conducting these studies is to randomize an aspect of assignment of formal titles and then to follow up with a sample of those who received a title and a sample of those who didn't. [14] claim to be the first evaluate a randomized allocation of land titles. Randomization occurred within a group of eligible villages, with approximately 300 selected to be eligible to receive titles. Follow-up surveys were conducted in 2011 [14] and 2015 [17]. In each case a sample of people whose land was mapped (and to whom land rights were attributed) were compared with a sample who were from control villages. Among land tenure studies this study was very well done, with strong randomization and multiple follow-up periods. However, it suffers (though not as badly as many) from several issues which make this type of study extremely difficult.

3.4.3 Benefits of a remote sensing and DML approach

In Benin two follow-up surveys were conducted–one in 2011 (the year the program roll-out concluded) and one in 2015. This is already better than most land-tenure studies in which only one follow-up survey is conducted [1], but will miss any effects which take more than five years to appear. Given that to date fewer than 30% of plots which received the treatment have had their titles retrieved it seems sensible to allow for further follow-ups (at the 10-year mark, for example). Using remote sensing imagery allows continuous measurement of the outcomes of interest. Additionally, using satellite imagery allows researchers to measure alternative outcomes retrospectively. In the Benin case I also evaluate the effect of land titles on cropland expansion, built up area, and natural forest.

The PFR program in Benin is one of the few studies which has randomized treatment with formal land titles [1, 17]. However, randomization occurred at the village level, but treatment occurred at the individual level, with many plots (and individuals) within a village failing to get their plots demarcated (see Figure 3.5 panel B). This means that a comparison between a random individual in a control village and a person whose plot was demarcated in a treated village is not a good comparison–we cannot be sure that a random individual from the control village would have had their plot demarcated had their village been selected for treatment. The probability of receiving a title is higher for the recipient individual from the treated village than the random individual from the control village–they each have the same probability of having their village selected for treatment, but the treated individual *had* their land titled, where we don't know if the individual from the control village would have. Aggregation to the village level eliminates this problem but makes it impossible to investigate heterogeneous treatment effects (for example, whether the effects vary by gender of the land owner) and drastically reduces power. The double machine learning approach allows me to model the likelihood of receiving a title for each location within villages, allowing me to account for confounders which might determine treatment at the sub-village level.

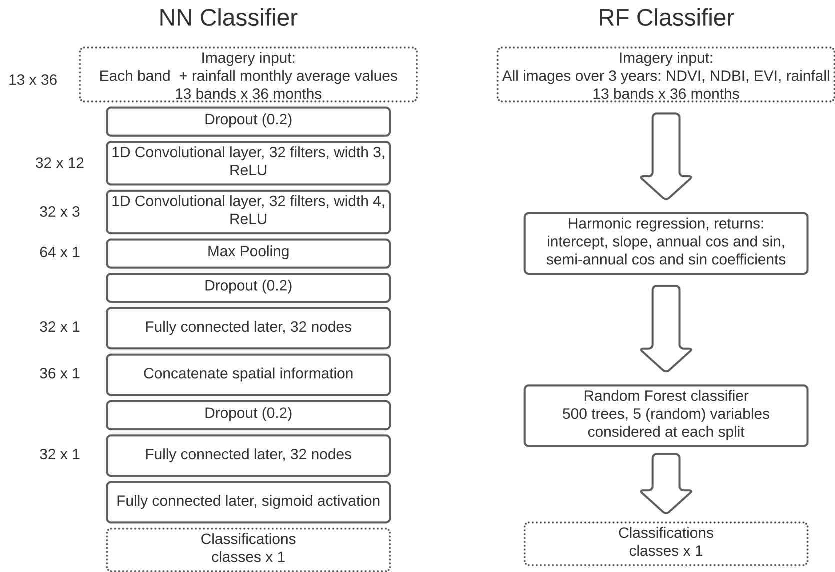
3.4.4 Empirical Strategy

The empirical strategy proceeds along the lines of the Figure 3.2. First I estimate a measurement model (using the neural network and random forest summarized in Figure 3.4 which are trained on hand-labeled 2019 landcover data and classifies any location's landcover based on the preceeding three years of satellite imagery along with location specific covariates. This is treated as the dependent variable. Then, two additional models are estimated. First, the probability of treatment for each unit is estimated using satellite imagery from 2006-2008 (before treatment was assigned) along with the other covariates. Seperately, the probability of each landcover class in 2019 is estimated using the same pre-treatment imagery and covariates. These predictions are subtracted from the true values, concentrating out the nuisance functions in each partially linear model. Finally, the residuals from the outcome model are regressed on the residuals from the treatment model, generating an estimate of the treatment effect. This is shown in Figure 3.2

3.4.5 Estimating land cover (measurement)

I regularly sample one out of every 100 pixels within 8km of any plot and inside the plots, excluding a buffer of 60m around all plots to prevent spectral contamination. Without knowledge of which villages were selected to be controls generating an ideal sampling strategy is difficult, but as far as I can tell the villages which were eligible for treatment are tightly clustered together. This sampling strategy likely misses some of the true control villages but more densely samples areas which are nearby the treated villages. Because many plots are smaller than 300 by 300 meters and unlikely to be sampled I also sample

Figure 3.4: Neural network (left) and random forest (right) structures. For more detail see the Supporting Information



the centroid of each plot to make sure that plots each have at least one sample, though this may affect the result if farmers act differently in the centers vs edges of plots. This process generates 10,520 points from 3,074 treated plots and 79,889 untreated points. The treated points are assigned a unique identifier for the point and also for the plot. The mean and median points per plot are 3.5 and 2 respectively. Untreated points are assigned an individual ID and a group ID, where groups are associated with square shaped areas 2km on a side. These groups have between 1 and 42 points in them with a mean and median of 28 and 30 respectively. These are combined with the plot IDs to generate geographic clusters which are used for cross-fitting and cluster-robust standard errors.

Data is extracted for each point (imagery and rainfall from 2005 to 2019 and location data from 2010) and the classifiers explained in the measurement section are applied. I save the class with maximum probability *and* the probabilities for each class.

While using the discrete classes is consistent with what one would do taking an off the shelf classification, I choose to use the class probabilities for two reasons. First, the discrete classes are sensitive to thresholding–by changing the threshold at which something is assigned to a class researchers can modify their classifiers to do better or worse at a particular class. Second, using only discrete classifications

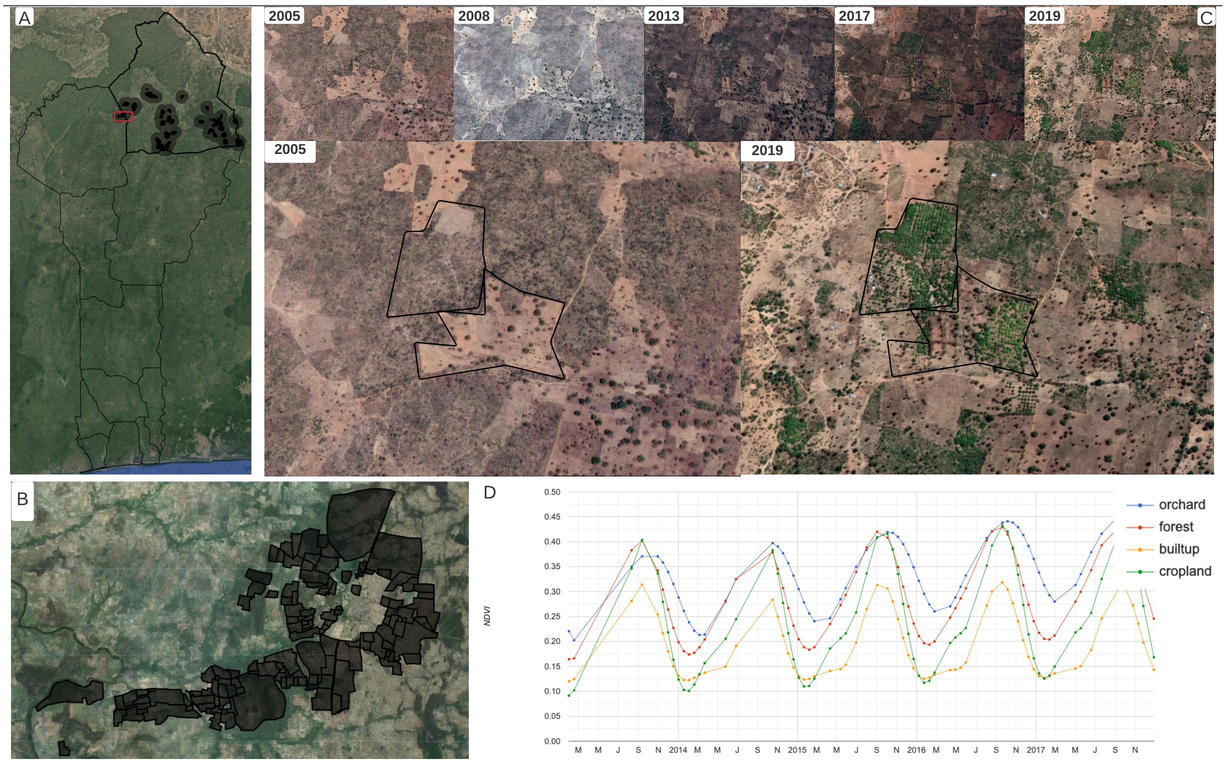


Figure 3.5: Panel A shows the outline of the Alibori department (dark black outline) in Benin (light black outline). The grey shaded areas are the regions from which untreated points are sampled, the black regions are the treated plots from which treated points are sampled. Panel B shows a village, located in the red box in Panel A. Note that it is difficult to predict which parts of the village will receive a title, even if the location of the village is known. Also note the plot-size discrepancies, and that plots cover a variety of different land classes. Panel C shows a series of high-resolution (1m) images of a region where some natural forest (dark brown) and some annual cropland (light brown) is converted to perennial tree-crops. Panel D shows how a time series of one of the bands (Normalized difference vegetation index, or NDVI) can differentiate between land cover classes. The neural network described here uses both the cross-sectional differences (orchard NDVI is almost always greater than the other classes) and time-series features (cropland has the largest seasonal swings) to measure the dependent variable and adjust for potential confounders.

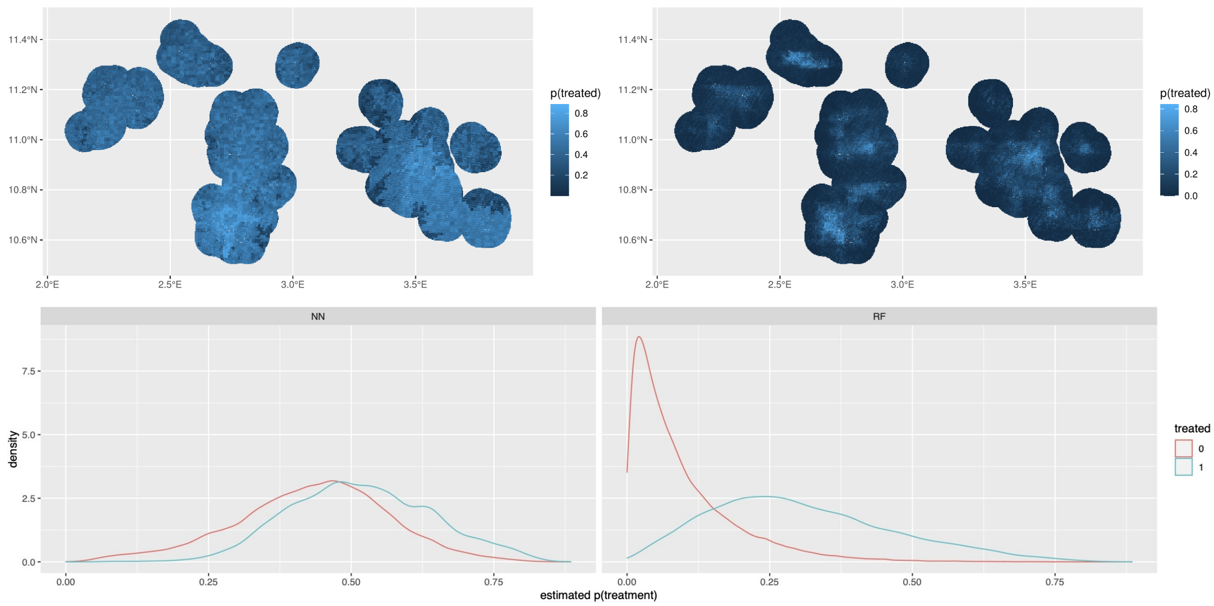


Figure 3.6: Estimates from neural network (NN) and random forest (RF) models predicting treatment from 2006-2008 imagery. Top left is the distribution of NN predictions. Top right is the distribution of RF predictions. Bottom row shows the density of probability estimates by true class. Note that because there are approximately eight times as many untreated as treated points, the count of (true) untreated points is higher than of (true) treated points except at very high probabilities (around 0.8) for both models.

needlessly throws away information and hides uncertainty.

3.4.6 Estimating $\hat{U}$

I estimate the treatment assignment model using satellite imagery from 2006-2008, before treatment was assigned in 2009-2011. Applying 10-fold cross fitting where assignment to each fold is done by cluster ID. This ensures that the out-of sample estimation of assignment is not done using the immediate neighbors of most points. Results from the two models are shown below in figure 3.6. Currently the RF model is has much higher accuracy and lower error than the NN model, though they are correlated ($R^2 = 0.44$). Residuals $\hat{U}$ are generated by subtracting the predicted probability from the true value ($D - \hat{D}$), or partialling out the part of treatment which can be predicted from how a place looks from space between 2006 and 2008. This corresponds to Figure 3.2 Panel C.

3.4.7 Estimating $\hat{V}$

I use the 2006-2008 imagery and covariates to estimate 2019 land cover classes using the same NN and RF models. I use the same 10-fold cross-fitting procedure using cluster ID. The RF is better able to predict landcover class probabilities than the NN, though both are able to perform the task with some accuracy. Not shown in the figure, there are positive correlations between RF and NN predictions for each class. Also note that the NN is predicting the classes from the NN measurement model and the RF is predicting classes from the RF measurement model. Residuals $\hat{V}$ are generated by subtracting estimated probabilities of each class from their measured probabilities ($Y_{landcover} - \hat{Y}_{landcover}$). This partials out what we can predict about 2019 landcover using 2006-2008 imagery. This corresponds to Figure 3.2 Panel B.

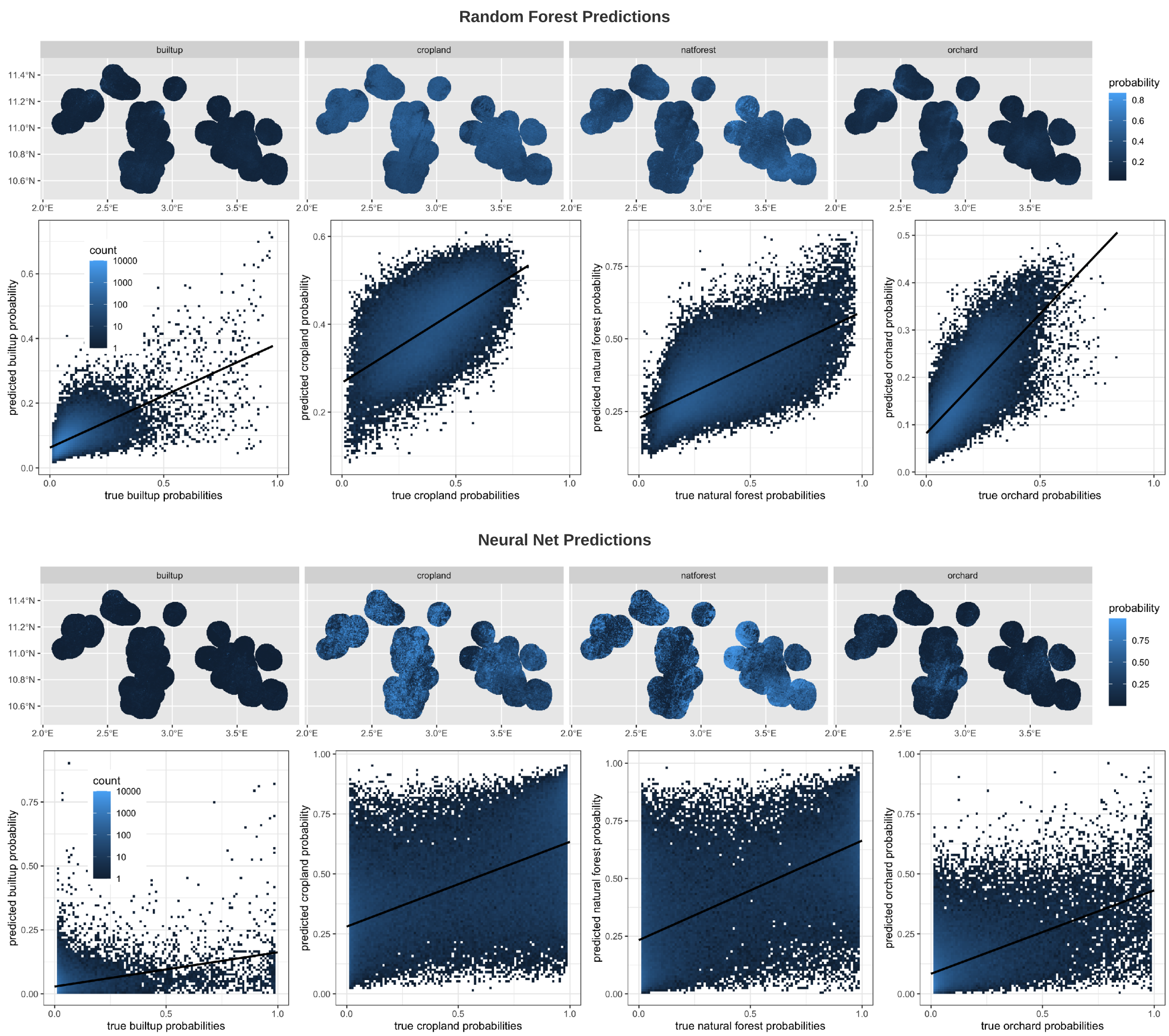


Figure 3.7: Estimates from neural network (NN) and random forest (RF) models predicting land cover class probability in 2019 using 2006-2008 imagery. Top panel shows spatial distribution of predictions for each landcover class for RF. Second panel shows measured probabilities of each class compared to predicted probabilities–color represents the number of scatter plot points in that bin. Third panel shows spatial distribution of predictions for each landcover class for NN. Fourth panel shows measured probabilities compared to predicted probabilities. Best fit lines are in black.

3.4.8 Results: Estimating $\hat{\beta}_1$

I estimate the equation $\hat{V} = \beta_0 + \beta_1 \hat{U} + \varepsilon$ once for each landcover class, with standard errors clustered at the cluster ID level. The results are shown in figure 3.8. The land titling program in Albori province resulted in an increase in cropland area in treated plots and a decrease in natural forest. There is no detectable effect for built-up areas or for orchard. Both the NN and RF models reach nearly exactly the same conclusion despite large differences in how they are constructed and the estimates they produce at each stage.

I compare the estimates obtained by the DML approach with three other approaches that researchers might choose and which do not account for confounders possibly present in satellite data. The first is a cross-sectional approach (Figure 3.9) using only classifications from 2019 and latitude, slope, elevation, and distance to city as covariates. Here the estimates produced with the two different sets of classifications diverge, with the NN model (more accurate at the measurement stage) producing much larger estimates. Though the direction and significance is the same, the point estimates are much larger. This also demonstrates a case when non-classical errors may be present: the differences in model predictions may be due to an error in the classification correlated with the probability of receiving treatment. This is a model which a researcher might use if they only had access to a single year of classifications and hoped to generate a suitable control group by sampling points nearby the treated areas.

The second estimator is a 1-1 matching estimator using the four location covariates: latitude, slope, elevation, and distance to city. Each treated point is matched to one unique untreated point based on euclidean distance in the covariate space. A researcher who wanted to non-parametrically control for the location covariates might use this approach. The estimates produced here are nearly identical to those produced in the cross-sectional estimator. Both of these approaches are likely biased because land which has crops is more likely to be assigned a title than land with natural forest.

The final estimator is a fixed-effects model where the researcher has access to a panel of classification probabilities for each landcover type. Two-way fixed effects are used to account for any time-varying confounders which affect all units equally, and unit fixed effects account for any location-specific confounders, including the four covariates included in the cross-sectional model. The coefficients are all

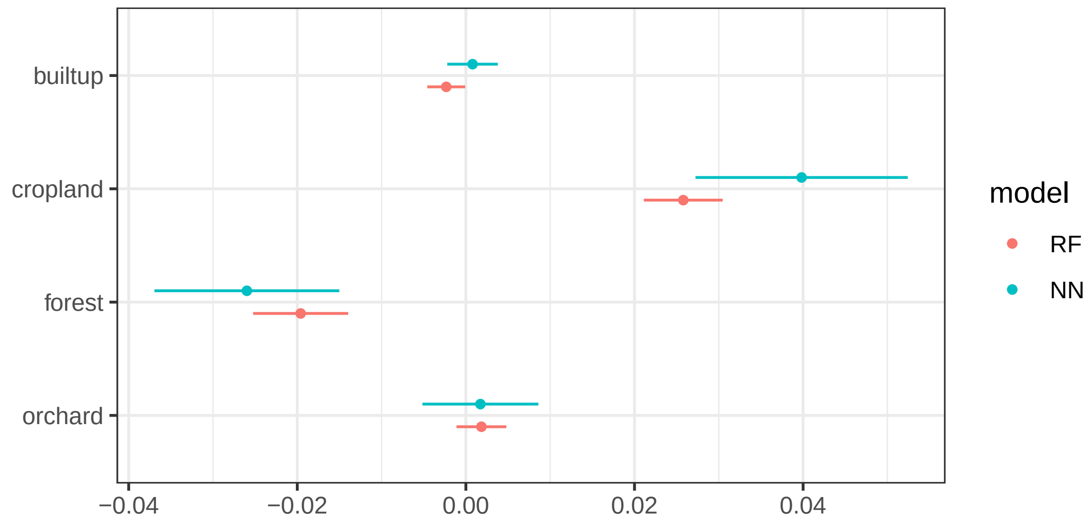


Figure 3.8: DML Method estimated effects and 95% confidence intervals of the Benin PFR program on the prevalence of four different landcover classes.

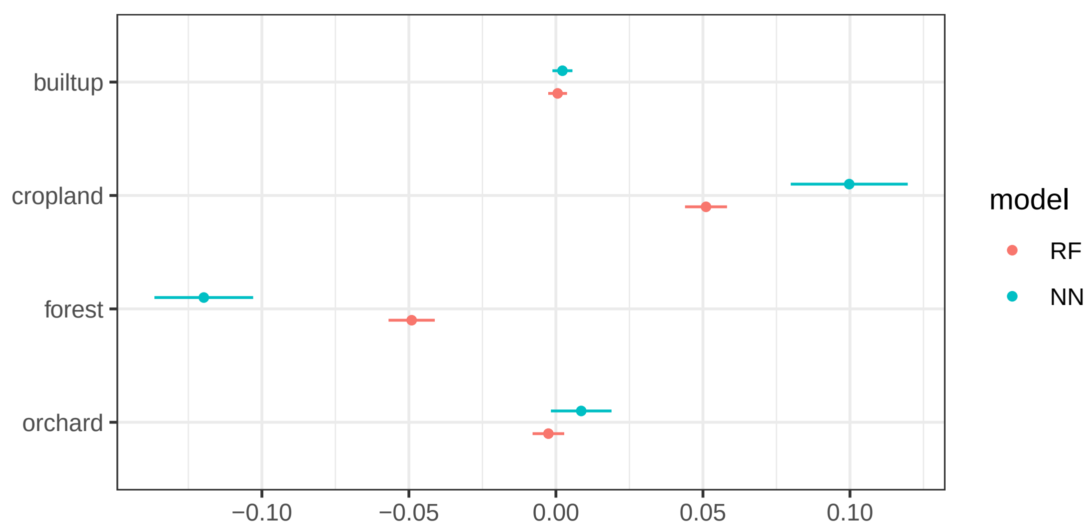


Figure 3.9: Cross-sectional estimated effects and 95% confidence intervals of the Benin PFR program on the prevalence of four different landcover classes using 2019 landcover and geographic covariates.

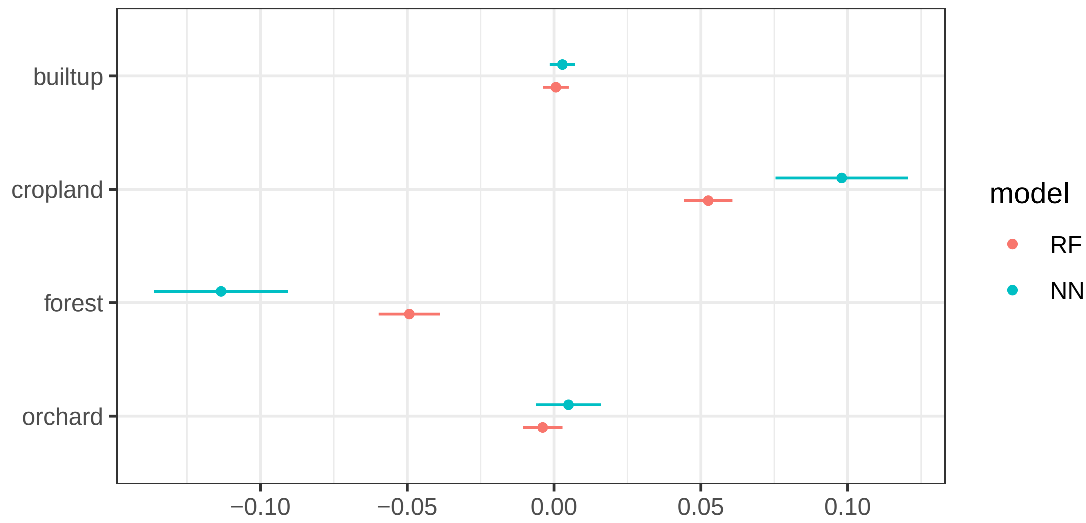


Figure 3.10: 1-1 Matching estimated effects and 95% confidence intervals of the Benin PFR program on the prevalence of four different landcover classes.

much closer to zero in this specification, but RF and NN measurement models generate very different conclusions. The NN model suggests an increase in orchard area and a decrease in cropland, while the RF measurement model suggests only an increase in cropland. Results are shown in Figure 3.11.

The consistency of the DML estimator despite differences in RF and NN classifiers at every stage suggests that the DML classifier is able to adjust for measurement error in the initial model. Though the RF measurement model most frequently mis-classifies cropland and natural forest and the NN model over-classifies orchards, the estimates of the PFR program are very similar. This suggests a DML approach when a researcher suspect non-classical measurement error [32].

3.5 Discussion

This differs from the findings of previous studies of the PFR project which estimated an increase in perennial orchard crops due to the titling project. In this very standard follow-up, between 4,000 and 5,000 households were surveyed across treated and control villages, and the researchers found an increase in perennial crops in both 1-year and 5-year follow ups. Neither of the effect sizes exceeded three percentage points, suggesting that there may have been more tree crop planting in about 2,100 of the 70,000 treated

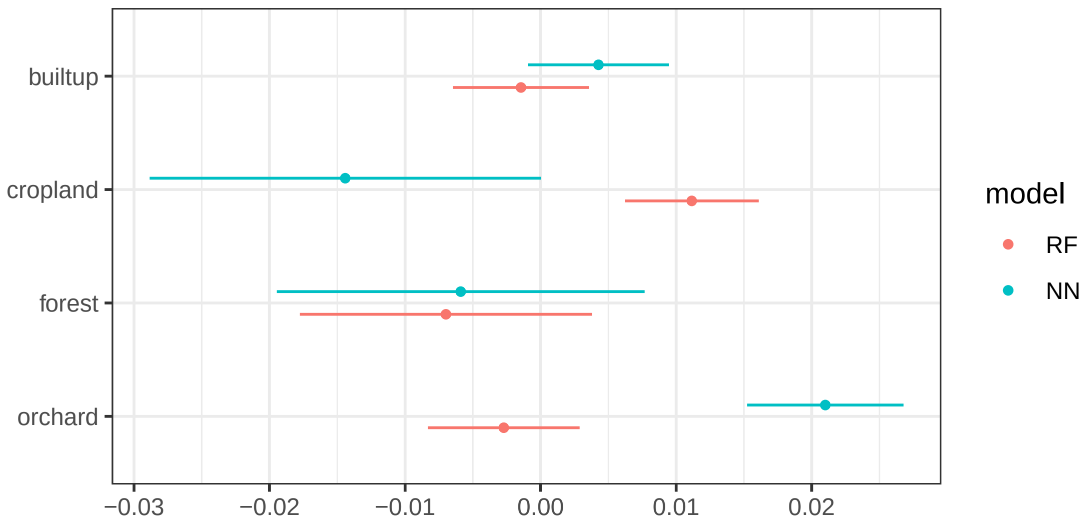


Figure 3.11: Two-Way Fixed Effects estimated effects and 95% confidence intervals of the Benin PFR program on the prevalence of four different landcover classes.

plots. Also note that these results are self-reported and at the parcel level and so cannot tell us what the change in landcover was.

There are several reasons why these findings might be different. First, previous studies treat the parcel as the unit of analysis meaning that even if a small portion of the parcel is converted to perennial crops, the whole plot is measured as having increased perennial crops. This paper uses a fixed area as the unit of analysis making it more likely to uncover a small effect size as even parcels with perennial crops tend to only have a small fraction of the total area with those crops. Second, my analysis only investigates approximately 3,000 treated parcels in one department, and the effects may vary across agro-climatic zone or ethnic composition of the region. A third possibility is there may be a discrepancy between what people report in surveys and what is visible in satellite imagery. Finally, the presence of one or more confounders which is unmeasured in previous studies but can be captured using remote sensing data might change the measured results.

More broadly, the methods employed in this study allow for the analysis of all 70,000 treated plots rather than a small sample of survey respondents, potentially yielding more power to test hypotheses about whether the effects land tenure formalization differ by gender. Similarly, Benin spans several agro-climatic zones from a wet, tropical climate in the south suitable for palm plantations to arid Sahel in the north where tree crops are cashews and shea trees. It also allows us to investigate differences across political characteristics of Benin's departments and ethnic characteristics of the farmers.

Outside of the Benin case, this methodology can be retroactively applied to any land tenure project where cadastral information was recorded, allowing re-analysis of many past projects or long-term follow up analyses. The potential is also not limited to land tenure security studies. Any intervention which has a geographically delineated treatment area and confounders which could be visible in satellite imagery could benefit from the strategy outlined above. In areas where there is not good data on why certain areas were selected for an intervention (environmental protection, agricultural assistance, etc.) this strategy can help to generate balance on any factor which changes how land is used (and thus appears in satellite imagery).

Like any procedure for observational data this procedure is not guaranteed to achieve an unbiased estimate of the treatment effect which one could achieve through randomization and a precise and unbiased measurement strategy. While some types of confounders can be adjusted for, others might still bias the results.

The network used to approximate the $\hat{g}()$ and $\hat{m}()$ functions should be able to detect a variety of possible confounders. These include the pre-treatment landcover class and variations in that class which predict treatment or the outcome. For example, if a location is covered by annual crops before treatment is assigned, that would make it more likely to be assigned treatment and more likely to be cropland in 2019 (compared to, for example, the large amount of publicly owned, natural forest in Alibori). However, among annual cropland areas, if those crops begin to grow earlier in the season than other areas it might mean that there is access to a water source, making that location both more likely to be owned (and thus receive treatment) and more likely to transition to perennial crops because of the suitable location. Similar remote sensing methods have been shown to detect crop type, crop yields, use of fertilizer or irrigation, income, and ecosystem health, among other things, all of which are potential confounders. More research is needed on how well the methods proposed here can account for various classes of confounding variables.

However, confounders which are not apparent in satellite imagery can still pose problems. I outline two examples of issues which could bias evaluations of land tenure which are unlikely to be resolved using the DML satellite imagery approach. A first example is savings: farmers who have some money saved are more likely to be able to afford the fee which is associated with obtaining a copy of the legal title *and* to afford the inputs necessary to switch some of their land to perennial crops. If those savings do not affect the farmer's behavior in a way which is visible in satellite imagery they will induce bias in the estimate of the treatment effect. A second example is someone who has the skills necessary to grow perennial crops but who has not done so because of low land tenure security. This individual is more likely to attempt to obtain a title and more likely to to switch to perennial crops. Because the farmer's skills are not visible until they switch to perennial crops this is not something which can be adjusted for using satellite imagery.

There are several other issues which should be addressed in future work. First, to my knowledge there is not yet a result which demonstrates that deep, convolutional networks achieve the necessary $n^{-1/4}$ rate of convergence necessary for valid confidence intervals. At present this is a necessary assumption, though it could be sidestepped by using a transfer learning approach. This would entail pre-training the convolutional part of the network and then freezing those weights but allowing the fully connected layers to update in the training. Then the network would just take a fixed transformation of the imagery input and run it through a deep neural network, which has been shown to have the $n^{-1/4}$ property[33, 43].

A second issue is obtaining proper standard errors for treatment estimates in this approach. This is complicated by data and measurement issues, propagation of errors across the DML approach, and more standard issues like spatial autocorrelation. While formal methods like Conley might be incorporated, a likely best-first-step approach would be boot-strapping the entire process over sub-samples of data.

A final potential area for future research is to further test whether the approach suggested here can adjust for the non-classical errors issue described in [32]. If the same model is used for both measurement and for de-biasing the estimate, then the bias in the measurement may be concentrated out by the $\hat{m}()$ function. For example, imagine that the geography of a region results in natural forest being frequently mis-classified as perennial crops in the measurement model. This will bias the estimated effect of land titling if that geography also plays a role in assignment of land titles. Because the geographic characteristics which result in mis-classification will likely also appear in pre-treatment satellite imagery they will automatically be removed from the outcome variable used in the estimation of the ATE. While theoretically this should be the case, further research is necessary to establish under which conditions this applies.

This is one of the first papers to show the potential for using satellite imagery to directly help with inference (rather than measurement). It is clear that features which can be measured in satellite images are potentially confounders in many applications in which the outcome is observable in satellite imagery, and the double machine learning approach is a way to control for those confounders using state of the art remote sensing techniques. I show how the can be used in the Benin land titling setting, showing that land titling resulted in increased cropland expansion in titled areas and an decrease in forest.

3.5.1 Dimensions of imagery

Satellite images are a matrix of pixels where the length of a pixel edge is the "spatial resolution" of of the image. Some commercially available imagery has a spatial resolution of 0.3 meters, but the most commonly used publicly available imagery has a maximum resolution of 10m (Sentinel 2) or as low as 1000 meters (MODIS). Spatial resolution determines the minimum size of an object which can be studied directly. Each pixel contains measurements for reflectance of different wavelengths of light. These include wavelengths which the human eye can percieve (450-520 nanometers is the color blue, 520-600 is green, 600-690 is red) as well as wavelengths which are both shorter and longer than we can perceive. "Spectral resolution" is the number of wavelengths which are measured. Different surfaces and materials have different spectral signatures, so higher spectral resolution can better distinguish the composition of the surface of an object in a pixel. "Temporal resolution" is the frequency with which a location is revisited and imaged by an instrument on a satellite. An object in a pixel may move or change over time so a higher temporal resolution allows researchers to identify the date at which changes occur. In particular, phenological patterns of vegetation over time can be useful for measuring land cover.

Different satellite-mounted instruments face tradeoffs in these three types of resolution. For example, the MODIS sensor has a temporal resolution of 1-2 days, a spectral resolution of 36 different wavelengths or "bands" and a spatial resolution of 250-1000 meters (varying by band). The Landsat 7 Thematic Mapper sensor has a temporal resolution of 14 days, a spectral resolution of 7 bands, and a spatial resolution of 30m. These tradeoffs determine the shape of the tensor which researchers can use to study a particular area. Take a 1km $\times$ 1km area in which a researcher wants to measure an outcome. Using MODIS a researcher would have a $365 \times 36 \times 1 \times 1$ array of input data for that location. In contrast, using Landsat TM, the researcher would have a $26 \times 7 \times 33 \times 33$ array of input data. While an in-depth discussion of the tradeoffs for measurement of each is beyond the scope of this paper, the remote sensing community has come up with sophisticated strategies for using particular sensors (or combinations of sensors) to measure outcomes of interest.

3.5.2 Machine Learning Models

The general set of algorithms I describe build upon state of the art methods for remote sensing using publicly-available imagery, making them more accurate and more robust to inter-annual variation and changes in geography. Additionally, these algorithms are designed to work across contexts and be useful for researchers in a variety of settings. The goal is to take at least one year of multi-spectral satellite images, potentially with missing data due to cloud cover, and accurately classify an outcome of interest to the researcher. To do this I compare two classifiers: a random forest which takes the coefficients from a harmonic regression on pre-defined remote sensing indices (NDVI, EVI, NDBI[10]) as inputs, and a deep convolutional neural network which takes the raw data as inputs.

Broadly, the challenge for any machine learning model is to take high-dimensional data, summarize it into useful "features" and then use those to solve a classification or regression problem. Here, the problem is to take a year's worth of satellite images of a particular location across multiple spectral bands to detect the type of landcover at that location. This may not seem like "high dimensional data" because there are nine spectral bands (measured by the Landsat ETM+ sensor) and 12 months (after monthly averaging, up to 26 images without), resulting in "only" 108 variables. However, the information which we know is useful for the classification problem is a complicated function of these variables. A traditional approach to this problem might be to use the difference between the maximum and minimum NDVI and some threshold values to estimate landcover where NDVI is $NDVI = \frac{(NIR-Red)}{(NIR+Red)}$, clearly a non-linear transformation as in [46]. This suggests that to be effective, a classifier must be able to discover and use the extremely large number of non-linear transformations of those 108 inputs.

The fundamental question in remote sensing measurement papers is how to approach this problem. There are two general approaches: generate researcher-designed features designed to capture some aspect of known physical or biological processes and use those in classification, or feed in raw data and let a

[10] These stand for "Normalized Differenced Vegetation Index", "Enhanced Vegetation Index", and "Normalized Differenced Built-up Index" and are all functions of the spectral bands which highlight certain land cover types or characteristics

machine learning algorithm "discover" features. In this paper I present two approaches. The first is a hybrid approach which makes use of several often-used indices and the knowledge that vegetation tends to follow an annual cycle in its growth resulting in a harmonic pattern. It summarizes that pattern in six variables representing mean value, overall trend, and annual and semi-annual harmonic frequencies of each index. These variables are then used as inputs in a random forest classifier (Figure 3.12). The second takes the raw data approach, by feeding the raw data (each year as a 13×12 matrix) into a deep, convolutional neural network (Figure 3.13).

Each of these classifiers is designed to target two sources of information. First, different types of surfaces reflect and absorb different wavelengths of light, resulting in a "spectral signature" which is useful for classification. Different types of vegetation have different spectral signatures, though differences between vegetation classes are often not large. Second, different species of plants undergo annual patterns of growth, responding to seasonal changes in climate. These phenological patterns affect the spectral signatures of vegetation over time. The classifiers used in this study are designed to make use of both cross-sectional differences in the spectral signatures of locations, and the time-series information in phenological patterns of vegetation. Both models have relatively few degrees of freedom, and do not require vast amounts of training data–the deep neural network has 7,912 trainable parameters (over two orders of magnitude fewer parameters compared to similar methods [45]) and converges quickly and consistently with the data used in this study which was hand-labeled in 12 hours by one researcher.

The models are augmented with several other sources of data designed to improve (nearby) out of sample prediction. Both use slope, elevation, monthly mean rainfall, and distance to major city as inputs. Preliminary evidence suggests that while these do not improve prediction on out of sample points in the same year, they do improve prediction when a model trained on some years is applied to other years.

As the application focuses on the PFR program as implemented in the Alibori department, all training data is sampled from that area. I used the high-resolution Google basemap from 2019 to hand-label 200 polygons of each of four landcover classes: builtup areas, annual cropland, perennial (tree) cropland, and natural forest. As tree crops are by far the most rare, this sampling was done by attempting to identify tree crops across the range of latitude and longitude spanned by Alibori, labeling those, and in the spatial vicinity of each parcel of tree crops also identifying an area of each of the other land classes. This stratified sampling approach maintains balance across classes in the training data. I sample 25 points from each labeled polygon from the year 2019 and assign each point a label corresponding to the landcover identified in that polygon. To augment the amount of training data I have, I assume that there has been minimal landcover change in the training data and use the 2019 labels for data from the years 2016-2019. While this introduces some noise into the training data it allows for a more complex machine learning model[47].

Using Google Earth Engine's javascript API I extract surface reflectances after cloud and cloud shadow pixels have been masked (using Landsat 7 ETM+ images), total monthly rainfall (using the CHIRPS dataset), as well as slope and elevation from a digital elevation model and distance to a major city (measured in the year 2000). The imagery and rainfall data is extracted for each year from 2014 to 2019.

Random Forest

Data proparation for the random forest classifier begins by generating NDVI[11], NDBI[12], and EVI[13]. For each year of labels (2016-2019) I take the three years up to the target classification year (for example, 2019 labels are associated with imagery and rainfall from 2017-2019). For each point, I run a harmonic regression with an intercept, a slope term, annual sin and cos terms, and semi-annual sin and cos terms on each of NDVI, NDBI, EVI, and rainfall. Then, for each point-year label there are six harmonic regression terms for each of the four time-series inputs, as well as the slope, elevation, longitude, and elevation of the point. These serve as the inputs to the random forest model.

The random forest model [48] contains 500 trees with five variables randomly sampled at each split. 75% of the data are used for training, with 25% held out for testing. The training/testing split is stratified on the sampled cluster, so each group of four polygons is assigned either training or testing to prevent data leakage. I use data from each year for training and testing, but also test on out of sample locations and years in table 3.3.

Table 3.1 shows the confusion matrix from the RF model with all inputs. The most common misclassification errors are across natural forest and cropland, potentially because some of the areas labeled

[11] $NDVI = \frac{(NIR-Red)}{(NIR+Red)}$

[12] $NDBI = \frac{(SWIR-NIR)}{(SWIR+NIR)}$

[13] $EVI = \frac{2.5*(NIR-Red)}{(NIR+6*Red-7.5*Blue+1)}$

cropland in 2019 were converted from natural forest since 2016. Table 3.2 shows the classification specifics across the four classes.

Table 3.1: Confusion matrix for the random forest model. 76% overall accuracy.

	builtup	cropland	natforest	orchard
builtup	5864	609	364	333
cropland	347	5422	1269	545
natforest	112	1219	4050	309
orchard	266	770	579	4582

Table 3.2: Statistics by class for the random forest model.

	Sensitivity	Specificity	Precision	Recall	F1	Balanced Accuracy
Class: builtup	0.89	0.93	0.82	0.89	0.85	0.91
Class: cropland	0.68	0.88	0.72	0.68	0.69	0.78
Class: natforest	0.65	0.92	0.71	0.65	0.68	0.78
Class: orchard	0.79	0.92	0.74	0.79	0.77	0.86

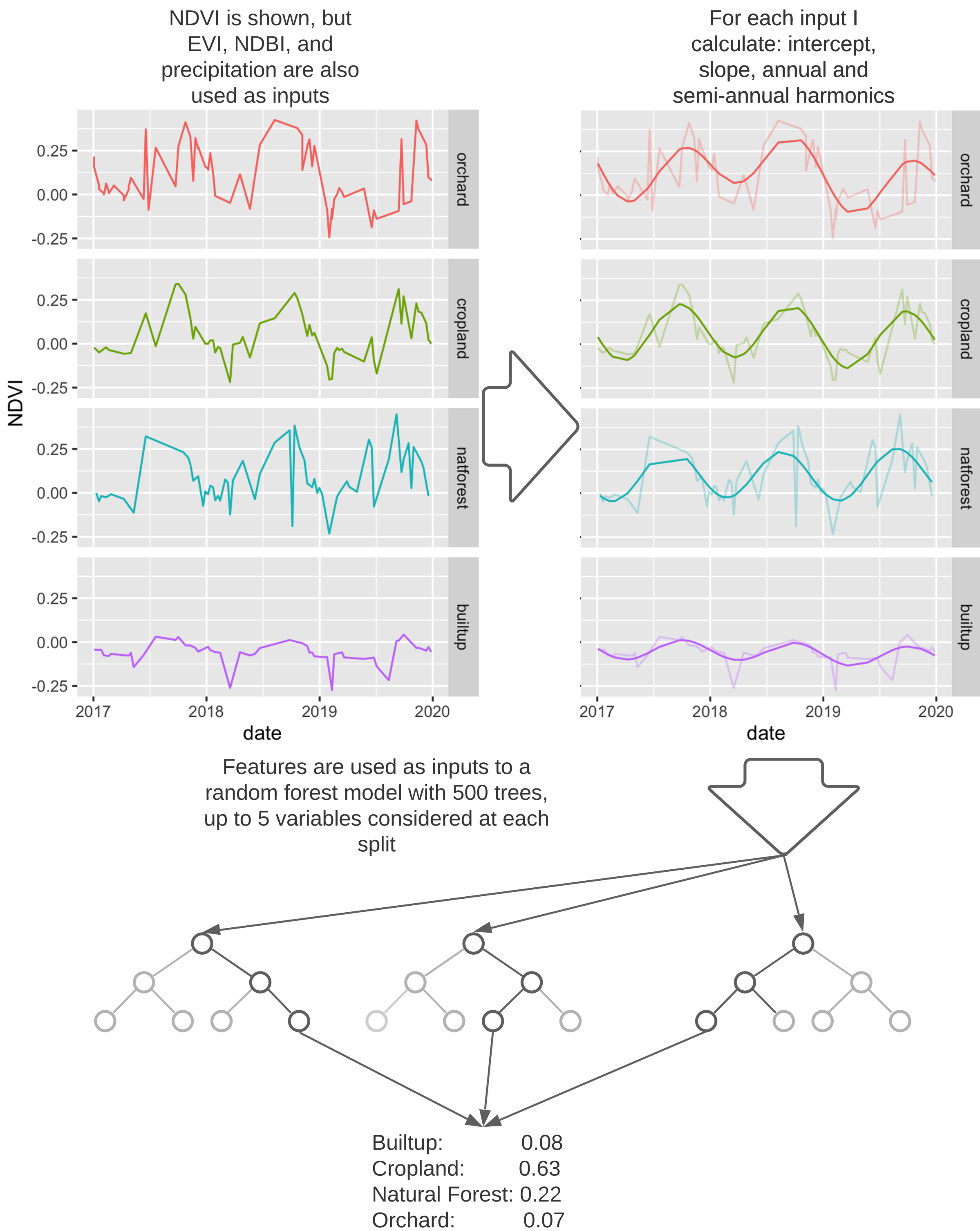
Table 3.3 shows six different combinations of input variables and their overall accuracy. Row 1 corresponds to the RF explained above. Row 2 does not include Longitude or distance to city, and loses 3% classification accuracy. Row 3 uses only imagery and rainfall and no location specific characteristics. Row 4 uses only imagery (no rainfall). Row 5 uses only NDVI, a common approach of many who use remote sensing data to measure vegetation. For 6 uses only location specific data. Rows 7-9 use 2016-2018 as training data and then predict on out of sample locations for 2019.

I expected geographic characteristics and rainfall to significantly improve performance on out of sample years, but there is a relatively small decrease in accuracy from a model which uses all inputs to one which uses only imagery, suggesting that at least in the context of Alibori there is not much to be gained from including slope, elevation, latitude and distance to city.

Table 3.3: Overall accuracy for random forest with various inputs and predictions on years which are in or out of sample. Imagery consists of NDVI, NDBI, and EVI. Location refers to slope, elevation, longitude, and distance to city.

	imagery	rainfall	location	year_out_of_sample	accuracy
1	yes	yes	yes	no	0.76
2	yes	yes	slope and elevation	no	0.73
3	yes	yes	no	no	0.72
4	yes	no	no	no	0.70
5	NDVI	no	no	no	0.60
6	no	no	yes	no	0.47
7	yes	yes	yes	yes	0.66
8	yes	no	no	yes	0.63
9	no	no	yes	yes	0.47

Figure 3.12: Random Forest classifier



Neural Network

Following [45] I use one dimensional convolutional neural network which takes one year's worth of satellite imagery across all available bands of the image. This data is augmented with location specific data about the slope, elevation, latitude, and distance to major city of each location.

The neural network is constructed to learn spectral time-series representations of phenological characteristics of vegetation. It learns to recognize how different types of vegetation in different conditions responds to changes in the seasons. Following the result in [45] for now I use only the time series and spectral variation, and not the spatial information to perform the classification. This means that the network cannot recognize spatial features represented by more than one pixel. [45] found that in crop classification such spatial features did not improve the performance of their network, even with very high resolution satellite imagery. The possibility of including spatial features, especially in built environments, is something which I hope to explore in later work.

Figure 3.13 shows the structure of the network which I use for all three machine learning tasks: measuring the outcome of interest, estimating the treatment nuisance function, and estimating the outcome nuisance function. Though the approach is similar to [45] the structure is slightly different. Notably, this network includes monthly rainfall as well as static covariates to improve classification. Because I expect most researchers to have less training data I estimate fewer parameters (by a factor of over 100) and attempt to make the network more robust to missing data from cloudiness and more targeted at plant phenology. Other ongoing work compares some of the features in neural networks for land cover classification.

Data preparation for the Neural Network (NN) follows these steps: 1) generate NDVI, NDBI, and EVI variables. 2) Aggregate all imagery to the monthly level. 3) Normalize all variables by subtracting the mean and dividing by the standard deviation. 4) For any missing values, replace with positive 10–a number outside of what is observed in the data. The NN learns to ignore these values. 5) Like the RF, associate labels with the previous three years of imagery and rainfall. The input to the neural network for each point-year is a 13×36 matrix of band-months and a 4×1 vector of location inputs.

The NN begins with a dropout layer, randomly deleting 20% of the input values to prevent over-training. Then are two 1-dimensional convolutional layers where the network learns to identify features which are useful for classification. The first 1-D Conv layer considers all of the bands, three months at a time. There are 32 filters which can each learn a representation, so this layer returns 32 activations for each "season," or a 32×12 matrix. The second 1-D conv layer does the same, but considers a year at a time, returning a 32×3 activation-year matrix. These are fed into a max-pooling layer, which just takes the maximum activation of each feature across the three years, returning a 32×1 vector of activations. These are fed through a dropout layer, and then a 32-unit fully connected layer, allowing the network to represent non-linear functions of those feature activations. The static variables are concatenated to the 32 unit dense layer, fed through another dense layer, and then to the final layer with four units, each of which represents one of the landcover classes. These are constrained to sum to 1 and can be interpreted as probabilities of the different classes.

For each location I get three years of monthly average reflectance data for the 9 bands that Landsat 7 ETM+ natively records as well as three calculated bands: Normalized difference vegetation index (NDVI), Enhanced vegetation index (EVI), and Normalized difference built-up index (NDBI) which are used to better distinguish vegetation characteristics and built-up areas. A 13th channel is added with monthly total precipitation. This means that each point is represented by a 13×36 matrix of band values. Each band is standardized to be mean 0 and standard deviation 1. Missing data (mostly due to cloud cover) is filled with the value 10, well outside the distribution of observed values after standardization.

Each matrix is first passed through a dropout layer where 20% of values are randomly deleted. This is to help prevent overfitting, or where the network learns to uniquely match certain patterns to specific units. The data is then passed through two consecutive one-dimensional convolutional layers. Each has 64 filters. The first has a filter width and stride of three, meaning it takes the first three months and tries to identify up to 64 distinct patterns across all 13 bands (called channels in the computer vision literature). For that first location it outputs 64 "activations" which reflect how well the first three months of input data match the patterns learned by the network. It then strides three months over, and looks for the same patterns over the months 4-6, again producing 64 activations for that location. This continues through all 36 months, resulting in 12 sets of 32 activations.

The second convolutional layer sees 12 time periods with 32 channels (corresponding to the 64 activations from the previous layer). The second layer has a width and stride of 4, corresponding to one year of inputs. This produces three sets of 64 activations from these filters, a measurement of how well each year fits one of 64 different patterns. These are then passed through a max pooling layer which just takes the maximum activation across the three years. These are the “features” which are used for downstream classification.

The information is passed through another dropout layer, and then a dense layer with 32 nodes, each of which produces a weighted average of the features (or node values) in the previous layer passed through the ReLU function. Then, the activations of those 32 nodes are concatenated with (standardized) static characteristics of the location of the unit (slope, elevation, latitude, and distance to major city). The data is then passed though another dropout layer, another dense layer, and finally a classification layer.

I use an RMSprop optimizer and a categorical cross-entropy loss function with a default learning rate of 0.01. I use a batch size of 256 and a maximum of 250 epochs, though the models tend to converge in fewer than 50 epochs in most cases. While these hyper-parameters can be tuned to particular applications, this setup should be suited to a broad range of tasks and is intended to work out of the box.

Table 3.4 shows the confusion matrix from the RF model with all inputs. The most common misclassification errors are when orchard is misclassified as cropland, or when cropland is misclassified as natural forest. Table 3.5 shows statistics by landcover class–builtup and orchard have the highest balanced accuracy, but orchard has the lowest precision, indicating that the network over-classifies orchard.

Table 3.4: Confusion matrix for the neural network model. 90% overall accuracy.

	builtup	cropland	natforest	orchard
builtup	5805	284	77	64
cropland	122	5139	481	128
natforest	21	332	5668	219
orchard	53	507	355	6320

	Sensitivity	Specificity	Precision	Recall	F1	Balanced Accuracy
Class: builtup	0.97	0.98	0.93	0.97	0.95	0.97
Class: cropland	0.82	0.96	0.88	0.82	0.85	0.89
Class: natforest	0.86	0.97	0.91	0.86	0.88	0.92
Class: orchard	0.94	0.95	0.87	0.94	0.91	0.95

Table 3.5: Statistics by class for the neural network model.

Table 3.6 displays a similar exercise, where the neural net is given various inputs and overall accuracy is assessed. Surprisingly, there is not much variation as long as the imagery is used. There is no decrease in accuracy with the removal of the location specific covariates or the rainfall data. Performance decreases when inputs are limited to the calculated indices, and especially when limited to NDVI.

Table 3.6: Overall accuracy for neural network with various inputs and predictions on years which are in or out of sample. Imagery consists of all nine ETM+ bands, NDVI, NDBI, and EVI. Location refers to slope, elevation, longitude, and distance to city.

	imagery	rainfall	location	year_out_of_sample	accuracy
1	yes	yes	yes	no	0.90
2	yes	yes	slope and elevation	no	0.89
3	yes	yes	no	no	0.89
4	yes	no	yes	no	0.89
5	NDVI,NDBI,EVI	yes	yes	no	0.84
6	NDVI	yes	yes	no	0.80
7	no	no	yes	no	0.29
8	yes	yes	yes	yes	0.87
9	yes	yes	slope and elevation	yes	0.85
10	yes	yes	no	yes	0.88

There is a much smaller decrease in performance on out of sample years using the NN than the RF model suggesting the features that the NN detects are more robust to temporal variation. Like in the RF model, introduction of rainfall and location specific characteristics weakly improve the performance of the model,

I compare these classifications to what one would obtain using off the shelf data. Copernicus's global land cover layers are 100m spatial resolution and have five years of coverage and a number of different landcover classes, making it perhaps the most suitable landcover dataset for this study [30]. I extracted labels from Copernicus 2019 landcover layer [49] and aggregated the classes into the most comparable classes, then compared those to the labels I assigned using the google basemap:

Table 3.7 shows that While there is impressive agreement in the builtup class, things are muddied afterwards. This demonstrates that one issue with using off the shelf satellite products is that even very high quality products may not contain the classes one wishes to study (in this case tree crops). A second issue is that global landcover datasets are impressive because of their global reach and consistency, but

Table 3.7: Confusion matrix of hand-labeled classification on the vertial axis and classifications from Copernicus on the horizontal axis.

	builtup	cultivated	forest	other veg	unknown
builtup	6675	340	15	140	0
cropland	110	5870	210	1160	233
natforest	0	1690	1575	2191	234
orchard	105	3010	1710	1122	250

often inaccurate in a specific context, even when that context is an entire country. I suspect this is one of the leading causes of the non-classical errors described in [32].

This suggests that each of these classifiers would improve upon an off the shelf measure of landcover and would extend the researcher's ability to measure land cover in the target region in years when global landcover datasets are not available.

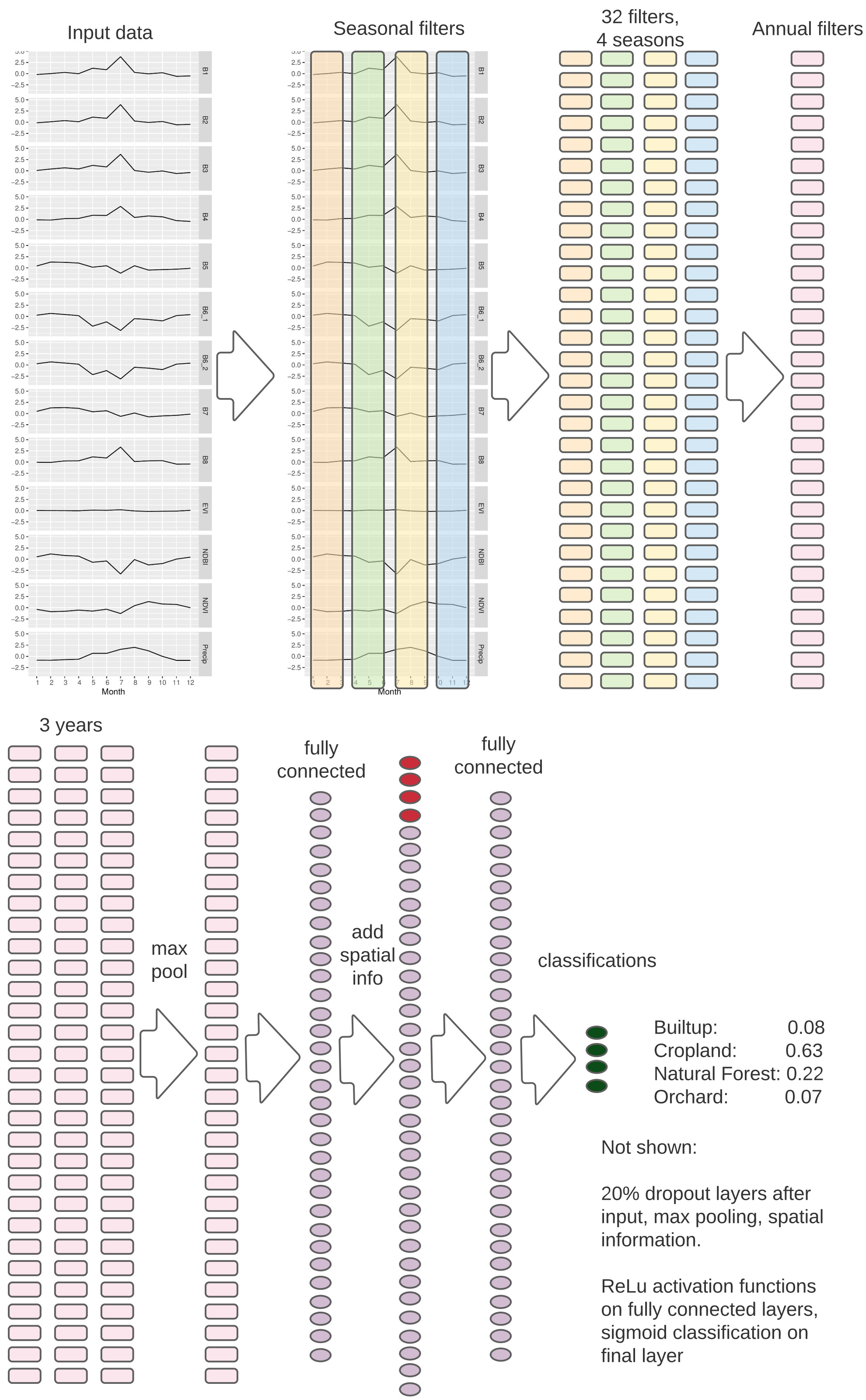


Figure 3.13: Convolutional neural network classifier

References

(1) Tseng, T.-W. J.; Robinson, B. E.; Bellemare, M. F.; BenYishay, A.; Blackman, A.; Boucher, T.; Childress, M.; Holland, M. B.; Kroeger, T.; Linkow, B.; Diop, M.; Naughton, L.; Rudel, T.; Sanjak, J.; Shyamsundar, P.; Veit, P.; Sunderlin, W.; Zhang, W.; Masuda, Y. J. *Nature Sustainability* **2020**, 1–10.

(2) Besley, T. *Journal of Political Economy* **1995**, *103*, 903–937.

(3) Lawry, S.; Samii, C.; Hall, R.; Leopold, A.; Hornby, D.; Mtero, F. *Journal of Development Effectiveness* **2017**, *9*, 61–81.

(4) Place, F. *World Development* **2009**, *37*, 1326–1336.

(5) Goldstein, M.; Udry, C. *Journal of Political Economy* **2008**, *116*, 981–1022.

(6) Fenske, J. *Journal of Development Economics* **2011**, *95*, 137–156.

(7) Ali, D. A.; Deininger, K.; Goldstein, M. *Journal of Development Economics* **2014**, *110*, 262–275.

(8) Bellemare, M. F. *Land Economics* **2013**, *89*, 272–290.

(9) Bambio, Y.; Bouayad Agha, S. *World Development* **2018**, *111*, 130–147.

(10) Deininger, K.; Jin, S. *European Economic Review* **2006**, *50*, 1245–1277.

(11) Deininger, K.; Ali, D. A.; Alemu, T. *Land Economics* **2011**, *87*, 312–334.

(12) Holden, S. T.; Deininger, K.; Ghebru, H. *The Journal of Development Studies* **2011**, *47*, 31–47.

(13) Place, F.; Otsuka, K. *Journal of Development Studies* **2002**, *38*, 105–128.

(14) Goldstein, M. P.; Houngbedji, K.; Kondylis, F.; O'Sullivan, M. B.; Selod, H. *Formalizing Rural Land Rights in West Africa : Early Evidence from a Randomized Impact Evaluation in Benin*; Policy research working paper series 7435; The World Bank, 2016.

(15) Lawin, K. G.; Tamini, L. D. *The Journal of Development Studies* **2019**, *55*, 177–190.

(16) Lawin, K. G.; Tamini, L. D. *Journal of Agricultural Economics* **2019**, *70*, 116–134.

(17) Goldstein, M.; Houngbedji, K.; Kondylis, F.; O'Sullivan, M.; Selod, H. *Journal of Development Economics* **2018**, *132*, 57–74.

(18) Wren-Lewis, L.; Becerra-Valbuena, L.; Houngbedji, K. *Science Advances* **2020**, *6*, eabb6914.

(19) Gavian, S.; Fafchamps, M. *American Journal of Agricultural Economics* **1996**, *78*, 460.

(20) Levin, N.; Kyba, C. C. M.; Zhang, Q.; Sánchez de Miguel, A.; Román, M. O.; Li, X.; Portnov, B. A.; Molthan, A. L.; Jechow, A.; Miller, S. D.; Wang, Z.; Shrestha, R. M.; Elvidge, C. D. *Remote Sensing of Environment* **2020**, *237*, 111443.

(21) Burke, M.; Driscoll, A.; Lobell, D. B.; Ermon, S. *Science* **2021**, *371*, DOI: 10.1126/science.abe8628.

(22) Hillson, R.; Coates, A.; Alejandre, J. D.; Jacobsen, K. H.; Ansumana, R.; Bockarie, A. S.; Bangura, U.; Lamin, J. M.; Stenger, D. A. *International Journal of Health Geographics* **2019**, *18*, 16.

(23) Jean, N.; Burke, M.; Xie, M.; Davis, W. M.; Lobell, D. B.; Ermon, S. *Science* **2016**, *353*, 790–794.

(24) Reba, M.; Seto, K. C. *Remote Sensing of Environment* **2020**, *242*, 111739.

(25) Chen, S.-W.; Wang, X.-S.; Xiao, S.-P. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2018**, *11*, 2657–2667.

(26) Alshehhi, R.; Marpu, P. R.; Woon, W. L.; Mura, M. D. *ISPRS Journal of Photogrammetry and Remote Sensing* **2017**, *130*, 139–149.

(27) Hansen, M. C.; Potapov, P. V.; Moore, R.; Hancher, M.; Turubanova, S.; Tyukavina, A.; Thau, D.; Stehman, S.; Goetz, S.; Loveland, T., et al. *science* **2013**, *342*, 850–853.

(28) Song, X.-P.; Hansen, M. C.; Stehman, S. V.; Potapov, P. V.; Tyukavina, A.; Vermote, E. F.; Townshend, J. R. *Nature* **2018**, *560*, 639.

(29) Jin, Z.; Azzari, G.; You, C.; Di Tommaso, S.; Aston, S.; Burke, M.; Lobell, D. B. *Remote Sensing of Environment* **2019**, *228*, 115–128.

(30) Kerner, H.; Tseng, G.; Becker-Reshef, I.; Nakalembe, C.; Barker, B.; Munshell, B.; Paliyam, M.; Hosseini, M. *arXiv:2006.16866 [cs, eess]* **2020**.

(31) Jain, M. *Review of Environmental Economics and Policy* **2020**, *14*, 157–169.

(32) Alix-Garcia, J.; Millimet, D. L. Remotely Incorrect?, 2020.

(33) Chernozhukov, V.; Chetverikov, D.; Demirer, M.; Duflo, E.; Hansen, C.; Newey, W.; Robins, J. *The Econometrics Journal* **2018**, *21*, C1–C68.

(34) Rubin, D. B. *Journal of Educational Psychology* **1974**, *66*, 688–701.

(35) Cunningham, S., *Causal Inference*; Yale University Press: 2021.

(36) Angrist, J. D.; Pischke, J.-S., *Mostly Harmless Econometrics: An Empiricist's Companion*, 1st edition; Princeton University Press: Princeton, 2009; 392 pp.

(37) Robinson, P. M. *Econometrica* **1988**, *56*, 931–954.

(38) Chernozhukov, V.; Chetverikov, D.; Demirer, M.; Duflo, E.; Hansen, C.; Newey, W. *American Economic Review* **2017**, *107*, 261–265.

(39) Semenova, V.; Chernozhukov, V. *arXiv:1702.06240 [stat]* **2020**.

(40) Belloni, A.; Chernozhukov, V.; Hansen, C. *Journal of Economic Perspectives* **2014**, *28*, 29–50.

(41) Knaus, M. C. *arXiv:2003.03191 [econ]* **2021**.

(42) Yang, J.-C.; Chuang, H.-C.; Kuan, C.-M. *Journal of Econometrics* **2020**, *216*, 268–283.

(43) Farrell, M. H.; Liang, T.; Misra, S. *Econometrica* **2021**, *89*, 181–213.

(44) Wager, S.; Athey, S. *Journal of the American Statistical Association* **2018**, *113*, 1228–1242.

(45) Wang, S.; Di Tommaso, S.; Faulkner, J.; Friedel, T.; Kennepohl, A.; Strey, R.; Lobell, D. B. *Remote Sensing* **2020**, *12*, 2957.

(46) Friedl, M.; Gray, J.; Sulla-Menashe, Damien MCD12Q2 MODIS/Terra+Aqua Land Cover Dynamics Yearly L3 Global 500m SIN Grid V006., 2019.

(47) Lobell, D. B.; Azzari, G.; Burke, M.; Gourlay, S.; Jin, Z.; Kilic, T.; Murray, S. *American Journal of Agricultural Economics* **2020**, *102*, 202–219.

(48) Breiman, L. *Machine Learning* **2001**, *45*, 5–32.

(49) Buchhorn, M.; Lesiv, M.; Tsendbazar, N.-E.; Herold, M.; Bertels, L.; Smets, B. *Remote Sensing* **2020**, *12*, 1044.

Chapter 4

Discovery of Influential Text in Experiments Using Deep Neural Networks

4.1 Abstract

We propose a method for discovering and testing influential concepts in media experiments. We apply a network with a recurrent and convolutional layer to experiments where unstructured text is given as treatment. Following existing models, our RCNN includes an intermediate layer specifically designed to make neural network classifications more interpretable. However we use this layer for a different purpose – to identify coherent phrases and sentences that are highly predictive of a human decision. We develop methods to interpret the filter layers of these RCNNs to facilitate the discovery of concepts within the text that are likely to have had the largest impact on the outcome. We validate our method by replicating a conjoint experiment on immigration using unstructured text in place of conjoint treatments. In addition, we apply the method to climate change communication where we discover which phrases that exert the most influence on learning and forming opinions about climate change. Last, we use our model to discover phrases that are predictive of censorship of Chinese social media posts.

4.2 Introduction

Political decision making is often influenced by writing and speech – humans might influence each other's vote choice through persuasion in conversation, make decisions about what political causes to donate to based on evidence put forward in documents or newspapers, or decide to turn out to vote or participate in a protest based on text messages, reminders, and social media posts. Experimental methods for estimating the impacts of text on human evaluation have been widely used in the social sciences. Most of these methods randomly assign texts to subjects, treating a random portion of subjects to a treatment text that is edited in particular way to be different from a control text. For example, framing experiments change one portion of a news article to estimate the effect on the way facts are presented on the opinions of the respondents [1]. Audit and correspondence experiments send the identical texts to decision makers, for example bureaucrats and legislators, varying only the name of the author of the text or some other attribute of the text to estimate the effect of the edited text has on the response of the person who receives the text (see [2] for a review). More complex experiments vary a number of aspects of the text than just one component, in what are commonly known as factorial or conjoint designs [3–5].

However, these methods have a set of well known drawbacks. First, concerns about power typically allow very few treatment texts to be used in media experiments, and therefore give researchers very few degrees of freedom in research design. Because they can use only a handful of treatment texts, researchers are required to know the types of texts that are likely to have the largest impact on the outcome before conducting the study. Further, even when researchers know the concepts they would like to test in a media experiment, they can only test a small number of incarnations of this concept in any given design. This creates questions about the generalizability of the experiment as well as about the interpretation of what aspect of the treatment text had an effect on the outcome [6].

A new literature in computational social science has sought to randomly assign unique texts to respondents and then discover treatments from these unstructured texts that have an effect on a particular outcome. [7] develop a groundbreaking model for identifying topics that affect an outcome, as well as an overarching framework for discovering latent treatments that have a causal effect on an outcome. [8, 9] follow their lead to use a neural network to identify the effect of a particular word on an outcome, controlling for covariates that might also be correlated with the text and the outcome. These efforts continue a long trend in statistics and computer science of mining texts for influential features that predict an outcome [10, 11] and other work that has sought to relate text to metadata [12–15].

We extend these efforts by offering an additional method of identifying concepts from texts that are predictive of an outcome. We posit that there are a number of applications in which human decision making pivots not on a set of topics or a single word, but on a section of text rather than the text in its entirety. These sections of text can be of varying length – sometimes just one word, sometimes a full sentence, sometimes an entire paragraph may influence the decision maker. Further, the same concept expressed in a pivotal section of text might be expressed in many different ways, complicating their discovery. Existing tools that focus on the full text to extract topics, on just individual words or fixed n-grams within the text may not be able to identify the pivotal components of decision making in these domains.

We develop a neural net with recurrent and convolutional structures that can use large numbers of human decisions to recognize sections of text of varying length that are predictive of a decision. Because our model relies on vector representations of words, these sections of text can be unique within the corpus and still be identified as predictive of the outcome. Our model draws inspiration from models for interpretable deep neural nets such as [16] by building an intermediate layer that identifies coherent phrases that are predicted to be most important to the outcome. It extends these methods by mapping identified phrases onto a vector that allows for clustering of the phrases into treatments and by allowing for the inclusion of non-textual metadata to enhance the predictive power of the model.

Last, we discuss how discovery of phrases predictive of an outcome with our method can be used to design follow up experiments to test their causality. Examining the text the model selects from documents can generate ideas of useful vignette experiments that could be used to test discoveries in a more controlled setting. In areas where follow up experiments are not possible, researchers can use our method to create categories and test the impact of these categories on an outcome in held out data [17]. Our model adds to a growing literature in computer science of using neural networks for causal inference [18–21].

We proceed as follows. First, we describe the set up of the problem. Second, we describe our method to extract phrases that predict human decisions. Third, we validate the method by showing that it can select treatments from a conjoint experiment where the treatments are expressed in unstructured text. Last, we apply our method to two datasets: an experiment of the influence of news article about climate change on learning and forming opinions about climate change and extracting phrases that are highly predictive of censorship decisions in Chinese social media posts.

4.3 Discovering Influential Phrases: The Problem

We consider the problem where a researcher has a set of N unique texts $W_1 \ldots W_N$ and a set of N outcomes $Y_1 \ldots Y_N$. For example, newspaper articles may have been randomly assigned in a survey experiment to respondents who are then asked their opinion of a policy, social media posts may be reviewed by censors who then make a decision of whether to allow them, or personal statements could be submitted to employers who then decide whether or not to hire. The researcher believes that the texts may have influenced the outcome, but does not know what *aspect* of the text influenced the outcome. The challenge for the researcher is to identify the latent categories that may have had an effect on treatment [6].

Researchers have a few tools available to them already to try to identify influential components of the text in this setting. First, they could use a set of tools to identify influential words or sets of words that are highly predictive of the outcome. A large set of methods in statistics have been developed to select influential features that are highly predictive of an outcome [22]. This is often done with some variant of regressing Y on the words within the texts with some regularization. Some of these methods have been adapted specifically for identifying influential words in social science problems [10, 11, 23] or when conditioning for other covariates [8]. Most of these methods can be generalized to look for effects of distinct two or three character sets of words on the outcome as well.

As an alternative to identifying single word or n-gram predictors of an outcome, [7] develop a topic model, the Supervised Indian Buffet Process (sIBP), that allows researchers to identify the latent topics (probability distributions over words) that may have had an effect on the outcome. This innovation allows researchers to discover how the amount the text talks about a particular topic may have an effect on an outcome, increasing the flexibility of the discovery of treatments within texts. [7] then use the model to estimate the Average Component Marginal Effect [4] of each individual topic on the outcome and discuss the conditions under which this estimate is causally identified.

While these new and exciting models estimate the impact of topics or particular words on an outcome, they focus mainly on the effects of single words or single topics on decision making, which are not relevant to all domains. The influence of a particular word might not be that meaningful, as even when it is replaced it might not change the meaning of a text that much. Individual words that are highly correlated with an outcome may not be particularly interpretable because they may not retain enough context to understand the nuance of the decision. Using n-grams is one way to model context, but in conventional methods this approach leads to parametric explosion.

Topic models employ all words within the text, but in some domains they may put unnecessary weight on portions of the text that are not relevant. There are domains where a particular phrase or section of the text may be influential and the remainder of the text irrelevant. In court, for example, the decision of the judge may turn on one particular fact of the case, while the rest of the text about the case may be largely irrelevant or procedural. Bureaucratic decisions might rely on one particular phrase, for example, a complaint may be routed to a certain institution because it contains a particular phrase or saying or a social media post might be removed from the Internet because it included a particular phrase that is sensitive to the government. While topic models may be able to pick up components of important phrases, it can be difficult to trace topics back to these phrases and other words within the document may overwhelm the model.

To this end, we develop a model that allows us to elicit phrases of varying length that are predictive of an outcome. Because long phrases are typically not repeated in text, we design the model to pick up even unique phrases that are correlated with an outcome and share similar ideas. Our model makes the following contributions:

1. Recognize phrases of varying length that are predictive of an outcome.

2. Allow for highlighted phrases to flexibly express the same idea.

3. Map identified phrases to a common vector for interpretability.

4. Allow for the inclusion of non-text covariates to aid in prediction.

4.4 Method for Extracting Phrases Predictive of Human Decisions

We design a neural network with recurrent and convolutional structures in order to address these requirements. We combine two major innovations into our RCNN. The first major innovation that we bring from [16] is writing a neural network that produces both an accurate classification, but also a concise rationale for that prediction. The second is extracting quantities of interest from the network which tell us the relationship between the network's learned representations and the dependent variable–a method which has been used extensively in understanding image-processing networks but rarely in text-processing [24].

The summary of the structure of this network is as follows, which we outline in more detail below. We extend [16] in passing the data first through a generator function (composed of two RNN layers and a FCN layer) which emphasizes certain words in the original text and de-emphasizes others. Unlike [16], however, we do not exclude less important words completely, but create a weight for each word between 0 and 1. This allows the model to consider important sections of the text most heavily in predicting the final outcome.

We then pass this weighted text through an encoder function (a CNN layer, a max-pooling layer and a fully connected layer) that uses the weighted text to predict the outcome. Similar to [16], we include a cost function that consists of a function of the prediction error[1] and penalties for more total words (enforcing sparsity) and more groups of words (enforcing coherency) in the phrase. However, the relative simplicity of our model in comparison to [16] allows us to extract document and word activations from the final CNN layer in order to identify phrases that have similar effects across the corpus.

Pre-processessing

We first tokenize each text convert it to lower case. We pad each text so that they are all of the same length. Texts are converted into a sequence of vectors using a set of word embeddings.[2] The padded portions of the text are represented by word embeddings of all 0's.

[1] We use binary crossentropy.

[2] In the main results, we use word embeddings of depth 300 from fastText in the main results: https://fasttext.cc/docs/en/crawl-vectors.html

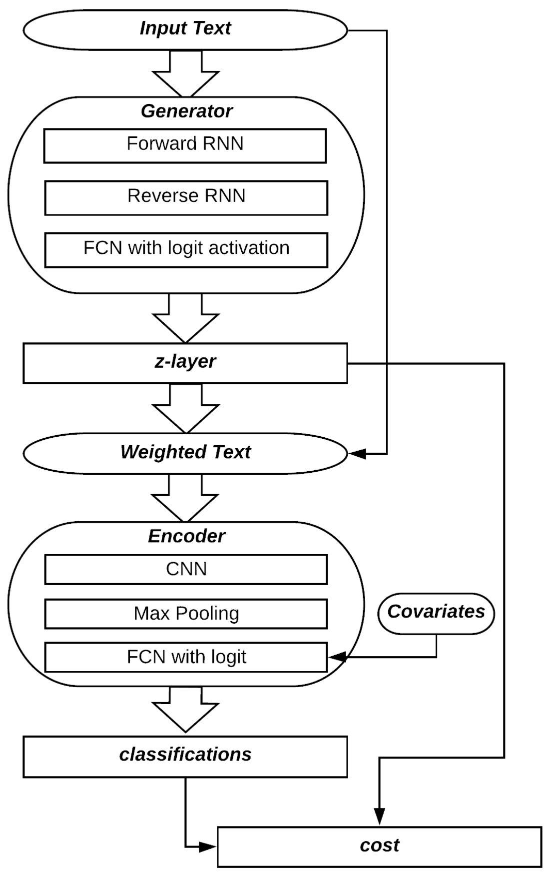


Figure 4.1: Outline of the model for extracting phrases. Round layers are composed of multiple parts explained in greater detail in following figures.

Network Detail

For the generator portion of the model, we pass each text through a recurrent layer which outputs a new length-300 vector at the location of each input word, generating another sequence of 100 depth 300 word embeddings. This forward recurrent layer is followed by a reverse recurrent layer of the same type. The objective of these recurrent layers is to condition the meanings (embeddings) of each word on the words that precede and follow it in the text. This is followed by a fully connected layer on the embedding dimension which takes a weighted sum of the new embeddings and passes it through a logistic function to generate a weight between 0 and 1. These weights represent the relative importance of each word for the downstream classification task.

We follow the notation in [25] to describe various layers and operations in the network. Allow $x_i \in \mathbb{R}^k$ to be the embedding vector of word i with depth k embeddings.[3] A text is a sequence (padded to length 100 if necessary): $x_{1:100} = x_1 \oplus x_2 \oplus ... \oplus x_{100}$. The recurrent unit which we use takes two inputs: the previous state of the hidden layer h_{i-1} and the embedding vector of the current word x_i and produces a new vector h_i which is passed through an activation function and then used for the next state. $h_i = f(Wh_{i-1} + W_x x_i)$ where $f(\cdot)$ is the hyperbolic tangent function. This produces a new sequence of vectors $h_{1:100}$ which are reversed and fed through a second recurrent layer producing a new set of $h_{1:100}$ Finally, each embedding vector h in the sequence is fed through a fully connected layer to generate a single weight z_i which is the output of the generator: $z_i = \sigma(Zh_i + b)$ where $\sigma(\cdot)$ is the logistic function and b is a bias term.

We then multiply the weights from the generator by the original word embeddings to create a text in which some words are more heavily emphasized than others, where at the extremes a weight of 0 essentially removes the word and a weight of 1 preserves the original intensity of the meaning. The weighted word $a_i = x_i z_i$ and the weighted text is $a_{1:100}$ is sequence of weighted original word embeddings.

The new weighted text is then passed to a convolutional layer with a flexible filter width and a number of filters which can be changed depending on the task we are performing. The convolutional layer essentially tests the degree to which some representation is present in the filter window. Phrases with similar meanings will produce similar activations for the same filter. The convolution operation uses a filter

[3] i refers to the position of that word in the input text throughout the network.

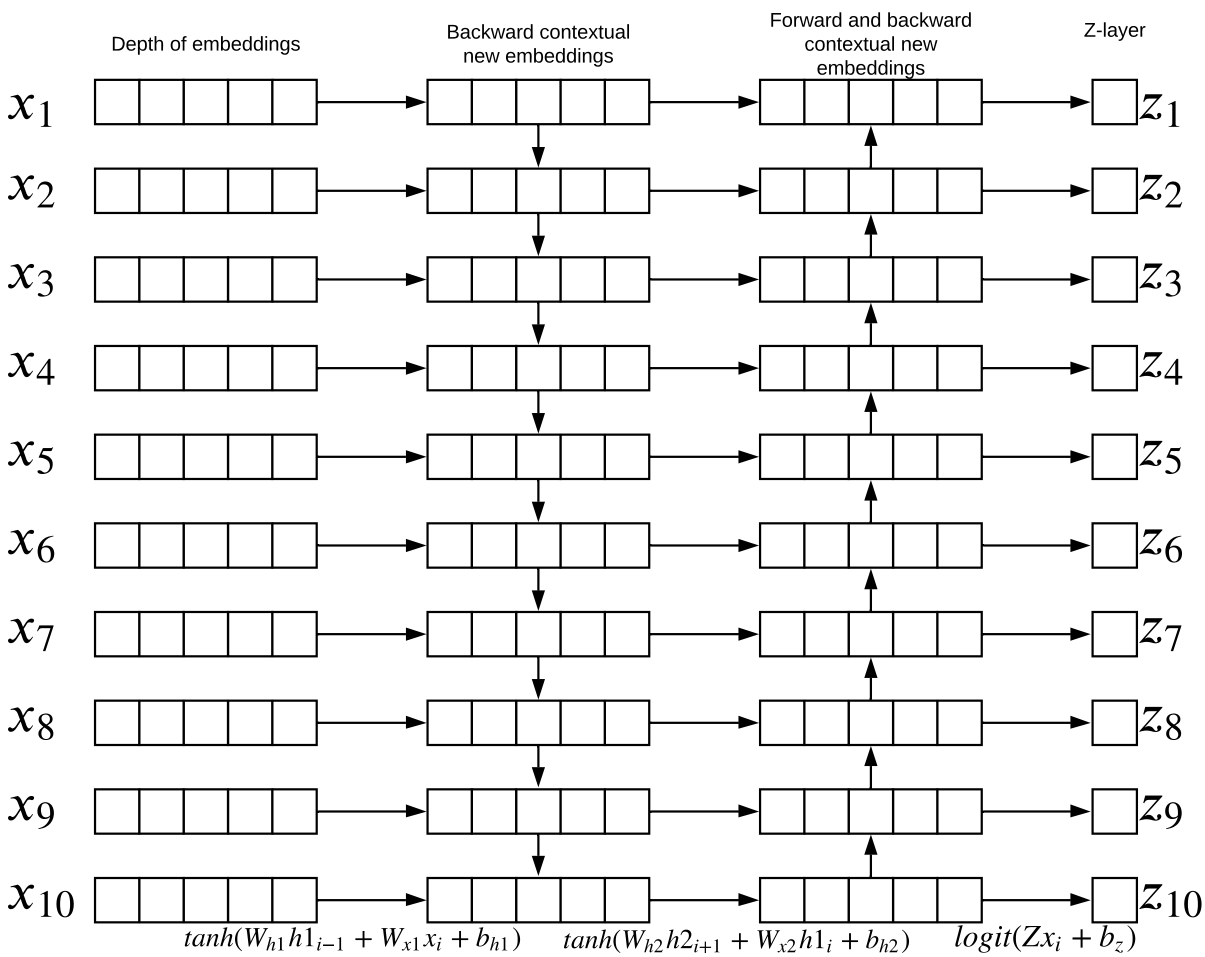


Figure 4.2: Detail view of how text passes through the generator. The inputs are word embeddings which pass through a forward and backward recurrent layer and then through a fully-connected layer. The final fully connected layer generates a value between 0 and 1 for each word based on the contextual embeddings.

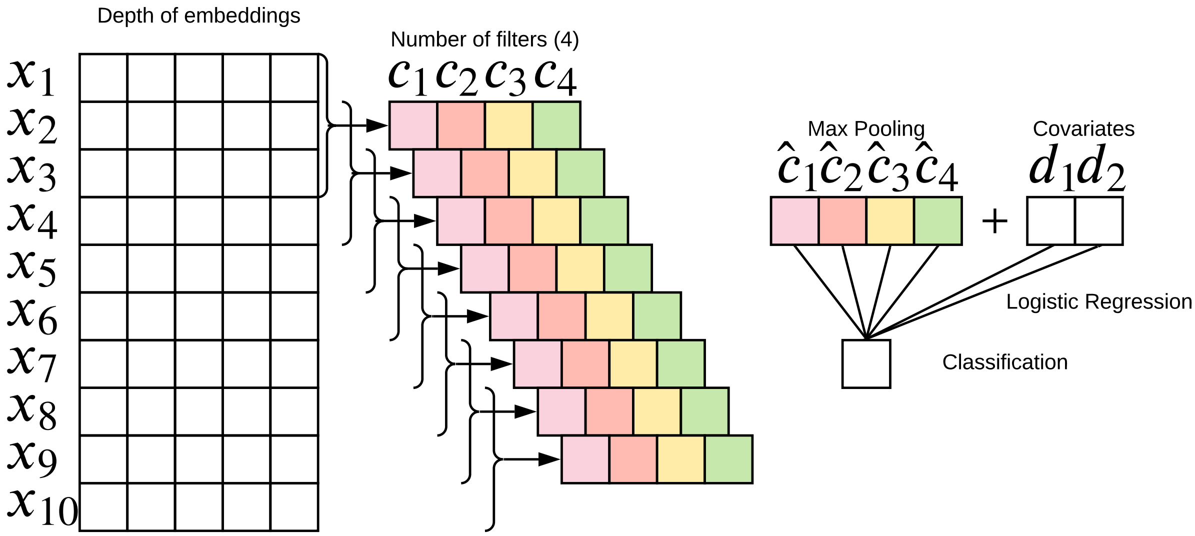


Figure 4.3: Detail view of how text passes through the encoder. The inputs are word embeddings weighted by the generator which compose the rationale. Here we have four filters and two covariates.

$p \in \mathbb{R}^{jk}$ which takes j weighted words as inputs and produces a new feature c_i: $c_i = g(p \cdot a_{i:i+j-1} + b)$ where $g(\cdot)$ is the ReLu function. This generates a feature map $\mathbf{c} = [c_1, c_2, ...c_{100-j+1}]$ for each filter-document combination. We refer to these $\mathbf{c}$ as "filter activations." This generates a matrix of filter activations $C \in \mathbb{R}^{m \times 98}$ where m is the number of filters.

Each document is then summarized in a max pooling layer which keeps only the highest activation of each filter for a document, $\hat{c} = \max(\mathbf{c})$ If our convolutional layer has n filters, this layer produces a vector of length n for each document where each element is the maximum activation of that filter for the given document. Finally these activations are concatenated with any covariates that the researcher wants to include and fed through a fully connected layer which takes a weighted average of these values and then pushes that average through a logistic function to generate a value between 0 and 1. $\hat{y} = \sigma(Y(\hat{c} \oplus a) + b_y)$ where Y is a weights matrix, a is a vector of covariates and b_y is a bias term. This value corresponds to the prediction of the binary dependent variable. This final layer is a logistic regression of the output on covariates and document filter activations.

Each of the many parameters in the above model are randomly initialized and updated with respect to a loss function using backpropagation and stochastic gradient descent. We use a loss function similar to [16]: a cross entropy loss for categorical outcome prediction and penalties which penalizes high weights

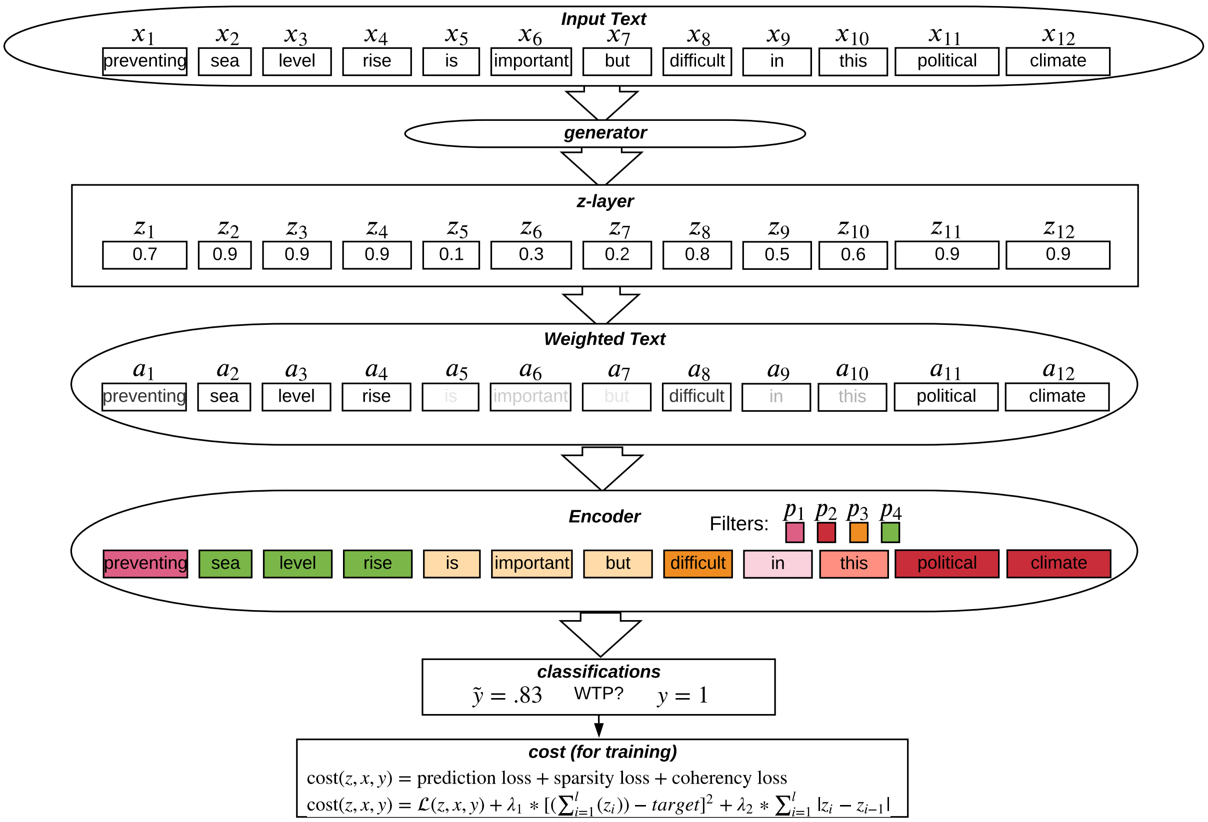


Figure 4.4: Detail view of how a text passes through the network. The z-layer tells the network which words to select, the rationale is a new text that the encoder uses to generate its classification. The loss function is composed of prediction error, a sparsity loss, and a coherency cost.

on individual words and encourages high weights to be consecutive within the text; encouraging sparsity and coherency in the word activations layer. Cross-entropy loss is $\mathscr{L}(z,x,y) = -\sum_{c=1}^{N} log(\hat{p})$ where $\hat{p}$ is the predicted probability of the true class, a function of (x,y,z). At the z-layer stage the network saves both $\sum_{i=1}^{l}(z_i)$ and $\sum_{i=1}^{l}|z_i - z_{i-1}|$ where z is the vector corresponding to the weights for each word, i indexes words in the original text and l is the maximum length of input text. These represent the sparsity and the coherence of the rationale respectively. We introduce a target sparsity *target* from which we penalize deviations on a batch-level. This encourages the average sparsity to be near the target, though some individual documents can have much higher or lower sparsity. The cost function for the entire network is the a weighted sum of the sparsity, coherency, and cross-entropy loss:

$$\text{cost}(z,x,y) = \mathscr{L}(z,x,y) + \lambda_1 * (\sum_{i=1}^{l}(z_i) - target)^2 + \lambda_2 * \sum_{i=1}^{l}|z_i - z_{i-1}|$$

The weights λ_1 and λ_2 and the target sparsity *target* can be adjusted to improve the quality of the rationale.

4.4.1 Interpretability

This structure provides two complementary ways to interpret latent factors that might contribute to the outcome. First, it allows for the identification of phrases of varying length that may be expressed differently across the text. For each document, the model provides a weight for the importance of each word within the text. Similar to [16], this cost function incentivizes the model to only select high weights when the word is very important and to cluster high weights together in coherent phrases. Phrases within each document that consistently have high weights are the parts of the original text which the network chooses to upweight because they have predictive power–meaning they are also the parts of the original text the researcher should pay attention to if she wants to understand how text influences the decision of the subject. Because the weights assign to each word vary between zero and 1, the model allows us to flexibly find longer or shorter phrases by setting a threshold to extract phrases consisting of words above a certain weight. It also allows the researcher to find words or phrases which consistently receive high (or low) weights across the entire corpus, suggesting that phrase is relatively important (unimportant) in the prediction.

We can then borrow from research on interpretability of deep neural networks (see [26] for a survey) to use word and document filter activations to cluster phrases with similar effects on the outcome. Remember that the encoder consists of a single convolutional layer, a max-pooling layer, and a fully connected layer. The fully connected layer is a logistic regression of the dependent variable on document-level filter activations and covariates. The max-pooling layer summarizes the matrix of filter activations with a vector of the maximum activation across each document. The convolutional layer learns matrices of word embeddings which are predictive of the outcome and produces an "activation" for each group of words (length depending on the filter width) which is higher if the words match the representation which the network has learned to recognize.

With these three layers we can identify the marginal effects of a filter activation from the weights in the fully connected layer much how one would interpret the coefficients in a logistic regression. However, because filters are amorphous representations which the network has learned to predict the outcome those coefficients are not directly interpretable. To help understand the concepts learned by each filter we can check which documents have the highest activation for any filter by examining $\hat{c}$ and read those documents to determine what the filter is capturing. Additionally we can examine the set of feature maps for each document $\mathbf{c}$ to see which phrases are maximally activating the filter. Applied across documents this method can find the set of phrases in the corpus which are most representative of the learned filter representation. We also generate plots which show filter activations for a single filter across all documents (all $\hat{c}_1$'s across the corpus) on the x-axis and observed outcome y on the y-axis to observe in what part of the activation space the differentiation occurs (if at all). We then compare phrases with high activations to those with middling and low activations to try to understand what the filter is capturing. We also examine the phrases which occur most frequently in the top 100 or 300 activations. With this information we can say that the increased presence of a concept captured by a filter is correlated with a higher or lower probability of a particular outcome. We also search over all $\mathbf{c}$ to find the phrases which generate the highest and lowest predicted outcomes when individually pushed through the final fully connected layer.

In all of the analysis we seperate training, validation and test sets. We train the model on the training data and use the validation set loss and accuracy to help determine when to stop training the model to prevent overfitting. We then use the test set to conduct the analysis described above. We find that if we use the training set to evaluate interpretability the model often overfits–all filters are significantly correlated with the outcome most of the time. In the test set, we find that in our experiments and simulated data that we are unlikely to discover a false positive, consistent with recommendations from [17].

Our method can help researchers test hypotheses in two types of research designs. First, it can be paired with observational data to learn how phrases are associated with an outcome of interest. Second, it can constitute the discovery portion of a larger experimental design where a researcher attempts to learn impactful treatments which "occur in the wild" and then test those treatments in a controlled experimental setting. In follow up experiments, research can isolate the treatments, eliminating the problems associated with correlations within the text of the treatment documents. We explain this process in more detail in the Climate Change Communication section.[4]

4.5 Replication of an Experimental Conjoint

To evaluate whether the model can pick up the components of human decision making, we first apply the model in a replication of a conjoint experiment. This evaluation tests whether the model is able to identify the conjoint treatments that have the largest effects on the outcome when these treatments presented as unstructured text rather than as experimental treatments. If the model works well it should be able to select the effective conjoint treatments from the unstructured text.

To do this, we replicate [27] who present 18,300 profiles of asylum seekers to 1,830 survey respondents from a handful of European countries. [27] randomize nine attributes of the asylum seeker's profile to these respondents, including the asylum seeker's gender, age, reason for migrating, country of origin, consistency of asylum testimony, religion, language skills, vulnerability, and previous occupation. They then ask respondents to separately evaluate each profile and decide between profiles. The randomization of attributes allows them to estimate the impact of any individual attribute on the respondents' ratings.

Because there are only nine attributes that are randomized in the experiment, [27] can identify the largest treatment effects through a simple linear regression. However, we assume as a hypothetical that the

[4]While the estimates from our model themselves could be interpreted causally with some rather stringent assumptions, we encourage researchers interested in causality to use our method to discover treatments that they can then test in a more controlled experiment in new data.

researcher did not know which aspects of the text were considered treatments and that the profiles were presented to respondents as unstructured text. Would the researchers have been able to use our model to identify the most promising treatments from the text itself?

To test this, we created text descriptions of the profiles based on the attributes assigned to them. For example, we would use the following text description for a respondent that was assigned to an asylum profile of an unemployed, Christian, physically handicapped, twenty-one year old woman from Kosovo with major inconsistencies in her application who is seeking better economic opportunities and speaks fluent English:[5]

> major inconsistencies. female. kosovo. twenty one years. unemployed. physically handicapped. seeking better economic opportunities. christian. speaks fluent english.

We randomize the order of the appearance of the treatments in the unstructured texts so that the model is unable to judge importance simply by placement. We then used the model with five filters to identify phrases within the text that it estimated to be the most predictive of the ultimate rating. We use the binned binary rating from [27] as the ultimate outcome we seek to predict.

If our method works, it should highlight words associated with the most effective treatments in the conjoint and be able to identify the direction of these treatments' impact on the outcome. In Figure 4.5, we show the results of the experiment. On the left, we include the nine categories of attributes and show the text that makes up each different level of each attribute. We highlight the phrases of the text that were predicted to most increase the outcome (in blue) and most decrease the outcome (in red). The right hand side of Figure 4.5 shows the point estimate of the effect size in the original study.

As you can see, the model easily separates out the phrases related to each individual treatment. In addition, it picks up only the phrases from the unstructured text that were most associated with treatment and correctly identified the direction of the effect. It identifies the most phrases in the treatment with the largest effect – seeking better economic opportunities as opposed to persecution. In other cases, it identifies the category of the treatment with the largest effect – for example, highlighting major inconsistencies and victim of torture as an influential phrases.

Figure 4.6 shows the relationship between each of the five filters and the outcome, as well as the

[5] Here we use English for convenience, but the original profiles indicated the language that was native to the respondent.

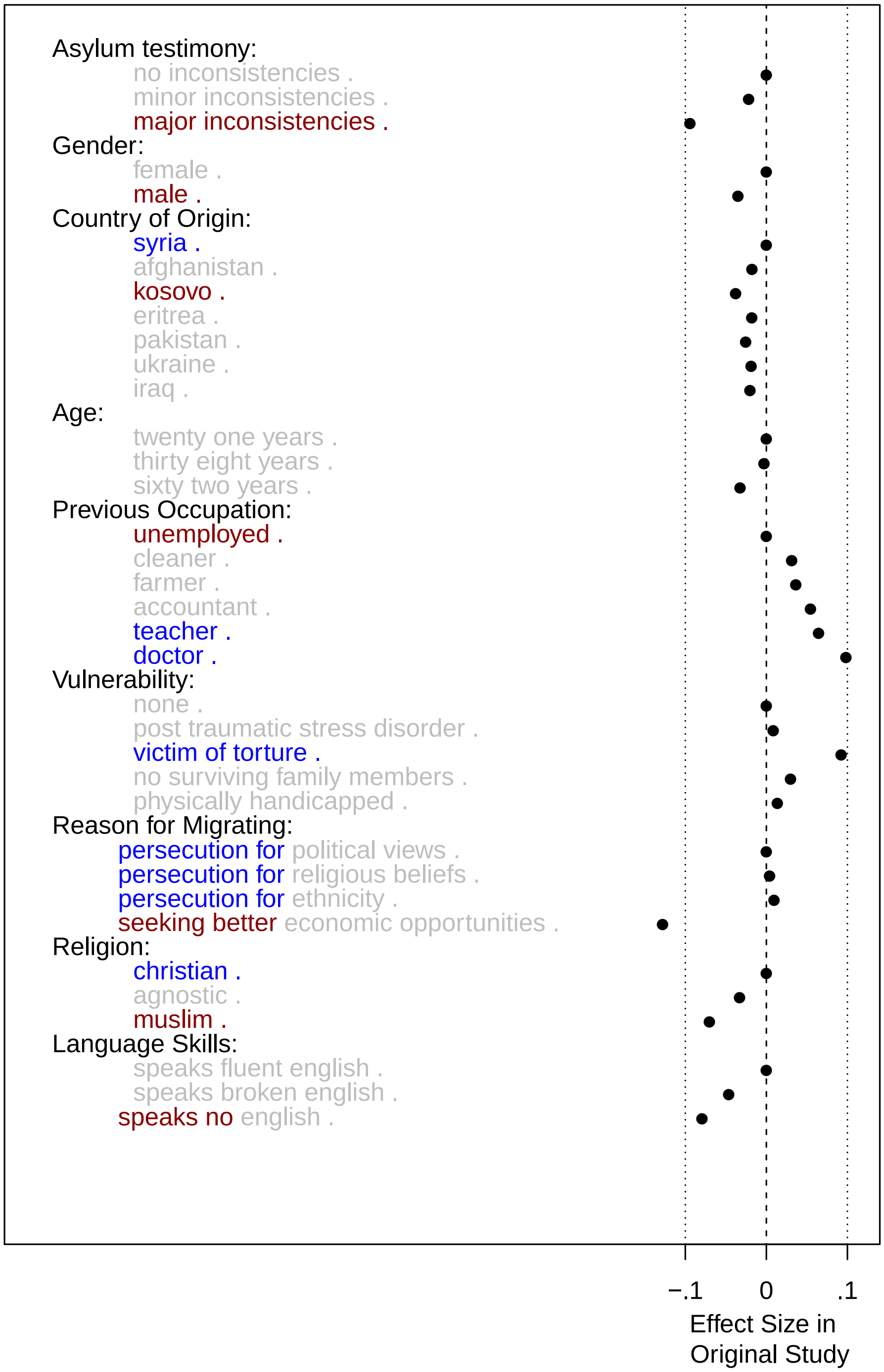


Figure 4.5: The left hand side of this figure shows the nine attributes randomized in the conjoint. We highlight the phrases that were predicted to have the most positive (blue) and negative (red) impact on the outcome. The right hand side of shows the point estimate of the effect size in the original study.

phrase most associated with each filter. The five filters of the model are respectively most associated with each of the phrases with the largest impact on the outcome, including seeking better economic opportunities, persecution, victim of torture, major inconsistencies, and speaks no (English). It also correctly identifies the direction of the impact of each of these filters on the outcome.

4.6 Censorship of Chinese Social Media Posts

We now apply our model to another example – which posts Chinese Internet companies decide to remove social media posts from the web based on directions from government censors [28–30]. Censors in China remove posts from the web using a combination of keyword and phrase searching and hand filtering [31]. In this section, we use the model to identify sets of phrases that are likely to have attracted the attention of the censors and therefore are highly predictive of censorship.

Previous work has sought to reverse engineer censorship events or discover online keyword lists that censors use to filter [28, 32, 33]. Here, we use large numbers of examples of posts to observationally reverse engineer phrases that are likely to have been used in filtering posts in order to decide what to censor. To do this, we use 10,000 examples of censored social media posts and 10,000 examples of uncensored social media posts taken from the Weiboscope data set [34], a project which followed 14,387,628 social media users on the Chinese social media website Weibo in 2012, recording which and when users' social media posts were removed from the Internet.

To extract phrases related to censorship, we used the network to predict censorship from the text of the post. Since Chinese texts generally do not include spaces, we first segmented the texts so that we could apply the model directly. We then used pre-trained Chinese word embeddings of dimension 300 trained on Wikipedia using fastText,[6] a filter width of 3 and 20 filters to extract both what phrases are likely to predict censorship and the direction in which they are likely to predict censorship. Our model resulted in an out of sample accuracy of 0.80.

The model resulted in 9 filters positively correlated with censorship, 10 filters negatively correlated with censorship, and 1 filter uncorrelated with censorship. Of the nine filters, three pick up on phrases

[6]See [35], Chinese word embeddings were retrieved from https://github.com/facebookresearch/fastText/blob/master/pretrained-vectors.md

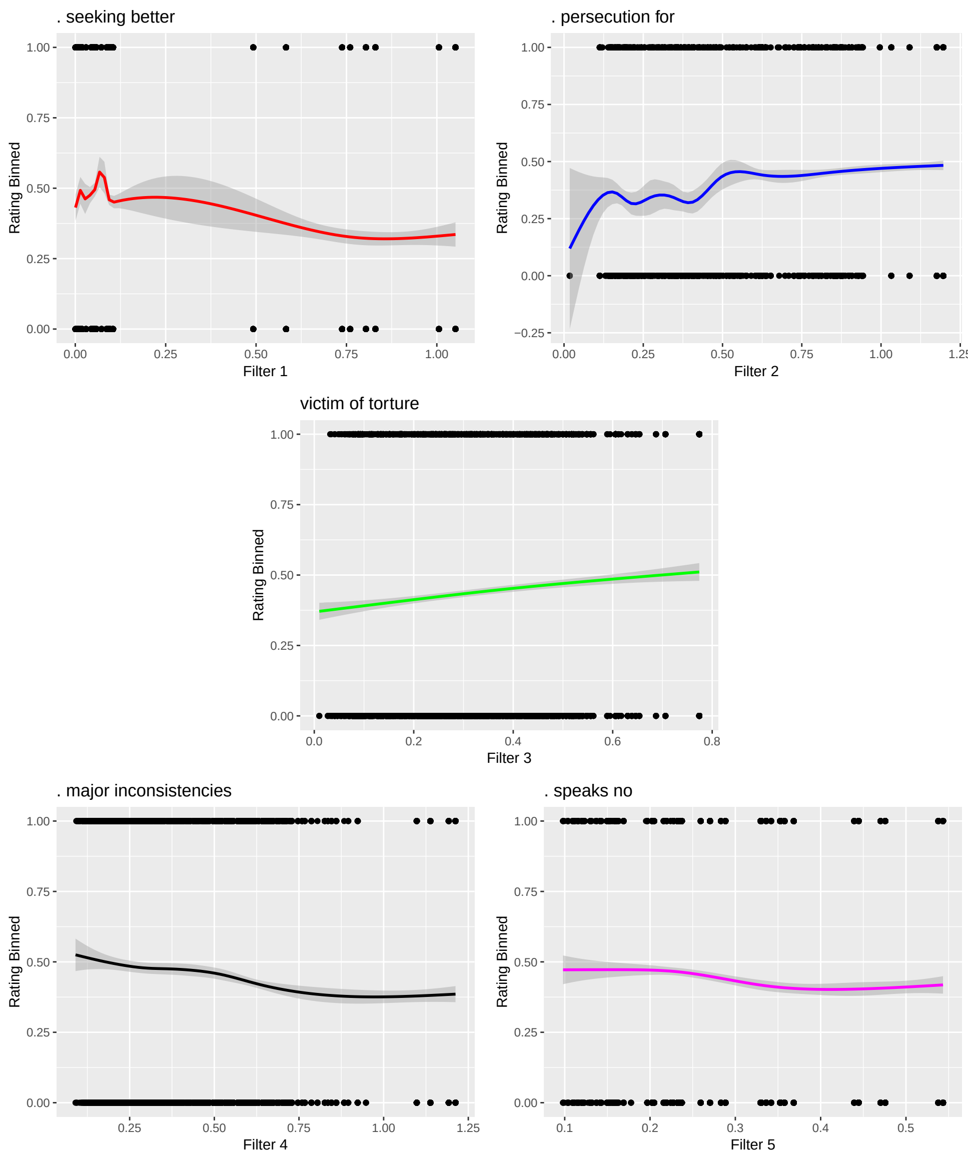


Figure 4.6: Each of five filters estimated by the model and the phrase that loads most highly on the filter. The model picks up the phrases with the largest impact on the outcome, including seeking better economic opportunities, persecution, victim of torture, major inconsistencies, and speaks no (English). It also correctly identifies the direction of the impact of each of these filters on the outcome.

Table 4.1: Strongest Filter Phrases Associated With Censorship

Phrase	Translation
饥荒 饿死 3600万	Famine that starved 36 million
被 暴徒 殴打	Beaten by a mob
甘肃 省委 常委	Gansu Provincial Party Committee Standing Committee
爱 国 主义 信仰	Patriotism
抢 劫 国有 资产	Robbery of state-owned assets
市委 书记 ——	Party secretary
变相 颠覆 国家	Disguised subversion
劳动 人民 在	Working People
人民 谋 和平	People want peace
人民 书记 人民	People's secretary, people
习近平 书记 领导	Xi Jinping Secretary leader
中央 经济 会议	Central economic conference
中国 民主 前途	China's democratic future

that indicate the post was retweeted, indicating that posts that are retweeted are likely to get censored. Three reflect emojis that are predictive of censorship; interestingly the "anger" emoji is strongly associated with censorship in this data. The last three (and strongest) filters reflect political phrases. Table 4.1 shows phrases from the filter most associated with censorship. These phrases are all what we would expect might be likely to be censored: phrases about the 1959 Great Famine, details of which are scrubbed from official history; phrases related to collective action, such as mobs, democracy, subversion and patriotism (2012 saw large numbers of "patriotic", anti-Japanese protests); and phrases related to the government, including Party Secretaries and Standing Committees. Importantly, many of these phrases occur only once within the corpus, but the model is able to flexible learn them and related phrases even though they are unique.

In contrast, the phrases that the model predicts are least likely to be censored (reported in Table 4.2) are very benign. They have to do with love, include a lot of exclamation points, or are related to advertisements. Other phrases that are predicted to be associated with low levels of censorship include wishes of happy birthday and encouraging phrases. These phrases are largely consistent with what we know from the current literature about what is censored in China [28, 30, 32, 36], a further validation of our approach.

Table 4.2: Phrases Associated with Lowest Censorship

Phrase	Translation
11 点 开 抢	Opens at 11
！ ！	!!
! ! 吧 !	!! right? !
! love your [爱	love your love
! your is so good	! your is so good
11 点 开	Opens at 11
$>>$ http :	$>>$ http :
! !	!!
美[爱	beautiful love
! !	!!
bb $<$ 3	bb $<$ 3
with bb $<$ 3	with bb $<$ 3
包邮 】11 点	Shipping 11 o'clock
! ! 吧	right?

4.7 How do phrases used by the media influence peoples' climate change opinions?

Last, we apply our model to climate change communication to see whether the model can identify the phrases from news articles about climate change that influence how respondents learn and form opinions about climate change. As the number of studies on public opinion about climate change ballooned in recent years some researchers have voiced concerns about how treatments (messages) are conceived of and used across studies [37]. In particular there seems to be no best practice for choosing what treatment to use in a survey experiment, with researchers generally relying on intuition, personal experience, literature suggestions, or baseline studies to develop treatments. While researchers generally settle on interesting and relevant treatments, their ultimate choice often causes problems for the generalizability of their findings–do conclusions about a particular treatment generalize to treatments with similar meaning but different wordings? How are people exposed to the tested treatment in texts that they might encounter outside of the lab?

In this section, we proceed in three parts. First, we describe the design of the survey and experiment.

Second, we describe a power analysis we conducted before running the survey to ensure that we would have sufficient data to pick up effects. We then apply our model to the data we collected from the survey to identify key clusters of phrases which occur "in the wild" and are associated with a shift in learning about climate change and climate change beliefs.

Design of Survey and Experiment

We used the Google News API to harvest 335 news articles about the a report detailing the risks of climate change released by the UN Intergovernmental Panel on Climate Change on October 18, 2018. We divided these articles into 1,323 paragraphs which constitute our treatments. We ran a survey on Mechanical Turk with waves on 05/29/2019, 06/27/2019 and 07/10/2019 where participants were asked to read a randomly selected paragraph drawn from the 1,323 paragraphs. Following reading the paragraph, respondents answered four questions about their climate change opinions in addition to a set of demographic and political questions and several attention checks. The four questions are representative of the most commonly used four outcome measures in the climate change literature (for example, see [38–44]):

1. Do you believe that climate change is being caused by human activities? (yes, maybe, no)
2. In your view, is climate change a very serious problem, somewhat serious, not too serious, or not a problem?
3. Do you think climate change should be a low, medium, high, or very high priority for the President and legislature of the United States?
4. Would you be willing to support a policy which reduces the effects of climate change but increases your taxes by $150? (yes, no) [7]

We also asked subjects to fill in the blank in the following sentence: "Temperatures are likely to rise by ___ degrees Celsius between 2030 and 2052 if global warming continues at its current pace and if the world fails to take rapid and unprecedented measures to stem the increase, a U.N. report said."

[7] [39] find that the inflection point for willingness to pay is $177 which we round down to $150.

Finally we ask questions about respondents' age, gender, education, race, marital status, state of residence, vote in 2016 presidential election, political donations, engagement in public affairs, party and party strength of affiliation. In total, we collected 4,108 responses.

Simulated Data and Power Analysis

Before collecting any responses, we simulated a variety of outcomes that were known functions of the text and some covariates to test whether our model would be able to identify the phrases which we used to generate the data. We report the results of one set of those tests here which we believe is most similar to the application in real data.

We created a dataset which consists of each treatment four times, or 5,292 total treatment units. For each unit we created two binary latent treatments for simulation purposes. The first is 1 if the phrase "climate change" appears in the text and 0 otherwise. The second is 1 if at least one of the phrases "degree celsius" or "degrees celsius" appears and 0 otherwise. We hope to be able to recover these phrases using our model. We also generate two binary covariates: gender and party, each is a draw from a binomial distribution with probability 0.5. We generate outcomes as a draw from a binomial distribution with probability $p = 0.2 + \alpha * \text{CC} + \alpha * \text{DC} + 0.3 * \text{party} + 0.2 * \text{gender}$ where CC is the "climate change" indicator and DC is the "degree(s) celsius" indicator. We set α and β equal to 0.2, 0.1, and 0.05 in three different tests to see whether under general circumstances which we estimate to be similar to our experimental conditions we are able to recover the known coefficients on CC and DC. We initialize the model with two different target generator selection rates: 20% selection and 100% selection. We find that in every case the network is able to recover a positive relationship between a filter which is maximally activated by one or both phrases in a test set.

We initialize the network with 5 filters of width 3, $n = 5, j = 3$, and let the model run until the development set loss is minimized. We then apply the model to a test set consisting of only units which the model has not yet seen. From the test set we observe several intermediate values which tell us which features the network is using to classify each unit. These include:

1. The words which receive the highest average weights in the rationale

Table 4.3: Frequency of top 100 filter activations for Filter 2 in 20% effect simulation (only presenting phrases that appear more than once.)

phrase	frequency
panel on climate	58
threat of climate	5
attractive ones, climate	3
1.5 degrees celsius	2
caused by climate	2
impacts of climate	2

2. The filter(s) which are most highly associated with a particular outcome

3. The phrases which most highly activate each filter

4. The texts which have the highest activations for each filter

5. The phrases which have the highest or lowest predicted outcome

With depth 300 word embeddings the generator part of the model contains 180300 parameters which makes the generator unable to effectively learn in a sample of less than 1% that size. As such, we use 100% generator selection for this application, unlike in the censorship and conjoint example, where we have enough data to extract the weighted text. To examine 2) we regress the outcome y on the document filter activations and covariates $\hat{c} \oplus a$ in the training set and the test set to identify which filters have a consistent effect on the outcome. We find that with 10% and 20% effect sizes the covariates consistently have the correct direction and are strongly correlated with the outcome and two filters emerge with strong correlations with the outcome. In the 5% case only one filter emerges with a strong and signficant correlation with the outcome.

Our model is able to pick up the simulated phrases. We find that in the 10% and 20% case one filter is consistently activated by the phrases with the word "climate" (including "panel on climate," "convention on climate," "avert global climate," "agreement on climate," "threat of climate" and others, see Table 4.3). The other is activated by a set of phrases which consist of either "1.5" or "2" followed by "degree(s) celsius." (Table 4.4). We plot the marginal effects of these filters in Figure 4.7.

In the 5% effect case one filter is consistently positive and is most frequently and most highly

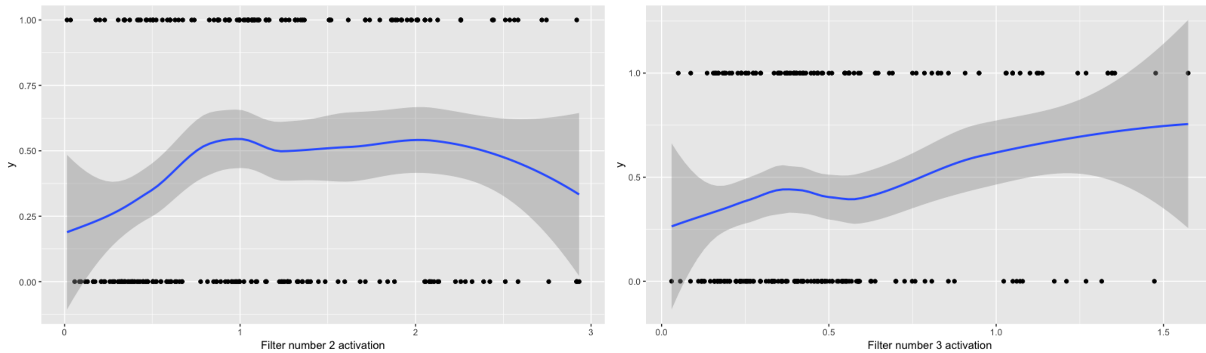


Figure 4.7: Plot of marginal effects of filters 2 and 3 in simulated data with 20 percent effect size

Table 4.4: Frequency of top 100 filter activations for Filter 3 in 20% effect simulations (only presenting phrases that appear more than once.)

phrase	frequency
1.5 degrees celsius	18
panel on climate	14
(2.7 degrees fahrenheit)	5
0.9 degrees fahrenheit	3
1 degree celsius	3
2 degree celsius	3
2 degrees target,	3
3.6 degrees fahrenheit	3
one degree celsius	3
1.5 degree celcius	2
1.5 degrees celsius,	2
2 degrees celsius	2
2 degrees of	2
2.7 degrees fahrenheit	2
2015 paris climate	2
3 degrees of	2

Table 4.5: Frequency of top 100 filter activations for Filter 1 in 5% effect simulations (only presenting phrases that appear more than once.)

phrase	frequency
panel on climate	6
not even on	5
temperature rise to	4
average temperatures were	3
framework convention on	3
limit warming to	3
paris agreement on	3
an increase of	2
from lack of	2
limiting warming to	2
pre industrial levels.	2
the costs and	2

activated by the phrases "intergovernmental panel on" and "1.5 degrees celsius" and their synonyms (Table 4.5).[8] In this case, the model picks up on a phrase that is almost always adjacent to climate change ("intergovernmental panel on"), but not climate change itself. Below a 5% effect size we are sometimes but not consistently able to identify effects of filters associated with our simulated treatments. On the other hand, we have not found any false positives where a filter consistently identifies an effect which we did not simulate. This suggests that we should be able to pick up phrases with a 5% effect size or larger.

This exercise demonstrates that we may be able to uncover treatments with effect sizes as small as 5% under a reasonable set of assumptions about our data. We follow this with a validation in the actual data.

Estimating the Impacts of Text on Learning: Survey Data

After collecting 4,108 responses, we first use the text to identify phrases associated with learning about climate change in our data. We use the fill in the blank question to measure learning:

> "Temperatures are likely to rise by ___ degrees Celsius between 2030 and 2052 if global warming continues at its current pace and if the world fails to take rapid and unprecedented

[8] We observe a similar pattern in the censorship and conjoint examples where the model sometimes picks up two different phrase clusters in one filter if they have a similar effect size.

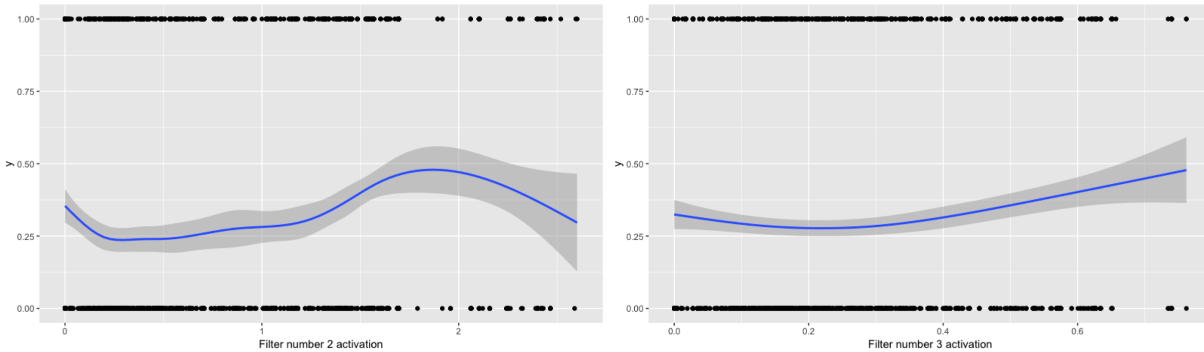


Figure 4.8: Plot of marginal effects of filter 2 and filter 3 in where the outcome variable is whether someone filled out the correct answer to the question about global temperature increases.

measures to stem the increase, a U.N. report said."

We create a binary outcome variable from the fill in the blank question above which is 1 if respondents responded with the correct answer, and 0 otherwise. The information necessary for the correct answer to this question appears in many of the news articles, so we expect that respondents who receive this information will fill in the "correct" answer. However, some news articles point out that temperatures have already increased by 1 degree and are likely to increase by another 0.5 degrees, or talk about the dangers of a 2 degrees world, which might confuse respondents. Our analysis should be able to point to both the text of the article that contributes to learning and the text of the article that might cause confusion.

After running our model, we find two filters which are strongly positively associated with reporting the correct answer. The first positive filter is most associated with the concept of global temperatures, indicating that if the paragraph the respondents received mentioned global temperatures, respondents were more likely to respond to the question correctly. This filter picks up phrases such as "average global temperature," "limit global temperature," and "in global temperatures" (Table 4.6). It also frequently picks up the phrase "1.5 degrees celsius," which is the correct answer to the question.

The second positive filter the most highly activating phrases are "paris agreement goal," "incheon, south korea," and "c and ipcc," indicating that if the respondent received a paragraph about the goals of the paris agreement or the IPCC report, they were more likely to respond correctly to the question (Table 4.7). We plot the marginal effects of these two filters on the outcome in Figure 4.8.

There is no filter which is strongly negatively associated with putting an incorrect answer, though

Table 4.6: Frequency of top 300 filter activations (only displaying frequency greater than 3)

phrase	frequency
of global warming	44
limiting global warming	38
limit global warming	30
on global warming	27
in global temperatures	17
the global temperature	13
in global temperature	9
limit global temperature	9
to limit temperature	8
1.5 degrees celsius	6
average global temperatures	6
a global temperature	5
help limit temperature	5
lower global warming	5
made global warming	5
with global temperatures	5
allowing global temperatures	4
considering global temperatures	4
course, global temperatures	4
limit global average	4
of global average	4

Table 4.7: Frequency of top 300 filter activations for Filter 2 in learning about climate change (only displaying frequency greater than 3)

phrase	frequency
_ the ipcc	21
incheon, south korea	17
scale," the un	15
in south korea	14
document, the un	11
_ the special	10
_ the un	7
"preventing an extra	6
2c, the un	6
society, the ipcc	6
_ "the ipcc	5
c an ipcc	5
a south korean	5
hitting this goal	5
rise. the ipcc	5
_ "a un	4
_ "the un	4
_ tropical nations	4
"a new un	4
c, an ipcc	4
2c. the ipcc	4
from south korea	4
monday, the ipcc	4
the associated un	4
will. the ipcc	4

all three of the other filters have negative weights. We examine the phrases which are most associated with putting an incorrect answer in Table 4.8. These phrases highlight several organizations that talked about temperature rises above and beyond 1.5 degrees. The phrase “the united nations” (as opposed to “the un”) tends to appear in documents which lead with the amount of warming which has already occurred, pointing out that the world is on track for a 3 degree rise in temperatures. The phrase “world meteorological organization” occurs in documents which contain information about several different scenarios but focuses on the previous 2 degree target. The phrase “oceans and unleash” occurs in paragraphs which talk about the dangers of an additional degree of warming. While these phrases are not direct evidence of a relationship they might suggest that hypothetical discussions about larger rises in temperatures may have confused respondents. The following are the texts with the identified phrases italicised:

> “incheon: avoiding global climate chaos will require a major transformation of society and the world economy that is unprecedented in scale, ***the united nations*** said monday in a landmark report that warns time is running out to avert disaster. earth s surface has warmed one degree celsius (1.8 degrees fahrenheit) — enough to lift ***oceans and unleash*** a crescendo of deadly storms, floods and droughts — and is on track toward an unliveable 3^{o}c or rise.”

> “u.n. ***world meteorological organization*** (wmo) secretary general ***petteri taalas told*** reporters in geneva: there is clearly need for a much higher ambition level to reach even a 2 degrees target, we are moving more toward 3 to 5 (degrees) at the moment.”

> “earth s surface has warmed one degree celsius (1.8 degrees fahrenheit) enough to lift ***oceans and unleash*** a crescendo of deadly storms, floods and droughts and is on track toward an unliveable 3c or 4c rise.”

Estimating the Impacts of Text on Opinions: Survey Data

We use the willingness to pay additional taxes for implementation of a policy that addresses climate change as the main dependent variable for this study, though we test the others as well. We find one filter that is marginally negatively associated with the outcome in the final layer of the model, though in the test set this relationship is inconsistent and the filter is unassociated with the outcome when we correlate it with the outcome independent of the other filters.

The words which have high activations for this filter are also not highly consistent (Table 4.9), though the most frequently appearing phrases in the top 300 activations seem to be related to efforts to

Table 4.8: Frequency of top 300 words negatively associated with learning (only presenting phrases that appear more than once.)

Phrase	Count
the united nations	95
world meteorological organization	23
petteri taalas told	16
oceans and unleash	15
the united states	13
nations to fight	10
overtime swiftly approved	9
reefs as soon	9
coral reefs would	8
body studying climate	5
miguel arias cañete	5
nations inter governmental	5
planet would not	5
climate protection department	4
corals from being	4
nations environment programme	4
new united nations	4
the united nation	4

Table 4.9: Phrases describing filter with negative association of willingness to pay for climate policy.

Phrase	Count
of heat could	45
bill hare, an	14
net zero emissions.	8
net zero by	7
paris agreement pledge	7
act than previously	5
climate action by	5
climate agreements endorsed	5
crops such as	5
100 year old	4
action to slash	4
an environmental economist	4
climate agreement in	4
fuels as an	4
in poland that	4
in poland to	4
packed. two deputy	4
so trump finds	4
been released today	3
clean energy company	3
climate action as	3
contradicting trump officials)but	3
donald trump declared	3
drastic action needed	3
fossil fuel industry,	3

counteract climate change and possible obstacles to them, including political agreements ("paris agreement pledge", "climate agreements endorsed"). This suggests that discussion of negotiations or stalemate may reduce people's willingness to try to counteract the problem. However, given the low correlation and inconsistency of the filter we caution against over-interpreting this result.

A different metric suggests that the text is predictive of decisions, but that we perhaps lack the sample size to identify how. If we replace each text with the word "a" and run the model including only covariates we find that both the loss increases and the prediction accuracy decreases. In the validation and test sets we achieve prediction accuracy of between 64% and 65% when we only include covariates, but 69% to 70% when we include the text. As such we believe that the model must be learning predictive features in the text but that the measurement and interpretation of the effect of those features is noisy.

Given this we turn to 5) from above: finding the phrases from the text which generate the most positive and most negative predictions in isolation. The top 100 phrases most associated with willingness to pay are all a number followed by "degrees fahrenheit" or "degrees celsius" or the phrase "the world meteorological organization." The phrases most associated with a lack of willingness to pay are less cohesive, but most include the terms "emissions," "fossil fuels" or "carbon dioxide." We view this as suggestive evidence that documents which are directly about the report may increase willingness to pay for mitigation while documents which focus on the changes society will have to make may decrease willingness to pay.

Several of the covariates which we include have strong effect sizes which mirror those found in the climate change communication literature: democrats are nearly 20% more likely to say they are willing to pay than independents while republicans are 12 percent less likely to to say they are willing to pay. People with a college education are about 7% more likely to say they are willing to pay.

We tested the model on two other outcome variables: the degree to which people thing global warming is a serious problem (dichotomized between those who think it is a "very serious problem" and those who think it is "somewhat serious," "not too serious" and "not a problem" to get the outcome as close to a 50/50 split as possible), and the level of priority that the government should commit to climate change (dichotomized at "very high" or "high" vs "low" or "medium"). In each case the filter activations were not significant predictors of the outcome. Particularly in the question about how serious of a problem climate change is the text did not seem to have much of an effect at all, with almost no increase in prediction accuracy from using the text and very low coefficients on the filter activations.

Because of the simulation exercises and ability to uncover phrases that are predictive of learning in the text, we are reasonably confident that the model should be able to uncover treatments which have an effect size of greater than 5%. The noise in our results suggests that the effects of the text which we presented to respondents are small or noisy or both, and that we need additional power if we want to be able to uncover those effects.

4.8 Future Directions and Conclusion

In this paper, we develop a new model to extract flexibly expressed and length phrases from text to evaluate their impact on the outcome. We incorporate ideas from other methods originally developed to understand machine classification can be used to understand how human decisions might be influenced by text, specifically by discovering influential phrases that are highly predictive of a human decision. Once we have discovered phrases we believe may influence a decision, we can test these hypotheses in follow up experiments. For example, we plan to use the clusters of phrases described above as treatments in conjoint or audit experiments to test whether these concepts and phrases indeed have a causal effect on decision makers. In the context of the censorship example, this could mean using the phrases that predict censorship in an audit experiment to determine the effect of each phrase on censorship without the presence of confounders. In the climate change example, we plan to design a follow-up experiment where survey respondents are treated with some of the influential phrases we identified within our analysis. Our method is uniquely suited to bridge observational and experimental methods because the outputs of the observational phase can be used directly in the experimental phase.

Appendix: Background on Neural Networks

Neural networks are an exciting subset of machine learning techniques, recently demonstrating the ability to perform tasks that experts deemed impossible just a few years ago. In [45] used a (now very simple) neural network to predict interstate conflict with some success, but it was not until recently that they began to really capture the imaginations of social scientists. While most of the advances in computer science have come in increasingly difficult classification problems, we seek to add to the growing literature that use neural networks to explain a phenomenon rather than simply predict it.

Neural networks (like other machine learning methods) are a method for classifying observations or predicting the outcome from set of training examples. Neural networks shine in their ability to learn complex representations in the data and their flexibility in fitting extremely complicated functional forms. More concretely, in their most simple form they are able to discover and fit high-order and interactive components of a classical regression. Because of this flexibility, they are extremely useful for modeling new data types, including text (where the meaning of a word's representation depends on the context), images (where the meaning of a pixel's value depends on that pixel's neighbors), sounds, and videos.

Neural networks exhibit great variety in their structure, with particular classes of networks designed to work well on particular problems or data sources. Similar to how a social scientist using regression might use ordinary least squares, logit, and negative binomial models to fit different data generating processes, a researcher using neural networks might use fully-connected network (FCN) layers, recurrent neural network (RNN) layers, convolutional neural network (CNN) layers, or some combination of the above to answer questions based on different types of data. However, all neural networks share some similarities.

First, neural networks take tensors as input variables, for example color intensities in red, green, and blue bands, n-dimensional word embeddings, or 1-hot word embeddings. Most commonly input values are scalars or vectors. These tensors are then passed through subsequent layers, with each layer composed of a number of "neurons." Each neuron is a linear combination of some or all of the neurons in the previous layer, generally with an "activation" function applied to scale the output of that function to the [0,1] or [0,inf) interval. The network ends with an output layer where there is one node for each continuous output variable and/or one node for each class if the output us categorical. Thought of in this way, a logistic regression can be expressed as a neural network with the number of inputs equal to the number of independent variables, and one output neuron in the second layer that takes a linear combination of all of the inputs and applies a logistic activation function to generate a value between 0 and 1. Similarly, OLS is the same network, but without the activation function to scale the output.

Most neural networks have more than simply an input and an output layer. To gain an intuition for how this works, imagine a structural model where a researcher uses independent variables to predict some intermediate regressors, and then uses those regressors to predict a final outcome. In that model, the intermediate regressors are a "hidden layer" in a "deep neural network." Today most neural networks consist of many hidden layers.[9]

Optimization of the parameters in a neural network is facilitated by an algorithm called back-propagation, whereby the derivative of the parameters of the last layer is taken with respect to the loss, those parameters are adjusted based on the gradient of the loss function, and then the loss is propagated backward into the network with the gradients adjusted at each layer. Because the number of parameters is very large for more complicated neural networks they tend to take significant computational power and many examples to train (although there is much research focusing on decreasing the data requirements of neural networks, see [47, 48]).

We now focus on neural networks designed for particular types of data. First, we consider RNNs, a type of neural network designed for data that has linear or temporal dependencies like text. A recurrent neural networks starts with the first observation in a time series, then employs a structure that takes a function of the first and second observations and outputs a tensor with the same dimensions as the input, then applies that function to the output of the previous round and the next observation, and so on. In this way, it uses previous observations to compute the effect of each subsequent observation on the outcome. For text, such a network would read in the first word in a string, then use the meaning of that first word to modify the meaning of the second word, then use that meaning to modify the meaning of the third word, and so on. RNNs frequently use stacked forward and backward layers to capture forward and backward dependencies in the text. This allows RNNs to use context to interpret the meaning of text, and RNNs are extremely effective at translation, summarization, and other NLP applications [49–53].

CNNs were invented for image processing and are designed to consider blocks of pixels at a time. These networks are able to learn recurring patterns (representations of certain shapes or objects across blocks of pixels) that are predictive of the outcome variable. Deep CNNs tend to learn low-level features like lines, curves and textures, gradually combine those to form higher-level representations of objects,

[9] An important result in the neural networks literature is the Universal Approximation Theorem – that a fully connected neural network with a single hidden layer with a finite number of neurons can approximate any continuous function with arbitrarily high precision [46].

and then use the presence of those objects to classify the image. More recently, CNNs have been used successfully in sentiment classification tasks in text [25, 54]. Instead of learning common representations in blocks of pixels, the network is able to learn representations contained in groups of words (n-grams). These networks only consider local order within the n-gram, but are better able to capture features that require multiple words to express.

Chapter 4 is coauthored with Sanford, L. Roberts, Margaret and Li, Kevin. The dissertation author was a lead author of this chapter.

References

(1) Druckman, J. N. *Journal of Politics* **2001**, *63*, 1041–1066.

(2) Bertrand, M.; Duflo, E. In *Handbook of Economic Field Experiments*; Elsevier: 2017; Vol. 1, pp 309–393.

(3) Wallander, L. *Social Science Research* **2009**, *38*, 505–520.

(4) Hainmueller, J.; Hopkins, D. J.; Yamamoto, T. *Political Analysis* **2014**, *22*, 1–30.

(5) Hainmueller, J.; Hangartner, D.; Yamamoto, T. *Proceedings of the National Academy of Sciences* **2015**, *112*, 2395–2400.

(6) Fong, C.; Grimmer, J. **2019**.

(7) Fong, C.; Grimmer, J. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2016; Vol. 1, pp 1600–1609.

(8) Pryzant, R.; Shen, K.; Jurafsky, D.; Wagner, S. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, 2018; Vol. 1, pp 1615–1625.

(9) Pryzant, R.; Basu, S.; Sone, K. In *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, 2018, pp 125–135.

(10) Miratrix, L.; Ackerman, R. *Statistical Analysis and Data Mining: The ASA Data Science Journal* **2016**, *9*, 435–460.

(11) Taddy, M. *Journal of the American Statistical Association* **2013**, *108*, 755–770.

(12) Grimmer, J. *Political Analysis* **2010**, *18*, 1–35.

(13) Roberts, M. E.; Stewart, B. M.; Airoldi, E. M. *Journal of the American Statistical Association* **2016**, *111*, 988–1003.

(14) Quinn, K. M.; Monroe, B. L.; Colaresi, M.; Crespin, M. H.; Radev, D. R. *American Journal of Political Science* **2010**, *54*, 209–228.

(15) Mimno, D.; McCallum, A. *arXiv preprint arXiv:1206.3278* **2012**.

(16) Lei, T.; Barzilay, R.; Jaakkola, T. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics: Austin, Texas, 2016, pp 107–117.

(17) Egami, N.; Fong, C. J.; Grimmer, J.; Roberts, M. E.; Stewart, B. M. *arXiv preprint arXiv:1802.02163* **2018**.

(18) Künzel, S. R.; Stadie, B. C.; Vemuri, N.; Ramakrishnan, V.; Sekhon, J. S.; Abbeel, P. *arXiv:1808.07804 [cs, stat]* **2018**.

(19) Shalit, U.; Johansson, F. D.; Sontag, D. *arXiv:1606.03976 [cs, stat]* **2016**.

(20) Johansson, F. D.; Shalit, U.; Sontag, D. **2017**, 10.

(21) Louizos, C.; Shalit, U.; Mooij, J.; Sontag, D.; Zemel, R.; Welling, M. *arXiv:1705.08821 [cs, stat]* **2017**.

(22) Hastie, T.; Tibshirani, R.; Friedman, J.; Franklin, J. *The Mathematical Intelligencer* **2005**, *27*, 83–85.

(23) Gill, M.; Hall, A. *Available at SSRN 2620781* **2015**.

(24) Arras, L.; Horn, F.; Montavon, G.; Müller, K.-R.; Samek, W. *PLOS ONE* **2017**, *12*, e0181142.

(25) Kim, Y. *arXiv preprint arXiv:1408.5882* **2014**.

(26) Zhang, Q.; Zhu, S.-C. *arXiv:1802.00614 [cs]* **2018**.

(27) Bansak, K.; Hainmueller, J.; Hangartner, D. *Science* **2016**, aag2147.

(28) King, G.; Pan, J.; Roberts, M. E. *American Political Science Review* **2013**, *107*, 326–343.

(29) Roberts, M. E., *Censored: Distraction and Diversion inside Chinas Great Firewall*; Princeton University Press: 2018.

(30) Bamman, D.; O'Connor, B.; Smith, N. *First Monday* **2012**, *17*.

(31) King, G.; Pan, J.; Roberts, M. E. *Science* **2014**, *345*, 1251722.

(32) Knockel, J.; Crete-Nishihata, M.; Ng, J. Q.; Senft, A.; Crandall, J. R. In *5th USENIX Workshop on Free and Open Communications on the Internet (FOCI 15)*, 2015.

(33) Knockel, J.; Ruan, L.; Crete-Nishihata, M., *Measuring Decentralization of Chinese Keyword Censorship via Mobile Games*; Munk School of Global Affairs: 2017.

(34) Fu, K.-w.; Chan, C.-h.; Chau, M. *IEEE Internet Computing* **2013**, *17*, 42–50.

(35) Bojanowski, P.; Grave, E.; Joulin, A.; Mikolov, T. *arXiv preprint arXiv:1607.04606* **2016**.

(36) Cairns, C.; Carlson, A. *The China Quarterly* **2016**, *225*, 23–49.

(37) Slater, D. M.; Peter, J.; Valkenburg, P. M. *Annals of the International Communication Association* **2015**, *39*, 3–31.

(38) Zhang, B.; van der Linden, S.; Mildenberger, M.; Marlon, J. R.; Howe, P. D.; Leiserowitz, A. *Nature Climate Change* **2018**, *8*, 370.

(39) Kotchen, M. J.; Turk, Z. M.; Leiserowitz, A. A. *Environmental Research Letters* **2017**, *12*, 094012.

(40) Nisbet, M. C. *Environment: Science and Policy for Sustainable Development* **2009**, *51*, 12–23.

(41) Gifford, R.; Comeau, L. A. *Global Environmental Change* **2011**, *21*, 1301–1307.

(42) Spence, A.; Pidgeon, N. *Global Environmental Change* **2010**, *20*, 656–667.

(43) Mildenberger, M.; Lubell, M.; Hummel, M. *Global Environmental Change* **2019**, *55*, 15–24.

(44) Bergquist, P.; Warshaw, C. *The Journal of Politics* **2019**, *81*, 686–691.

(45) Beck, N.; King, G.; Zeng, L. *The American Political Science Review* **2000**, *94*, 21.

(46) Hornik, K. *Neural Networks* **1991**, *4*, 251–257.

(47) Zheng, J.; Yang, W.; Li, X. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp 2362–2366.

(48) Yosinski, J.; Clune, J.; Bengio, Y.; Lipson, H. *arXiv:1411.1792 [cs]* **2014**.

(49) Cho, K.; Van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. *arXiv preprint arXiv:1406.1078* **2014**.

(50) Bahdanau, D.; Cho, K.; Bengio, Y. *arXiv preprint arXiv:1409.0473* **2014**.

(51) Nallapati, R.; Zhou, B.; Gulcehre, C.; Xiang, B., et al. *arXiv preprint arXiv:1602.06023* **2016**.

(52) Gambhir, M.; Gupta, V. *Artificial Intelligence Review* **2017**, *47*, 1–66.

(53) Chung, J.; Gulcehre, C.; Cho, K.; Bengio, Y. *arXiv preprint arXiv:1412.3555* **2014**.

(54) Ju, H.; Yu, H. In *Proceedings of the 12th International Conference on Ubiquitous Information Management and Communication*, 2018, p 9.

So long and thanks for all the fish.

—Douglas Adams