London School of Economics and Political Science

Department of Economics

Conservation Policies in General Equilibrium: Evidence from Brazil

Verónica Salazar Restrepo

A thesis submitted to the Department of Economics of the London School of Economics for the degree of Doctor of Philosophy. London, April 2024

Abstract

Deforestation and the subsequent use of deforested land for agricultural activities account for roughly 20% of the global CO2-equivalent emissions in the past two decades. Despite the global scope of the consequences of deforestation, public policies and private initiatives to reduce deforestation are often spatially targeted: they intensify environmental protection in specific ecosystems, making agricultural land scarcer. While potentially effective at a local level, their global effectiveness may be attenuated in general equilibrium, due to resulting increases in the demand for agricultural land in non-targeted areas, i.e. deforestation leakage. To quantify leakage, build a quantitative spatial equilibrium model of the Brazilian economy where agricultural land is the output of a costly process of deforestation, firms produce goods that are differentially land-demanding, and there is costly trade and migration. Our main findings are that (i) targeting the regions with highest deforestation levels can be an effective tool to curb aggregate deforestation in Brazil, and (ii) leakage increases significantly when considering a longer time-horizon. After one year, 2-3% of the deforestation reductions are outdone by leakage. Simulating the model forward for 10 years, this number goes up to 10%. The relatively small leakage is driven by agricultural intensification, including more crop farming, increased worker and cattle density per pasture, and shifts of production towards more productive regions.

Acknowledgements

I would like to give an immense thank you to my supervisors, Robin Burgess, Oriana Bandiera, Gharad Bryan, and Daniel Sturm for their unwavering support. For trusting my ideas even when they seemed a little crazy, but mostly for challenging me to discipline and explain them. Over and over.

Robin, thank you for taking me under your wing after I decided to change fields and felt a considerable degree of confusion; for encouraging me to be an Environmental Economist even when there were not many others around in the department. And for growing the beautiful EEE community at exponential speed.

Thank you Gharad for the meetings about ideas even at their earliest stage; for being an excellent example of a researcher, showing us that a healthy work-life balance is possible; for always asking the difficult question and always daring to say "why should we care?".

Thank you Daniel for the generosity with your time and your meticulous attention to details. But even more for your relentlessly encouraging attitude.

Thank you Oriana for the childish enthusiasm and the sharp questions.

For your friendship, coauthorship, emotional support, and for the maniac laughs; for being my regression-addicts sponsor always available for an emergency call, thank you Gabriel!!!

I also want to thank Clare Balboni, Tim Doberman, Matthias Doepke, James Fenske, Allan Hsiao, Xavier Jaravel, Per Krusell, Isabella Manelici, and Julien Wolfersberger for their generous comments and insightful discussions on this work.

For all their comments and questions, thank seminar audiences at LSE's Development Economics Work in Progress, LSE's Energy & Environmental Economics seminar, LSE's International Trade Workshop, and conference audiences at STEG-CEPR 2023, LSE's Environment Week 2023, Bocconi (EMUEA 2023), Cyprus (EAERE 2023), and Oxford (OxDev 2023).

Thank you to those who guided me while I was still and aspiring theorist: Balazs, Tomas, Alkis, I'll always be a theorist at heart. Thank you for all the rigour and the George nights.

Thank you for everyone at LSE who helped me survive the job market. And above all thank you to my colleagues who are the most brilliant and lovely people I have ever come across. Thank you for always putting kindness first, while being sharp as a razor. 3.12, best office ever, I will miss every single one of you.

Amen, thank you for being so dedicated and passionate, without losing sight of what matters, always seeing the best in people.

Pol, thank you for sharing your curiosity, intelligence, and awe at absolutely everything. But mostly, thank you for the dancing and the climbing that held my mind together.

Xavier, thanks for being a breeze of ocean air every time.

Juli, thank you for Kimberley Gardens, for being an unconditional friend, and for showing me how to work with love, letting go of the fruits of your efforts time and time again.

Iacopo, thank you for being a soft rock.

Tash, you know I owe you all my sanity (and all my insanity) of the last couple of years. Thank you for deciding to merge your colourful life with mine.

Dad, thank you for the most productive weeks of my life. But mostly thank you for making forests an integral part of my childhood. For teaching me love and discipline, always leading by example.

Mom, thank you for inspiring me from wherever you are.

Declaration

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it). The copyright of this thesis rests with the author. Quotation from it is permitted, provided that full acknowledgement is made. This thesis may not be reproduced without my prior written consent. I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party. I declare that my thesis consists of 15761 words.

Statement of Co-authored Work

I confirm that Chapters 2 and 3 were jointly co-authored with PhD candidate Gabriel Leite-Mariante and I contributed 50% of this work.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Tropical deforestation is among the human activities with the highest environmental impact. Forest clearing and the subsequent land use for agricultural activity are responsible for about one-fifth of global CO2-equivalent emissions over the past two decades. In addition to direct instantaneous carbon emissions, deforestation permanently destroys carbon sinks, causes the extinction of species, degrades native soil, alters weather patterns, and negatively impacts the livelihoods of millions of people living in forest dwelling communities. Given its enormous environmental damage, combating deforestation is a key component of emission reduction pathways in the global fight to mitigate the effects of climate change, as established by the Intergovernmental Panel on Climate Change (IPCC) (IPCC, 2019)[1]. Nevertheless, strategies to mitigate deforestation are excluded from major green finance mechanisms, such as the United Nations Clean Development Mechanism (CDM), due to concerns about leakage of locally targeted policies into untargeted areas, i.e. the reduction of deforestation in a specific area might be outdone by increases in deforestation elsewhere, undermining the global effects of local policies[2].

Since approximately three-quarters of global deforestation is driven by agriculture, quantifying the general equilibrium effects of anti-deforestation policies requires a model of how agriculture is redistributed across space in response to

[1] In its 2019 Special Report on Climate Change and Land, the IPCC mentions the word "deforestation" a total of 493 times, and states that *"Reducing deforestation and forest degradation lowers GHG emissions, with an estimated technical mitigation potential of 0.4–5.8 GtCO2/yr"*

[2] The UN defines deforestation leakage as *"The unexpected loss of anticipated carbon benefits due to the displacement of activities in the project area to areas outside the project, resulting in carbon emissions."*. In other words, locally targeted policies might not decrease overall deforestation, but rather displace it to untargeted areas.

local policies. To that end, I build a quantitative model that explicitly embeds deforestation as an economic sector that supplies land as a factor of production for agriculture. Beyond the direct policy relevance of assessing the global efficiency of localised policies, this framework can guide the understanding of broader trade-offs in environmental policy (promoting conservation vs curtailing agricultural production), and the value of mitigating global externalities of localised deforestation activities[3].

Deforestation in Brazil

With over 5 million km^2 of rainforest area (MapBiomas Collection 6.0), Brazil accounts for about one third of all remaining rainforests and 13% of all forests in the planet, far more than any other country. It also accounts for a third of yearly tropical deforestation, with an average net yearly forest loss of about 20,000 km^2 between 1985 and 2020[4] (roughly equivalent to the size of Wales, or the American state of Connecticut). As such, outcomes of policies aimed at reducing deforestation in Brazil will have significant impact in the global forest coverage and, consequently, carbon emissions.

Deforestation has played an important role in the Brazilian economy over the past 35 years as an input for the country's agricultural sector, currently responsible for approximately one-quarter of the Brazilian GDP (CNA and CEPEA, 2023). Since 1985, the first year for which reliable high resolution satellite data on land use is available, total agricultural area has increased by more than 45%, or 840,000 km^2 (MapBiomas Collection 6.0), of which 92% were previously forests and 8% were previously tropical savannas. This activity is highly spatially concentrated along the so-called "arc of deforestation", in the fringe of the Amazonian forest, especially in the states of Pará, Mato Grosso and Rondônia. The large municipality of São Félix do Xingu, in the state of Pará, with a total area of around 84,000 km^2 (roughly the size of Austria), is an excellent illustrative example. In 1985, it had a forest cover of over 80,000 km^2, more than 95% of its total area, and agricultural activity occupied only 410 km^2 (0.5%). In 2021, it had just under 63,000 km^2 (under 75% of its area) of forest cover, and over

[3]This framework could be used, for example, to implement a compensation policy for specific regions or for the country as a whole, a policy that has been suggested by a number of researchers and politicians as the only viable way to sustainably avoid deforestation (Brner and Wunder, 2008).

[4]The gross loss of primary forest has been higher, over 25,000 km^2 per year, but has been considerable forest regrowth.

19,000 km^2 (well over 20%) of agricultural land, most of which pastures.

Over the past three decades, the Brazilian government has enacted a wide array of policies aimed at tackling deforestation. These policies include country-level actions, such as changes in the national legislation (e.g. Forest Code) or the use of satellite monitoring (SIVAM - from the 90s), but also localised policies aiming at protecting specific areas of the Brazilian territory. The latter group, which is the focus of this study, has typically been implemented either as increased enforcement in areas with particularly high deforestation, or as the designation of specific areas in the territory as conservation units or protected indigenous land. Such localised policies have been the subject of past evaluation which has generally found them to be successful in the targeted areas (Assunção and Rocha, 2019).

Typically, studies evaluating such policies use an event study approach to measure their effectiveness in the targeted regions vis-à-vis comparable non-targeted regions (Assunção and Rocha, 2019). However, this approach does not account for potential relocation of deforestation activities into non-targeted areas, which can significantly attenuate the effect of such policies on overall country-wide deforestation. To illustrate this idea, suppose that regions targeted by a certain policy see a decrease of 5,000 km^2 in their annual deforestation, whereas comparable non-targeted regions see an increase of 5,000 km^2 in annual deforestation rates – i.e. there is perfect relocation. A simple difference-in-differences analysis would suggest that the policy decreased deforestation by 10,000 km^2 when in fact it only changed where it occurred, and the net global effect is zero. This issue, referred to in the climate policy literature as leakage, is often discussed (Pfaff and Robalino, 2017) but rarely measured when evaluating specific anti-deforestation policies[5].

The Model

Our model considers a multi-region economy with two main sectors: agriculture and non-agriculture. The agricultural sector is further split into different types of crops and pastures that demand different amounts of land and labour as factors

[5] It has become increasingly common to measure spillovers by looking at the changes in land use in the vicinity of protected areas, for example. While a valuable empirical exercise, it might not appropriately account for global spillovers, as the most suitable substitute for the land being protected may be in an entirely different part of the country - in this approach, the choice of which areas can be the subject of spillovers needs to be made *ex-ante* and is arbitrary.

of production, whereas the non-agricultural sector has only labour as input. Crucially, I model deforestation as an intermediate sector which endogenously supplies land as a factor of production for agriculture. Each location differs in their sectoral productivity (agricultural, non-agricultural and deforestation), amenities and trade links. I allow workers to migrate between regions subject to frictions, and goods to be traded across regions subject to iceberg costs.

In the model, a local anti-deforestation policy is modelled as an exogenous negative shock to land supply in a targeted region. Leakage happens due to relocation of agricultural activity via the markets for goods and labour. In the goods market, a local decrease in supply of land will decrease the local supply of agricultural goods. Consumers then substitute these goods with non-agricultural goods, or with agricultural goods produced elsewhere, increasing demand for land in other regions, which creates leakage. Analogously, in the labour market, a decrease in supply of land will decrease the local demand for agricultural labour – workers will then partly change sectors, and partly migrate increasing the supply of agricultural labour elsewhere, which also creates leakage. The extent to which reductions in forest loss are outdone by leakage ultimately depends on the elasticity of demand for agricultural goods, and on the substitutability of agricultural land across space.

Literature

The widespread availability of satellite data has contributed to a wave of empirical microeconomic research on the economics of deforestation (Balboni et al., 2023). This research has explored the key role of conservation policies, agricultural prices, trade, roads, property rights, and conflict, amongst others.

First, let us focus on the literature evaluating the effectiveness of conservation policies (Burgess et al., 2012; Jayachandran et al., 2017; Szerman et al., 2022). I conceptualise these as “supply-side” drivers of deforestation. The main policies considered are: command-and-control place-based restrictions (the focus of this work), payments for ecosystem services (PES), taxes (often hypothetical) and tariffs. Existing evidence suggests that command-and-control policy instruments have played a crucial role in the slowdown of deforestation in Brazil. Burgess et al. (2019) shows that Brazil-wide policies lowered it by as much as 56% from what it would have been in the absence of policies that started to be implemented around 2004. Assunção and Rocha (2019) focuses on the Priority List of municipalities, also examined in this thesis, and shows reductions of around 50% in the treated municipalities. Further work by Assunçao and coauthors on the Priority List (discussed in the paragraph below) has also done some empirical estimation of spillovers to neighbouring municipalities. They find that neighbouring municipalities seem to have reductions in deforestation as they neighbours join the list. Mangonnet et al. (2022) analyse how national politics influences the allocation of Protected Areas. The establishment of Protected Areas is a decision of the federal government with important economic costs at the local level in the form of foregone resource extraction and agricultural land. The authors document that municipalities with mayors from opposition parties were more likely to have protected areas established within their boundaries over the period spanning 1997 to 2012.

Second, there is the empirical literature has also explored the response of deforestation to changes in demand for agricultural land, driven by commodity prices, rural credit access, and market access, amongst others. These are important contributions and inspiration for my work as they highlight how deforestation is a process that is deeply embedded in the agricultural economy (Assunção et al., 2015; Berman et al., 2023).

This thesis adds to these strands of the reduced-form empirical literature in two ways. First, I build on and confirm with slightly different methodologies some of its findings. In particular I look at the reduced-form effects of Priority List municipalities with erences in differences and I propose an alternative geographical unit. Second, I jointly consider the many driving forces of deforestation: regional comparative advantage in agricultural activities, market access, residential amenities, and deforestation productivity. This helps to unpick their relative importance.

There is also a strand of the empirical micro literature that looks at spillovers from conservation policies in reduced-form. Pfaff and Robalino (2017) summarise the theory and evidence on the spillovers of conservation programs which, at the time, lacked any structural quantitative analysis of spillovers due to general equilibrium effects. Instead, it looked at spillovers to neighbouring areas outside conservation zones. This has the advantage of measuring all the different types of spillovers from conservation policies. Some of these may be positive enforcement spillovers driven either by the spread of pro-conservation practices and attitudes or by agglomeration externalities in conservation efforts. The mechanism that I will be exploring and quantifying through the model is the price effect driven by decreased availability in new agricultural land. This is interesting in isolation as it is the one that is more problematic for the stated goals of conservation policies. Moreover, reduced-form approaches have the disadvantage of requiring a simple assumption on the structure of spillovers: that they decay with distance to the protected areas and that places far away are valid controls. The findings of this literature are mixed. Fuller et al. (2019) systematically review the empirical evidence on deforestation leakage from Protected Areas. Their work highlights the leakage is a potential problem, they find evidence of leakage in 12% of the protected areas where there is evidence of deforestation reduction, but the magnitude of its incidence seem to be highly context dependent. Alix-Garcia et al. (2012) investigate a Mexican PES scheme. They estimate that deforestation was reduced by 50% in enrolled parcels, although the baseline deforestation risk was low. They also empirically evaluate leakage within owners (to other plots of land of the same farmer) and within markets (to other property within a market with high PES participation). They find evidence of both, and estimate that leakage undoes about 4% of the deforestation reductions. Robalino et al. (2017) analyse the heterogeneity in spillovers in the 0-5km and 5-10 km rings around new national parks in Costa Rica. This is one step closer to a structural analysis of it, as their heterogeneity analysis helps disentangle various theoretical mechanisms for leakage and positive spillovers. The authors look at the heterogeneity of spillovers by distances to roads and distances to park entrances, both of which are of economic importance, given critical local roles for transport costs, which affects the demand for new agricultural land, and tourism, which increases the demand for conservation. They find large and statistically significant leakage close to roads but far from park entrances, consistent with their theory. They also find that parks facing greater threats of deforestation show greater leakage. This is interesting in light of the results of the analysis by Jop (2009). They find a bias in the location of protected areas globally, they are typically located in places that do not face land conversion pressures, even in the absence of protection. Countries' PA network tends to be biased in elevation, slope, distances to roads and cities, and suitability for agriculture in a way that makes them significantly less at risk of deforestation.

In this thesis, I aim to quantify one specific source of conservation spillovers: the general equilibrium effects caused by conservation policy, which lead to leakage into unprotected areas. By looking locally targeted reductions in deforestation in a general equilibrium framework I can make predictions about the likely size of anti-deforestation policies' leakage and its spatial distribution. This can help assess whether it is actually likely to happen only nearby or if there are more distant regions likely to be affected by leakage, perhaps because of market access or greater agricultural suitability. This modelling approach could be used both to inform targeting decisions in future conservation policy and to correct biases in retrospective reduced-form policy evaluation.

A recent strand of the trade and IO literature looks at deforestation and its general equilibrium implications through structural models (Souza-Rodrigues, 2018; Hsiao, 2021; Domiguez-Iino, 2021). Copeland and Taylor (2004) provide a theoretical trade model to think about the tradeoffs environmental conservation. A related literature analyses how comparative advantage in agricultural activities shapes the spatial distribution of different land uses, generally without considering deforestation explicitly (Cui, 2020; Pellegrina and Sotelo, 2021). I contribute to this literature by adding labour and migration to general equilibrium models of land use change, and by explicitly linking land use changes with patterns of economic growth and structural change in space (Eckert and Peters, 2022; Farrokhi and Pellegrina, 2020; Bustos et al., 2016; Herrendorf et al., 2014; Boppart, 2014) . My model builds on the spatial equilibrium model proposed by Eckert and Peters (2022). I build on their framework by (i) adding a deforestation sector that endogenously produces new agricultural land, and (ii) considering different agricultural sectors that can have heterogeneous shares of land and labour as in Farrokhi and Pellegrina (2020).

Chapter 2

Data and institutional background

This chapter describes the data used in the analysis and provides some institutional background on the main policies of interest. To begin with, I describe the main sources of data that will be used in the economic model and in the reduced-form analysis. I then document some motivating facts. First, the scale, speed, and spatial distribution of the transformation of land use in Brazil from 1985 onwards. I then show that regions with high levels of deforestation tend to engage in less intensive forms of agriculture, both in terms of output per hectare and in terms of workers per hectare. These regions also tend to have lower populations, higher shares of migrants, and a larger share of the workforce in agriculture. Finally, I describe the two main policies that will be analysed in this paper, the establishment of Protected Areas and the Priority List municipalities. In the next chapter I will estimate their effects in reduced-form.

2.1 Data

The data used in this dissertation is available at different levels of spatial granularity. For the reduced-form analysis, I overlay the extension of the Brazilian territory with a hexagonal grid with 10 km width, and I take each hexagonal grid cell to be the unit of analysis. The whole country of Brazil has 100,173 10-km-wide hexagons, each with an area of 779.42 km^2. For comparison, the average municipality in Brazil has an area of 1,571 km^2 and the average municipality in the Legal Amazon[1], where the population is much more sparse, has an area of 6,369 km^2. This is done for two reasons. First, in the case of the analysis of the Protected Areas, their boundaries do not coincide with the boundaries of municipalities, and this requires either looking at a continuous treatment, the fraction of municipality protected, or choosing an arbitrary threshold after which regions would be considered treated, introducing measurement error. Although it is still true that some grid cells will be intersected by Protected Area boundaries, this is much less of an issue than with municipality-level data, as shown in the data appendix section A.1. Second, grid cells are a more comparable units of analysis and increase the statistical power of the analysis. Grid-cells all have the same area whereas the municipalities range from 3.8 km^2 to 161,104 km^2 and the 90th percentile is 23 times the 10th percentile. Thus, for the evaluation of the Priority List municipalities, I will also look at the grid cell level as a robustness check.

For the quantitative model, I need a unit for which there is economic and demographic data as well. The lowest geographical level at which this can be obtained is the municipality level. I choose instead the micro-region level, of which there are 558 in Brazil. At a practical level, this reduces the computational requirements of this analysis significantly. It also has an advantage in terms of interpretation, which is that the micro-regions correspond more closely to local labour markets or commuting zones. The micro-region corresponding to the greater Rio de Janeiro, for example, is comprised of 16 municipalities - one of which is the municipality of Rio de Janeiro. Micro-regions are on average comprised of 10 municipalities, though this ranges from 1 to 41. The median micro-region has 8 municipalities. In the Legal Amazon, since municipalities are larger, each micro-region is on average comprised of 7.5 municipalities, the median is 6, and they range from 2 to 25. Another attractive feature of micro-regions when compared to municipalities is that there is less dispersion in the distribution of their area for Amazonian municipalities. The interquartile ratio of surface areas of municipalities in the Legal Amazon is 6.7, compared to 3.2 for micro-regions.

2.1.1 Land use and land use change data

Data on land use and land use change comes from the MapBiomas project, a multi-institution collaborative initiative that processes satellite images into

[1]The Legal Amazon is the region of Brazil comprised by the states of Acre, Amapá, Amazonas, Maranhão, Mato Grosso, Pará, Rondônia, Roraima, and Tocantins.

publicly available datasets on land use for the Brazilian territory. The first dataset used contains land use data for Brazil between 1985 and 2020 at 30 m resolution. Each 30 m pixel in this dataset belongs to a land use category.

Land use

The original data have 28 land use categories. For most of the analysis, I disregard the distinction between (i) types of forest, (ii) types of other non-forest natural ecosystems, (iii) types of perennial crops, (iv) types of temporary crops, and (v) types of non-agricultural and non-vegetated areas such as urban, mining, and bodies of water. In table 2.1, I summarise the main sub-categories within each of these six broader categories. For the reduced-form analysis and the descriptive statistics, I will aggregate these data at the hexagon level, so that for each hexagon I observe the area it contains in each of these categories for every year. For the quantification of the model, the data is aggregated at the micro-region level.

Land use change

When aggregating at the hexagon-year level, however, some information is lost regarding the details of the land use changes that occurred. If, for example, a 10 km^2 region of a hexagon is deforested and, simultaneously, a 15 km^2 area is reforested, I would only be able to see the net change, that is a 5 km^2 increase in forest area. In order to overcome this challenge, I use MapBiomas' Land Use Change product, which, for every 30 m pixel for 1988-2019, identifies several types of deforestation and several types of reforestation. I will focus on three main types of transitions: (i) loss of primary vegetation to anthropic use, (ii) loss of secondary vegetation to anthropic use, (iii) regrowth, that is from anthropic use to secondary vegetation. It is worth noting, however, that everything that has been forest since 1985 is classified as primary vegetation, and all vegetation in pixels that have been classified as anthropic from two consecutive periods in the past (starting in 1985) is classified as secondary. Figure 2.1 in the following section summarises Brazil-wide trends on land use transitions.

Table 2.1: Fine categories of Land Use 1985-2021

Land Use Class	1985	1994	2003	2012	2021
Forest	5950.86	5685.43	5392.86	5228.32	5083.67
% Forest Formation	73.36	73.84	73.95	74.1	74.57
% Savanna Formation	23.07	22.46	22.14	21.87	21.32
% Mangrove	.17	.18	.2	.2	.21
% Floodable Forest	3.27	3.4	3.59	3.69	3.76
Non Forest Natural	598.68	575.42	572.45	551.8	531.8
% Wetland	33.07	29.7	32.21	29.89	29.16
% Grassland	63.75	67.07	64.49	66.69	67.22
% Salt Falt	.1	.1	.11	.1	.1
% Rocky Outcrop	2.5	2.62	2.64	2.73	2.92
% Herbaceous Sandbank	.55	.47	.51	.55	.58
% Other non Forest	.05	.03	.03	.03	.03
Agriculture	1986.57	2292.09	2593.83	2773.22	2937
% Pasture	54.71	61.97	64.79	61.5	57.54
% Temporary Crops	10.19	13.73	14.55	18.52	21.24
% Soybean	24	42.07	52.69	54.44	65.46
% Sugar Cane	11.33	9.1	10.5	16.79	15.27
% Rice	3.95	.88	1.16	1.51	2.65
% Cotton	0	0	.1	.25	.37
% Other Temporary Crops	60.72	47.95	35.55	27.01	16.25
% Permanent Crops	.43	.37	.54	.68	.85
% Coffee	81.74	72.57	62.21	59.93	56.23
% Citrus	6.3	8.2	10.44	11.12	9.96
% Palm oil	2.91	3.39	3.08	2.59	7.31
% Other Perennial Crop	9.05	15.84	24.27	26.36	26.51
% Forest Plantation	.81	1.51	1.62	2.63	3.22
% Mosaic of Uses	33.87	22.42	18.5	16.67	17.15
Non vegetated area	56.15	44.98	52.37	57.05	65.66
% Beach, Dune and Sand Spot	8.5	10.66	8.36	7.64	6.14
% Urban Area	22.83	46.41	54.67	59.73	59.46
% Mining	1.05	3.21	3.58	4.46	6.36
% Other non Vegetated Areas	67.62	39.72	33.39	28.18	28.04
Water	189.69	184.02	170.42	171.46	163.76
% River, Lake and Ocean	99.82	99.79	99.72	99.67	99.64
% Aquaculture	.18	.21	.28	.33	.36

Note: This table shows the distribution of land use across Brazil in a nested set of categories for the years 1985, 1994, 2003, 2012, and 2021. For the coarsest land use classes (forest, non forest natural vegetation, agriculture, non vegetated area, and water) I display the total area in thousands of square kilometers. For the second tier, I show the percentage of the first-tier category in that land use. For instance 3.27% of the 5.95 million km^2 of forests in 1985 were floodable forests. Similarly, for the third tier (present for temporary crops and permanent crops) I display the % of that land use out of the second tier category it belongs to. For instance the area in coffee in 2021 was 56.23% $\times$ 0.85% of the 2.9 million km^2 in agriculture.

2.1.2 Geographic data

Administrative boundaries

Data on the administrative boundaries of municipalities, micro-regions and states comes from the Brazilian Institute of Geography and Statistics (IBGE) and is publicly available. Brazil is a federation comprised of 27 federative units: 26 states and 1 federal district, the capital city of Brasilia. The current division in states has remained the same since 1988. States are in turn divided into municipalities. As of 2023, there are 5,570 municipalities. For the quantitative model I will select a higher scale of aggregation, the micro-region, of which there are 558 including Brasilia, which more closely resemble commuting zones.

Conservation Policy data

The geo-coded information on Indigenous Territories comes from the website of the National Foundation of Indigenous Peoples (FUNAI)[2]. The FUNAI also provides information on the year in which they have undergone the five stages of approval (study, delimitation, declaration, homologation, and regularization). I consider the homologation as their year of creation, as this is the stage of judicial approval officially recognising the territory as an indigenous land.

Geo-coded information on the Conservation Units comes from the website of the Ministry of the Environment[3]. They also contain data on the year of creation of the various conservation units.

Environmental data

I gather publicly available data on temperature, precipitation, soil moisture, and the carbon density of ecosystems from a variety of sources. Temperature data comes from TerraClimate. The data gathered is the minimum and maximum monthly temperature at 4638.3 m resolution for years 1985-2019. Precipitation data comes from ERA5. The data gathered is the monthly aggregate precipitation in mm at 11132 m resolution for years 2000-2019. Soil moisture data also comes from TerraClimate. It is measured in monthly mm and is derived using

[2]https://www.gov.br/funai/pt-br/atuacao/terras-indigenas/geoprocessamento-e-mapas

[3]https://antigo.mma.gov.br/areas-protegidas/cadastro-nacional-de-ucs/dados-georreferenciados.html

a one-dimensional soil water balance model. The spatial resolution and time period are the same as for temperature. Above and Below Ground Biomass Carbon Density comes from the UN Environment Programme World Conservation Monitoring Centre (UNEP-WCMC) carbon biomass dataset. This dataset estimates a snapshot of biomass carbon density for 2010. I use these data, alongside land use data for 2010, in order to assign an average carbon density to the natural ecosystems in different micro-regions and that way approximate the carbon emissions under different simulated scenarios.

2.1.3 Economic data

This paper uses economic data for four purposes. First, as ancillary outcomes of the reduced-form analysis of Priority List municipalities in chapter 3 [4]. Second, to calibrate some of the structural parameters of the model in 5.1. Third, to structurally estimate the productivities at the micro-region-year level in 5.2. And fourth, to validate the model estimation by checking the correlation of model-derived moments and unmatched moments in A.3. The precise way in which the data is used will be described in the relevant sections in greater detail. All price data is deflated to 1994 Reais.

Agricultural data

Our main sources of agricultural data are: the 2006 Agricultural Census, the Municipal Survey on Agricultural Production (PAM) for 2000-2019, and the Municipal Survey on Livestock Production (PPM). The variables of interest from the agricultural census, at the municipality level and for each agricultural activity (pastures, temporary crops, and permanent crops), are: (i) agricultural area, (ii) number of workers employed in that activity, and (iii) revenues. From the PAM I have, at the crop-municipality-year level, estimates of: (i) area planted, (ii) production (kg), (iii) revenues. The PPM, for livestock, does not provide area not revenues, but it has an estimate of the number of cows per year per municipality which I combine with other data to estimate revenues and intensity of production.

The main agricultural data I need for the model is a yearly panel of the revenues from each of the three primary agricultural activities modelled. The Municipal

[4] Since the most granular level at which this is observed is the municipality level, this analysis is done at the municipality instead of the grid-cell level.

Survey on Agricultural Production (PAM), provides yearly estimates of the revenues of 71 different crops. For the model, I aggregate crops into two categories: temporary and permanent, in a way that is consistent with their classification in the agricultural census. These estimates yield very similar values to those in the agricultural census in 2006.Instead of relying on the PAM alone, I construct an region-level index of the PAM revenues that equals one in the 2006 and multiply them by the 2006 Agricultural Census data as follows

$$\widehat{Reveue}^{k}_{rt} = Revenue^{k,AgCensus}_{r2006} \times \frac{Revenue^{k,PAM}_{rt}}{Revenue^{k,PAM}_{r2006}}$$

for each $k \in \{\text{permanent crops}, \text{temporary crops}\}$. This ensures that I use the cross-sectional variation in levels coming from the agricultural census, which is likely to be more accurate, and then use the PAM to estimate the relative yearly variation in each municipality.

No data on yearly revenues from cattle ranching at the municipality level exists for Brazil, so I leverage cross-sectional data on other sources that cover the dimensions governing yearly revenue changes: the number of cows, the weight of each cow, and the price of beef. I obtain yearly estimates of the number of cows in each municipality from the Municipal Survey on Livestock Production (PPM). I obtain data on yearly national variation in the price of beef in each year from a daily time series curated by the School of Agricultural Studies of the University of São Paulo[5]. I obtain yearly state-level data on the average weight of slaughtered cows from the national statistical office's Quarerly Survey of Salughtered Animals. Given the average age of a cow at the time of slaughter is 3 years, I estimate yearly cattle revenues by multiplying the 3-year lagged change in weight, number of cows and price of beef by the 2006 revenue data from the Agricultural Census[6], so that for region r in state $s(r)$:

$$\widehat{Reveue}^{\text{pasture}}_{rt} = Revenue^{k,AgCensus}_{r2006} \times \frac{p^{\text{beef}}_{t}}{p^{\text{beef}}_{2006}} \times \frac{kg/cow_{s(r)t-3}}{kg/cow_{s(r)2003}} \times \frac{\#\, cows^{PPM}_{rt-3}}{\#\, cows^{PPM}_{r2003}}$$

[5]Available for public consultation at: https://www.cepea.esalq.usp.br/br/indicador/boi-gordo.aspx

[6]This extrapolation over time requires the assumption that the yearly changes in the price of beef can be considered as constant across all municipalities, and that yearly changes on average slaughtered weight can be considered constant across all municipalities within a given state.

Other economic data

For the validation of the model and as ancillary outcomes of the municipality-level reduced-form analysis of the Priority List municipalities, I use data on municipal GDP and wages.

I leverage data on Gross Domestic Product and gross value added by major sector (agriculture, manufacturing, services, and public administration) at the municipality level for the years 2002-2019. These data are estimated by the Brazilian Institute of Geography and Statistics (IBGE) in partnership with State Statistical Organisations, State Government Departments, and Free Trade Zones.

Wage data comes from the population census of 2000 and 2010. In both of these individuals are asked to report their current monthly income. I aggregate these data so that I have, for each micro-region and each year, the average wage in agriculture and the average wage in non-agriculture. These data are also used in order to calibrate the share of labour in agricultural production in the model.

Internal trade

Data on interstate trade flows comes for 1999 comes from (de Vasconcelos, 2001), as cited by Morten and Oliveira (2018). This is a matrix of the inter-state trade in goods and services of Brazil. These data come from administrative records collected by the author from each state's office. At the time the records were not centralised nor digitised and it is likely that there were more gaps and inconsistencies. For 2017, the data comes from the National Council of Fiscal Policy (CONFAZ), which relies on data from the Electronic Invoice (NF-E) to establish the interstate commercial balance. The NF-E contains details of goods departure, including its destination and value, as well as information about the entry of merchandise. CONFAZ aggregates these data to calculate interstate exports and imports. Given potentially large gaps in trade flow records (in 1999, five states, mostly Amazonian, did not have any records of inter-state trade) I use these data to estimate a distance-elasticity of trade rather than to match observed trade flows. Moreover, this analysis is at the micro-region level, not at the state level, so I will use the state-level-estimated distance elasticity of trade costs to approximate the iceberg trade costs that decays with distance. More details can be found in subsection 5.1.5.

Household expenditures

Finally, I use the 2017-2018 Household Expenditure Survey (POF) to calibrate the preference parameter that governs the non-homotheticity in individual utility. This survey contains individual-level data on incomes and itemised expenditures which are used to see the correlation between individual income levels and the share of their expenditure that goes to food. I use household expenditure shares on food items as a proxy for consumption of agricultural goods.

2.1.4 Demographic data

Data on micro-region level labour markets comes from 10% samples of the 2000 and 2010 National Census, which is representative at the municipal level (smaller than the micro-region), made publicly available by IBGE. For each municipality, I compute the share of individuals who report working in the agricultural sector, as well as the average municipal earnings for agricultural and non-agricultural activities. I get the number of workers in agricultural and non agriculture for both of these years and then, for each micro-region, I interpolate and extrapolate these values linearly to approximate their time series from 2003 to 2019. Given the static nature of the equilibrium concept, it is not too important to accurately measure relative yearly changes in population and labour shares, but to get the levels right. Therefore, I think that the errors introduced by this approximation are unlikely to change the main results. I also use the 2010 census to obtain information on internal migration, as respondents are also asked to report in which municipality they lived 5 years prior to data collection.

2.2 Deforestation over time and space

This section discusses some of the main characteristics of deforestation in Brazil over time and space. there are four main points that are worth highlighting. First, deforestation had a persistent decrease around 2003-2008 across Brazil. Second, most of what I will call deforestation is a process that begins with the conversion of natural ecosystems towards pastures. Third, deforestation is highly concentrated around what I will define the agricultural frontier. Fourth, deforestation happens in regions that are more sparsely populated, poorer, and less productive than the rest of Brazil.

Decreasing trends in deforestation

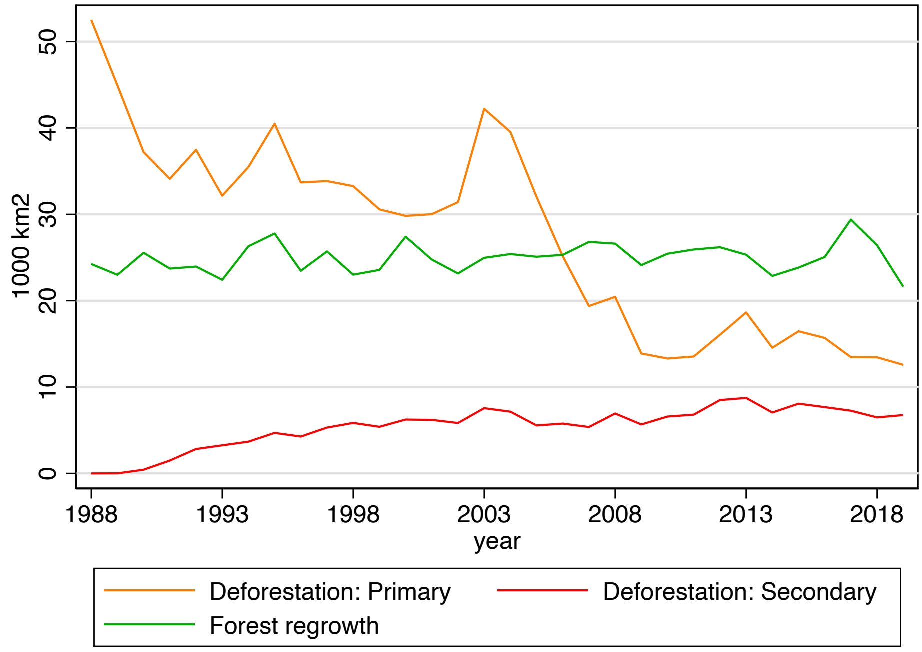
A first glance at the aggregate trends of deforestation in Brazil shows that it has decreased dramatically since the 1980s and 1990s. Figure 2.1 (a) show how the yearly loss of primary vegetation loss to anthropic use has gone from around 35,000 km^2 until 2003 to on average 15,000 km^2 from 2008 onwards. Meanwhile, the rate of forest regrowth has remained roughly constant at around 25,000 km^2 per year. Most of the 2003-2008 deforestation decline is coming from primary forest loss in the legal Amazon.[7]

Deforestation is mostly the conversion of forests and savannas into pastures

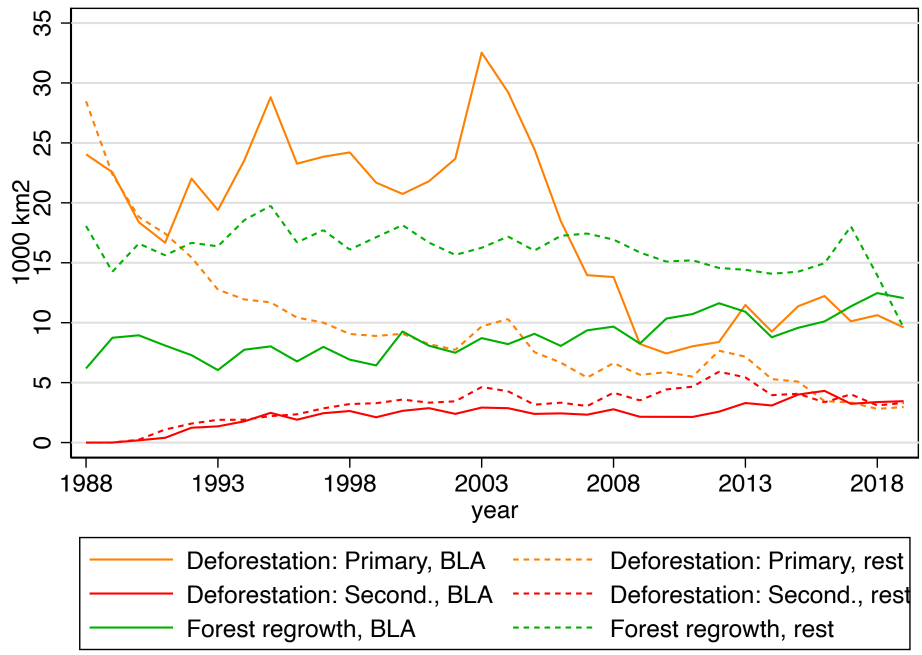
Another fact the emerges from the data is that the majority of the loss of natural ecosystems is a transformation of forests and savannas into pastures. Figure 2.2 decomposes the loss of natural ecosystems (including both primary and secondary vegetation) into the share of the various transitions across the finest categories reported by MapBiomas, as shown in table 2.1. These shares are averaged out by decade, except that the 1980's start with 1985. All transitions comprising at least 0.5% of the loss of natural ecosystems are included in a row, and the rest are aggregated under "Other". The graph shows that around 40% of the loss of natural ecosystems is conversion of forests to pastures, followed by the conversion of savannas to pastures (15%). Next, is the conversion of natural forests, savannas, and grasslands to a mosaic of uses, that is a combination of crops and pastures. Together these three transitions account for between 20% and 35% of primary ecosystem loss. The direct conversion of forests or other natural ecosystems to "pure" crop farming is almost non-existent and that almost all conversion of land towards agricultural use is accompanied by cattle. The natural ecosystems that seem to be msot converted into cropland directly are grasslands (into soybean) and savannas. It is also worth noting that conversions

[7]The loss of secondary forest cover seems to have increased, but this is partly a mechanic result of the fact that a the beginning of the series there are no forests that are classified as secondary to cut down. When looking at deforestation as a percentage of the standing forests of each type, there seems to be a decrease in both trends (See figure A.3). It is also interesting to note that, although loss of secondary forests accounts for a relatively small amount of deforestation, a much larger percentage of secondary forests are deforested each year (3-5% compared to .under 0.2-0.8%) This is consistent with some persistence in the costs of deforestation. That is, that having been deforested in the past reveals lower costs of deforestation in a given location, which to some extent persists over time. It is also consistent with the fact that secondary forests may be less valuable as standing forests.

Figure 2.1: Yearly land use change transitions



(a) all of Brazil



(b) Brazilian Legal Amazon vs. the rest of Brazil

Note: This figure illustrates the Brazil-wide yearly trends in the gross loss of primary forest cover, secondary forest cover, and in forest regrowth. Source: MapBiomas Collection 8, 2024.

of natural ecosystems into permanent crops, urban use, and mines, are negligible in terms of area. This does not preclude the possibility that, after having been converted into pasture, land gets planted with crops or is converted into any other use.

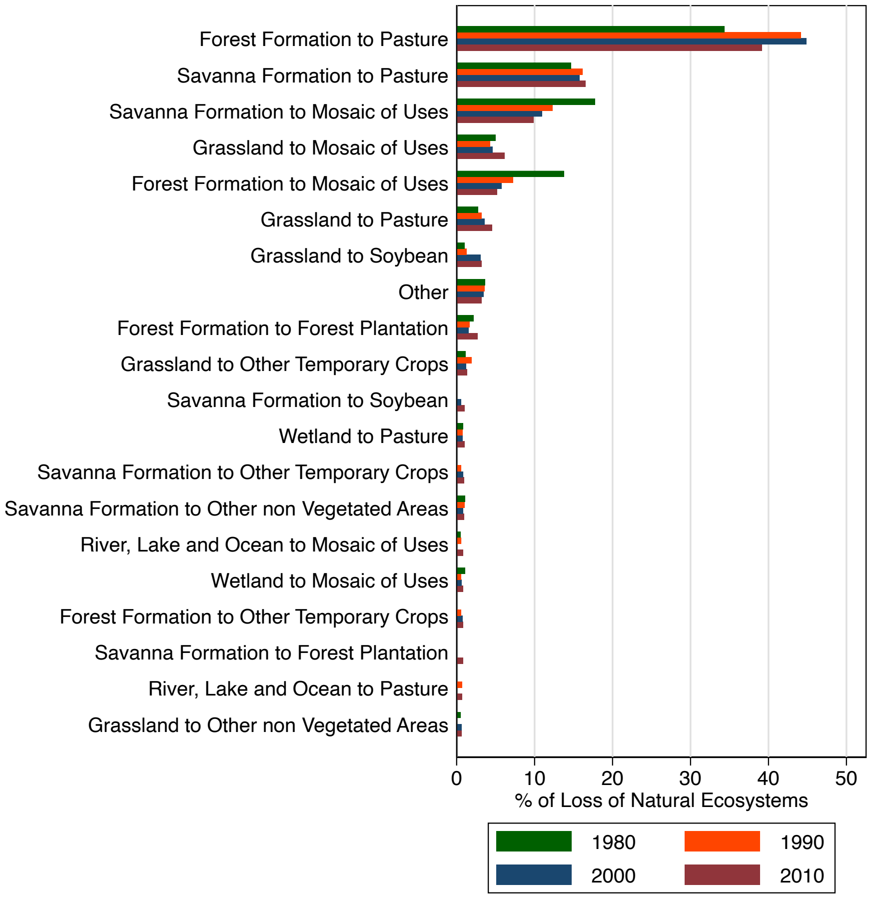
The agricultural frontier

Third, deforestation in Brazil over the past few decades is strikingly concentrated in space. The map in figure 2.3 shows the spatial patterns of the conversion of natural vegetation. To construct the “forest edge”, I do the following. First, I categorise each 10 km hexagon as being natural or anthropic depending on whether at least 50% of its surface area is in natural ecosystems. Then, I dissolve all contiguous natural hexagons into contiguous polygon. The edges of these polygons are then what I consider the “forest frontier”. A stark pattern emerges. Most of the area that has been converted from vegetated to non-vegetated is concentrated around the outer borders of the Brazilian Legal Amazon, mainly in the northernmost part of the Center-West region (in the states of Goiás and Mato Grosso) and in the southernmost and easternmost borders of the Northern region (in the states of Rondônia and Pará. This area of concentrated deforestation coincides with the area that has seen the strongest expansion of agricultural activities in the country, and is often referred to as the “agricultural frontier”. Figure 2.4 illustrated how the vast majority of forest loss in Brazil over the past few decades has occurred at a relatively short distance from the frontier. In particular, 40% of forest loss (both primary and seconary) has occured in the 10 km band that lies just inside the agricultural frontier. Morover, from the map in figure 2.3, it is very clear that the frontier has been shifting inwards over time. In the 1980’s and 1990’s, deforestation was more concentrated towards the outermost borders of the frontier, and has been gradually shifting inwards, and reaching deeper parts of the Amazonian region.

Socio-economic characteristics of high-deforestation areas

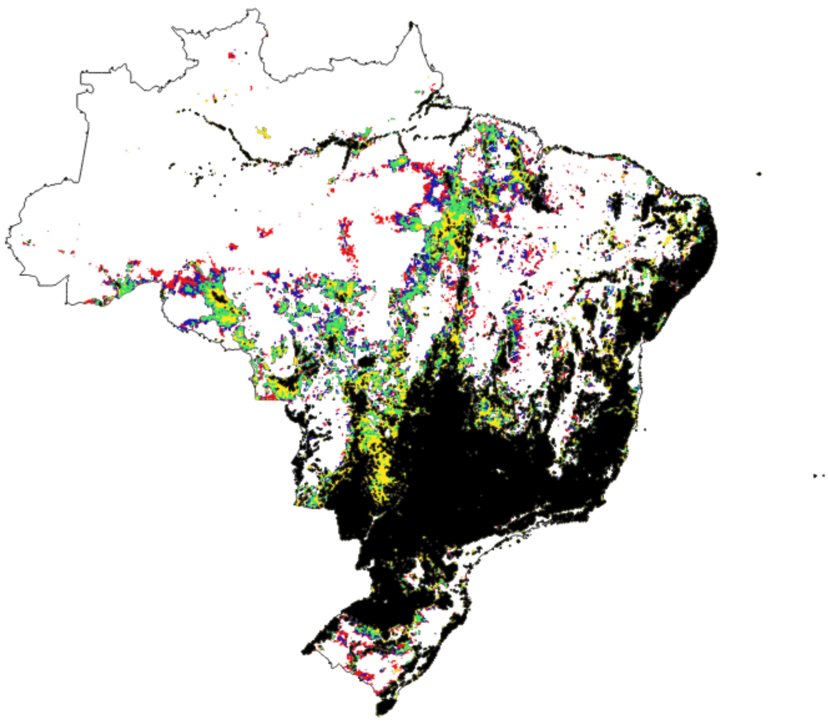
Another way of highlighting the spatial concentration is by ranking administrative units by levels of deforestation, and comparing the concentration with that of other variables such as area, population, and GDP. Out of the 558 micro-regions that make up Brazil, the top 20 (100) micro-regions account for almost 40% (80%) of the total yearly deforestation, while only accounting for much

Figure 2.2: Loss of natural ecosystems: decomposition (% contribution)



Note: This figure decomposes the broad land use change category of "loss of natural ecosystems", which refers to all conversion of natural ecosystems to anthropic uses, sometimes referred to throughout this thesis as "deforestation" into finer categories of land use change categories, specifically the categories in table 2.1. The different colors of the bars indicate different years for which this decomposition is done. Source: MapBiomas Collection 8, 2024.

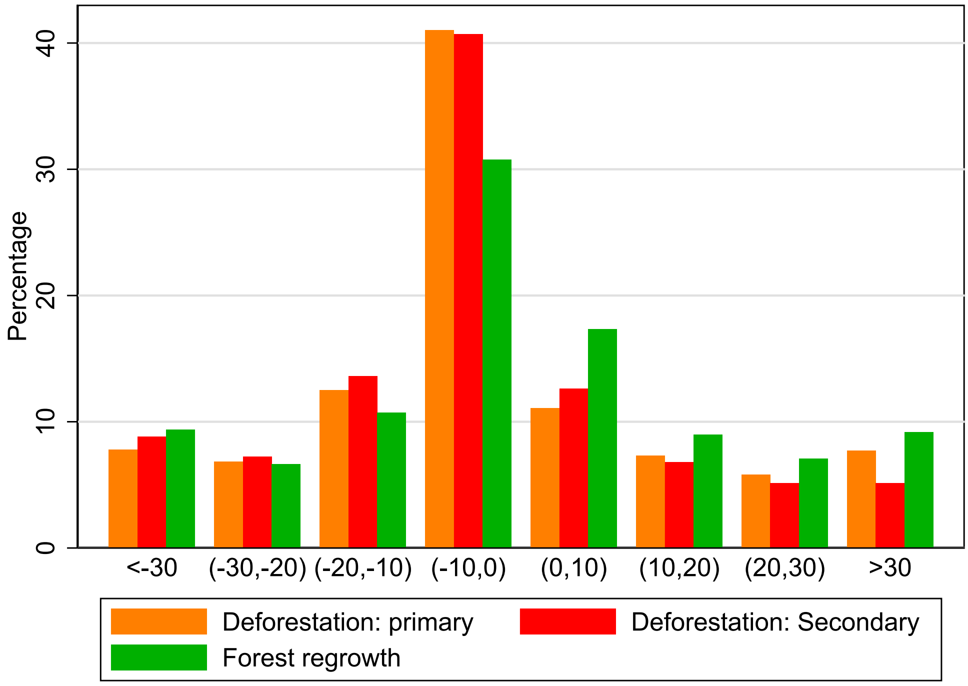
Figure 2.3: Shifting frontier in Brazil 1985-2020



Note: This map shows the gradual loss of natural vegetation from 1985 to 2021. In black, we can see the areas that, as of 1985, were not covered by natural vegetation. To make the data more visually clear, we coarsened the MapBiomas 30 m resolution land use data to the level of the 10 km-wide hexagonal grid cells and assigned a binary indicator (natural vegetation or not) to each of them. That is, they were either agricultural, non vegetated areas, or water bodies. In yellow, we see those that changed to the status of non-vegetated between 1985 and 1994. In green, 1994-2003; in blue, 2003-2012; and in red, 2012-2021.

smaller shares of the surface area, value added in agriculture, population, and GDP of Brazil. This is shown in figure 2.5, which shows the cumulative share of (i) deforestation, (ii) area, (iii) value added in agriculture, (iv) population, and (v) GDP. As expected, the micro-regions with highest deforestation are on average larger, but they deforest more even in proportion to their area. The more striking facts shown by this graph are that micro-regions with high levels of deforestation have lower population (despite having larger area) and lower GDP. The 100 micro-regions accounting for 80% of Brazil's deforestation only account for under 30% of its value added in agriculture, well under 20% of its

Figure 2.4: Spatial concentration around the "deforestation frontier"



Note: This figure illustrates what percentage of three land use transitions (loss of primary forest, loss of secondary forest, and forest regrowth) happens within 10 km-wide hexagons that are at varying 10 km distance bins from the forest edge. Negative distances are within majority anthropic-use hexagons, (or within the agricultural frontier), and positive distances are outside of it, or in hexagons with majority "natural" ecosystems.

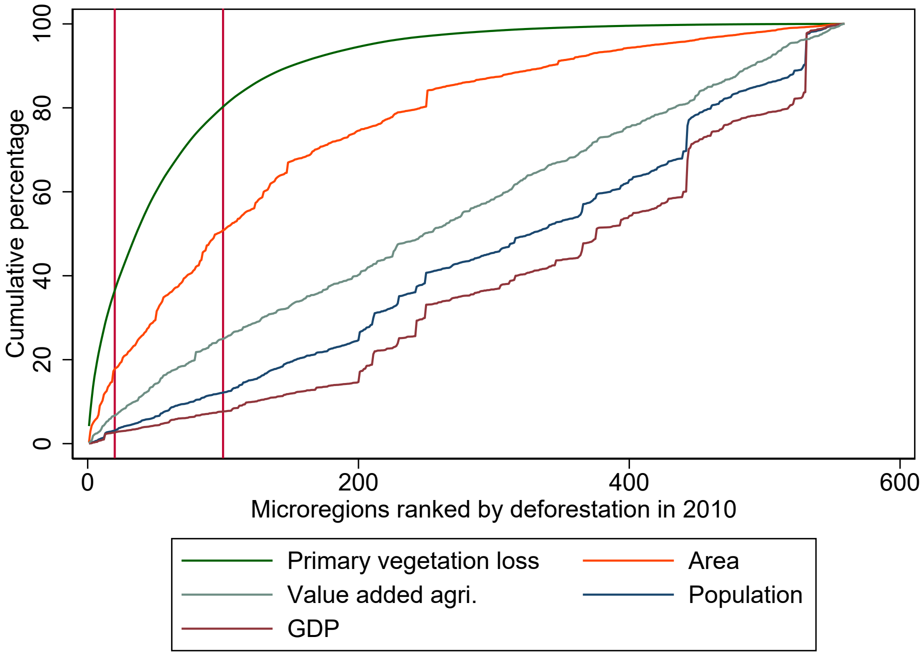
population, and well under 10% of its GDP.

Regions with high levels of deforestation are very different from the rest of Brazil in socio-economic terms. Deforestation happens in poorer and more remote areas, where land is dedicated to activities that are more land-intensive and less productive.

2.3 Locally targeted conservation policies

Existing research suggests that Brazil's approach to tackling deforestation over the first two decades of the 21st century has been successful in reducing country level deforestation (Burgess et al., 2019; Assunção et al., 2015). During this

Figure 2.5: Cumulative distribution of deforestation relative to other socio-economic variables, at the level of micro-regions



Note: This figure displays the cumulative percentage of several variables (primary vegetation loss, area, value added in agriculture, population, and GDP) across Brazil's 557 continental microregions ranked by their level of deforestation (more specifically, primary vegetation loss) in 2010. The red vertical lines indicate the top 20 micro-regions and the top 100 micro-regions in terms of primary forest loss.

period, a vast array of policies were implemented by the federal government. Some of these affect the entire Brazilian territory or the Brazilian Legal Amazon, such as changes in laws increasing legal limits for deforestation in private land (DOU, 2012) and the adoption of a unified *Action Plan for the Prevention and Control of Deforestation in the Legal Amazon* (PPCDAm). Others, however, were locally targeted policies. That is, the federal government established specific regions that meet the required criteria and increase the anti-deforestation efforts in those regions. This approach raises the concern that there may be leakage, that is, that part of the forest loss avoided in these areas has moved elsewhere.

Institutional background

Among the local anti-deforestation policies enacted by the Brazilian government, two have been especially prominent. The first is the establishment of "Priority Municipalities" policy, enacted by the Brazilian government in 2007 (DOU, 2007). A set of municipalities in the Amazon region where deforestation rates were among the highest in the country were selected to be subject to extra enforcement actions. In a first round, 36 municipalities accounting for around 45% of the previous year's deforestation were included in the Priority List, which has been updated on a yearly basis ever since. Selected municipalities were subject to increased law enforcement activities such as fines, embargoes on private farms, political agreements with local leaders, and credit incentives from the federal government. After the start of the policy, yearly deforestation in targeted municipalities significantly decreased relative to the non-targeted municipalities in the Brazilian Amazon (Assunção and Rocha, 2019). The second prominent example of local conservation policies is the continuous establishment by the Brazilian Government of specific areas where deforestation is completely banned. Such areas are established either for wildlife and biodiversity conservation – the so-called Unidades de Conservação (Conservation Units, from now on UCs), or for preservation of land that has been traditionally inhabited by native indigenous people – the so-called Territórios Indígenas (Indigenous Territories, from now on ITs). These policies began with the country's return to democracy in 1985 and, as of 2022, a total area of approximately 3,500 km^2 has been granted one of these two statuses. Due to their practical similarities, both types of policies will hereafter be referred to as Protected Areas. Unlike the Priority Municipalities policy, Protected Areas typically have high forest coverage and little deforestation, and the goal is to conserve the natural biome and prevent future potential increases in deforestation rather than to crack down on ongoing deforestation. Priority Municipalities, Conservation Units, and Indigenous Territories are all chosen so that they meet specific criteria. This means that there will be potential selection, not only on level of deforestation and forest cover, but also on their trends. In the next chapter I will discuss the methods used to identify their causal effect. The preferred identification strategy for the priority list municipalities will be a synthetic difference-in-differences approach in order to match the pre-treatment trends. For the Protected Areas, the preferred approach will be a regression discontinuity design (RDD).

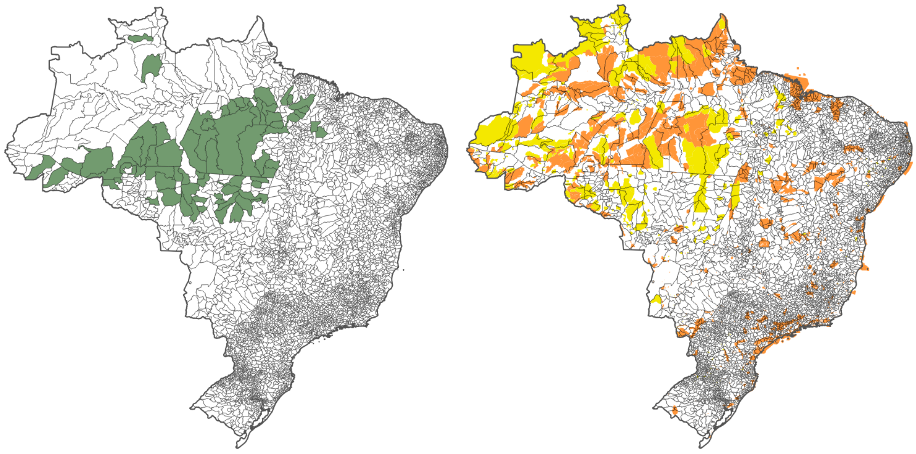
Why targeting?

Before turning to a quantitative analysis of leakage in place-based policies from Chapter 4, it is important to discuss why the need to target only some parts of the country in the first place. Firstly, the enforcement of conservation policies is costly – research estimates that, in order to place 80% of the Brazilian Amazon under some form of currently existing policy, the federal government would need to spend at least 1.7 Billion USD per year (da Silva et al., 2022). Secondly, from a social welfare perspective, the conversion of forested land into agriculture generates private profits, which could result in a non-zero optimal level of deforestation from the perspective of Brazil. Thirdly, and related to the privately optimal argument, there is a political cost of restricting deforestation activities, since the proceeds of those are often captured by local political elites. Fourthly, from a public revenues perspective, modern agriculture that results from forest removal is more easily measured and, consequently, taxed, than alternative sustainable economic activities.

Descriptives

Figures 2.6 and 2.7 illustrate the evolution of place-based anti-deforestation policies in Brazil between 1985 and 2020. From Figure 2.6, it can be seen that both policies are largely focussed on areas in Amazonian region. The Priority List (in green) targets exclusively municipalities within the Brazilian Legal Amazon, most of which located in the so-called “deforestation arc” covering the south of the states of Amazonas and Pará and the north of the state of Mato Grosso.

Figure 2.6: Spatial distribution of Priority List municipalities (green), Conservation Units (orange), and Indigenous Territories (yellow) as of 2021.



Note: This figure illustrates the spatial targeting of the three relevant types of protected areas implemented by the Brazilian government over the past decades. The left panel shows, in green, the municipalities that have been added to the Priority List since its start in 2008. The right panel shows the location of protected areas (which do not necessarily coincide with municipal borders) according to its type: orange areas are Conservation Units, yellow areas are demarcated Indigenous Territories

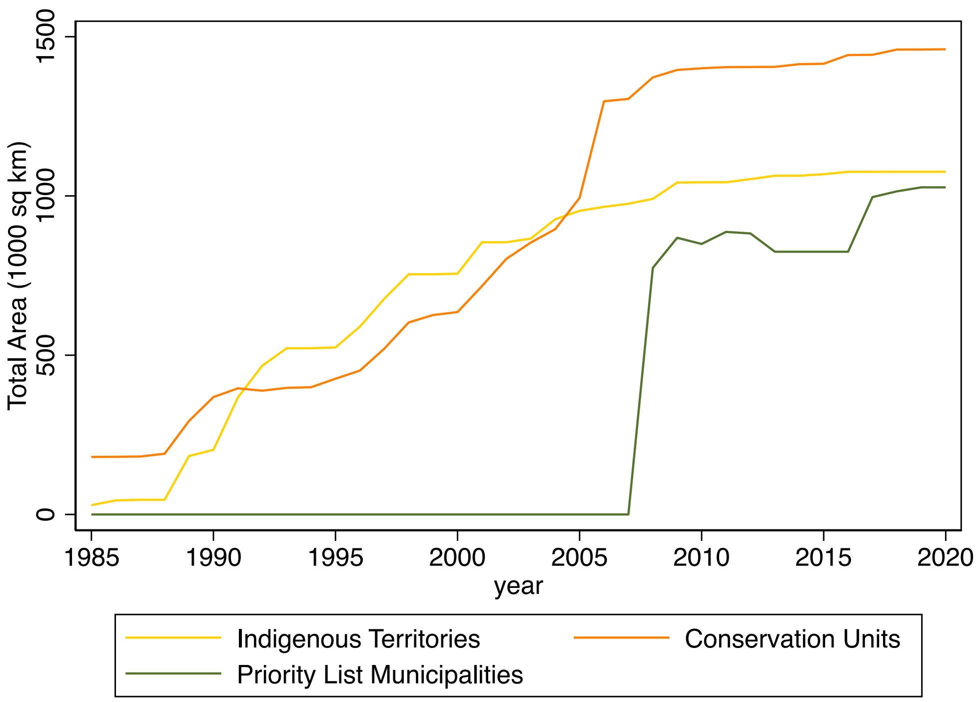
Although the Protected Areas are somewhat more spread across the country[8], they are still spatially concentrated in the Amazon biome, north of the municipalities in the Priority List, in areas with more forest cover. Figure 2.7, shows the temporal evolution of these policies. Protected Areas have been gradually established since the country's re-democratisation in 1985, whereas the Priority List policy was created in 2007, with most municipalities being added to the list in that 2008.

Figure 2.8 below illustrates trends in natural vegetation cover and rates of natural vegetation loss in regions defined by their conservation status. Attention is restricted to the Legal Amazon given that it is where most of the deforestation has been happening over the study period. The Legal Amazon is then divided in ITs, CUs, Priority List municipalities, and the rest of the Legal Amazon which is neither[9]. Between 1985 and 2020, Priority List municipalities and the

[8]Mainly due to territories historically occupied by indigenous peoples in the Northeast and Southeast regions

[9]To keep the area constant and avoid conflating land use changes with protected status changes I take the ITs and CUs as of 2020 and for the Priority List I consider all municipalities ever in the list. Note that there is considerable overlap between Indigenous Territories, Conservation Units, and Priority List municipalities

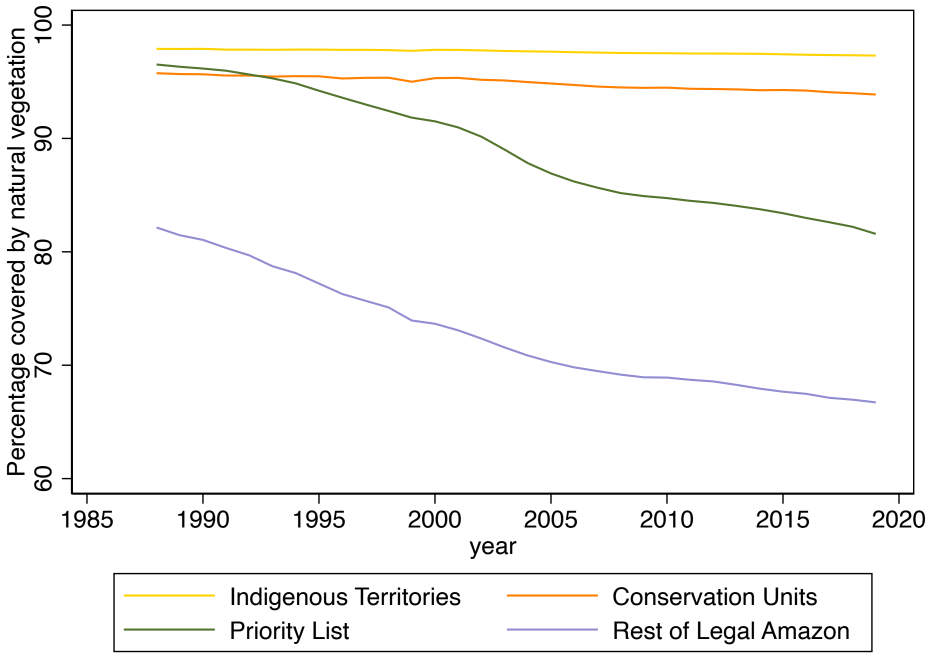
Figure 2.7: Evolution of spatially targeted conservation policies



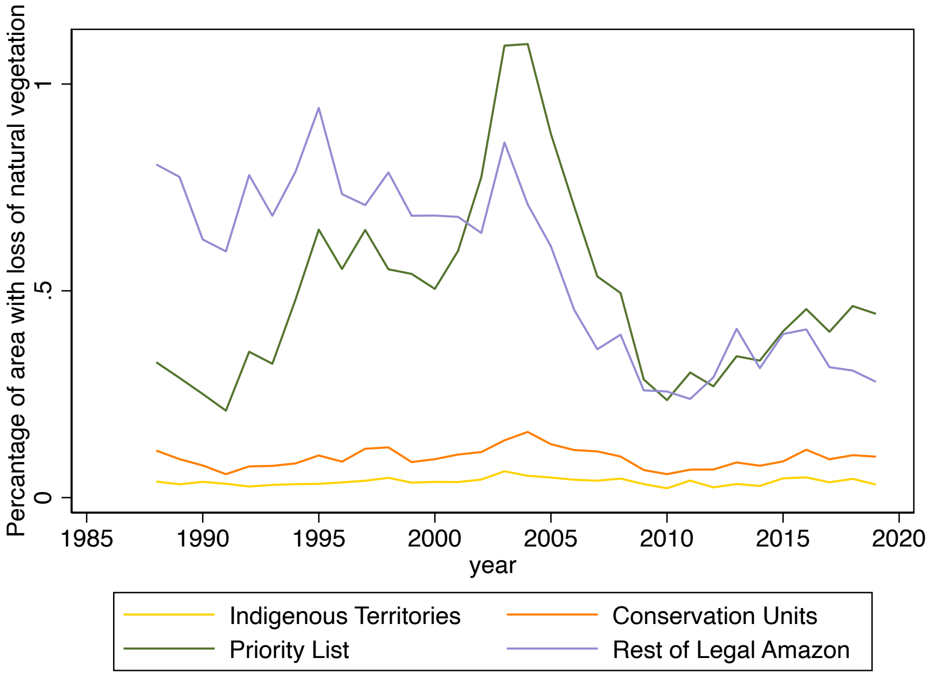
Note: This figure illustrates the evolution over time of the total area under different types of locally targeted policies by the Brazilian government. The yellow line represents the total area of demarcated Indigenous Territories, the orange line represents the total area under Conservation Units, and the green line represents the total area of all municipalities included in the Priority List policy.

rest of the Legal Amazon lost much more natural vegetation as a proportion of their area than Protected Areas. Throughout the period, forest cover in ITs and CU has remained roughly constant, around 97% and 95% respectively. The rate of deforestation has been also roughly constant, although higher in CUs, at 0.1% of their area per year, as opposed to under 0.04% in ITs. The areas that were chosen as Priority List municipalities starting in 2008 used to be as vegetated as Protected Areas until the early 1990s. These places, however, had increasing rates of forest loss from the early 1990s until the early 2000s, peaking at over 1% of their total area (around 15,000 km^2) in 2003 and 2004, going down dramatically to near 1990 levels in 2010 (around 0.24% of their area, or 3,300 km^2) and they have increased again in the period 2010-2019, nearly doubling. The rest of the Legal Amazon has had a significant loss in forest cover as well, although starting from a lower baseline of 82% in natural vegetation in 1988 and going down to 67% in 2019. The rates of natural vegetation loss there were around 0.74% of the total area (between 15,000 km^2 and 20,000 km^2) from 1988 until 2003, when they declined over the course of four years to around 0.32% for the rest of the period.

Figure 2.8: Land use trends by locally targeted conservation areas



(a) Natural vegetation, Legal Amazon



(b) Primary forest loss, Legal Amazon

Note: This figure shows the evolution over time of (a) the coverage of natural vegetation and (b) the loss of primary forest, both as a share of total area, for regions under the different types of locally targeted policies implemented the Brazilian government, restricting attention to the states that belong to the Brazilian Legal Amazon. The yellow line corresponds to Indigenous Territories, the orange line to Conservation Units, the green line to municipalities from the Priority List, and the lavender line to the rest of the rest of the Legal Amazon.

Chapter 3

Local Effects of Targeted Conservation

Figure 2.8 in the previous chapter illustrates the main aggregate facts regarding locally targeted conservation in the Brazilian Legal Amazon. Protected Areas have always had very high forest cover and minimal deforestation. Priority List municipalities had a sharp increase in deforestation until 2003, when deforestation began to decrease sharply across all of the Legal Amazon. This decline was more pronounced in Priority List municipalities than in the rest of the Legal Amazon.

These facts alone are not convincing evidence that locally targeted conservation has worked as an effective halt on deforestation. There are two main concerns regarding the validity of a comparison of treated and untreated regions. Firstly, the facts above are strong evidence of selection. There is selection in levels of forest cover, in trends of forest cover, that is, in levels of deforestation, as well as in trends of deforestation. I will rely on a synthetic difference-in-differences strategy to evaluate the effectiveness of the Priority List (section 3.1) and a Regression Discontinuity Design for the effectiveness of Protected Areas (section 3.2).

Secondly, when evaluating the effectiveness of any local conservation policy is the possible presence of violations of the Stable Unit of Treatment Variable Assumption (SUTVA). I will attempt to quantify the magnitude of this effect, leakage, through a spatial general equilibrium model in the following chapters.

3.1 Priority List

3.1.1 Econometric specification

To evaluate the effectiveness of placing municipalities in a Priority List, my preferred approach is synthetic differences in differences using as unit of analysis the municipality (Arkhangelsky et al., 2021). The main regression equation is

$$\text{(log) Forest Loss}_{mt} = \delta_t + \gamma_m + \beta \text{Priority}_{mt} + \epsilon_{mt} \tag{3.1}$$

and the dynamic version with different coefficients for different years relative to treatment, or event study, is

$$\text{(log) Forest Loss}_{mt} = \delta_t + \gamma_m + \sum_{\tau=-N_L}^{N_F} \beta_\tau \text{Priority}_{mt-\tau} + \epsilon_{mt} \tag{3.2}$$

where (log) Forest Loss_{mt} is the logarithm of the loss of natural vegetation observed in municipality m at year t, δ_t are year fixed effects, γ_m are municipality fixed effects, $\text{Priority}_{m,t-\tau}$ is a dummy variable equal to one if municipality m has been added to the Priority List exactly τ years ago, and β_τ are the coefficients of interest. I consider forest loss data between years 1995 and 2019.

Parallel trends

The validity of the event study approach relies on the assumption of parallel trends, i.e., treated and untreated (or not-yet-treated) municipalities followed parallel trends in deforestation rates in the years leading up to their treatment year, which implies $\beta_\tau = 0 \; \forall \; \tau \; \in [-N_L, 0)$. Given that the Priority List policy was explicitly targeted at municipalities considered "deforestation hotspots", treated municipalities followed different trends by design. The synthetic difference-in-differences model employed relaxes the parallel trends assumption and weights observations so that pre-periods exhibit parallel trends. The assumption is therefore that parallel pre-trends imply counterfactual trends after the implementation of the policy. The method used also adjusts for the problems arising from staggered adoption so that already treated units are never used as controls for units treated in later years. It also weights observations so

Table 3.1: Synthetic differences in differences (all outcomes in logs)

	Nat. veg. loss	Primary loss	Secondary loss	Regrowth
	(1)	(2)	(3)	(4)
Priority List	-0.445***	-0.443***	0.161***	0.210***
	(0.060)	(0.074)	(0.058)	(0.046)
Observations	19,575	19,375	17,025	19,375

Note: This table shows the results of the municipality-year-level the synthetic differences in differences regression of deforestation and reforestation outcomes (specifically: total loss of natural vegetation, loss of primary natural vegetation, loss of secondary vegetation, and gain of secondary vegetation) in logarithms on the onset of the Priority List. I only include municipalities in the Brazilian Legal Amazon, as they are more likely to resemble Priority List municipalities. Municipalities are dropped whenever there is any year in which the outcome is missing, which occurs in years with zero forest loss or regrowth. Standard errors, in parenthesis, are calculated via bootstrap with 50 repetitions. * p<0.10, ** p<0.05, *** p<0.01.

that the parallel trends assumption is closer to being satisfied[1].

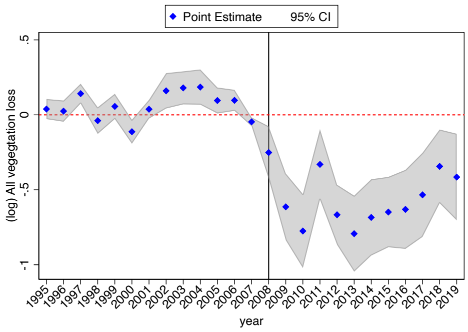
3.1.2 Results

Table 3.1 shows the average effects of the treatment effect on four main outcomes variables: (i) all loss of natural vegetation, (i) primary vegetation loss, (iii) secondary vegetation loss, and (iv) secondary vegetation gain or forest regrowth. These results come from looking at the Legal Amazon, where municipalities are more comparable to Priority List municipalities, and looking at all Priority List cohorts. Figure 3.1 below shows a visual representation of the event study estimates. For clarity of exposition, the event study restricts the sample to those municipalities that are treated in 2008 or those that are never treated. Panel (a) show the effects on the logarithm of the total loss of area in natural vegetation in the municipality. This is a sum of the loss of primary vegetation and the loss of secondary vegetation. The effects on the (log) primary vegetation loss and the (log) secondary vegetation loss are shown in panels (b) and (c) respectively. While panel (d) shows the effects on the (log) regrowth of secondary vegetation.

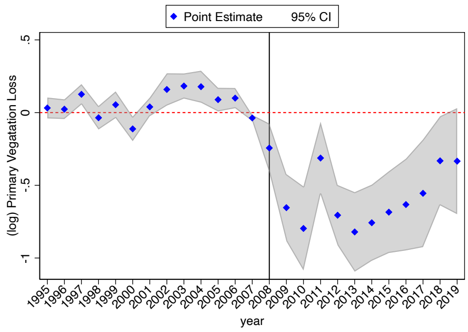
From Table 3.1, it can be seen that there is a large and statistically significant decrease in total forest loss of 0.44 log points, or 35%. This is the same as the observed reduction in primary vegetation loss, which is the vast majority of the

[1]This is implemented using the sdid command in Stata developed by Clarke et al. (2023). Standard errors are generated with the bootstrap method with 50 repetitions and they are clustered at the municipality level.

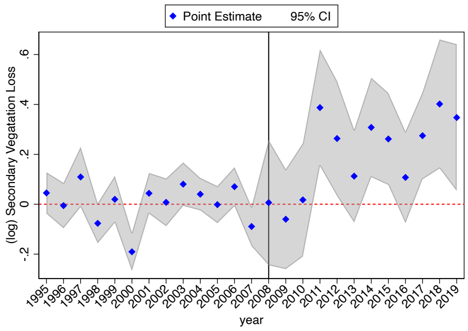
Figure 3.1: Dynamic effects of Priority List on Forest Cover Changes



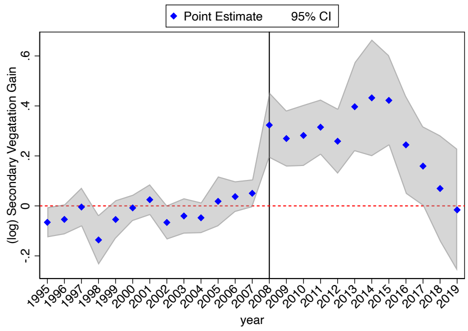
(a) (log) All loss of natural vegetation



(b) (log) Loss of primary vegetation



(c) (log) Loss of secondary vegetation



(d) (log) Gain of secondary vegetation

Note: This figure shows the municipality-year-level event study of the synthetic differences in differences regression of deforestation and reforestation outcomes (specifically: total loss of natural vegetation, loss of primary natural vegetation, loss of secondary vegetation, and gain of secondary vegetation) in logarithms on the onset of the Priority List in 2008. Since treatment is staggered, I restrict attention to the municipalities that join the Priority List in 2008 and those that never join as a pure control group. I only include municipalities in the Brazilian Legal Amazon, as they are more likely to resemble Priority List municipalities. Standard errors are calculated via bootstrap with 50 repetitions. In the regression results in table 3.1 I include municipalities that enter the list in other years and rely on Clarke et al. (2023), which does not use already-treated units as controls.

deforestation observed. Secondary vegetation loss, on the other hand, went up by 17%, which is partially explained by the fact that the gains in secondary vegetation also went up, by 23%, so there is more forest classified as secondary available to cut down. In terms of the timing of the effects, there seems to be a large instantaneous decrease of around 50% in the first couple of years, gradually disappearing over the years. This is consistent with the change of administration towards the governments of Temer in 2016 and Bolsonaro in 2019, who openly declared their intention to reverse Lula and Rousseff's environmental policies and reduce the budget allocated to environmental agencies. The positive effects on forest regrowth also seem to disappear completely by 2019 (Burgess et al., 2019).

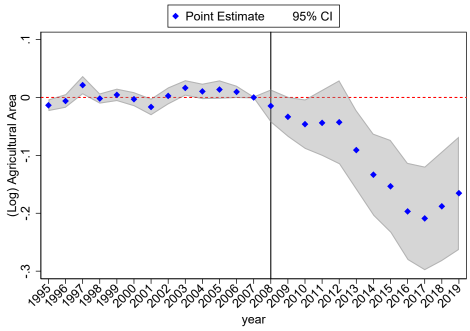
Effects on other outcomes

I look at other agricultural outcomes in order to understand the relationship of the Priority List to the local agricultural economy more broadly.

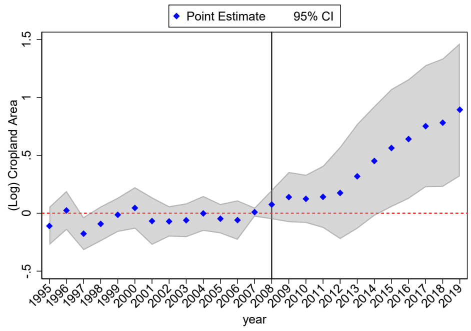
This is informative of the mechanisms via which deforestation is reduced. In particular, I find that deforestation is reduced by switching towards less land intensive activities. There is a persistent decrease in the area in pastures while there is an persistent increase in the total area in crops. There is not only an increase in the share of area in crops in the municipality, which I might expect if deforestation was simply halted and no other changes took place, since deforestation is mostly a conversion of forests to pastures. The increase is in the total area in crops. That is, although there is less agricultural area, there are more crops. Surprisingly, but consistently with the land use facts, municipalities that get added to the Priority List see an increase in their agricultural value added as measured in the Regional Accounting system. This seems to be primarily due to three phenomena: (1) a shift away from pasture towards crops (especially soy and maize) as seen in Figure 3.2, (2) an increase in the yields of some crops, and (3) an increase in the number of cattle heads per area in pasture.

Beyond informing how deforestation is reduced, these results help make sense of the effects of restriction in the supply of agricultural land on a wider set of economic decisions, which is informative of their general equilibrium effects. As will be shown in chapter 6, conservation policies lead to increases in the area in crops and to intensification in all agricultural activities. In this model, the only margin for such intensification is more labour per hectare. With higher

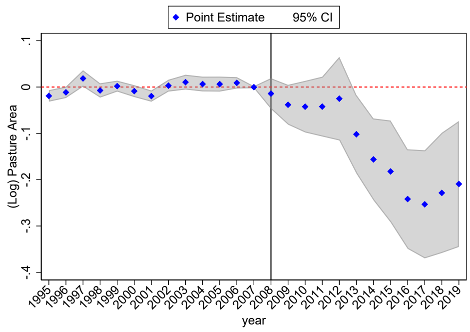
Figure 3.2: Dynamic effects of Priority List on Agricultural Land Use Changes



(a) (log) Agricultural Area



(b) (log) Cropland Area



(c) (log) Pasture Area

Note: This figure shows the municipality-year-level event study of the synthetic differences in differences regression of agricultural area stocks (specifically: total agricultural area, cropland area, and pasture area) in logarithms on the onset of the Priority List in 2008. Since treatment is staggered, I restrict attention to the municipalities that join the Priority List in 2008 and those that never join as a pure control group. I only include municipalities in the Brazilian Legal Amazon, as they are more likely to resemble Priority List municipalities. Standard errors, in parenthesis, are calculated via bootstrap with 50 repetitions. * p<0.10, ** p<0.05, *** p<0.01

frequency data on agricultural inputs, future research could assess empirically the margins along which the intensification of production occurs.

3.1.3 Robustness

As a robustness check, I do this analysis taking 10 km-wide hexagons as the unit of analysis. Municipalities that get selected to join the Priority List are different in a few fundamental ways: they are larger, closer to the forest frontier, and had different forest cover and deforestation trends than the rest of the Legal Amazon. The advantage of using the grid cell as opposed to the municipality as unit of

Table 3.2: Effects of Priority List on Crop-specific Agricultural Outcomes

	(log) Revenue	(log) Area	(log) Price	(log) Yields	N observations
Ag Value Added	0.190***				
	(0.036)				13,685
Coffee	0.252	0.274	0.088**	-0.034	
	(0.255)	(0.291)	(0.037)	(0.129)	1,780
Cassava	-0.137	-0.116	0.022	-0.027	
	(0.098)	(0.097)	(0.036)	(0.026)	14,140
Beans	-0.495***	-0.383**	0.076**	-0.130***	
	(0.178)	(0.157)	(0.034)	(0.050)	8,660
Maize	0.515***	0.390**	-0.056***	0.150***	
	(0.181)	(0.166)	(0.020)	(0.045)	13,280
Sugar	-0.307	-0.392	0.133	-0.053	
	(0.381)	(0.358)	(0.118)	(0.058)	3,460
Soy	0.363**	0.396**	-0.003	0.004	
	(0.175)	(0.165)	(0.020)	(0.034)	1,920
Orange	-0.084	-0.110	0.110	-0.068	
	(0.283)	(0.177)	(0.074)	(0.065)	3,780
Banana	0.312*	0.154	0.185***	-0.033	
	(0.161)	(0.142)	(0.049)	(0.044)	10,200
Cocoa	0.450	0.567***	-0.045	-0.082	
	(0.310)	(0.192)	(0.035)	(0.067)	1,760
Cotton	-0.390	-0.282	-0.024	0.026	
	(0.266)	(0.255)	(0.044)	(0.060)	480
Rice	-0.100	-0.193	0.026	0.092**	
	(0.173)	(0.139)	(0.029)	(0.038)	10,300

Note: This table shows the results of the municipality-year-level the synthetic differences in differences regressions of agricultural value added and crop-specific outcomes (specifically: revenue from crop, area harvested, average farm-gate price, and average yields) in logarithms on the onset of the Priority List. The first row is a regression as (3.1) but with outcome the log of agricultural value added as obtained from the system of regional accounts. The second row onwards are regressions as (3.1) with outcomes calculated from the yearly municipal agricultural survey PAM. Standard errors are calculated via bootstrap with 50 repetitions. I only include municipalities in the Brazilian Legal Amazon, as they are more likely to resemble Priority List municipalities. Years in which the crop is not grown (according to the data) are dropped. Subsequently, municipalities that have not grown a crop (in a year that was not dropped in the previous step). Standard errors in parenthesis. * p<0.10, ** p<0.05, *** p<0.01.

analysis is that it allows us to better match the areas inside Priority Listed municipalities to areas outside. Having a much larger sample size of comparable units with the same area, I can more accurately reweight and match pre-policy trends.

Another advantage of considering grid-level outcomes is that I can control for two variables that might be confounding the estimation. The first is the distance to the forest edge of a hexagon. As shown in chapter 2, the vast majority of deforestation happens near the forest edge. It could be that as the forest edge moves further North and West, places are simultaneously more likely to be deforested and to become part of the Priority List. Second, I control for whether a hexagon is part of a protected area, since the timing of protected area establishment and Priority Listing might be correlated. The left panel of 3.3 shows the dynamic results of a synthetic diff-in-diff analysis on the total natural forest loss without controls.[2] The right panel adds controls for: (i) the log of the absolute value of the distance to the forest edge, (ii) a dummy for being in an indigenous territory, (iii) a dummy for being inside a protected area. Reassuringly, the grid-level results are very similar in significance and magnitude to the municipality-level analysis. The only difference is that the grid-level analysis shows better matched pre-trends and does not show a reversal in the deforestation reduction.

3.2 Protected Areas

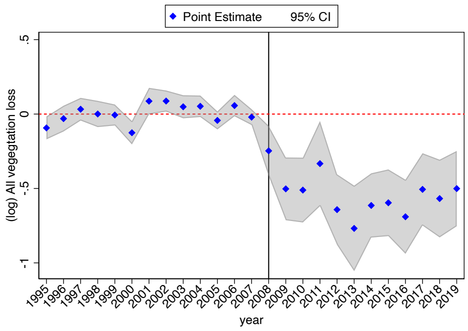
3.2.1 Econometric specification

Unlike the Priority List policy, the establishment of protected areas (both Conservation Units and Indigenous Territories) does not necessarily coincide with municipal borders. Hence, I split the Brazilian territory in hexagons with 10 km width, and classify each of them as being or not part of a protected area if at least 50% of its surface falls within the demarcated boundaries.

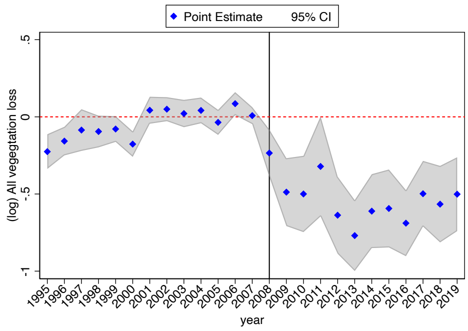
I use the hexagon-level data to estimate a Regression Discontinuity Design around the border of the conservation unit. The baseline specification is:

[2] To speed up the analysis I do it on a 10% random sample of all the hexagons in the Legal Amazon.

Figure 3.3: Dynamic effects of Priority List on Forest Cover Changes (grid-level)



(a) (log) All loss of natural vegetation, no controls



(b) (log) All loss of natural vegetation, with controls

Note: This figure shows the 10 km hexagon grid cell level event study of the synthetic differences in differences regression of agricultural area stocks (specifically: total agricultural area, cropland area, and pasture area) in logarithms on the onset of the Priority List in 2008. Since treatment is staggered, I restrict attention to the municipalities that join the Priority List in 2008 and those that never join as a pure control group. I only include hexagons in municipalities in the Brazilian Legal Amazon, as they are more likely to resemble Priority List municipalities. Standard errors are calculated via bootstrap with 50 repetitions.

$$\text{(log) ForestArea}_{ht} = \delta_t + \gamma_{m(h)} + f(\text{Distance}_{ht}) + \beta D_{ht} + \epsilon_{ht} \tag{3.3}$$

Where (log) ForestArea$_{ht}$ is the logarithm of the total forest area observed in hexagon h, at year t, δ_t are year fixed effects, $\gamma_{m(h)}$ are municipality fixed effects, $f()$ is a continuous function, Distance$_{rt}$ is the running variable measuring the distance in km to the border of the nearest conservation unit where negative (positive) values mean that the hexagon falls inside (outside) the protected area, and D_{ht} is a dummy variable indicating whether hexagon h falls within a protected area at time t. $D_{ht} = 1\{\text{Distance}_{ht} \leq 0\}$. The coefficient of interest in β, identifying the effect on deforestation of being inside the conservation area. I use forest area instead of deforestation as an outcome because protected areas are typically established in regions with very low levels of deforestation, so the numeric interpretation of the effect on forested area is clearer.

Validity

The validity of the RD design relies on the assumption of continuity of the outcome variable with respect to the running variable in the absence of treatment. In other words, in the absence of the protected areas, deforestation does not see a discontinuous spatial jump at $D_{ht} = 0$. A potential concern that arises with this specification is that, given the size of Brazil, the data was coarsened to the 10 km hexagon level and hence the distance to the border (running variable) is discrete. Therefore I cannot consider an arbitrarily small neighbourhood of the cut-off. To test the continuity assumption in this setting, I estimate the same Regression Discontinuity design considering only hexagons located in future protected areas. For the results of these exercises, see section 3.2.3.

3.2.2 Results

Table 3.3 show the estimated β from Equation 3.3 where f is a quadratic spline, adding different sets of fixed effects: (1) has only year fixed effects, (2) has only municipality fixed effects, and (3) has municipality-year fixed effects. The results suggest that there is an increase of between 15% and 27% in the forest coverage of hexagons inside of a protected area as compared to those just outside. The preferred specification is (3), as it controls for the fact that the process of establishing a protected area is a political one and changes in the municipal government or administrative bureaucracy may correlate with changes in protected area status.

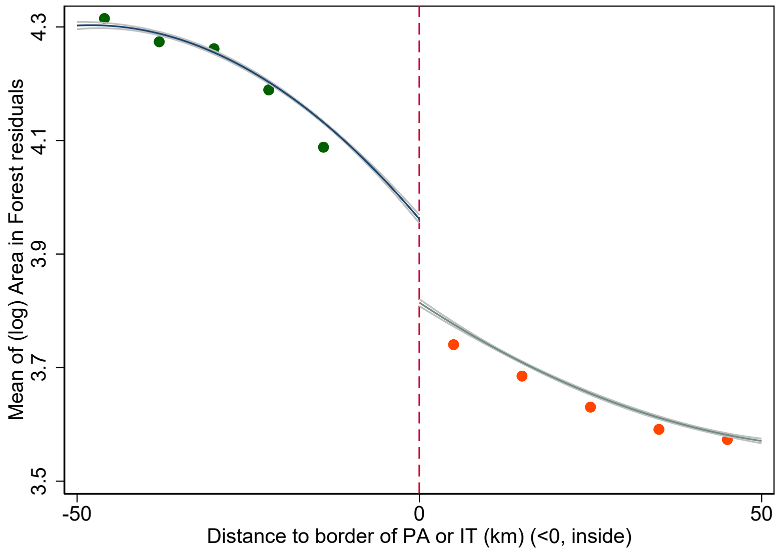
Figure 3.4 illustrates the regression discontinuity results graphically by showing average (log) forest area in bins of multiples of 10 km of distance from the border of a protected area. Panel (a) on the left shows the simple regression discontinuity specification with a quadratic fit and no fixed effects, whereas panel (b) on the right has as dependent variable the (log) forest area residualised by municipality-year fixed effects. Both specification show a clear jump in forested area of around 0.2 log-points (or a 22% increase) at the demarcated border. In panel (a) I can see that hexagons deeper inside a protected area typically have higher forest cover than those closer to the border, however, this relationship seems to change discontinuously around the border.

Table 3.3: Regression discontinuity estimates for the effect of Protected Areas on forested area

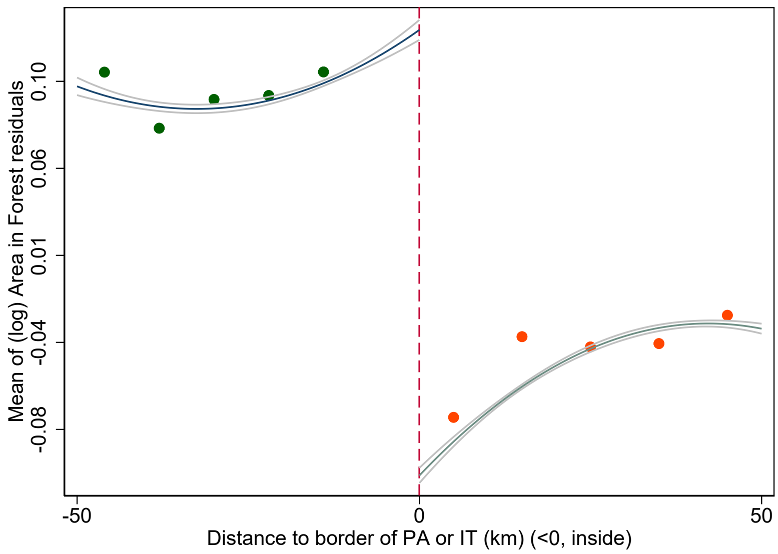
	(Outcome: log-forested area)		
	(1)	(2)	(3)
Estimated Gap (PA: CU or IT)	0.1430***	0.2396***	0.2378***
	[0.0425]	[0.0343]	[0.0356]
Quad spline	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Mun FE	No	Yes	Yes
Year X mun. FE	No	No	Yes
R2	0.0691	0.6081	0.6259
Observations	2.16e+06	2.16e+06	2.16e+06

Note: This figure shows the results of the hexagon-year-level regression discontinuity of (log) forest area on the boundaries of Protected Areas (PA). PAs include both Conservation Units (CU) and Indigenous Territories (IT). The regression result presented displays coefficient β in (3.3) under three different specifications. The regression includes years 1985-2022 and all hexagons within 50km of a PA boundary. In all of them $f(\cdot)$ is a quadratic spline, split around the threshold (0), of the distance to the PA's boundary, which is negative inside and positive outside. Column (1) controls for year fixed effects only to control for common trends. Column (2) includes year and municipality fixed effects to control for fixed characteristics at the municipality level. Column (3) includes municipality-year fixed effects to control for municipality-specific time trends. Standard errors in parenthesis. *p<0.1, **p<0.05, *** p<0.01.

Figure 3.4: Discontinuity in forested area around borders of protected areas.



(a) No controls



(b) Municipality-Year FEs

Note: This figure illustrates the discontinuity in forested area around the borders of protected areas. Each dot represents the mean total log-forested area within bins of 10 km around the border of all protected areas, from 50km inside of 50km outside. The x-axis shows distance to the border, with negative (positive) values indicating the area inside (outside) the protected area. On top of the dots are the quadratic fits of both distance trends. The top panel illustrates the raw averages. The bottom panel illustrates averages of the residuals of a regression of log-forested area on municipality-year fixed effects, which controls for differential trends at the municipality level. Since there are observations for multiples years (1985-2022) the snapshots with varying different protected area-borders are pooled together so that a hexagon that in 1985 is 50km away from the border could be 10 km inside a border in 2020. For each hexagon this distance can only decrease over time.

3.2.3 Robustness

In order to test the validity of the regression discontinuity design, I investigate the plausibility of the smoothness assumption by comparing what happens at the border a future protected area before it is established. To do so I run the following placebo test.

$$\text{ForestArea}_{ht} = \delta_t + \gamma_{m(h)} + f(\text{MinDistance}_h) + \tilde{\beta}\tilde{D}_h + \epsilon_{ht}, \tag{3.4}$$

where $\text{MinDistance}_h \equiv \min_t \text{Distance}_{ht}$ is the time-invariant minimum distance that a hexagon h ever has from a Protected Area, which is also the last one because they never get dismantled and hence Distance_{ht} can only decrease, and $\tilde{D}_h = 1\{\text{MinDistance}_h \leq 0\}$. The sample is restricted to observations of hexagons h at times t for which that minimum distance to a protected area, MinDistance_h, is achieved in a certain number of years. That is, for each hexagon, I can define an event time $T(h)$ from which the distance does not further decrease and for the placebo test I consider only observations (h, t) such that $t < T(h) + k$ where k is the number of years before establishment that I want to consider. Table 3.4 below shows the corresponding estimates for the RD parameter $\tilde{\beta}$ for $k = 10$ for columns (1) and (2) and for $k = 5$ for columns (3) and (4). Columns (1) and (3) have only year fixed effects whereas (3) and (4) have municipality-year fixed effects. While no effects are found more than 10 years before, there is a discontinuity 5 years before. A discontinuity around the border of a future protected area before its implementation can be due to a number of factors, for example: (i) lengthy legal disputes over the demarcation of new protected areas, during which deforestation decreased before the final settlement and establishment [3] or (ii) historical occupation of the area by indigenous people whose livelihood depends on the preservation of the forest before legal demarcation of the Protected Area.

Figure 3.5 illustrates this graphically. The four plots in panel (a) the same RD graph as in figure 3.4, panel (a), but for different event times relative to the establishment of the nearest protected area. Clockwise from the top-left: (i) more than 10 years before establishment, (ii) up to 10 years before, (iii) up to 10 years after, and (iv) more than 10 years after. Panels (b) show the exact same comparison, but now including year and municipality fixed effects. Panel (b) shows a significant discontinuity for all event time groups but the gap seems

[3] This is particularly likely in the case of Indigenous Territories. Their establishment follows a judicial process that includes several stages: study, delimitation, declaration, homologation, and regularisation, that can take decades

Table 3.4: Placebo Regression Discontinuity

	10+ years before		5+ years before	
	(1)	(2)	(3)	(4)
Estimated Placebo Gap	0.0387	0.0678	0.2509***	0.0777**
	[0.0620]	[0.0450]	[0.0583]	[0.0336]
Quad. spline	Quad. spline	Quad. spline	Quad. spline	Quad. spline
Year FE	Yes	Yes	Yes	Yes
Year X mun. FE	No	Yes	No	Yes
Years before	10+	10+	5+	5+
R2	0.0801	0.6756	0.0513	0.6742
Observations	3.22e+05	3.22e+05	2.82e+05	2.82e+05

Note: This figure shows the results of the hexagon-year-level regression discontinuity of (log) forest area on the boundaries of Protected Areas (PA). PAs include both Conservation Units (CU) and Indigenous Territories (IT). The regression result presented displays coefficient β in (3.4) under four different specifications. The regression includes years 1985-2022 and all hexagons within 50km of a PA boundary in 2022. In all of them $f(\cdot)$ is a quadratic spline split around the threshold (0) of the distance to the PA's boundary, which is negative inside and positive outside. Columns (1) and (2) include hexagons only 10+ years before they reach their minimum distance to a PA, that is, 10+ years before the closest PA to them (or in which they are contained) is established. Columns (3) and (4) include hexagons 5+ years before that occurs. Columns (1) and (3) control only for year fixed effects. Columns (2) and (4) control for municipality-year fixed effects. Standard errors in parenthesis. *p<0.1, **p<0.05, *** p<0.01.

very small more than 10 years before and it seems to be clearly widening over time.

3.3 General Equilibrium Effects and SUTVA

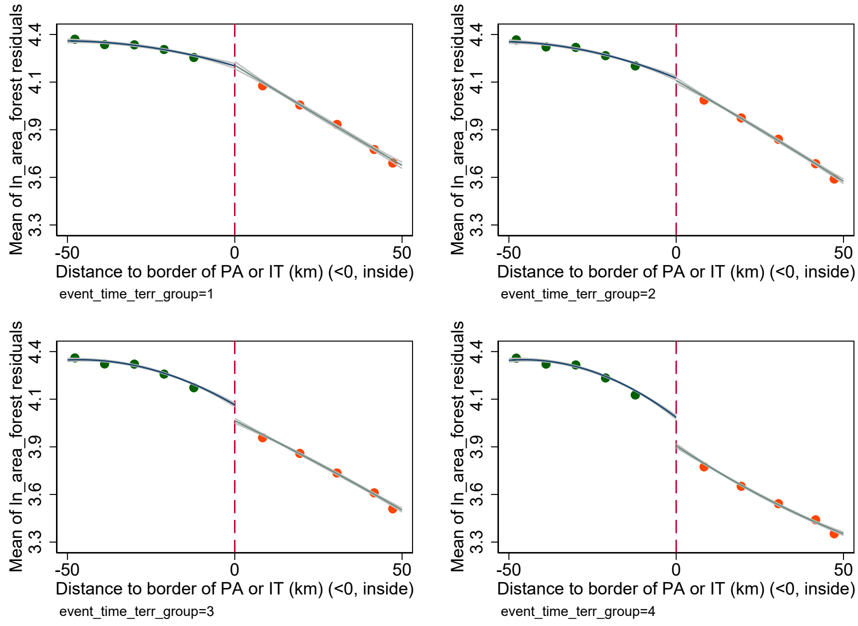
As previously discussed, an assumption required for the validity of both reduced-form methods discussed above is Stable Unit of Treatment Variable Assumption (SUTVA). Consider the treatment D_i a binary variable equal to 1 if a place is subject to a conservation policy, 0 otherwise. Assume that the potential outcome of i depends on two factors, its own conservation status, and the price of land, which depends on the conservation statuses in all regions $Y_i(D_i, P(\vec{D}))$. Even if treatment was as good as randomly assigned, but SUTVA did not hold, I would be estimating a combination of the desired treatment effect and leakage as shown below

$$\begin{aligned}
\widehat{\beta} \rightarrow & \mathbb{E}\big[Y_i(1, P(\vec{D}_{policy}) - Y_i(0, P(\vec{D}_{policy})\big] \\
= & \underbrace{\mathbb{E}\big[Y_i(1, P(\vec{D}_{nopolicy}) - Y_i(0, P(\vec{D}_{nopolicy})\big]}_{\text{Pure effect of policy on treated}} \\
+ & \underbrace{\mathbb{E}\big[Y_i(1, P(\vec{D}_{policy}) - Y_i(1, P(\vec{D}_{nopolicy})\big]}_{\text{Leakage on treated}} \\
- & \underbrace{\mathbb{E}\big[Y_i(0, P(\vec{D}_{policy}) - Y_i(0, P(\vec{D}_{nopolicy})\big]}_{\text{Leakage on untreated}} .
\end{aligned}$$

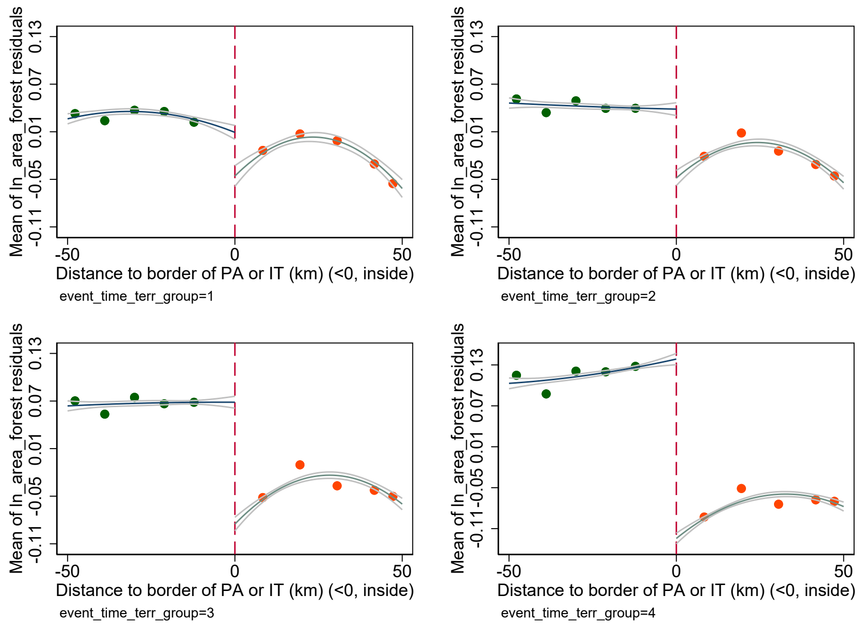
A reduced-form estimate, therefore, would include both the desired ATET and potential leakage onto untreated areas. Therefore, it requires an assumption about the nature of the second right-hand side term in the equation above. Typically, the implicit assumption is that there is no leakage, i.e. $Y_i(d, P_1) = Y_i(d, P_2)$, which results in an overestimation of the true treatment effect.

Even if leakage is explicitly considered, a reduced-form estimate requires assuming some arbitrary *ex-ante* structure determining which regions are prone to leakage, and which regions can be considered a "pure control". One common approach is to assume that neighbouring areas are prone to leakage, whereas places further away from treatment are not affected in any way. Whereas this might be plausible from a purely spatial perspective, it has two problems: (i) it still requires some arbitrary distance cut-off that separates areas that are susceptible

Figure 3.5: Discontinuities in forested area around borders of Protected Areas, by: period relative to the introduction of conservation policy



(a) No controls



(b) Municipality- Year FEs

Note: This figure illustrates the discontinuity in forested area around the borders of protected areas, before and after they are protected. More specifically: the top-left panel considers 10+ years before the protection of the nearest area, the top-right 0-10 years before, the bottom-left 0-10 years after, and the bottom-right 10+ years after. Each dot represents the mean log-forested area within bins of 10 km around the border of all protected areas, from 50km inside of 50km outside. The x-axis shows distance to the border, with negative (positive) values indicating the area inside (outside) the protected area. Here the hexagons are classified according to the minimum (and final, since it can only decrease) distance they ever have from a Protected Area across all periods in the sample, but their forest area is only considered for years in which (i) they area 10+ years from reaching that minimum distance, (ii) they are 0-10 years away from reaching that minimum distance, (iii) they reached it 0-10 years ago, (iv) they reached it 10+ years ago.

to leakage from areas that are not, and (ii) it ignores other spillover mechanisms, such as price changes, transportation connections and local amenities that might also be important determinants of the location of spillovers.

Chapter 4

Deforestation in Spatial Equilibrium

I build a spatial multi-sector general equilibrium model that explicitly captures the key features of the spatial distribution of agriculture, the economic forces at play in the market for land, and frictions to movement of goods and labour via internal trade and migration.

Mechanisms for leakage

In this framework, a local anti-deforestation policy is interpreted as an exogenous shock to the local supply of agricultural land. To quantify spatial leakage, I consider two key mechanisms through which a negative supply shock generates increased deforestation elsewhere. The first mechanism operates through the market for agricultural goods. A negative shock to the local supply of agricultural land decreases the supply of agricultural goods. The extent to which this avoided deforestation leaks elsewhere depends on the extent to which goods are substitutable across space and can be produced by clearing forest elsewhere. The second mechanism operates through the market for labour. A local reduction in deforestation coming from a shock decreases the demand for agricultural labour, changing workers' migration incentives. Inflows of workers to regions without anti-deforestation policies will increase the demand for agricultural land, raising incentives to deforest[1].

[1] A useful benchmark is a flat economy where all land is made equal, workers and goods can move freely, and demand for agricultural land is perfectly inelastic. In this case, leakage

To capture these channels, the model considers deforestation as an intermediate economic sector that supplies land as a factor of production for agriculture. Regions differ in the sectoral productivities for agricultural and non-agricultural production, as well as in their productivity in producing land via deforestation[2].

Model features

I conceptualise deforestation as an investment in agricultural land that accumulates over time. Brazil is modelled as a closed economy with domestic trade and migration. The model considers Brazil's 558 microrregions[3] indexed by r, which differ on their sectoral productivities, land endowments, and amenities. The economy is composed of $K+1$ sectors: K agricultural commodities that use land and labour as inputs and have different labour shares, and one non-agricultural sector that uses only labour. Additionally, I consider deforestation as a sector that uses a composite investment good in order to produce agricultural land for the K agricultural sectors. There is trade between municipalities subject to iceberg costs. Consumer preferences are non-homothetic, represented by the Price-Independent Generalized Linear preference formulation (Boppart, 2014). Final goods in each sector are a composite of regional varieties aggregated with constant elasticity of substitution σ and a final agricultural good is a composite of the various agricultural commodities aggregated with constant elasticity of substitution θ.

The model features a sequence of static spatial equilibria linked by the laws of motion of land and labour. The law of motion of land is determined by deforestation and the law of motion of labour is determined by migration and population growth.

would be 100%: banning deforestation entirely in one region would have no global effect, as it would be perfectly leak to other regions - the demand for agricultural goods would be the sole determinant of the amount of agricultural land, and hence of the level of deforestation. The model departs from all of those assumptions in ways that are consistent with the data.

[2]Deforestation productivity can be thought of as region-specific factors that govern how suitable a particular region is for forest cutting. It can include natural factors (forest density, type of vegetation, altitude, weather patterns, geographical features), infrastructure and accessibility, the political environment and the level of enforcement of anti-deforestation laws.)

[3]This is typically considered the aggregation level that is most closely associated with a local market, see for example Dix-Carneiro and Kovak (2017)

4.1 Land use dynamics: the endogenous accumulation of agricultural land

Initially, each region r is endowed with L_{r0} workers, T^A_{r0} units of agricultural land area, and T^N_{r0} units of terrestrial natural ecosystems which can be converted into agricultural land.

For most of this section, I omit the time subscript and treat the equilibrium as static. However, since agricultural land accumulates, it is a time-varying quantity. Assuming a fixed forest regrowth rate ρ, at letting T^D_{rt} be the level of deforestation, agricultural land evolves according to

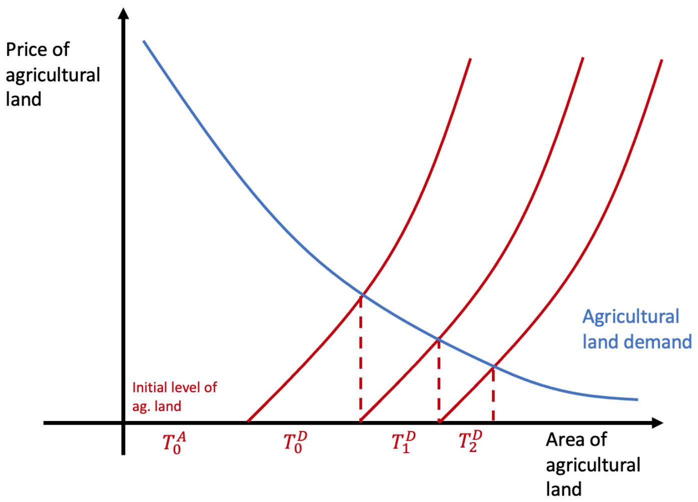
$$T^A_{rt+1} = T^A_{rt}(1-\rho) + T^D_{rt}.$$

The following simple agricultural land market graph helps visualise how land evolves over time. In this simple graph, land demand is fixed, there is only one region, there is no forest regrowth, and the deforestation supply curve does not change over time. In particular, the scarcity of forest does not change its value relative to agricultural land. The graph illustrates in what sense the supply for deforestation is conceptualised as yearly. This means that the decreasing returns to scale of the deforestation production function, discussed below, reflect how deforesting more land within a single year is increasing costly, which could reflect resource constraints of the deforesting agents. A decade-long deforestation supply curve would be flatter. With the assumptions in the graph (fixed demand for land, fixed yearly deforestation supply, no forest growth) eventually the whole forest would be converted to agriculture. However, in practice, I will relax all of these assumptions: the demand for and will change in response to population growth, migration, and productivity shocks, and past land accumulation; there will be a rate of forest regrowth; and the costs of deforestation will go up as there is less forest area left in a region.

4.2 Population dynamics: Migration

As in Eckert and Peters (2022), individuals born in a location of origin o can choose to live in destination d according to the migration costs μ_{od}, the utility at the destination $V(e_d, p_d)$, the amenities at destination d, B_d, and an idiosyncratic

Figure 4.1: The market for agricultural land over time



Note: This figure illustrates how agricultural land would accumulate over time. In this graph, land demand is fixed, there is only one region, there is no forest regrowth, and the deforestation supply curve does not change over time.

shock $\nu_d(i)$, drawn from a Frechet distribution with parameter ϵ. The origin-destination specific migration utility $\mathcal{U}_{od}$ is given by:

$$\mathcal{U}_{od}(i) = V(e_d, p_d)B_d\mu_{od}\nu_d(i).$$

Therefore, from the Frechet nature of the shock, the share of people who move from o to d is given by the following expression[4]

$$\rho_{od} = \frac{(V(e_d, p_d)\mu_{od}B_d)^\epsilon}{\sum_{r=1}^{R}(V(e_r, p_r)\mu_{or}B_r)^\epsilon}.$$

Hence the law of motion of population will be given by the following equations

$$N_{dt} = \sum_{o=1}^{R} \rho_{rd}N_{ot}, \quad N_{ot} = g_{ot-1}^N N_{ot-1}$$

where g_{ot-1}^N is the growth rate of population in origin region o at time t.

[4]This includes ρ_{oo}, i.e. the share of people from region o who choose to stay in region o.

4.3 The market for deforestation

Deforestation is modelled as a costly investment in the production of agricultural land, which is a factor of production for the agricultural sector. I can think of deforesters as atomistic agents operating in a perfectly competitive market where they access forested land, pay the fixed cost of clearing it, and sell it at the price of agricultural land. I model the aggregate deforestation production function so that it has decreasing returns to scale. This motivated by the fact that, within each time period, the forest that is closer to the edge is cheaper to access and clear. This means that each additional dollar spent deforesting is less productive. The aggregate deforestation production function is such that if an amount I_r^D of the final good, bought at price p_r, is invested in deforestation, it delivers T_r^D units of agricultural land according to

$$T_r^D = Z_r^D (I_r^D)^{\delta}, \tag{4.1}$$

where Z_r^D is the region-specific "deforestation productivity", and $\delta \in (0,1)$ governs the returns to scale of the production function. The returns of each square kilometre of deforested land equal the value of agricultural land in a given region, q_r. I can interpret Z_r^D and δ as the productivity parameters that dictate the aggregate (convex) costs of deforesting, $C_r^D(T_r^D, p_r; Z_r^D, \delta) = p_r (Z_r^D)^{-\frac{1}{\delta}} (T_r^D)^{\frac{1}{\delta}}$.

Our model of the market for deforestation has two defining features. First, land in natural ecosystems is open access, which means that the future value of the forest is fully discounted[5]. Second, there is free entry of deforesters. Accordingly, they enter the deforestation market as long as there are non-negative marginal profits, so that the equilibrium level of deforestation is when marginal costs, $MC_r^D = \partial C_r^D(T_r^D, p_r; Z_r^D, \delta)/\partial T_r^D$, equal marginal revenues, $MR_r^D = q_r$,

$$T_r^{D*} = \underbrace{(Z_r^D)^{\frac{1}{1-\delta}}}_{\text{Local Factors}} \underbrace{\left(\frac{\delta q_r}{p_r}\right)^{\frac{\delta}{1-\delta}}}_{\text{Equilibrium Effects}} . \tag{4.2}$$

[5]Around 50% of deforestation in the Amazon over the past few years has happened in untitled public lands. From the remaining 50%, about half has happened in rural settlements where land was and half in private properties. Although the amount of deforestation in private properties is not negligible, about 25% of the total, rights over forested land are insecure even within private property. This is due to it historically being often regarded as "unproductive" and thus subject to the ownership claims of squatters. See (Alston et al., 1999) for an in depth description of property rights in the Brazilian Amazon frontier.

Supply side of deforestation

The curve $T_r^{D*}(q_r/p_r)$ is the region-specific deforestation supply curve: how deforestation responds to the local relative price of land. It depends on two parameters δ, the returns to scale of deforestation in one year, and Z_r^D, the location-specific deforestation productivity.

The returns to scale of deforestation δ is taken to be the same for all of Brazil, and it governs the supply elasticity of deforestation to real land prices, which equals $\frac{\delta}{1-\delta}$. That is, for a 1% increase in the value of agricultural land relative to that region's price index, deforestation goes up by $\frac{\delta}{1-\delta}$%. See section 5.1.1 in the appendix for the estimation of δ.

Z_r^D is the source of regional heterogeneity in deforestation costs. It reflects differences in characteristics such as, (i) local environmental conditions that influence how difficult it is to deforest, such as rainfall and temperature, (ii) the level of enforcement of anti-deforestation policies, (iii) revenues obtained from the act of deforestation itself, for example through the sale of wood, and (iv) the option value of keeping land as forest. While most of these things are difficult to observe and quantify, I correlate them to the structurally estimated Z_r^D's and find that the current area of unprotected forest in a region, T_r^F is a strong predictor of Z_r^D. A log-log model does a remarkably good job at describing their empirical relationship in cross-sectional data. Therefore, I let

$$Z_r^D = \overline{Z_r^D}(T_r^F)^{\psi}.$$

It is estimated so that it perfectly explains observed differences in levels of deforestation in regions that cannot be explained by agricultural rents and market access, which are reflected by q_r and p_r respectively. In turn, prices come from the inversion of a spatial equilibrium model in each time period, as described in 5.2. Intuitively, I estimate the demand for agricultural land from farmers, which in turn depends on the demand for agricultural products from consumers, and then, using the observed data on deforestation, I calculate the productivity of deforestation as a residual that rationalises their spatial distribution. To be more concrete, for a given level of observed deforestation, a region with lower market access and lower agricultural productivity will have a higher estimated Z_r^D.

4.4 Technology: local demand for workers and land

In order to estimate the demand for land in each region, I first need to impose some structure on the firms operating in each region. This will determine how they demand different factors of production, land and workers, given productivities and prices. Productivities will be treated as exogenous fundamentals to be backed out from the model and prices will depend on the full (static) equilibrium which takes into account consumer preferences and trade costs.

There are four broad sectors: three agricultural sectors with varying land intensities, and non-agriculture which uses no land for production. Regions produce a differentiated variety of each of these four goods as in Armington (1969) that consumers combine with Constant Elasticity of Substitution as in Anderson (1979). The market in each of the four sectors is composed of perfectly competitive firms with constant returns to scale. The local non-agricultural goods are a product of only labour with regional productivity Z_r^{NA}, so that $Y_r^{NA} = Z_r^{NA} L_r^{NA}$. The agricultural goods, indexed by k, are a Cobb-Douglas function of land and labour with constant returns to scale and regional productivities Z_r^{Ak}, and a share of land equal to α_k, so that

$$Y_r^{Ak} = Z_r^{Ak} (L_r^{Ak})^{1-\alpha_k} (T_r^{Ak})^{\alpha_k}.$$

Goods of sector s are produced to be sold at origin prices p^{so}.

In a competitive equilibrium, the rental rate of agricultural land v_r equals the marginal product of agricultural land and the wages in each sector equal the marginal product of labour in each sector. Assuming simple adaptive expectations (i.e. agents assume the future rental rate of land equals today's) and a discount rate of β, land should be priced at its expected present value,

$$q_r = \frac{1}{1-\beta(1-\rho)} v_r.$$

Our data allows us to get agricultural revenues (see section 5.2) and use them to get v_r as

$$v_r = \sum_k \alpha_k \frac{p^{Akr} Y_r^{Ak}}{T_r^{Ak}}.$$

4.4.1 Wage gaps and occupational choice

In order to allow for a gap between agricultural and non-agricultural wages, as I consistently find in the data, I rely on a model like the one introduced by Lagakos and Waugh (2013) and applied to the context of Brazil in Alvarez (2020). In this model, individuals draw idiosyncratic productivities for agriculture and non agriculture from a joint distribution $F(z_i^A, z_i^{NA})$ and given their observed productivities for each sector[6]. Workers choose sector in order to maximise their wage income, so that they work in non-agriculture if and only if $z_{ri}^{NA} w_r^{NA} \geq z_{ir}^A w_r^A$. Firms set wages per efficiency unit (w_r^A, w_r^{NA}) that equal the marginal product of a worker with unit productivity in that sector. By L_r^s I refer to the total labour efficiency units in sector $s \in \{A, NA\}$ in region r, which is equal to the number of workers multiplied by the expected productivity of those who choose to work in sector s. Since the idiosyncratic productivities in each sector are not independent draws, there will be income gaps, the average wage income ($\overline{y}_r^{LA}, \overline{y}_r^{LNA}$) will not be equalised across sectors. See appendix section B.1 for a more explicit mathematical description of the distribution of incomes that results for an arbitrary joint distribution of productivity shocks.

4.5 Preferences: demand for goods

Having characterised the factors that determine the demand for agricultural land given the prices of agricultural goods, let us now turn to the preferences that governs the demand side of agricultural goods markets.

Non-homothetic preferences between agricultural and non-agricultural goods

Consumers have PIGL preferences as in Boppart (2014) over agricultural and non agricultural goods. These preferences are represented by the following indirect utility function:

$$V(e, \vec{p}) = \frac{1}{\eta} \left(\frac{e}{(p^A)^{\phi} (p^{NA})^{1-\phi}} \right)^{\eta} - \nu \left(\frac{p^A}{p^{NA}} \right). \tag{4.3}$$

[6]The joint distribution is taken to be, as in Alvarez (2020), a Frank copula or two Frechet distributions with shape parameter χ^A and χ^{NA} and correlation ι

Where p^A and p^{NA} are prices of agricultural and non-agricultural sector, e is total expenditure, and η, ϕ, and ν are exogenous parameters. By Roy's identity, after relabelling the price of the composite consumption good $p \equiv (p^A)^\phi (p^{NA})^{1-\phi}$, the expenditure share in agricultural goods equals

$$\vartheta^A(e, \vec{p}) = \phi + \nu \underbrace{\left(\frac{p^A}{p^{NA}}\right)}_{\text{relative price effect}} \underbrace{\left(\frac{e}{p}\right)^{-\eta}}_{\text{Income effect}} .$$

CES aggregation between goods

The agricultural good is, in turn, a CES aggregate of agricultural goods $k \in \{\text{beef, temporary crops, permanent crops}\}$ with elasticity of substitution θ, so that the agricultural price index p^A is equal to $p^A \equiv \left(\sum_{k=1}^{K} (p^{Ak})^{1-\theta}\right)^{\frac{1}{1-\theta}}$ and the share of agricultural good k in the overall agricultural expenditure is given by $\left(\frac{p^{Ak}}{p^A}\right)^{1-\theta}$.

CES aggregation between origins

Each of the 3+1 final goods (the three agricultural goods and the non-agricultural good) is, in turn, a CES composite of differentiated regional varieties produced in region r, with a constant elasticity of substitution σ. Trade is taken to have symmetric iceberg trade costs $\tau_d^o \geq 1$ that do not vary by good so that a consumer from region d pays the origin price of a good from region o scaled by the bilateral iceberg trade cost. Thus, the share of goods g (which can be beef, temporary crops, perennials, or non-agricultural goods) consumed in region d, that is produced in region o, is given by[7]

$$\pi_d^{go} = \frac{(\tau_d^o)^{1-\sigma} (p^{gr})^{1-\sigma}}{\sum_{r=1}^{R} (\tau_d^r)^{1-\sigma} (p^{gr})^{1-\sigma}}. \tag{4.4}$$

[7] To differentiate between origin and destination prices I shall indicate the region of origin with a superscript and the region of destination with a subscript. Then the farm-gate price of temporary crops from o is $p^{A,temp,o}$, the price of temporary crops from o faced by consumers in d is $p_d^{A,temp,o} = p^{A,temp,o}\tau_d^o$, and the consumer price index of temporary crops in destination d equals $p_d^{A,temp} = \sum_{o=1}^{R} (\tau_d^o)^{1-\sigma} (p^{A,temp,o})^{1-\sigma}$.

I take σ from Eckert and Peters (2022) to be 9.[8] Having a value for σ, I use inter-state trade flow data from 1999 and 2017 to estimate a gravity equation derived from 4.4 along with the assumption that the iceberg trade costs depend on distance according to $\tau_d^o = (\text{distance}+1)^\kappa$.

4.6 Market clearing: closing the model

In order to close the model, I equalise the revenues from producers of good g from origin o to the sum of the expenditures of consumers in that good from across all regions d of Brazil. These expenditures equal the trade share of d to o times the total consumption of region d in good g,

$$p^{go} Y_o^g = \sum_{d=1}^{R} \pi_d^{go} X_d^g. \tag{4.5}$$

Aggregate consumption in a region

The overall expenditure of a region in each good can be calculated in three steps. First, I estimate the aggregate share of consumer expenditure in agriculture ϑ_r^A, which will depend on the PIGL preference parameters, the relative price of agricultural goods, the aggregate expenditure, and the distribution of incomes as dictated by wages and the joint distribution of productivity shocks[9]. Then, given the fact that the investment good is a Cobb-Douglas aggregate of the agricultural and non-agricultural goods with share ϕ, the total agricultural and non-agricultural expenditures equal

$$X_r^A = \vartheta_r^A E_r + \phi p_r I_r^D, \quad X_r^{NA} = (1-\vartheta_r^A) E_r + (1-\phi) p_r I_r^D \tag{4.6}$$

where E_r is the total consumer expenditure in a region and I_r^D the total investment in deforestation. Because of the CES preferences between agricultural

[8] Given that this context is closer to Domiguez-Iino (2021), I will also invert the model and simulate counterfactuals with values up to $\sigma = 15$, given his estimated elasticity of substitution between "counties" in Brazil and Argentina of around 13.

[9] A mathematical expression for the region-wide agricultural share can be found in the appendix section B.1. Given the non-homothetic nature of preferences for agricultural goods and the fact that there is within-region inequality due to the productivity shocks introduced to generate a wage gap, the formula for ϑ_r^A is not very pretty. Intuitively, in the face of higher inequality, the share of expenditure in agriculture will be lower because a higher share of earnings will accrue to those who spend proportionally less in agriculture.

goods, the expenditure in agricultural good k equals $X_r^{Ak} = X_r^A \left(\frac{p_r^{Ak}}{p_r^A}\right)^{1-\theta}$. Finally, the total consumer expenditure equals the total income, labour income plus land rental rate payments, net of deforestation investments,

$$E_r = \overline{y}_r^L L_r + v_r T_r^A - p_r I_r^D .$$

where $\overline{y}_r^L$ is the average wage income in region r across all worker, equal to $s_r^A \overline{y}_r^{LA} + (1 - s_r^A)\overline{y}_r^{LNA}$.

Equilibrium definition

Consider the economy described above. Let the initial agricultural land area in each region $\{T_{r0}^A\}_r$, the initial level of area under terrestrial natural ecosystems $\{T_{r0}^N\}_r$, and the distribution of workers across space $\{N_r\}_r$ be given as exogenous. A competitive equilibrium is a set of prices $\{p_r^s\}_{r,s}$, wages $\{w_r^s\}_r$, land rental rates $\{v_r\}_r$, occupational choices $\{N_r^s\}_{r,s}$, regional deforestation levels $\{T_r^D\}_r$, and regional expenditure shares $\{\vartheta_r^A\}_r$ such that:

1. consumers' choices maximize utility (equation (B.1));
2. the demand for regional varieties follows equations (4.4);
3. firms' factor demands maximize firms' profits;
4. marginal product of labour is equalised across sectors
5. local markets clear and there is trade balance (4.5).

Chapter 5

Model calibration and estimation

In order to use the theoretical model in the previous chapter to inform the magnitude of leakage in Brazil. I apply two data-driven strategies. First, in section 5.1, I calibrate the parameters of the model through empirical analyses that are consistent with the model structure but do not rely on its equilibrium solution. [1]. Second, in section 5.2, I invert the model to structurally estimate the main regional fundamentals - the total factor productivity of the various economic activities, including deforestation, and the amenities that rationalise observed migration rates. This estimation relies on the equilibrium equations of the model matched with selected moments observed in the data. In appendix section A.3, I show the correlation of the data with model outcomes for some non-targeted moments.

5.1 Calibration of model parameters

Table 5.1 summarises the calibration of the various structural parameters of the model. All parameters are either out-of-model estimates, or set exogenously from similar exercises in the literature. This section is composed of five parts. It begins by describing the calibration of the parameters governing the aggregate deforestation function: how the deforestation production function responds to scale

[1] The literature considers this type of exercise as an out-of-model estimation, even though it partially relies on the model, for example, to impute yearly regional land prices given that these data do not exist

Table 5.1: Summary of structural parameters calibration

Parameter		**Value**	**Source/Method**
	Deforestation Function		
δ	Deforestation returns to scale	0.5	Two-way fixed effect of deforestation on land prices
ψ	Natural area elasticity of deforestation TFP to	0.32	Regression derived from steady-state deforestation
ρ	Forest regrowth rate	0.01	Observed reforestation rates
	Final Goods Production Functions		
$\{\alpha_k\}_k$	Land share in ag. activities	(0.36;0.54;0.71)	From 2006 ag. census
χ_A, χ_{NA}, ι	Workers' sectoral productivity shocks distribution	(2;1.6;12.8)	Alvarez (2020)
	PIGL preference parameters		
ϕ	Ag. share in price index	0.1	Eckert and Peters (2022)
ν	PIGL Preference parameter	0.5	Eckert and Peters (2022)
η	Engel elasticity	0.506	Expenditure Survey Data (2017/18)
	Elasticities of substitution		
σ	Between origins	9	Eckert and Peters (2022)
θ	Between ag. goods	2	Costinot et al. (2016), Dominguez-Iino (2023)
	Trade and Migration parameters		
$\{\mu_d^o\}_{o,d}$	Bilateral migration utilities	residuals	Migration flows 2005-2010
ϵ	Dispersion of tastes	1.25	Migration flows and incomes from 2010 census
κ	Trade costs distance elasicity.	0.11	Gravity equation of trade flows
	Other Parameters		
β	Discount rate	0.9	NA

within a year (δ), how the total factor productivity of deforestation decreases as the remaining forest area in a region decreases (ψ), and the rate at which forests regenerate (ρ). Second, it describes the calibration of parameters governing the production of final goods: the parameters governing the joint distribution of the workers' productivity in agriculture and non-agriculture (χ_A, χ_{NA}, ι), and the land shares of the various agricultural activities (α_k). Third, it describes the calibration of the PIGL preference parameters for the consumption of final agricultural and non-agricultural goods. Of these, I take two from the literature, and calibrate the "Engel elasticity" (η) from consumer expenditure survey data. Fourth, it derives the gravity equation of migration flows and shows how it is used to calibrate the bilateral migration costs (μ_d^o) and the dispersion of idiosyncratic taste for different locations (ϵ). Fifth and last, it describes the gravity equation governing trade flows and uses its structure, alongside state-level trade flows from 1999 and 2017 to estimate the distance-elasticity of trade costs (κ).

5.1.1 Deforestation parameters

Returns to scale: δ

For the estimation of the parameter δ from the Cobb-Douglas deforestation production function, I rely on the fact that the supply-elasticity of new agricultural land or the deforestation elasticity, as it is sometimes referred to in the literature, is $\frac{\delta}{1-\delta}$. Recall that the optimal deforestation level T^D_{rt} (from the perspective of the deforesters) given land price q_{rt}, price index p_{rt}, and aggregate deforestation TFP $Z^D_r t$ is equal to:

$$T^D_{rt} = (Z^D_{rt})^{\frac{1}{1-\delta}} \left(\frac{\delta q_{rt}}{p_{rt}}\right)^{\frac{\delta}{1-\delta}}.$$

Taking logarithms, the equation becomes:

$$\log(T^D_{rt}) = \frac{1}{1-\delta}\log(Z^D_{rt}\delta^\delta) + \frac{\delta}{1-\delta}\log q_{rt} - \frac{\delta}{1-\delta}\log p_{rt}. \tag{5.1}$$

From this equation, I observe deforestation levels, T^D_r, and I can approximate yearly land rental rates v_{rt} (assumed to have a simple linear relationship to land prices so that $q_{rt} = \frac{1}{1-\beta(1-\rho)} v_{rt}$) from the yearly data on agricultural revenues described in section 2.1.3 and the Cobb-Douglas production function assumption[2]. Ideally I would like to estimate the coefficient on $\log(v_{rt})$ from shocks that come exclusively from changes in the demand for land and not from changes in the costs of deforestation. Given the fact that this is an equilibrium equation, simultaneity is a concern. Just as increases in the demand for land increase both land prices and deforestation in equilibrium, increases in the supply of new agricultural land (for example due to less strict conservation policies) increase deforestation but lower land prices. In the absence of an instrument that proxies for an exogenous shock to supply or demand separately, I run a regression with time and municipality fixed effects as in the equation below:

$$\boxed{\log(T^D_{rt}) = \alpha^1_r + \alpha^2_t + \frac{\delta}{1-\delta}\log v_{rt} + \epsilon_{rt}}. \tag{5.2}$$

The municipality and year fixed effects help to control for constant differences between municipalities and common macroeconomic shocks to anti-deforestation

[2] In each region, the land rental rate is calculated as $v_{rt} = \left(\sum_k \alpha_k \text{Revenue}_k\right)/\text{Area Agri}_{rt}$ for all k

policy and to agricultural markets that are common to all regions. Panels (1) and (2) of Table 5.2 show the results of this two-way fixed effects regression without weights and weighting by the total area in agriculture. In panels (3) and (4), I instrument for $\log v_{rt}$ with $\log v_{rt-1}$ and $\log v_{rt-2}$, following Anderson and Hsiao (1982). This helps us to lessen simultaneity concerns. Another reason why I may be less worried about simultaneity in this context relative to standard supply-elasticity estimations, is that land accumulates over time. Hence, the yearly rate of deforestation is likely to be too small to have a significant effect on the price of agricultural land in a region, so that the movements observed in agricultural land prices are most likely due to changes in the demand for land. The resulting elasticities of deforestation range between 0.25 and 1.1, implying δ between 0.2 and 0.52. For the main specification of the model, δ is set to be 0.5. This is for two reasons. First, because it is in line with a deforestation elasticity of 1, which is very close to what is calibrated in Farrokhi et al. (2024) for global deforestation (0.9) and to the crop-price elasticity of pan-tropical deforestation in Berman et al. (2023). Moreover, I choose to err on the side of overestimating δ, which will in turn overestimate leakage. This is because, as I shall see in the next chapter, I find relatively low rates of leakage, and therefore I want to put their seemingly small magnitude to test.

Returns to unprotected forest left: ψ

The second parameter of interest governing the deforestation technology is ψ. This parameter dictates how the productivity of aggregate deforestation goes down as there is less area left in unprotected natural ecosystems. While $\delta < 1$ (yearly decreasing returns to scale) makes it so that it is optimal to space out deforestation over time because each additional hectare is harder to cut down in a given year, $\psi > 0$ makes it so that the costs of deforesting the area of forest over the same extension of time is higher when there is less forest left. This could be either because the forest that is easier to cut down is chosen first or because the value of the standing forest, and hence the opportunity cost of deforestation, increases as it becomes scarcer. ψ can be calibrated in two ways, both of which deliver similar results. One requires doing the model inversion first and estimating deforestation TFPs Z^D_{rt} for each region r in each year t and then regressing that on the forest area left in each micro-region in a log-log regression. This will be described in the next section. The out-of-model alternative to calibrate ψ is to derive it from the steady-state condition that

Table 5.2: Estimation of δ

	Outcome: log(Deforestation)			
	(1)	(2)	(3)	(4)
	OLS	OLS	IV	IV
$\log(v_r)$	0.391***	0.253***	0.817***	1.065***
	[0.025]	[0.038]	[0.165]	[0.283]
Mun. FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Instrument	-	-	First Difference	First Difference
Weight	None	Agri. Area	None	Agri. Area
R2	0.240	0.346	0.002	-0.001
Observations	89072	89072	77938	77938

Note: This table shows the results of a two-way fixed effects regression of the log of deforestation on the log of estimated land rental rates at the municipality level, (5.2). The coefficient estimate displayed is $\frac{\delta}{1-\delta}$, where δ governs the returns to scale of the deforestation production function. Columns (1) and (2) show the results of the OLS regression and columns (3) and (4) show the results of the IV regression that instruments for land rental rates with its lagged values following Anderson and Hsiao (1982). Columns (1) and (3) weigh all municipalities equally and columns (2) and (4) weigh them by their total area in agricultural use. Standard errors in parenthesis. *p<0.1, **p<0.05, *** p<0.01.

would make the amount of forest in a region constant by equating forest loss and forest regrowth rates,

$$T^D_{r,ss} = \rho T^A_{r,ss}.$$

Substituting in the optimal level of deforestation and $Z^D_{r,ss} = \overline{Z}^D_{r,ss}\left(T^F_{r,ss}\right)^{\psi}$, I can derive an equation that relates the ratio of area in forest over the area in agricultural use on the left hand side to the total area in forest to the power of $1 - \psi/(1-\delta)$ on the right hand side.

$$\frac{T^F_{r,ss}}{T^A_{r,ss}} = \rho(\overline{Z}^D_{r,ss})^{\frac{1}{\delta-1}}\delta^{\frac{\delta}{\delta-1}}\left(\frac{q_{r,ss}}{p_{r,ss}}\right)^{\frac{\delta}{\delta-1}}(T^F_r)^{1-\frac{\psi}{1-\delta}}$$

Take logs

$$\begin{aligned}\log\left(\frac{T^F_{r,ss}}{T^A_{r,ss}}\right) =& \log\left(\rho(\overline{Z}^D_{r,ss})^{\frac{1}{\delta-1}}\delta^{\frac{\delta}{\delta-1}}\right)\\ &- \frac{\delta}{1-\delta}\log\left(\frac{q_{r,ss}}{p_{r,ss}}\right)\\ &+ \left(1 - \frac{\psi}{1-\delta}\right)\log(T^F_{r,ss})\end{aligned} \tag{5.3}$$

This is governing the way in which the ratio of forest to agricultural depends on scale in steady state. More precisely, how regions with larger steady-state area in forest have a higher (if $\psi < 1-\delta$; lower if $\psi > 1-\delta$) fraction of their area in forest relative to agricultural use. If $\psi = 0$, the optimal amount of agricultural area (cumulative deforestation) would be completely independent on the available forest (assuming that I only have interior solution), and hence the ratio of forest area to agricultural area would increase proportionally to forest area (with elasticity 1). For conciseness let us refer to $1-\psi/(1-\delta)$ as the long-run scale elasticity of forest cover, since it dictates how the total area in forest cover relates to the relative area in forest cover. As ψ increases, this elasticity decreases: having a larger area in forest is associated with having a lower relative forest cover. This is because ψ dictates how the yearly deforestation productivity increases with the extent of forest left, and hence with scale. The parameter δ is a compounding force influencing the long-run scale elasticity of forest cover. As δ increases, deforestation has more strongly decreasing returns to scale, and hence a larger area is harder to convert to agriculture, so that the steady state agricultural area becomes relatively smaller as a function of scale. If $\psi > 1-\delta$ this long-run scale elasticity of forest cover becomes negative, so that larger regions would end up with higher fractions in agricultural land relative to smaller regions. I consider this scenario implausible. Given that the only variables in the equation above that I observe are v_{rt} (proportional to q_{rt}), T^F_{rt}, and T^A_{rt}, I run the regression below to estimate $\delta/(1-\delta)$ and $1-\psi/(1-\delta)$ simultaneously, and then use this estimated δ (or the chosen 0.5) to back out ψ. The regression equation is

$$\boxed{\log\left(\frac{T^F_{rt}}{T^A_{rt}}\right) = \alpha_{s(r)t} - \frac{\delta}{1-\delta}\log(v_{rt}) + \left(1-\frac{\psi}{1-\delta}\right)\log(T^F_{rt}) + \epsilon_{rt}}, \tag{5.4}$$

where $s(r)$ is the state of region r. I control for state-year fixed effects to look only at within state-year differences in forest-to-agricultural areas. The source of variation that I want to use is cross-sectional rather than a diff-in-diff because I are interested in the long-run differences across regions with differences sizes. Since regions with high deforestation are less likely to have reached their steady state, in column (3) I consider exclusively those regions were deforestation rates are below the regrowth rate, i.e. 1% of the agricultural area. The results, shown in Table 5.3, are consistent with a δ between 0.26 and 0.29. Taking this value, I get a ψ between 0.3 and 0.32. If instead I took $\delta = 0.5$ I get ψ between 0.17 and 0.22. In order to err on the side of underestimating leakage, again, I opt for a ψ on the larger side. This is also more consistent with the results of the model-based method.

Table 5.3: Estimation of ψ

	Outcome: log((Natural Area)/(Agri. Area))		
	(1)	(2)	(3)
log(Unprotected Natural Area)	0.668***	0.589***	0.571***
	[0.004]	[0.004]	[0.004]
$\log(v_r)$		-0.419***	-0.357***
		[0.005]	[0.007]
State X Year FE	Yes	Yes	Yes
Weight	Agri. Area	Agri. Area	Agri. Area
Sample			Low deforest.
R2	0.735	0.793	0.805
Observations	28244	27875	16899

Note: This table shows the results of a regression of the log of the ratio of land in forest over land in agriculture on the log of the total area in forest at the municipality-year level, following (5.3). The first coefficient estimate displayed is $1 - \frac{\psi}{1-\delta}$, where ψ governs the "returns to total unprotected forest left" and δ governs the returns to scale of the deforestation production function. All columns controls for state-year fixed effects and weigh municipalities by agricultural area. Column (1) does not have any additional controls. Columns (2) and (3) control for the log of the estimated rental rate of agricultural land. Column (3) restricts he sample to observations with less than 1 square kilometer of forest loss so that it is more plausibly looking at a steady state. Standard errors in parenthesis. *p<0.1, **p<0.05, *** p<0.01.

Regrowth rate: ρ

The regrowth rate ρ is calculated from MapBiomas forest transitions data on growth of secondary forest divided by land under anthropic use in the previous year. Calculating an average regrowth rate over the whole of Brazil for the model period (2002-2019), I get 1%, ranging between 0.81% (in 2019) and 1.1% (in 2017), with a slightly decreasing trend over time. I consider ρ to be the same for all regions, but in reality it has significant spatial variation. The average micro-region has a regrowth rate of 1.5%, the median micro-region of 0.8%. The distribution of regrowth rates is shown in Figure 5.1, panel (a) for four different years in the model period. I can see that the mode of the regrowth rate distribution is around 0.3%, and there is a long right tail, with the 99th percentile having 13% and the largest regrowth rate being 60%. For simplicity, I abstract away from this heterogeneity and take one regrowth rate for all of

Brazil. To inform where this heterogeneity comes from and the extent to which regrowth rates are changing over time, in panel (b) I plot the regrowth rates by Brazil's large regions: North (N), North-East (NE), South (S), Center-West (CW), and South-East (SE).

5.1.2 Production Parameters

To estimate α_k for each agricultural commodity k, I rely on the 2006 Agricultural Census, a census of the universe of farming establishments. Using data on land prices, land in each agricultural activity, labour in each agricultural activity, and agricultural wages, I estimate how the ration of land income over rental income varies between pastures for cattle grazing, temporary crops, and perennials.

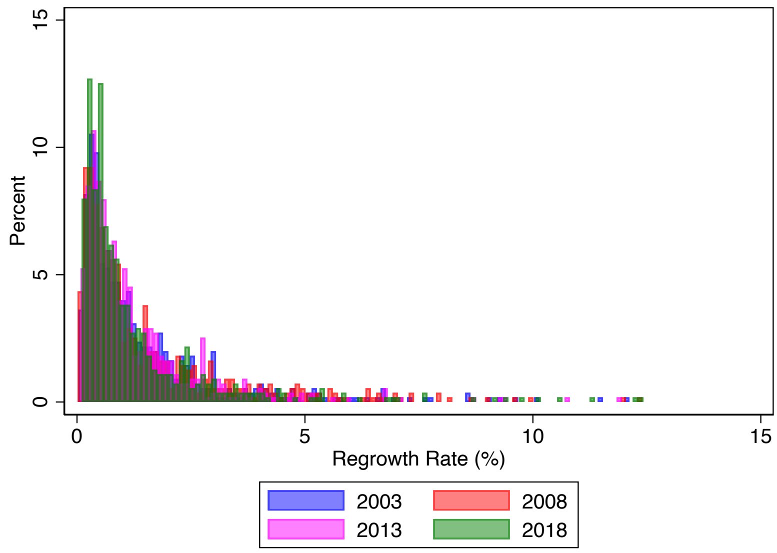
Let the area planted in each type of agricultural activity T_r^{Ak}, workers in type of agricultural activity N_r^{Ak}, labour incomes per capita in agriculture $\overline{y}_r^{LA}$, and land value q_r.

The Cobb-Douglas functional form of the agricultural production function and the zero-profit condition together imply that

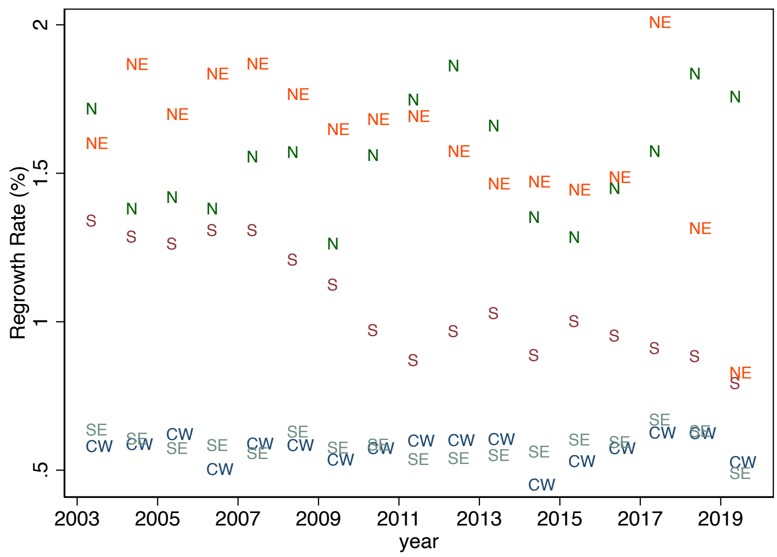
$$\frac{\alpha_k}{1-\alpha_k} = \frac{\sum_{r=1}^{R} v_r T_r^{Ak}}{\sum_{r=1}^{R} \overline{y}_r^{LA} N_r^{Ak}}.$$

Doing this I find that the land share for pastures, temporary crops, and perennials vary significantly and they are, respectively: 0.71, 0.54, and 0.36. These differences make it so that there is greater substitutability between land and labour in agricultural production as a whole. This has important implications for leakage. Consider a region where the supply of agricultural workers increases. This could be due to more migration from regions where conservation policies have been enacted. Higher supply of agricultural workers increases the demand for agricultural land. In a model with multiple agricultural sectors, however, it might also shift land use away from very land intensive activities, for example cattle ranching, towards more labour intensive ones, such as perennial crops. The extent to which this happens in equilibrium would depend mainly on consumers willingness to substitute consumption between goods and on the slope of the supply of deforestation. If deforestation is very cheap and consumers very reluctant to substitute towards less land-intensive goods, cattle-ranching will remain preferable and the inflow of workers will lead to much deforestation. By

Figure 5.1: Forest regrowth rates



(a) Distribution of forest regrowth rates of microregions



(b) Trends of forest regrowth rates by 5 regions

Note: The figures above illustrate the regional and temporal variation of forest regrowth rates across Brazil. Forest regrowth is defined as the percentage of non-forest area that becomes forest in a given year. Panel (a) above shows the distribution of regrowth rates by micro-region for 4 different years: 2003, 2008, 2013, and 2018. Panel (b) shows the yearly regrowth rates of Brazil's five large macro-regions: the North East (NE), the North (N), the South (S), and the Centre-West (CW) for every year in the study period of the counterfactual analysis (2003-2020).

contrast, in a region where deforestation is very costly and consumers readily substitute beef with soy, the increased supply of labour will mean farmers may opt to convert pastures to soy fields and demand very little deforestation.

5.1.3 Preference parameters

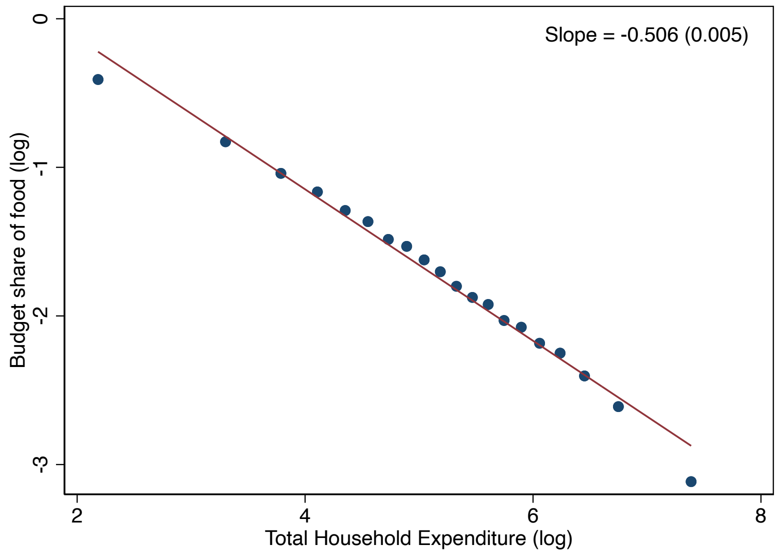
For PIGL parameters ϕ and ν, I take the values estimated by Eckert and Peters (2022). The parameter governing the non-homotheticity of preferences, η, also referred to as Engel elasticity, captures the rate at which higher-income consumers shift their budget shares away from agricultural goods (i.e. food items). I estimated by regressing the logarithm of the share of expenditure in food on the logarithm of total expenditure using the household-level data from the 2017/2018 consumer expenditure survey (POF). The table below shows the coefficients from that regression. For robustness, I consider both food expenditure as a share of total non-durables expenditure (column 1 - the preferred specification) and as a share of total household income (column 2).

Table 5.4: Estimation of η

	Outcome: log-budget share on food items	
	(1)	(2)
Log(Non-durable Expenditure)	-0.506***	
	[0.005]	
Log(Income)		-0.575***
		[0.008]
Constant	0.914***	0.632***
	[0.025]	[0.056]
R^2	0.309	0.195
Dep. Var. Mean	-1.730	-3.229
Observations	45322	45322

Note: This table displays the estimated Engel curve for food: the relationship between income and budget share of food, which illustrates the non-homotheticity in consumer preferences. The dots represent the average share of total household expenditure dedicated to food items for 20 bins of total household expenditure. Standard errors in parenthesis. *p<0.1, **p<0.05, *** p<0.01.

Figure 5.2: Engel curve for food consumption



Note: This figure displays the estimated Engel curve for food: the relationship between income and budghet share of food, which illustrates the non-homotheticity in consumer preferences. The dots represent the average share of total household expenditure dedicated to food items for 20 bins of total household expenditure.

5.1.4 Migration parameters

Recall that the utility of a migrant moving from origin o to destination d depends on: (i) consumption utility V_d, (ii) residential amenities B_d, (iii) an origin-destination migration cost μ_{od}, and (iv) an idiosyncratic preference shock with a Frechet distribution with shape parameter ϵ. These shocks are drawn every year when making migration decisions so that each year, the share of migrants from o that pick d as destination equals

$$\text{Share of Migrants}_d^o = \frac{\left(V_d B_d \mu_d^o\right)^\epsilon}{\sum_j \left(V_j B_j \mu_j^o\right)^\epsilon}.$$

The above is referred to as the migration gravity equation. To estimate it, first I take logs so that it becomes the following equation with origin and destination fixed effects

$$\log(\text{Share of Migrants}_d^o) = \delta_o + \delta_d + \epsilon \log\left(\mu_d^o\right). \tag{5.5}$$

Using bilateral migration flows between the 558 micro-regions according to the 2010 census, I use the equation above to estimate $\epsilon\mu_d^o$ and the correlation of (log) income and destination fixed effects to estimate ϵ. My preferred approach will be to estimate the full matrix of origin and destination bilateral migration utilities μ_d^o from the residuals of the regression of (log) migration rates of origin and destination fixed effects. Alternatively, I could estimate μ_d^o to be a function of the linear distance between micro-regions, $\mu_d^o = (\text{dist}_d^o + 1)^{\upsilon}$ from a regression of migration shares on log of distance plus 1 with origin and destination fixed effects as reported in table 5.5, column (1). I add one so that $\mu_o^o = 1$ and $\mu_d^o(\text{dist}_d^o)$ is a decreasing function of distance as long as $\upsilon < 0$. The distance elasticity of migration is estimated to be -1.55.

To estimate ϵ, note that the destination fixed effect, according to the structure of the model, depends on income at destination and amenities as follows,

$$\delta_d = \epsilon \log(V_d) + \epsilon \log(B_d). \tag{5.6}$$

And $\log(V_d)$, up to an approximation of equation 4.3, is $\log(V_d) \approx \eta \log(e_d/p_d) - \log(\eta)$. Thus, following Buggle et al. (2023), I estimate ϵ from a regression of the fixed effect on the estimated fixed effects on the logarithm of income in the 2010 census. The income used in this regression is the income of those people who were in-migrants to destination d. Although average income and real expenditure (e) are not the same thing, as long as they are proportional, a regression of destination fixed effects on income can help us estimate ϵ. Column (2) below show results of PPML regression, where I get $\epsilon\eta = 0.633$, which implies $\epsilon = 1.25$ for the estimated $\eta = 0.506$. Two concerns with this method are (i) endogeneity: income may be correlated with amenities, which would bias the estimation and (ii) measurement error: I use income instead of expenditure and I do not divide incomes by their regional price index, which is unobserved. A potential solution to be implemented include using a panel of migration flows from older and newer census data, and including micro-region and year fixed effects, as in Buggle et al. (2023).

5.1.5 Trade parameters

Iceberg trade costs are estimated to fit trade costs in 1999 and 2017 from the residuals in the regression equation of (log) trade flows with origin and destination fixed effects. Taking logarithms from equation 4.4, and sending the volume of expenditure of consumers in region d to the right hand side so that the data

Table 5.5: Estimation of migration parameters

	(1) Share of Migrants	(2) Destination F.E.	(3) Destination F.E.
log(Distance+1)	-1.550***		
	[0.002]		
log(Mig. income)		0.633***	
		[0.044]	
log(Avg. income)			0.646***
			[0.043]
Origin FE	Yes	No	No
Dest. FE	Yes	No	No
Pseudo R2	0.805	0.25	0.27
Observations	3.11e+05	558	558
Method	PPML	PPML	PPML

Note: This table displays the results of the regressions used for the estimation of migration parameters. Column (1) is a Poisson Pseudo Maximum Likelihood (PPML) regression of bilateral migration shares in the past 5 years at the micro-region level on the log of their distance plus 1. Migration data comes from the 2010 population census. The regression includes origin and destination fixed effects Column (2) is also a PPML regression. It has as dependent variable the destination fixed effect estimated from a PPML as in column (1) but without the fixed effects only, and not the log of distance plus 1, as specified in (5.5). The coefficient displayed is the coefficient on the log of the average income of migrants according to the 2010 population census. Column (3) is as column (2) but considers the average income of all people living in the destination micro-region. Standard errors in parenthesis. *p<0.1, **p<0.05, *** p<0.01.

on trade flow rather than shares can be used, yields

$$\log X_d^o = \log p^o - \log p_d X_d + (1-\sigma)\log(\tau_d^o)$$

Notice that these data do not include the observations for which the origin and the destination are the same, as it is data from inter-state customs. Thus, if I assume that $\tau_d^o = (\text{dist}_{od} + 1)^\kappa$, then the coefficient of the log-log regression of trade flows on distance (+1) with origin and destination fixed effects equals $\kappa(1-\sigma)$. Given the results from the regressions reported in the table below, $\kappa \approx \frac{-0.9}{1-9} \approx 0.11$.

Table 5.6: Estimation of trade parameters

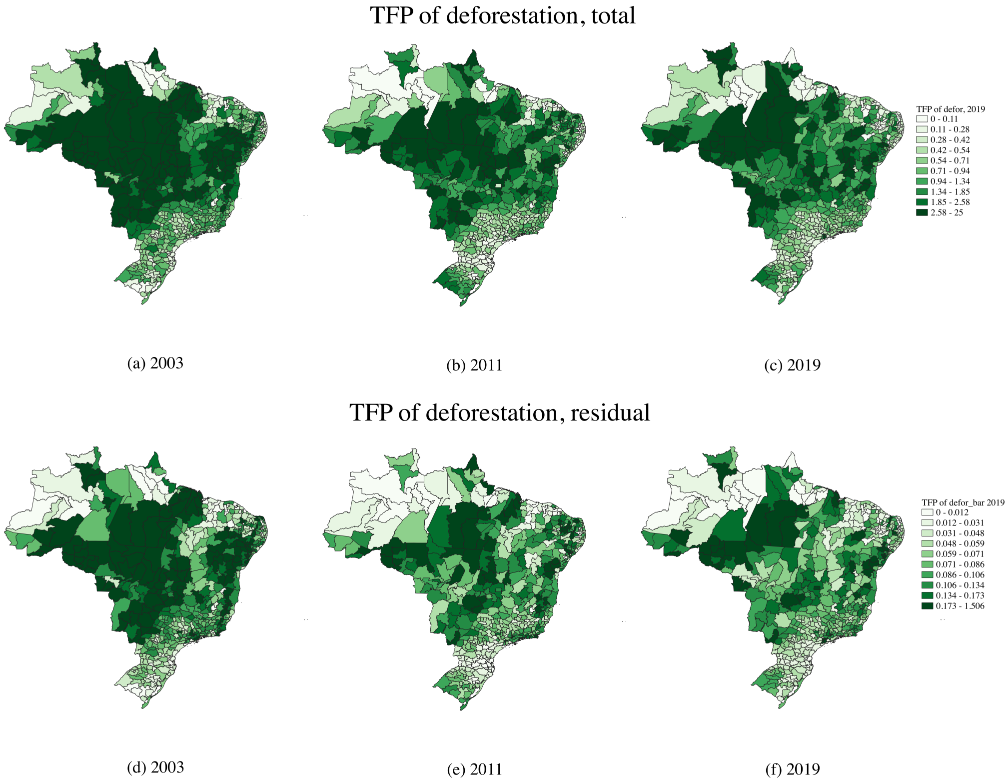
	Outcome: log-trade flows (1)	(2)
log(Distance+1)	-0.871***	-0.916***
	[0.053]	[0.036]
Year	1999	2017
Origin FE	Yes	Yes
Dest. FE	Yes	Yes
Pseudo R2	0.966	0.957
Observations	702	702

Note: This table displays the results of the regressions used for the estimation of trade parameter κ that governs the distance-elasticity of trade. They are OLS regressions of the bilateral trade flows between different states on the log of their distance plus with origin and destination fixed effects. The structural interpretation of the coefficient is that it equals $(1-\sigma)\kappa$. Column (1) uses 1999 state-level trade flow data and column (2) uses 2017 trade flow data. Standard errors in parenthesis. *p<0.1, **p<0.05, *** p<0.01.

5.2 Model inversion

In order to invert the model and back out the total factor productivities of each sector and the deforestation productivity (Z_{rt}^{Ak}, Z_{rt}^{NA}, Z_{rt}^{D}) I assume that each year in the period 2003-2019 the economy is in equilibrium as described above and I use data on some observed features. First, I use data on the endowments that are treated as exogenous within each equilibrium. That is, working population, land already in agriculture, and land in natural ecosystems. Second, I use data on the following equilibrium quantities in each region for that year: (i) the total amount of deforestation, T_{rt}^D, (ii) the agricultural labour share (s_{rt}^A), (iii) the

Figure 5.3: Estimated Deforestation TFPs

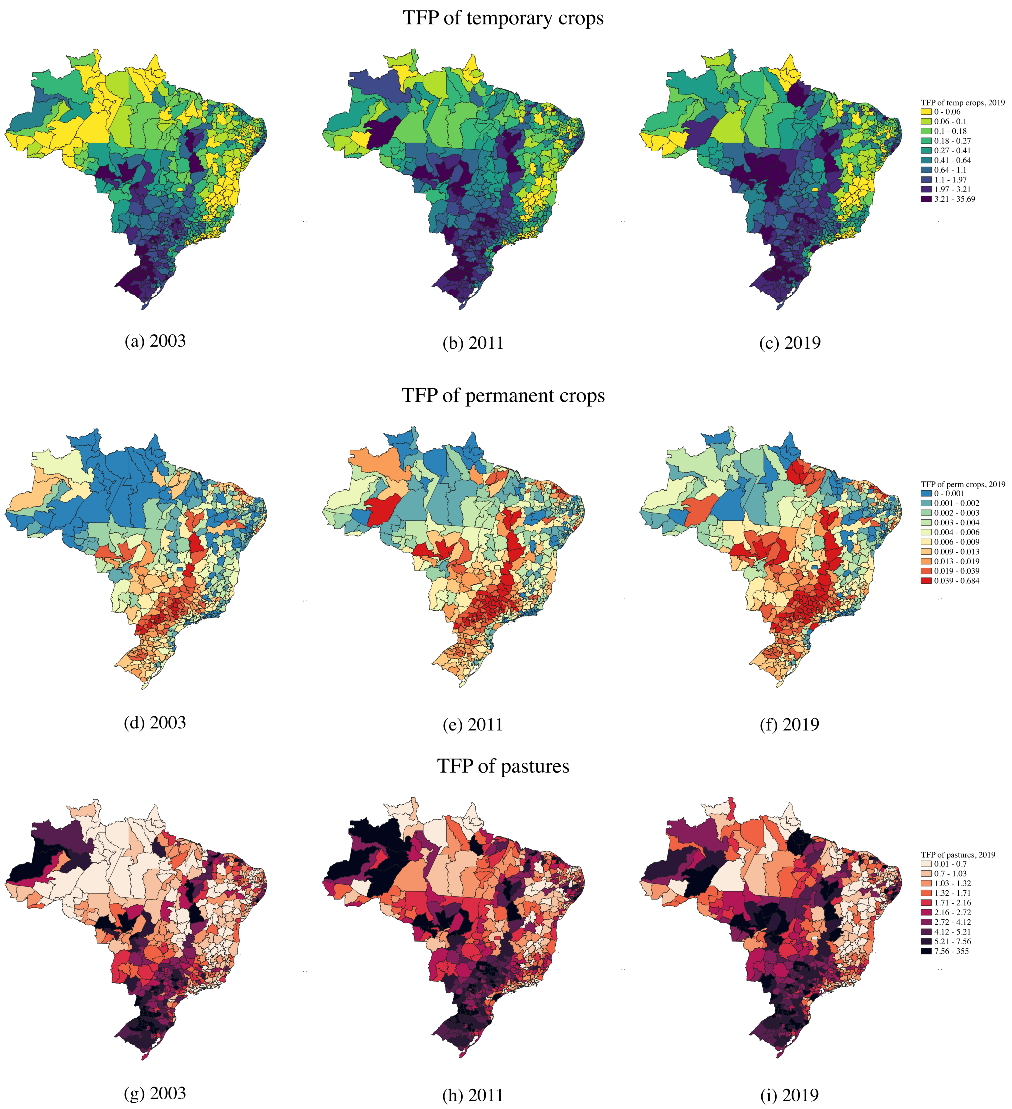


Note: These maps illustrate the spatial distribution of the deforestation productivities across Brazil for the years 2003, 2011, and 2019. The maps at the top, (a)-(c), show the total deforestation TFP $Z^D_{rt} = \overline{Z}^D_{rt}(T^F_{rt})^\psi$ which includes the dependence of the TFP on unprotected forest area. The legend is consistent across these three maps. The maps at the bottom, (d)-(f), show the residual TFP after taking out the dependence on unprotected forest area left, $\overline{Z}^D_{rt}$. The legend is consistent across these three maps.

share of agricultural land in each land use (T^{Ak}_{rt}), and (iv) agricultural revenues ($p^{Akr}_t Y^{Ak}_{rt}$). Through the algorithm described in appendix section B.3, I back out the TFPs. the residential amenities are assumed to be constant and are backed out from equation 5.6 (i) the observed level of consumption utilities V_d, the destination fixed effects estimated from 5.5, and the calibrated ϵ.

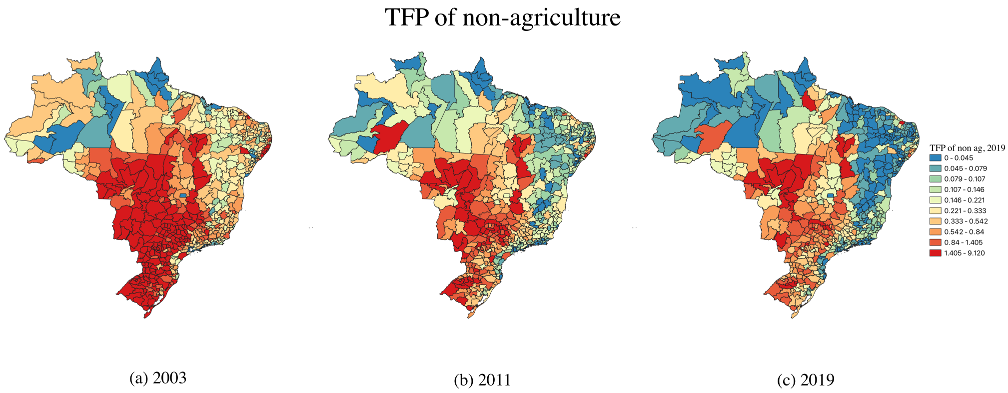
The maps below show the spatial distribution of the regional fundamentals and how they change over the 2003-2019 period.

Figure 5.4: Estimated Agricultural TFPs, by activity



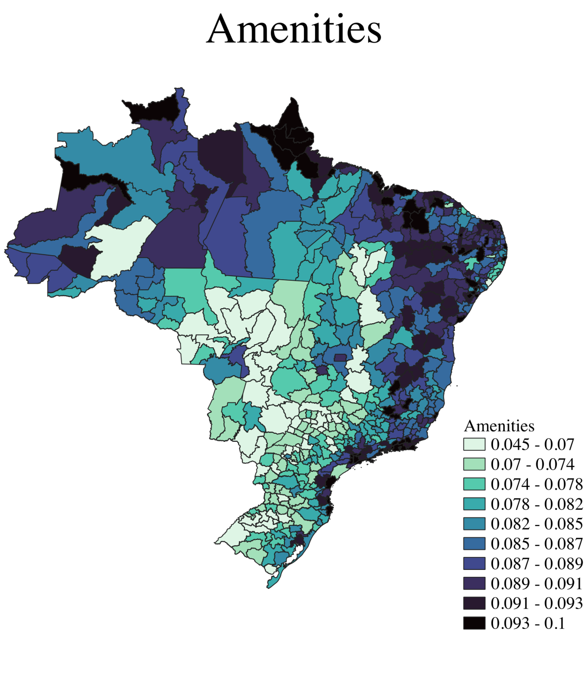
Note: These maps show the spatial distribution of the agricultural productivities of the three agricultural sectors considered (permanent crops, temporary crops, and pastures) across Brazil for the years 2003, 2011, and 2019. The estimation of each year's productivities is done through an exact inversion of the model in that year, assuming equilibrium conditions.

Figure 5.5: Estimated Non Agricultural TFPs



Note: These maps show the spatial distribution of the agricultural productivities of the non-agricultural productivity across Brazil for the years 2003, 2011, and 2019. The estimation of each year's productivities is done through an exact inversion of the model in that year that assumes equilibrium conditions.

Figure 5.6: Estimated Amenities



Note: These maps show the spatial distribution of the residential amenities across Brazil for the years 2003, 2011, and 2019. The estimation is done through an exact inversion of the model, which rationalises observed migration flows given estimated bilateral migration costs and region-specific yearly utilities.

Correlation between fundamentals and deforestation levels

In this theoretical framework, the deforestation is rationalised by the combination of two market forces: the demand for agricultural land and the supply of deforestation. The supply-side regional fundamental governing deforestation is Z_r^D. The demand for agricultural land is given, in turn, by the equilibrium in the market for agricultural goods. From the agricultural goods supply side, I expect to see more deforestation in regions which have higher agricultural TFPs, especially for the land-intensive agricultural activities (i.e. cattle grazing). From the demand side, the regions that have higher market access should experience greater demand for agricultural goods, and hence for agricultural land, as discussed in Donaldson and Hornbeck (2016).[3] In summary, there are three sets of regional fundamentals driving the spatial allocation of deforestation, each of which have different implications for the level of leakage: (1) the deforestation TFP, (2) the TFP of agricultural activities, especially the most land-intensive ones, and (3) market access. Table 5.7 presents the estimated coefficients of a simple regression of the baseline levels of (log) deforestation on the (log) productivities of various sectors and the log of a measure of market access approximated by $MA_r \approx \sum_d (\tau_d^r)^{\sigma-1} L_d$ as in Donaldson and Hornbeck (2016). In columns (1) and (2), which do not include the deforestation TFP, the coefficients go in the opposite direction as we would expect them to affect deforestation. Column (3) shows that a large share of the variation in deforestation levels is explained by the deforestation TFP, the immobile factor that is shocked for the counterfactual policy simulation ($R^2 = 0.811$). Column (4) shows that deforestation is well explained ($R^2 = 0.987$) and that once the deforestation TFP is added to the regression in column (2) all the coefficients flip sign to align with the model-based predictions. That is, deforestation increases with: (i) higher agricultural productivity, (ii) lower non-agricultural productivity, and (iii) higher market access.

[3]Here σ corresponds to the elasticity of substitution between different regional varieties of the final goods.

Table 5.7: Correlates of observed deforestation

	Outcome: log-observed yearly deforestation			
	(1)	(2)	(3)	(4)
(log) temp. crops TFP	-0.54***	-0.21***		0.10***
	[0.04]	[0.03]		[0.00]
(log) perm. crops TFP	-0.39***	0.03		-0.00
	[0.03]	[0.03]		[0.00]
(log) pasture TFP	-0.16***	-0.61***		1.14***
	[0.05]	[0.04]		[0.01]
(log) non-ag TFP	1.08***	0.67***		-0.38***
	[0.03]	[0.03]		[0.00]
(log) Market acces		-1.19***		0.26***
		[0.02]		[0.00]
(log) Deforestation TFP			1.39***	1.90***
			[0.01]	[0.00]
R^2	0.122	0.306	0.811	0.987
Dep. Var. Mean	2.69	2.69	2.69	2.69
Dep. Var. SD	1.95	1.95	1.95	1.95
Observations	8587	8587	8587	8587

Note: This table displays the results of four OLS regressions at the micro-region-year level. The dependent variable is the log of the yearly deforestation rate. The regressors are a variety of combinations of model-estimated regional fundamentals (all in logarithms). Column (1) includes the productivities of all sectors producing final goods. Column (2) includes market access too. In columns (1) and (2) the coefficients go in the opposite direction as we would expect them to affect deforestation. Column (3) regresses log deforestation only on the log of the estimated deforestation TFP and column (4) includes all the regressors in column (2) and the estimated deforestation TFP.

Chapter 6

Counterfactual Analysis

The theoretical framework developed in the previous section allows us to simulate counterfactual scenarios of local deforestation policy, and understand how deforestation would have evolved in general equilibrium in these alternative scenarios. In particular, I can apply the model to analyse the national impact of the two types of local, yet large scale, anti-deforestation policies implemented by the Brazilian government over the past decades (Priority List and Protected Areas), taking into account general equilibrium effects.

6.1 Defining counterfactuals and leakage

For each of the two policies, I simulate two counterfactual scenarios for the evolution of nation-wide deforestation between the years 2003 and 2018.

> **Counterfactual A: No policy** – a scenario in which, between 2003 and 2018, the policy was never enacted.
>
> **Counterfactual B: Policy, no leakage** – a scenario in which, between 2003 and 2018, the policy was implemented, but it has no general equilibrium effects

The difference between counterfactual A and the observed evolution of deforestation measures the overall effect of the policy, including both the direct effect via banned deforestation in targeted areas, and the indirect effect via potential leakage to non-targeted areas. The difference between the observed evolution of deforestation and counterfactual B measures the amount of leakage caused by the localised policies. The difference between counterfactuals A and B measures the total effect of the localised policies under no deforestation leakage, i.e. only considering the direct effect via banned deforestation in targeted areas.

For each of the two policies, I construct the counterfactuals as follows: First, I calibrate the baseline regional fundamentals $\{Z^s_{rt}, B_{rt}\}$ to fit the observed deforestation between the years of 2003 and 2018. I construct $(Z^D_{rt})^{nopolicy}$, the deforestation productivity of each region in a counterfactual scenario where the policy was not implemented as described in the subsections 6.1.1 and 6.1.2 below.

For counterfactual A, I simulate total land cleared for agriculture in each year between 2003 and 2018, solving for the optimal deforestation level $(T^D_{rt})^{nopolicy}$ according to Equation 4.2 using the calculated regional deforestation productivity $(Z^D_{rt})^{nopolicy}$. Total agricultural land at each period is then given by equation 6.1.

$$(T^A_{rt})^{nopolicy} = (1-\rho)(T^A_{rt-1})^{nopolicy} + (T^D_{rt})^{nopolicy} \tag{6.1}$$

For counterfactual B, I calculate a "no leakage" scenario that has the implied prices coming from counterfactual A (i.e. no-policy prices), and baseline region deforestation productivities estimated through the model Z^D_r (with policy). Fixing the prices under the no-policy scenario effectively shuts down the channel through which deforestation could leak to other areas, as both labour market and goods market relocation mechanisms operate through price channels.

$$(T^D_r)^{noleakage} = \left(\delta \frac{(q_r)^{nopolicy}}{(p_r)^{nopolicy}}\right)^{\frac{\delta}{1-\delta}} (Z^D_r)^{\frac{1}{1-\delta}} \tag{6.2}$$

6.1.1 No Priority List deforestation productivities

In order to estimate the counterfactual deforestation productivities of regions in the Priority List, I do the following. First, I estimate the effect of the policy on the values of Z^D_{rt} obtained from the model inversion. The preferred model is a following Poisson quasi-maximum likelihood model with region and time fixed effects, as shown below

$$\mathbb{E}(\overline{Z}_{rt}^{D}|r,t,\text{Priority}_{rt}) = \exp(\alpha_r + \delta_t + \beta\text{Priority}_{rt} + \epsilon_{rt})$$

Having estimated β, I construct a no-priority-list $(\overline{Z}_{rt}^{D})^{nopolicy}$ that equals

$$(\overline{Z}_{rt}^{D})^{nopolicy} = \overline{Z}_{rt}^{D}/\exp(\widehat{\beta}). \tag{6.3}$$

Using this new productivity of deforestation, I simulate the model forward for 17 periods starting in 2003. The resulting timeseries corresponds to counterfactual A for the Priority List. The overall deforestation productivities $(Z_{rt}^{D})^{nopolicy}$ will be different both because of the changes in $\overline{Z}_{rt}^{D}$ in Priority List regions that are external to the simulation and because of the endogenous changes in T_{rt}^{F} that will happen in all regions. In particular, Priority List municipalities will have more remaining areas in forest, and the rest will have less.

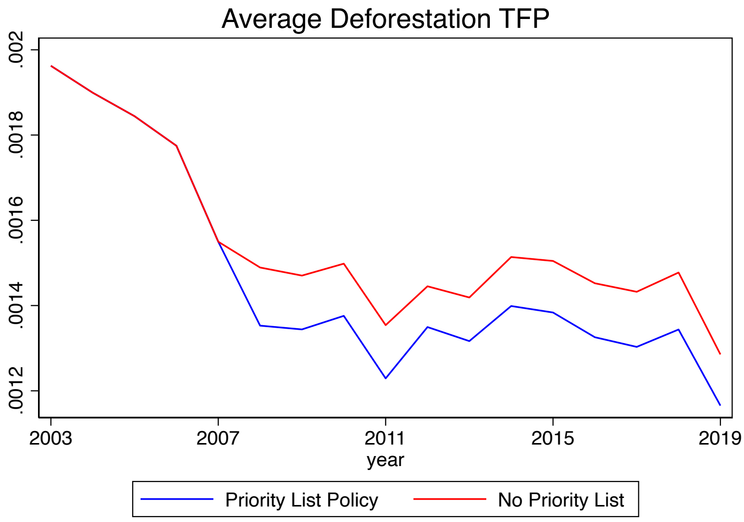
$$(Z_{rt}^{D})^{nopolicy} = (\overline{Z}_{rt}^{D})^{nopolicy}((T_{rt}^{F})^{nopolicy})^{\psi} \tag{6.4}$$

Figure 6.1 below shows these changes in deforestation productivities. In panel (a) I plot the change in the yearly average over all regions, and in panel (b) I plot three maps, for 2008, 2011, and 2019, of the resulting changes in deforestation productivity Z_{r}^{D} of each region.

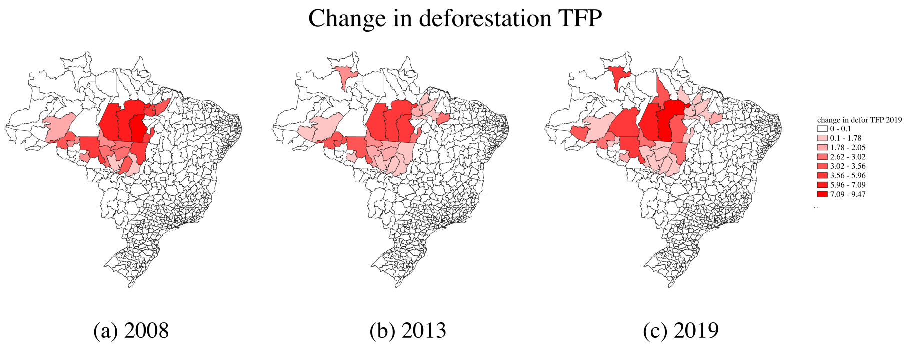
6.1.2 No Protected Areas deforestation productivities

The procedure to estimate the counterfactual deforestation productivities as if there were no new protected areas established between 2003 and 2008 is as follows. I assume that part of the deforestation productivity that changes is not $\overline{Z}_{rt}^{D}$ but T_{rt}^{F}, the area of r under unprotected forest. Given the extremely low rates of deforestation inside protected areas, I assume that they have perfect enforcement and that none of them can be converted to agricultural. Let T_{rt}^{P} be the area of region r that gets protected in year t. In order to simulate the reversal of new Protected Areas, I increase the area in available land for deforestation by

Figure 6.1: Changes in deforestation productivity: Priority List



(a) Changes in average deforestation productivity



(b) Map of changes in deforestation productivity

Note: This figure illustrates the difference in estimated deforestation TFPs between (i) the estimated value from the inversion of the model using observed data, and (ii) the simulated scenario in which there is no Priority List following equations (6.3) and (6.4). Panel (a) on top, shows the trends in deforestation TFP Z^D_{rt} for both scenarios over the study period 2003-2019. Panel (b) at the bottom show the spatial distribution of the differences in Z^D_{rt} between scenarios (i) and (ii).

T^P_{rt}. So that,

$$(Z^D_{rt})^{nopolicy} = \overline{Z}^D_{rt}((T^F_{rt})^{nopolicy} + T^P_{rt})^{\psi}, \tag{6.5}$$

where $(T^F_{rt})^{nopolicy}$ is the remaining observed area of unprotected forest in region r at time t resulting from the simulation of counterfactuals with no new protected areas in previous periods. This is an iterative definition, as $(T^F_{rt})^{nopolicy}$ depends on $(Z^D_{r\tau})^{nopolicy}$ for $\tau < t$.

6.2 Counterfactual deforestation and leakage

Now I turn to the results of the counterfactual simulations on the levels of deforestation. Given the counterfactual levels of deforestation for counterfactuals A, $(T^D_{rt})^{nopolicy}$, and B, $(T^D_{rt})^{policy,noleakage}$, I can also calculate the total percentage of deforestation reduction that is undone by general equilibrium effects:

$$\text{Leakage}_t = \sum_r \Big(\underbrace{(T^D_{rt})^{policy}}_{\text{Data}} - \underbrace{(T^D_{rt})^{noleakage}}_{\text{Counterfactual B}} \Big) \Big/ \Big(\underbrace{(T^D_{rt})^{nopolicy}}_{\text{Counterfactual A}} - \underbrace{(T^D_{rt})^{noleakage}}_{\text{Counterfactual B}} \Big) \tag{6.6}$$

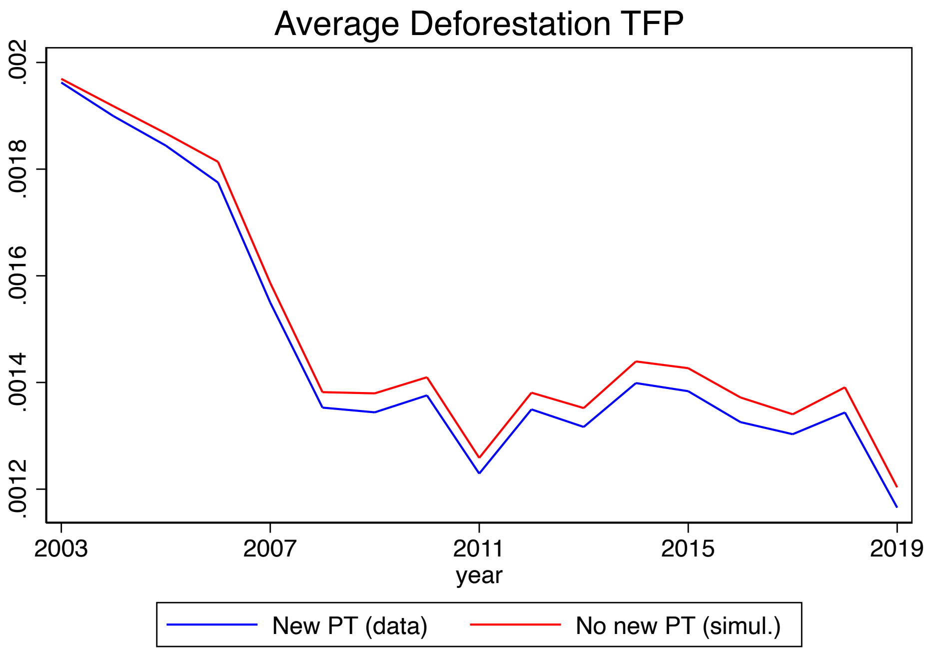
I will then convert the rates of deforestation from surface area to carbon emissions (in C tonnes) using data on the heterogeneous carbon density of the natural ecosystems of the various regions, shown in appendix figure A.10.

Priority List

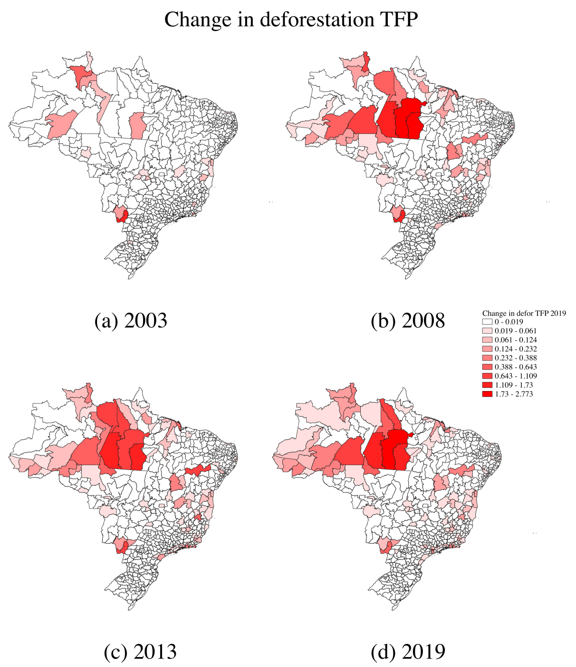
Figure 6.3 shows the evolution in total observed deforestation across Brazil, as well as the results from simulating counterfactuals A and B for the Priority List Policy. The line in blue depicts the reduction in deforestation in the data. That is, in the scenario where the policy is implemented and has GE effects. The line in red shows the reversal of the policy (Counterfactual A), and the line in green shows the rates of the policy without GE effects (Counterfactual B).

Figures 6.4 and 6.5 show the differences between observed data and the two counterfactuals. The former shows the overall changes in deforestation between the data and the scenario with no policy. This corresponds to the aggregate effects of the policy combining direct and GE effects. Overall one can see that the microregions that are in the Priority List have substantial decreases in deforestation that are about one order of magnitude greater than those in the

Figure 6.2: Changes in deforestation productivity: Protected Areas



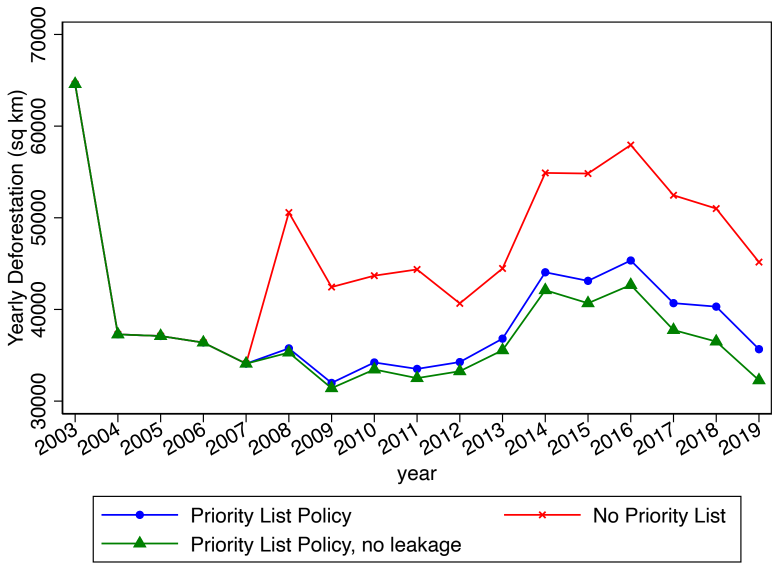
(a) Changes in deforestation productivity: Protected Areas



(b) Map of changes in deforestation productivity

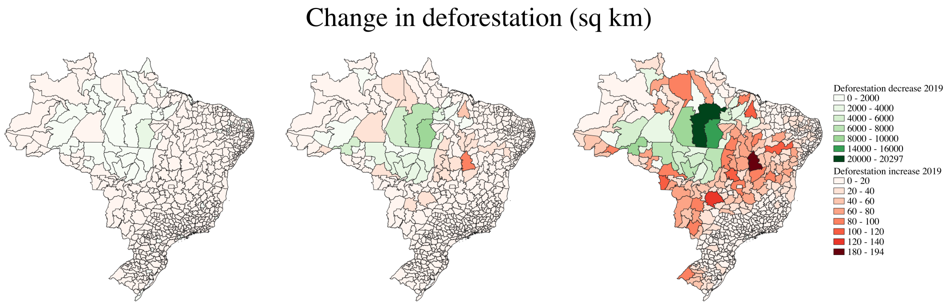
Note: This figure illustrates the difference in estimated deforestation TFPs between (i) the estimated value from the inversion of the model using observed data, and (ii) the simulated scenario in which there are no new Protected Areas from 2003 following equation (6.5). Panel (a) on top, shows the trends in deforestation TFP Z_{rt}^{D} for both scenarios over the study period 2003-2019. Panel (b) at the bottom show the spatial distribution of the differences in Z_{rt}^{D} between scenarios (i) and (ii).

Figure 6.3: Counterfactual deforestation trends: Priority List



Note: This figure shows the trends in yearly deforestation levels (in thousands of squared kilometers) across Brazil between the years of 2003 and 2019 under counterfactual simulations based on the Priority List policy. The blue curve represents observed deforestation, the red curve represents a counterfactual scenario in which the policy was never implemented, and the green curve represents a counterfactual in which the policy was implemented and GE effects, and hence leakage, are shut down by considering the prices under the no-policy scenario.

Figure 6.4: Counterfactual deforestation maps: Priority List



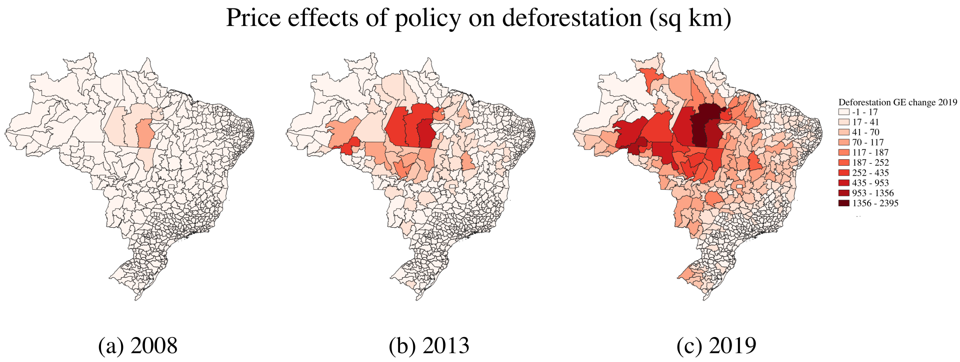
(a) 2008 (b) 2013 (c) 2019

Note: This figure shows the Priority-List-driven changes in cumulative deforestation (in thousands of squared kilometers) across the various microregions of Brazil for years 2008, 2013, and 2019. Specifically, it maps the differences between deforestation in the data (which includes the effect Priority List policy) and the no-policy counterfactual simulation. The regions where the Priority List policy is implemented have some large decreases in deforestation, while the rest of the country has increases, albeit much smaller in magnitude.

non-Priority List regions. In the second set of maps, it can be seen that the price effects on deforestation (difference between the data and counterfactual B) are higher where deforestation productivites are higher. Since the Priority List does not fully ban deforestation, some of the general equilibrium effects, which I still call leakage, are *within* priority-listed regions.

Figure 6.6 illustrates the evolution of the cumulative leakage to non-treated areas expressed as a share of the avoided deforestation, as defined in equation 6.6. Leakage at the start of the considered period is low, around 3%, but increases gradually over time. By 2019, approximately 15% of the deforestation avoided by the policy in the targeted areas is outdone by general equilibrium effects. About 50% of these general equilibrium effects happen within Priority List regions. In-situ leakage would not cause biases of reduced-form estimates. It would instead be a mechanism that could explain why effects are lower than they might otherwise be if the conservation policy did not make agricultural land scarcer. This means that only about 7% of reductions in deforestation are outdone by displacement to non-Priority List regions. When looking at these results for Carbon emissions rather than total area the size of leakage to non-Priority Listed regions is even smaller, only 3%. This makes sense given that: (i) Priority List regions have the highest productivities of deforestation, even after implementing the policy, and leakage is expected to be higher in places with highest

Figure 6.5: General Equilibrium Effects maps: Priority List



Note: This figure shows the spatial distribution of the part of Priority-List-driven changes in cumulative deforestation (in thousands of squared kilometers) that is due to GE effects. Namely, it is the difference between the scenario with Priority List policy (data) and the scenario with TFPs as in the data, but prices as in the no-policy simulation. The spatial units are the various microregions of Brazil for years 2008, 2013, and 2019. GE effects have consequences in all of Brazil, but they are more pronounced in the Priority Listed micro-regions and their neighbours.

deforestation productivity, (ii) Priority List regions exhibit some of the highest carbon densities. There are regions with higher carbon densities, especially in the northern side of the Amazon River basin, but these are at lower risk due to a combination of lower market access and lower agricultural productivities.

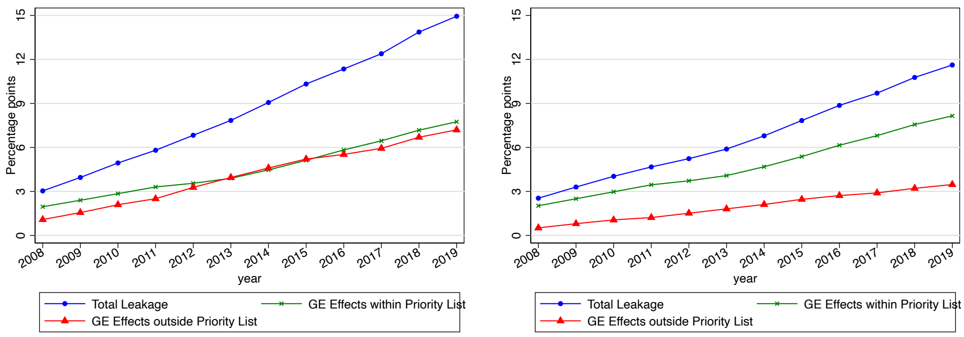
Protected Areas

Now let us look at the same results but for the Protected Areas. Figure 6.7 shows the trends in deforestation for the same three scenarios: Control: policy with GE effects (i.e., as in the observed data); Counterfactual A: no policy (i.e., no new Protected Areas established from 2003 in equilibrium); and Counterfactual B: policy, no leakage (i.e., with prices as if there were no new Protected Areas).

In the maps below, one can see the changes in the total (from 2003) area deforested as a result of the policy. Figure 6.8 shows the difference between the data and counterfactual A while figure 6.9 shows the difference between the data and counterfactual B, which isolates the price-effects of the policy.

In this case, there is no meaningful distinction between GE effects in-situ and displacement. This is because I assume Protected Areas to lead to zero deforestation within them. There will be some leakage towards micro-regions with

Figure 6.6: Cumulative leakage: Priority List

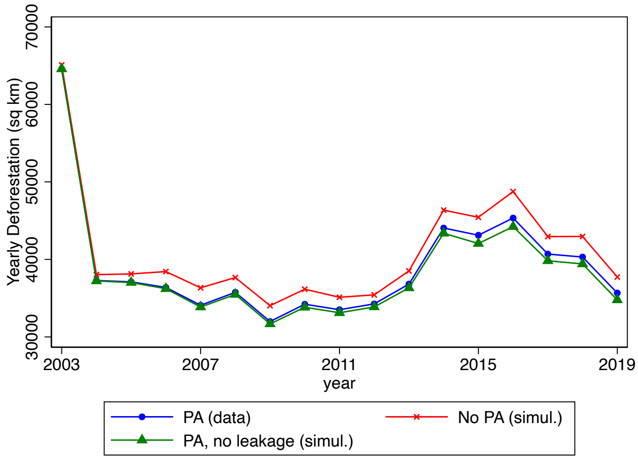


(a) Leakage of area deforested

(b) Leakage of Carbon emissions

Note: This figure shows the evolution of the leakage of the Priority-List policies over time. The blue line indicates the total % of (cumulative) avoided deforestation that is outdone by increases due to GE effects ("leakage"), following formula (6.6). Some of the GE effects, however, are experienced within Priority List regions, as shown in map (6.5) because GE effects counteract deforestation efforts within the targeted region. The green line shows the fraction of this leakage that is due to deforestation reduction in Priority-List regions. The red line shows the fraction of leakage that is "true leakage": that is, % of avoided deforestation outdone be increases in deforestation in no-policy regions. Panel (a) shows leakage in terms of area deforested while panel (b) shows leakage in terms of carbon emissions. Carbon emissions are calculated using the micro-region level average of the carbon density (including both above- and below-ground) of natural ecosystems in 2010.

Figure 6.7: Counterfactual deforestation trends: Protected Areas

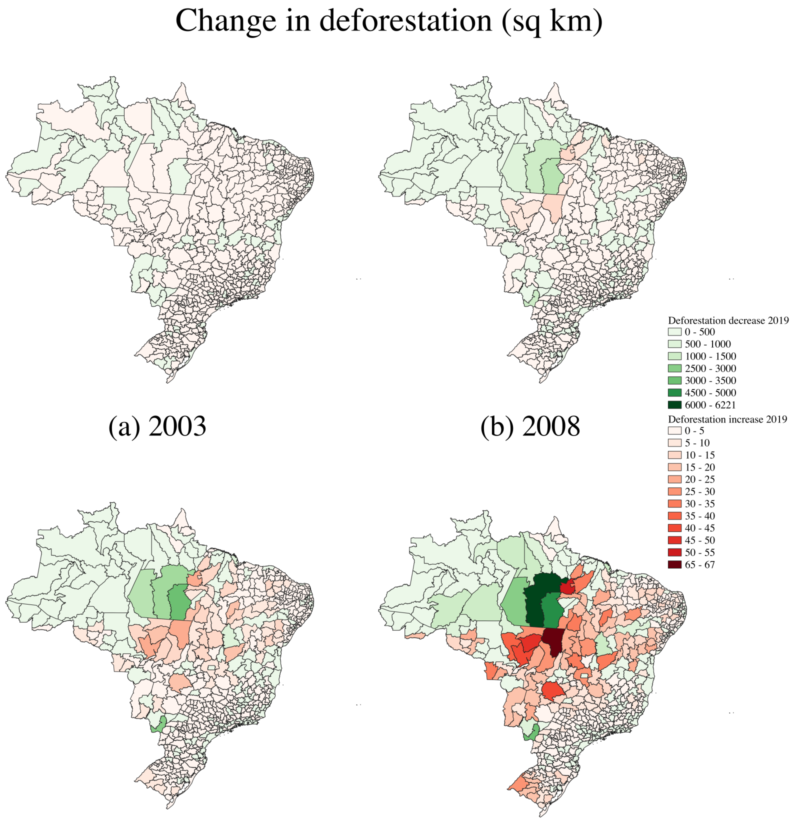


Note: This figure shows the trends in yearly deforestation levels (in thousands of squared kilometers) across Brazil between the years of 2003 and 2019 under counterfactual simulations based on the establishment of Protected Areas. The blue curve represents observed deforestation, the red curve represents a counterfactual scenario in which no new territories are protected from 2003, and the green curve represents a counterfactual in which new Protected Areas are established and GE effects, and hence leakage, are shut down by considering the prices under the no-policy scenario.

Protected Areas and some towards micro-regions without, but I do not consider this distinction meaningful as it depends arbitrarily on the way in which Protected Areas and micro-regions overlap. The graph below shows the percentage of leakage caused by Protected Areas. Interestingly, although the total deforestation reduction caused by the new Protected Areas is lower than that of the Priority List, their leakage is higher. This is because they do not target the regions with the highest fundamental productivity of deforestation and hence they remain available to absorb the increased demand for land caused by Protected Areas.

Overall, the conclusion from the simulation exercise is that, even though there is a detectable amount of leakage to non-targeted areas in localised anti-deforestation policies, it does not mean that such policies are ineffective in reducing global deforestation. Both localised policies implemented by the Brazilian government over the past decades were, not only effective in reducing deforestation locally, but were able to retain at least 80% of this effect when considering national deforestation levels for the following 15 years. This finding suggests that concerns about leakage outdoing the majority of gains in localised anti-deforestation policies might be unwarranted, and that these policies may be effective in reducing

Figure 6.8: Counterfactual deforestation maps: Protected Areas

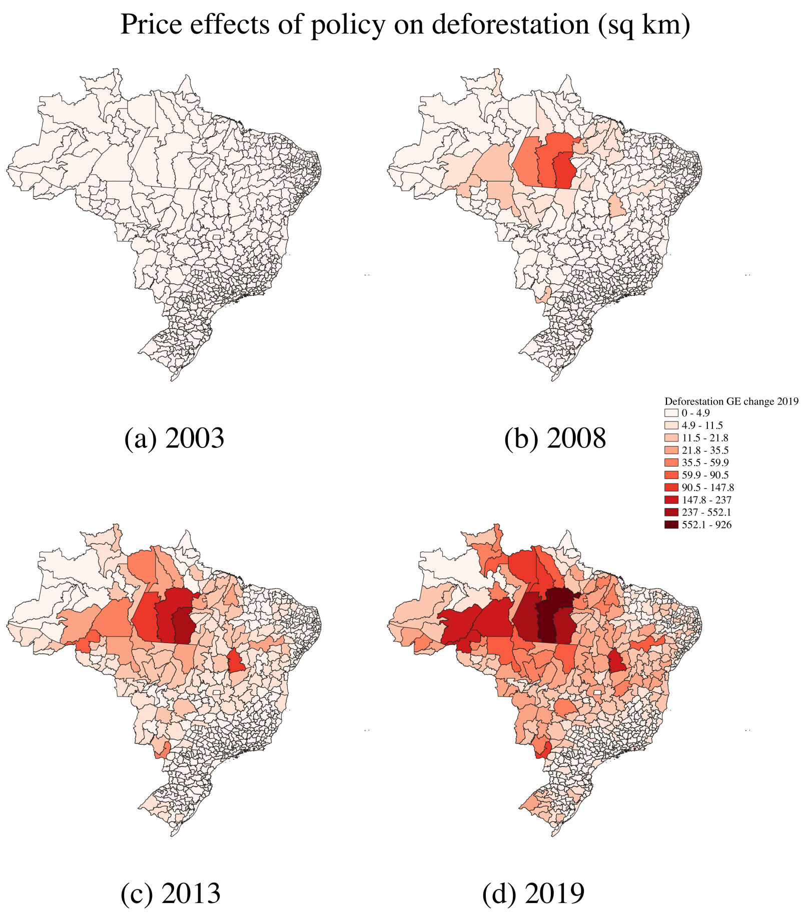


(a) 2003 (b) 2008

(c) 2013 (d) 2019

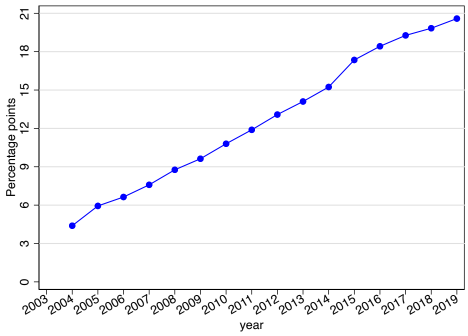
Note: This figure shows the new-Protected-Area-driven changes in cumulative deforestation (in thousands of squared kilometers) across the various microregions of Brazil for years 2003, 2008, 2013, and 2019. Specifically, it maps the differences between deforestation in the data (which includes the effect Protected Area policy) and the no-policy counterfactual simulation. The establishment of a Protected Area in part of a municipality leads to a direct reduction in deforestation by decreasing the area available for deforestation (T^F_{rt}), which mechanically reduces $Z^D_{rt} = \overline{Z}^D_{rt}(T^F_{rt})^{\psi}$. However, it also leads to an increase in deforestation via GE effects. Increases in deforestation seem to be of a much smaller magnitude overall.

Figure 6.9: General Equilibrium Effects maps: Protected Areas

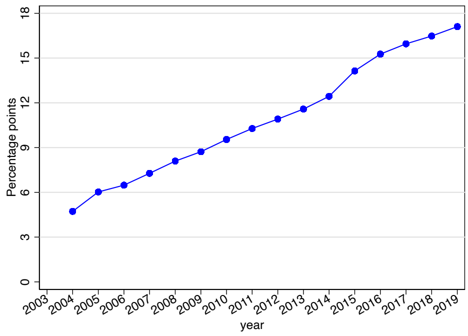


Note: This figure shows the spatial distribution of the part of new-Protected-Area-driven changes in cumulative deforestation (in thousands of squared kilometers) that is due to GE effects. Namely, it is the difference between the scenario with new Protected Areas established (data) and the scenario with TFPs as in the data, but prices as in the no-new-Protected-Areas simulation. The spatial units are the various microregions of Brazil for years 2003, 2008, 2013, and 2019.

Figure 6.10: Cumulative leakage: Protected Areas



(a) Leakage of area deforested



(b) Leakage of Carbon emissions

Note: This figure shows the evolution of the leakage of the establishment of Protected Areas from 2003 until 2019. The blue line indicates the total percentage of (cumulative) avoided deforestation that is outdone by increases due to GE effects ("leakage"), following formula (6.6). Panel (a) shows leakage in terms of area deforested while panel (b) shows leakage in terms of carbon emissions. Carbon emissions are calculated using the micro-region level average of the carbon density (including both above- and below-ground) of natural ecosystems in 2010.

global forest loss. Even more importantly, this exercise teaches us that targeting the areas with highest rates of deforestation is not only ideal to maximize deforestation reductions but also to minimize GE implications which may (i) undermine the goals of the policy, (ii) bias identification, and (iii) lead to economic losses. The figures in appendix section A.5 show the effects of the policies in general equilibrium on other outcomes, specifically, on the distribution of agricultural area amongst different agricultural activities, the price of land, the share of labour working in agriculture, and the total working population.

Chapter 7

Conclusion

To what extent do spatially targeted policies are globally effective when considering the possible geographical displacement of environmental damage as a response to the policy? The answer to this question has deep implications for policy design, given the global public good nature of ecosystem conservation. I address this issue in the context of tropical deforestation - an activity that has been responsible for one-fifth of global CO2 emissions in the past two decades - in Brazil - home to a third of the world's remaining rainforests.

The issue of conservation leakage is intrinsically a general equilibrium problem in space: a policy crackdown on deforestation in a spatially delimited region changes economic incentives through a shock to the price of deforested land, generating both sectoral and spatial reallocation. This reallocation depends on the costs of deforestation, the productivity of industries with varying land-intensity, consumers' substitution elasticities between goods and between origins, trade costs, regional amenities and migration frictions. Consequently, reduced-form evaluations conflate deforestation reduction in targeted areas with leakage to non-targeted ones. To separate these two effects, I develop a multi-sector spatial economic model of the Brazilian economy. Given the tight link between deforestation and agriculture, I model agricultural land as the endogenous output of a deforestation sector intermediate to the production of agricultural goods.

The theoretical framework is employed to quantify the global effects on national deforestation of two spatially targeted policies implemented by the Brazilian government over the past decades: the establishment of priority regions with high levels of illegal deforestation that receive extra resources for command and control, and the delimitation of protected areas with high forest coverage where deforestation is entirely banned. I find that both policies are highly effective in reducing global deforestation. Over a period of 12 years, leakage outside of targeted areas undoes between 10% and 15% of the policy impact in targeted areas.

Overall, I show that localised policies can be effective in reducing overall deforestation. This paper's theoretical framework provides a useful starting point to answer other questions of interest that require a spatial general equilibrium model of the relationship between deforestation and agricultural goods. Future research can build on this framework to study, for example, other indirect consequences of deforestation such as its impact on rainfall in nearby areas (Leite-Filho et al., 2021), or the role of technological change in agriculture on environmental conservation.

Bibliography

High and far: Biases in the location of protected areas. *PLoS ONE*, 4(12): e8273, December 2009. URL `https://doi.org/10.1371/journal.pone.0008273`.

J. M. Alix-Garcia, E. N. Shapiro, and K. R. E. Sims. Forest conservation and slippage: Evidence from mexico's national payments for ecosystem services program. *Land Economics*, 88(4):613–638, 2012. ISSN 0023-7639. doi: 10.3368/le.88.4.613. URL `https://le.uwpress.org/content/88/4/613`.

L. J. Alston, G. D. Libecap, and B. Mueller. *Titles, Conflict, and Land Use: The Development of Property Rights and Land Reform on the Brazilian Amazon Frontier*. University of Michigan Press, 1999. ISBN 9780472110063. URL `http://www.jstor.org/stable/10.3998/mpub.16208`.

J. A. Alvarez. The agricultural wage gap: Evidence from brazilian micro-data. *American Economic Journal: Macroeconomics*, 12(1):153–73, January 2020. doi: 10.1257/mac.20170436. URL `https://www.aeaweb.org/articles?id=10.1257/mac.20170436`.

J. E. Anderson. A theoretical foundation for the gravity equation. *The American Economic Review*, 69(1):106–116, 1979. ISSN 00028282. URL `http://www.jstor.org/stable/1802501`.

T. Anderson and C. Hsiao. Formulation and estimation of dynamic models using panel data. *Journal of Econometrics*, 18(1):47–82, 1982. ISSN 0304-4076. doi: https://doi.org/10.1016/0304-4076(82)90095-1. URL `https://www.sciencedirect.com/science/article/pii/0304407682900951`.

D. Arkhangelsky, S. Athey, D. A. Hirshberg, G. W. Imbens, and S. Wager. Synthetic difference-in-differences. *American Economic Review*, 111(12):4088–4118, December 2021. doi: 10.1257/aer.20190159. URL `https://www.aeaweb.org/articles?id=10.1257/aer.20190159`.

P. S. Armington. A theory of demand for products distinguished by place of production (une théorie de la demande de produits différenciés d'après leur origine) (una teoría de la demanda de productos distinguiéndolos según el lugar de producción). *Staff Papers (International Monetary Fund)*, 16(1):159–178, 1969. ISSN 00208027. URL `http://www.jstor.org/stable/3866403`.

J. Assunção and R. Rocha. Getting greener by going black: the effect of blacklisting municipalities on amazon deforestation. *Environment and Development Economics*, 24(2):115–137, 2019. doi: 10.1017/S1355770X18000499.

J. Assunção, C. Gandour, and R. Rocha. Deforestation slowdown in the brazilian amazon: prices or policies? *Environment and Development Economics*, 20(6): 697–722, 2015. doi: 10.1017/S1355770X15000078.

C. A. Balboni, A. Berman, R. Burgess, and B. A. Olken. The economics of tropical deforestation. Working Paper 31410, National Bureau of Economic Research, June 2023. URL `http://www.nber.org/papers/w31410`.

N. Berman, M. Couttenier, A. Leblois, and R. Soubeyran. Crop prices and deforestation in the tropics. *Journal of environmental economics and management*, 119:102819–, 2023. ISSN 0095-0696.

T. Boppart. Structural change and the kaldor facts in a growth model with relative price effects and non-gorman preferences. *Econometrica*, 82(6): 2167–2196, 2014. doi: https://doi.org/10.3982/ECTA11354. URL `https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA11354`.

J. Brner and S. Wunder. Paying for avoided deforestation in the brazilian amazon: from cost assessment to scheme design. *International Forestry Review*, 10(3):496–511, 2008.

J. Buggle, T. Mayer, S. O. Sakalli, and M. Thoenig. The refugee's dilemma: Evidence from jewish migration out of nazi germany. *The Quarterly journal of economics*, 138(2):1273–1345, 2023. ISSN 0033-5533.

R. Burgess, M. Hansen, B. A. Olken, P. Potapov, and S. Sieber. The political economy of deforestation in the tropics. *The Quarterly Journal of Economics*, 127(4):1707–1754, 2012. ISSN 00335533, 15314650. URL `http://www.jstor.org/stable/41812147`.

R. Burgess, F. Costa, and B. A. Olken. The brazilian amazon's double reversal of fortune. 2019.

P. Bustos, B. Caprettini, and J. Ponticelli. Agricultural productivity and structural transformation: Evidence from brazil. *American Economic Review*, 106(6):1320–65, June 2016. doi: 10.1257/aer.20131061. URL https://www.aeaweb.org/articles?id=10.1257/aer.20131061.

D. Clarke, D. Pailañir, S. Athey, and G. Imbens. Synthetic difference in differences estimation, 2023.

CNA and CEPEA. Pib do agronegócio. 2023.

B. R. Copeland and M. S. Taylor. Trade, growth, and the environment. *Journal of Economic Literature*, 42(1):7–71, March 2004. doi: 10.1257/002205104773558047. URL https://www.aeaweb.org/articles?id=10.1257/002205104773558047.

X. Cui. Climate change and adaptation in agriculture: Evidence from us cropping patterns. *Journal of Environmental Economics and Management*, 101:102306, 2020. ISSN 0095-0696. doi: https://doi.org/10.1016/j.jeem.2020.102306. URL https://www.sciencedirect.com/science/article/pii/S0095069620300292.

J. M. C. da Silva, L. C. F. Barbosa, J. Topf, I. C. G. Vieira, and F. R. Scarano. Minimum costs to conserve 80% of the brazilian amazon. *Perspectives in Ecology and Conservation*, 20(3):216–222, 2022.

J. R. de Vasconcelos. Matriz do fluxo de comércio interestadual de bens e serviços no brasil – 1999. Job market paper, August 2001.

R. Dix-Carneiro and B. K. Kovak. Trade liberalization and regional dynamics. *American Economic Review*, 107(10):2908–2946, 2017.

T. Domiguez-Iino. Efficiency and redistribution in environmental policy: An equilibrium analysis of agricultural supply chains. 2021.

D. Donaldson and R. Hornbeck. Railroads and american economic growth: A "market access" approach. *The Quarterly Journal of Economics*, 131(2):799–858, 02 2016. ISSN 0033-5533. doi: 10.1093/qje/qjw002. URL https://doi.org/10.1093/qje/qjw002.

DOU. Lei nº 6.321, de 21 de dezembro de 2007. *Diário Oficial [da] República Federativa do Brasil*, 2007.

DOU. Lei nº 12.651, de 25 de maio de 2012. *Diário Oficial [da] República Federativa do Brasil*, 2012.

F. Eckert and M. Peters. Spatial structural change. Working Paper 30489, National Bureau of Economic Research, September 2022. URL `http://www.nber.org/papers/w30489`.

F. Farrokhi and H. S. Pellegrina. Global trade and margins of productivity in agriculture. Working Paper 27350, National Bureau of Economic Research, June 2020. URL `http://www.nber.org/papers/w27350`.

F. Farrokhi, E. Kang, S. Sotelo, and H. Pellegrina. Deforestation: A global and dynamic perspective. March 2024.

C. Fuller, S. Ondei, B. W. Brook, and J. C. Buettel. First, do no harm: A systematic review of deforestation spillovers from protected areas. *Global Ecology and Conservation*, 18:e00591, 2019. ISSN 2351-9894. doi: https://doi.org/10.1016/j.gecco.2019.e00591. URL `https://www.sciencedirect.com/science/article/pii/S2351989419301143`.

B. Herrendorf, R. Rogerson, and Ákos Valentinyi. Chapter 6 - growth and structural transformation. In P. Aghion and S. N. Durlauf, editors, *Handbook of Economic Growth*, volume 2 of *Handbook of Economic Growth*, pages 855–941. Elsevier, 2014. doi: https://doi.org/10.1016/B978-0-444-53540-5.00006-9. URL `https://www.sciencedirect.com/science/article/pii/B9780444535405000069`.

A. Hsiao. Coordination and commitment in international climate action: Evidence from palm oil. 2021.

IPCC. Ipcc, 2019: Climate change and land: an ipcc special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems. 2019.

S. Jayachandran, J. de Laat, E. F. Lambin, C. Y. Stanton, R. Audy, and N. E. Thomas. Cash for carbon: A randomized trial of payments for ecosystem services to reduce deforestation. *Science*, 357(6348):267–273, 2017. doi: 10.1126/science.aan0568. URL `https://www.science.org/doi/abs/10.1126/science.aan0568`.

D. Lagakos and M. E. Waugh. Selection, agriculture, and cross-country productivity differences. *American Economic Review*, 103(2):948–80, April 2013. doi: 10.1257/aer.103.2.948. URL `https://www.aeaweb.org/articles?id=10.1257/aer.103.2.948`.

A. T. Leite-Filho, B. S. Soares-Filho, J. L. Davis, G. M. Abrahão, and J. Börner. Deforestation reduces rainfall and agricultural revenues in the brazilian amazon. *Nature Communications*, 12(1):2591, 2021. doi: 10.1038/s41467-021-22840-7. URL `https://doi.org/10.1038/s41467-021-22840-7`.

J. Mangonnet, J. Kopas, and J. Urpelainen. Playing politics with environmental protection: The political economy of designating protected areas. *The Journal of Politics*, 84(3):1453–1468, 2022. doi: 10.1086/718978. URL `https://doi.org/10.1086/718978`.

M. Morten and J. Oliveira. The effects of roads on trade and migration: Evidence from a planned capital city. 2018.

H. S. Pellegrina and S. Sotelo. Migration, specialization, and trade: Evidence from brazil's march to the west. Working Paper 28421, National Bureau of Economic Research, January 2021. URL `http://www.nber.org/papers/w28421`.

A. Pfaff and J. Robalino. Spillovers from conservation programs. *Annual Review of Resource Economics*, 9(1):299–315, 2017. doi: 10.1146/annurev-resource-100516-053543. URL `https://doi.org/10.1146/annurev-resource-100516-053543`.

J. Robalino, A. Pfaff, and L. Villalobos. Heterogeneous local spillovers from protected areas in costa rica. *Journal of the Association of Environmental and Resource Economists*, 4(3):pp. 795–820, 2017. ISSN 23335955, 23335963. URL `https://www.jstor.org/stable/26544527`.

E. Souza-Rodrigues. Deforestation in the Amazon: A Unified Framework for Estimation and Policy Analysis. *The Review of Economic Studies*, 86(6):2713–2744, 12 2018. ISSN 0034-6527. doi: 10.1093/restud/rdy070. URL `https://doi.org/10.1093/restud/rdy070`.

D. Szerman, J. Assunção, M. Lipscomb, and A. M. Mobarak. Agricultural productivity and deforestation: Evidence from brazil. Working Paper 1091, Yale Economic Growth Center, January 2022. URL `https://elischolar.library.yale.edu/egcenter-discussion-paper-series/1091`.

Appendix A

Data appendix

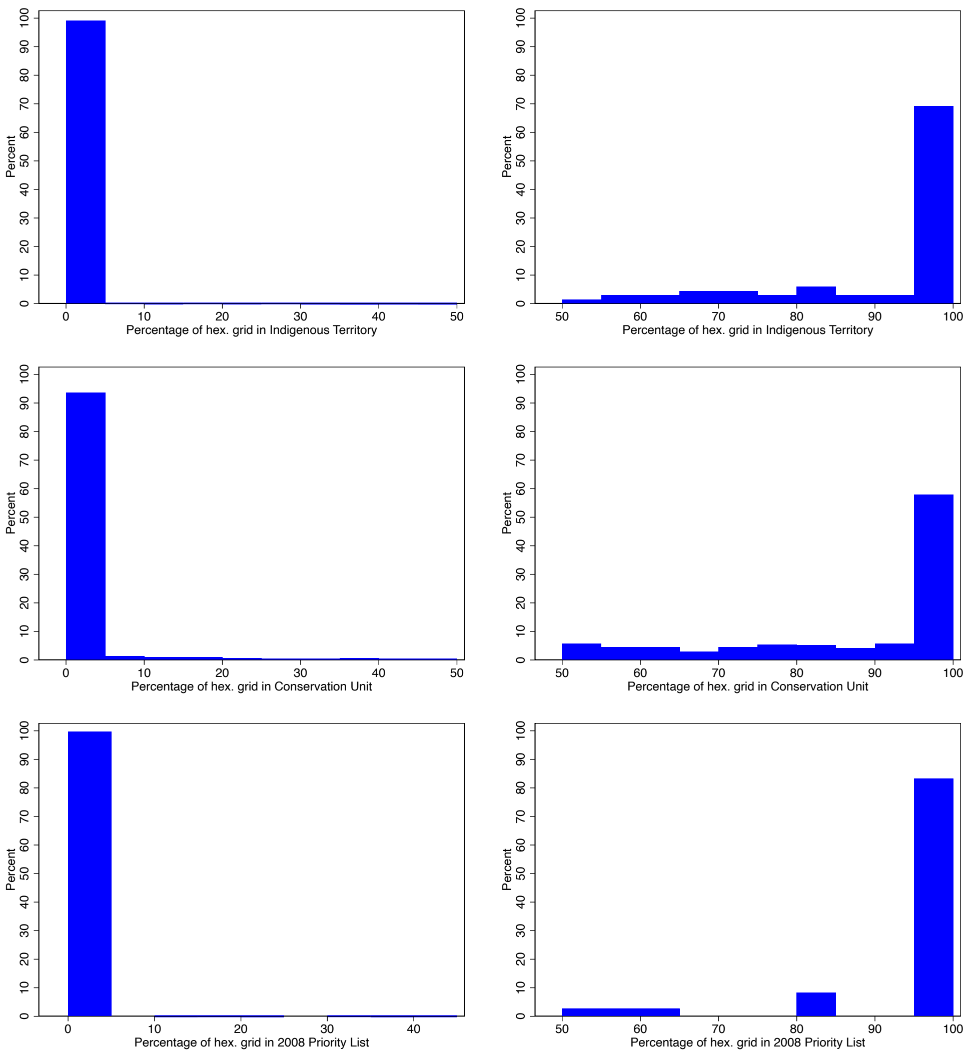
A.1 Hexagonal grid approximation

For the purposes of the reduced-form analysis, the boundaries of the protected areas and municipalities are discretised so that the grid cells can be categorised as within or outside an Indigenous Territory, Conservation Unit, or Priority List municipality. In order to see the extent to which this leads to imprecise assignments, I look at the distribution of the percentage of hexagons that in either of these three categories. Since these categories change over time, I have to consider a specific time period for this analysis. For the priority municipalities I consider the list in 2008, the results are very similar when looking at the set of municipalities in other years. For the protected areas, I consider the latest set of protected areas in my data only. What the table shows is that the vast majority of the grid cells are either entirely inside or entirely outside of the conservation policy area, for all three policies.

The histograms in A.1 and A.2 show the distributions conditional on being categorised as treated or untreated for each of the 3 treatments. Given the higher precision when restricting attention to the Legal Amazon, especially when it comes to the potential false positives, I prefer the analysis restricted to the Legal Amazon.

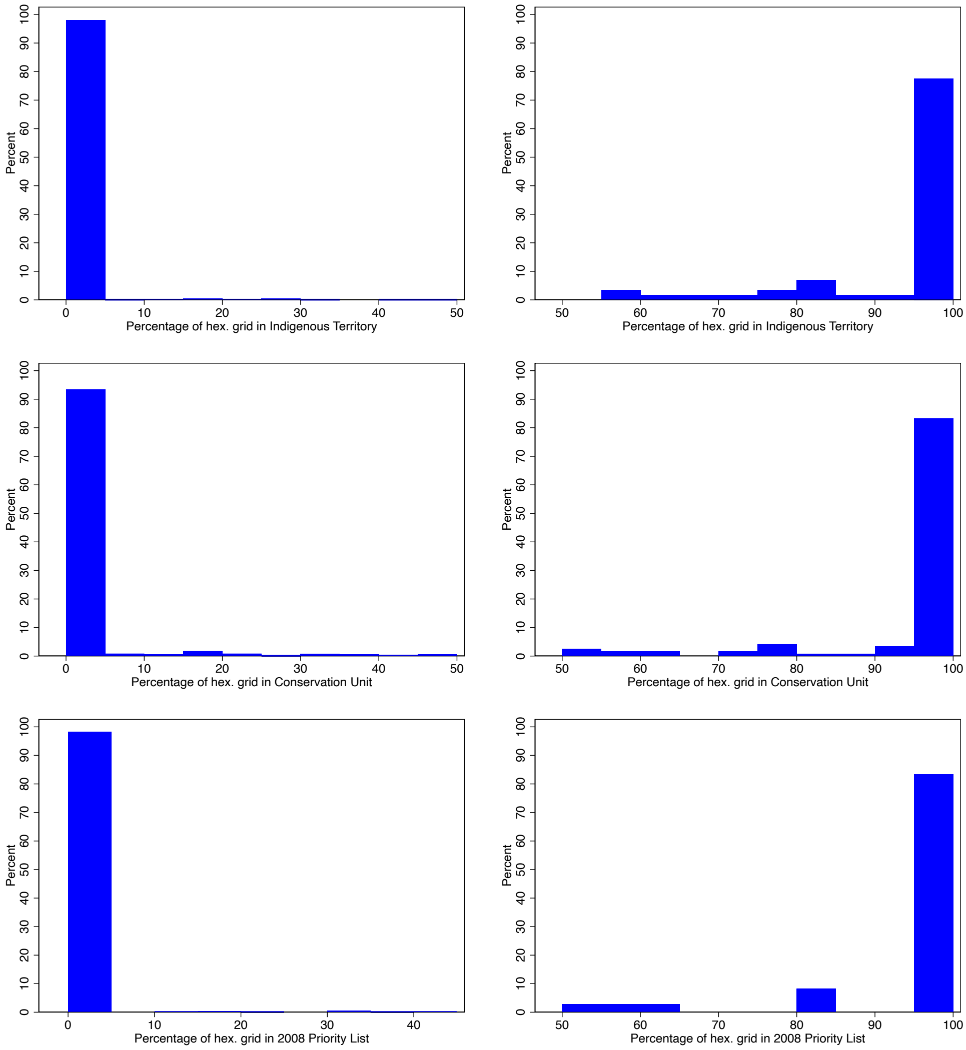
A.2 Land use data

Figure A.1: Overlap analysis of hexagonal grid and conservation policies, by assigned treatment status, considering all of Brazil



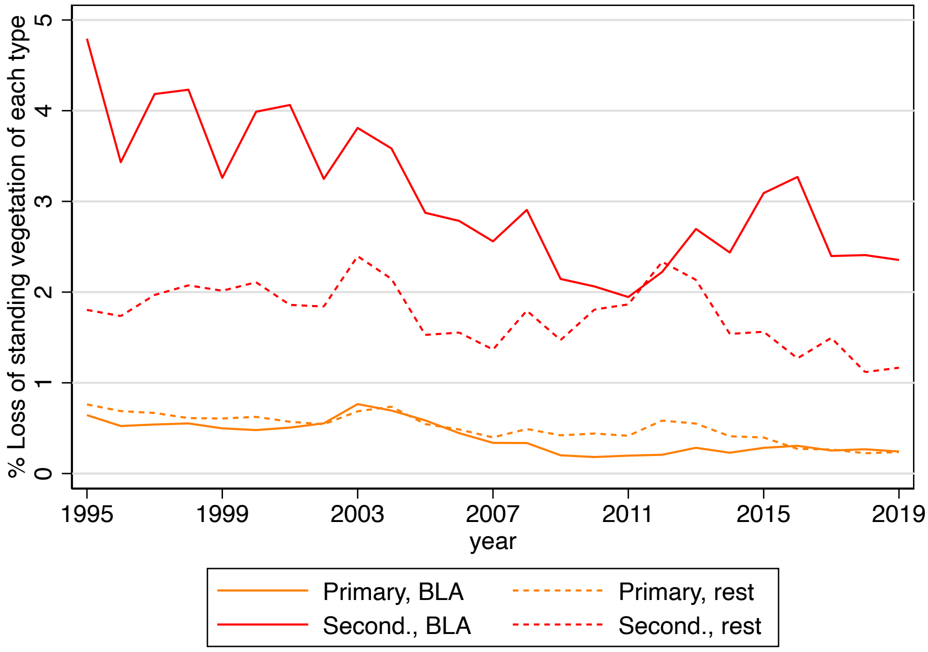
Note: This figure displays the distribution of the percentage of hexagonal grid cells that falls within: (i) an Indigenous Territory (first row), (ii) a Conservation Unit (second row), and (iii) a Priority List municipality. The first column considers those that are labelled as outside each of those three types of polygons and the second column considers. Since I will categorise them as either within (if the percentage within exceeds 50%) or outside (otherwise), this figure helps to understand how accurate the 10 km-wide hexagon discretisation is.

Figure A.2: Overlap analysis of hexagonal grid and conservation policies, by assigned treatment status, considering only the Brazilian Legal Amazon



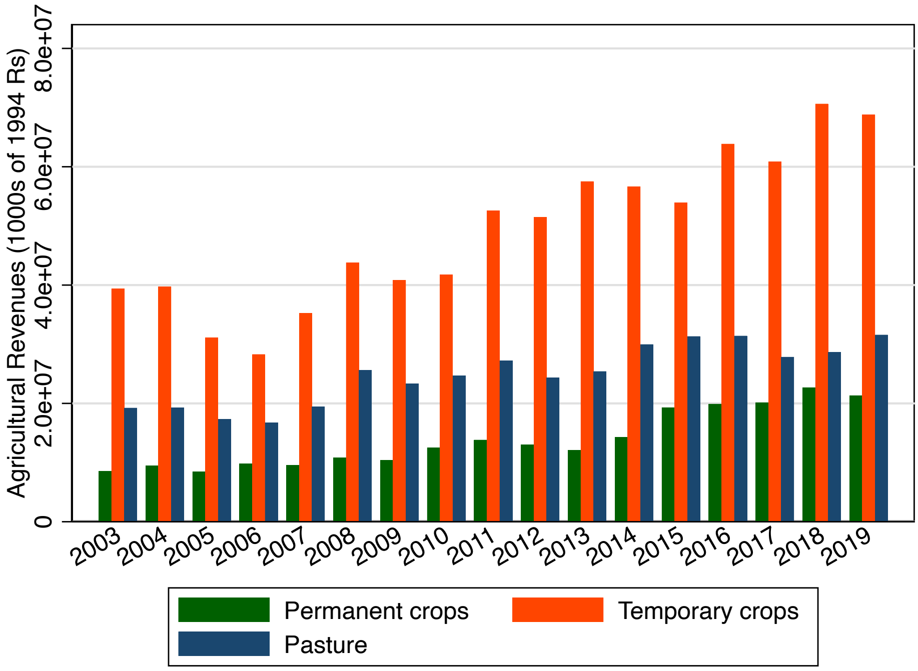
Note: This figure displays the distribution of the percentage of hexagonal grid cells that falls within: (i) an Indigenous Territory (first row), (ii) a Conservation Unit (second row), and (iii) a Priority List municipality. The first column considers those that are labelled as outside each of those three types of polygons and the second column considers. Since I will categorise them as either within (if the percentage within exceeds 50%) or outside (otherwise), this figure helps to understand how accurate the 10 km-wide hexagon discretisation is. This figure considers only hexagons in municipalities within the Legal Amazon.

Figure A.3: Yearly land use change transitions (percentage of remaining)



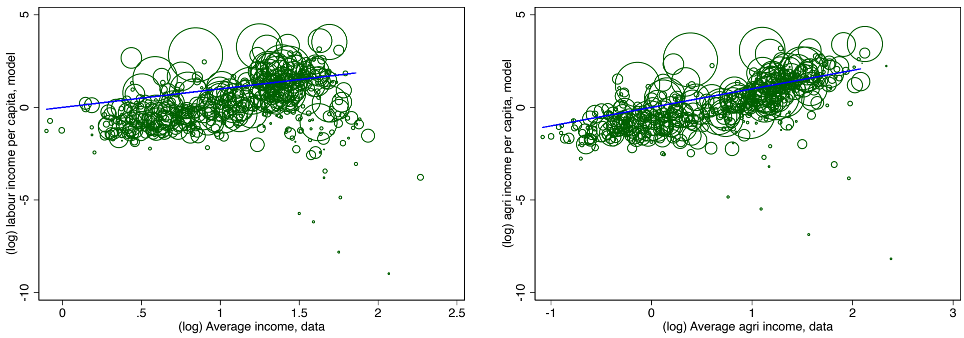
Note: This figure shows the evolution of yearly primary and secondary vegetation loss as a percentage of the standing vegetation of each type split by Brazilian Legal Amazon and the rest of Brazil. The orange-yellow lines indicate the percentage losses of primary vegetation cover loss each year and the red lines indicate the percentage of secondary vegetation cover loss. The continuous lines indicate the vegetation loss in the Legal Amazon and the dotted lines in the rest of Brazil.

Figure A.4: Yearly agricultural revenues by ag. activity



Note: This figures shows Brazil-wide agricultural revenue from 2003 to 2019 decomposed into: revenue from permanent crops, revenue from temporary crops, and revenue from pastures (assuming they are all cattle-grazing. The methodology used in described in section 2.1.3.

Figure A.5: Model validation: wages



(a) All sectors

(b) Agriculture only

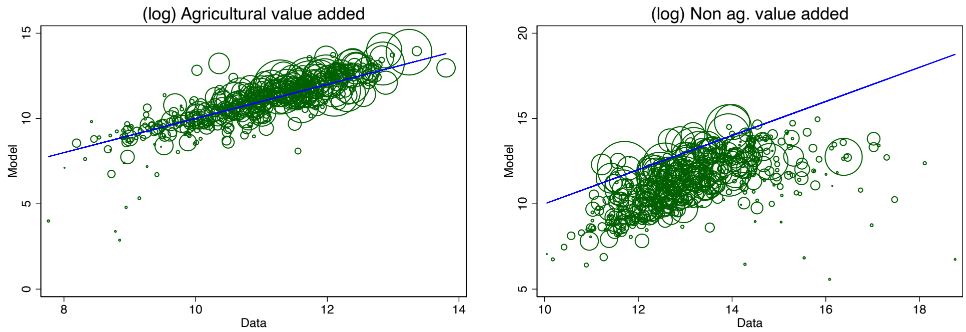
Note: This figure plots the log of average labour income per capita as estimated in the model in 2010 against the log of the average income per capita as measured in the 2010 population census. Panel (a) includes all incomes and panel (b) focuses exclusively on the comparison of agricultural incomes. Each green circle represents a municipality and the size of the circle indicates the level of deforestation in 2010. This is so that we can assess the accuracy of the model for the regions that are most relevant to our main outcome of interest. The blue line is the 45 degree line.

A.3 Model validation

A.4 Counterfactual Analysis

A.5 Other outcomes

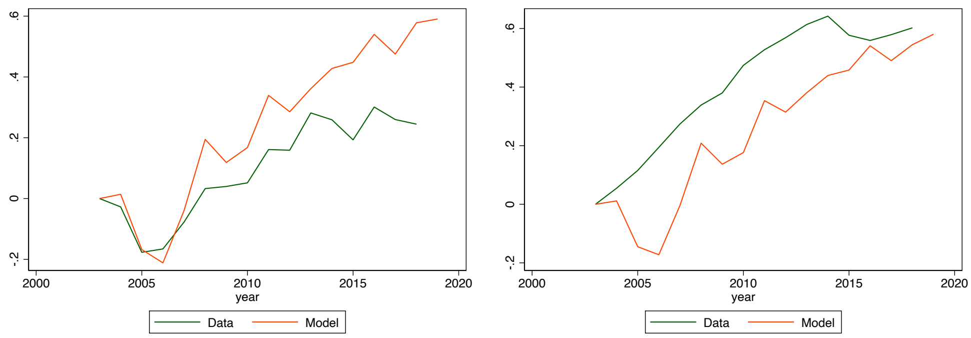
Figure A.6: Model validation: value added



(a) Agriculture (b) Non Agriculture

Note: This figure plots the log of total revenue as estimated in the model in 2010 against the log of the value added per sector and micro-region as measured in the system of regional accounts described in section 2.1.3. Panel (a) looks at total agricultural revenue/value added and panel (b) looks at the non-agricultural sectors. Each green circle represents a municipality and the size of the circle indicates the level of deforestation in 2010. This is so that we can assess the accuracy of the model for the regions that are most relevant to our main outcome of interest. The blue line is the 45 degree line.

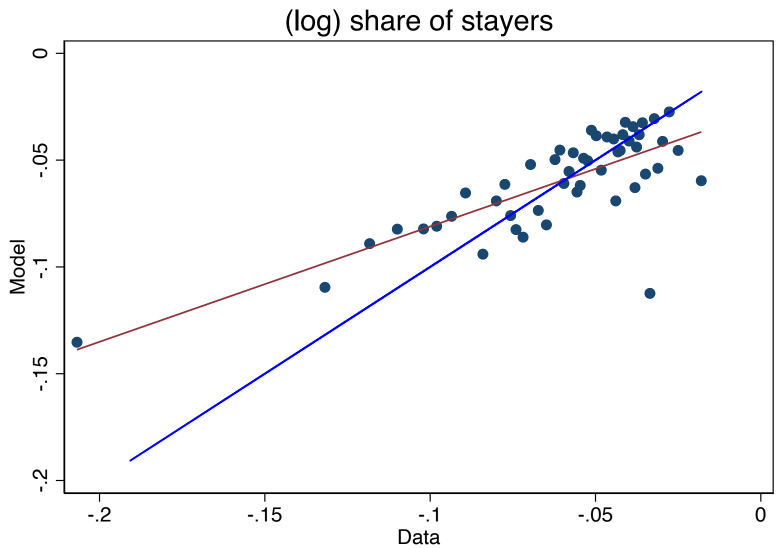
Figure A.7: Model validation: value added changes



(a) Agriculture (b) Non Agriculture

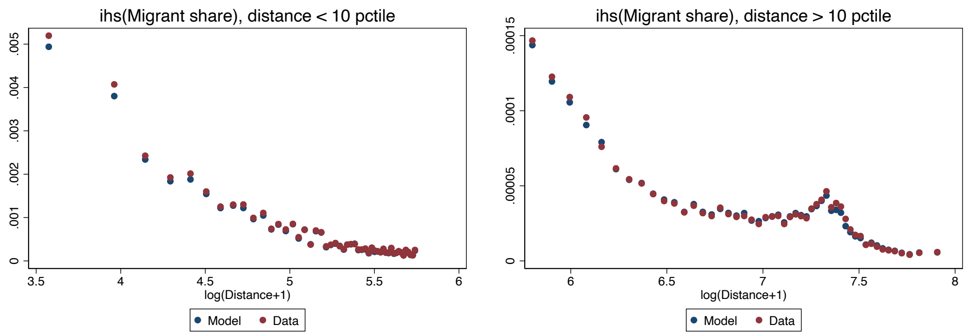
Note: This figure plots the percentage changes in revenue as estimated by the model (orange line) alongside the percentage changes in value added as measured in the system of regional accounts described in section 2.1.3. Panel (a) looks at total agricultural revenue/value added and panel (b) looks at the non-agricultural sectors.

Figure A.8: Model validation: migration share



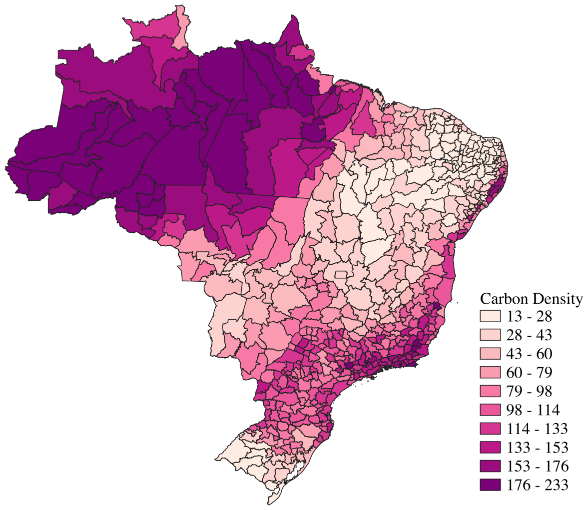
Note: This figure plots a binned scatter plot (with 50 bins) of the log of the share of people who did not migrate in the past years in the model against the analogous in the data. The data used is the 2010 census, which asks for the municipality of residence 5 years ago. We coarsen the data to the micro-region level and calculate the share of people who, 5 years ago, lived in the same micro-region as at the time of the census.

Figure A.9: Model validation: bilateral migration shares as a function of distance



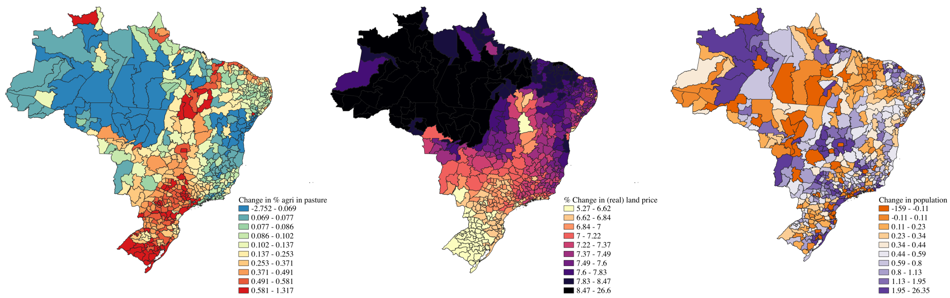
Note: This figure plots the bilateral migration shares (in the model and in the data) between micro-regions against the distance between the municipalities. they are overlayed binned scatter plots (with 50 bins) at the micro-region-pair level. More precisely, they plot the inverse hyperbolic sine of migration shares against the log of distance. The red dots represent the migration shares in the data and the blue dots in the model. the graph on the left restricts attention to micro-region-pairs whose distance is below the 10th percentile, which have much larger migration shares, and the figure on the right restricts attention to the micro-region-pairs with distances above the 10th percentile.

Figure A.10: Carbon Density



Note: This figure maps the distribution of the storage of above- and below-ground biomass carbon density of areas classified as natural ecosystems across Brazil. The unit of analysis is tons of carbon per hectare. Source is described in 2.1.2.

Figure A.11: Change in other outcomes: Priority List



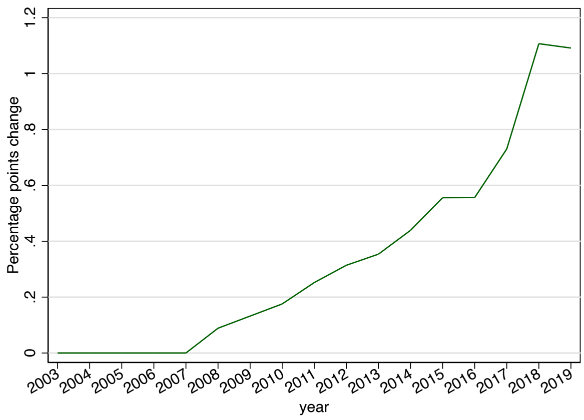
(a) Change in % of ag land in pasture (b) % Change in price of land (c) Change in population

Note: This figure maps the effect of the Priority List policy three outcomes: percentage of the area in pastures, the price of agricultural land, and population. This is calculated by comparing the data to the no-policy counterfactual simulated in the model. Panel (a) shows the spatial distribution of changes in the percentage of agricultural area in pastures as opposed to crops. There seems to be a displacement of pasture land away from the North and the North East and towards the South, where pastures are more productive (both in terms of yields and profitability). Panel (b) shows the changes in the price of agricultural land. Overall land prices increase, consistent with a reduced supply of new agricultural land, but they do so more near Priority-Listed regions. Panel (c) shows changes in population. Here the pattern is less clear but it seems like population decreases in Priority-Listed regions.

Figure A.12: Change in % area in ag. activities: Priority List

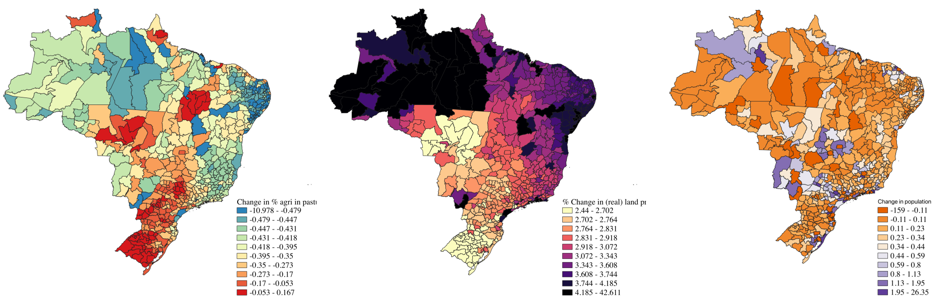
Note: This figure displays the % changes in the total area dedicates to different agricultural activities caused by the Priority List policy. This is calculated by comparing the data to the no-policy counterfactual simulated in the model. Overall there are only small percent increases in the area in permanent crops, a more considerable but modest increase in the area in temporary crops (they increase by 0.4% by 2019) and a similar percentage decrease in the area in pastures.

Figure A.13: Change in % ag. labour: Priority List



Note: This figure displays the changes in the percentage of workers in the agricultural sector caused by the Priority List policy. This is calculated by comparing the data to the no-policy counterfactual simulated in the model.

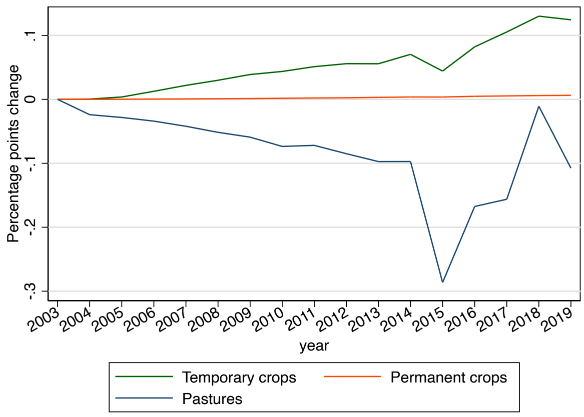
Figure A.14: Change in other outcomes: Protected Areas



(a) Change in % of ag land in pasture (b) % Change in price of land (c) Change in population

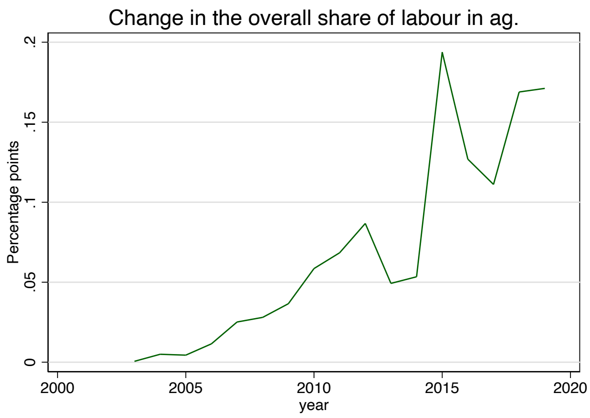
Note: This figure maps the effect of the establishment of Protected Areas (2003-2019) on three outcomes: percentage of the area in pastures, the price of agricultural land, and population. This is calculated by comparing the data to the no-policy counterfactual simulated in the model. Panel (a) shows the spatial distribution of changes in the percentage of agricultural area in pastures as opposed to crops. Panel (b) shows the changes in the price of agricultural land. Panel (c) shows changes in population.

Figure A.15: Change in % area in ag. activities: Protected Areas



Note: This figure displays the % changes in the total area dedicates to different agricultural activities caused by the establishment of Protected Areas (2003-2019) policy. This is calculated by comparing the data to the no-policy counterfactual simulated in the model. Overall there are only small percent increases in the area in permanent crops, a more considerable but modest increase in the area in temporary crops (they increase by 0.1% by 2019) and a similar percentage decrease in the area in pastures.

Figure A.16: Change in % ag. labour: Protected Areas



Note: This figure displays the changes in the percentage of workers in the agricultural sector caused by the establishment of Protected Areas (2003-2019). This is calculated by comparing the data to the no-policy counterfactual simulated in the model.

Appendix B

Mathematical appendix

B.1 Deriving agricultural expenditures

Let the joint distribution of pairs of individual productivities (z_{ir}^A, z_{ir}^{NA}) is given by the Frank copula as in Lagakos and Waugh (2013) with parameters $(\chi^A, \chi^{NA}, \rho)$. Then there are no simple closed-form expressions for the share of employment in agriculture and the labour income in each sector. Instead,

$$s_r^A = Pr\big[z_{ir}^{NA}/z_{ir}^A \leq w_r^A/w_r^{NA_r}\big] = \int_0^\infty \int_0^{z^A w_r^A/w_r^{NA}} f(z^A, z^{NA})\, dz^{NA} dz^A$$

$$\overline{y}_r^{LA} = w_r^A \mathbb{E}\big[z_{ir}^A \big| z_{ir}^{NA}/z_{ir}^A \leq w_r^A/w_r^{NA_r}\big] = w_r^A \frac{1}{s_r^A} \int_0^\infty \int_0^{z^A w_r^A/w_r^{NA}} z^A f(z^A, z^{NA})\, dz^{NA} dz^A$$

$$\overline{y}_r^{LNA} = w_r^{NA} \mathbb{E}\big[z_{ir}^{NA} \big| z_{ir}^{NA}/z_{ir}^A > w_r^A/w_r^{NA_r}\big] = w_r^{NA} \frac{1}{s_r^{NA}} \int_0^\infty \int_{z^A w_r^A/w_r^{NA}}^\infty z^{NA} f(z^{NA}, z^{NA})\, dz^{NA} dz^A$$

Expenditure share in agriculture. The share of consumer expenditure in agriculture of a household with expenditure e_i equals

$$\vartheta_{ir}^A = \phi + \nu\Big(\frac{p_r^A}{p_r^{NA}}\Big)\Big(e_{ir}\Big)^{-\eta} p_r^\eta. \tag{B.1}$$

Thus, the total agricultural expenditure in region r equals

$$X_r^A = \phi p_r I_r^D + \int \vartheta_{ir}^A e_{ir}\, dG(i) = \phi(E_r + p_r I_r^D) + \nu\Big(\frac{p_r^A}{p_r^{NA}}\Big) p_r^\eta \int \big(e_{ir}\big)^{1-\eta} dG(i),$$

Assume that people with higher labour incomes get proportionately higher land rents and also spend proportionately more on deforestation. Define the ratio between land rents and labour income from the aggregates, a_r, and the ratio of the total income spent on deforestation investments, b_r.

$$a_r \equiv \frac{v_r T_r^A}{\int_i y_{ir}^L dG(i)}, \quad b_r \equiv \frac{p_r I_r^D}{\int_i y_{ir}^L dG(i) + v_r T_r^A}.$$

Then the assumption of proportional land rents and deforestation expenditures can be expressed formally as:

$$(1)\, y_{ir}^T = a_r y_{ir}^L, \quad (2)\, x_{ir}^d = b_r y_{ir} \implies (3)\, \frac{e_{ir}}{y_{ir}^L} = (1+a_r)(1-b_r) \equiv m_r$$

$$e_{ir} = \underbrace{\max\{w_r^{NA} z_{ir}^{NA}, w_r^A z_{ir}^A\}}_{y_{ir}^L} m_r,$$

where

$$m_r = \frac{E_r}{L_r \overline{y}_r^L}, \quad \overline{y}_r^L = s_r^A y_r^{LA} + s_r^{NA} y_r^{LNA} = \mathbb{E}(y_{ir}^L)$$

For the second term of the total agricultural expenditure equation, we compute the following integral

$$\begin{aligned}
\int (e_{ir})^{1-\eta} dG(i) =& (m_r)^{1-\eta} L_r \mathbb{E}\big((y_{ir}^L)^{1-\eta}\big) \\
\mathbb{E}\big((y_{ir}^L)^{1-\eta}\big) =& \Bigg(\int_0^\infty \int_0^{z^A w_r^A / w_r^{NA}} (w_r^A z^A)^{1-\eta} f(z^A, z^{NA})\, dz^{NA} dz^A \\
&+ \int_0^\infty \int_{z^A w_r^A / w_r^{NA}}^\infty (w_r^{NA} z^{NA})^{1-\eta} f(z^A, z^{NA})\, dz^{NA} dz^A \Bigg)
\end{aligned}$$

Thus, the aggregate local share of consumer expenditure in agriculture equals

$$\vartheta_r^A = \phi + \nu \Big(\frac{p_r^A}{p_r^{NA}}\Big) \Big(\frac{E_r}{p_r L_r}\Big)^{-\eta} \frac{\mathbb{E}\big((y_{ir})^{1-\eta}\big)}{\big(\mathbb{E}(y_{ir})\big)^{1-\eta}}, \tag{B.2}$$

the total agricultural expenditure equals

$$X_r^A = \vartheta_r^A E_r + \phi p_r I_r^D \tag{B.3}$$

Because of the CES preferences between agricultural goods, the expenditure in agricultural good k equals

$$X_r^{Ak} = X_r^A \left(\frac{p_r^{Ak}}{p_r^A}\right)^{1-\theta}. \tag{B.4}$$

B.2 Equilibrium equations for simulation

Start with a guess (w_r, v_r).
Get prices of regional varieties at origin

$$p^{NAr} = \frac{w_r^{NA}}{Z_r^{NA}}, \quad p^{Akr} = \frac{1}{Z_r^{Ak}} \left(\frac{w_r^A}{1-\alpha_k}\right)^{1-\alpha_k} \left(\frac{v_r}{\alpha_k}\right)^{\alpha_k} \tag{B.5}$$

Destination prices and trade shares for final goods

$$p_d^s = \left(\sum_{r=1}^{R} (\tau_d^r p^{sr})^{1-\sigma}\right)^{\frac{1}{1-\sigma}} \quad \forall s \in \{A \times \{1, \ldots, K\}, NA\} \tag{B.6}$$

$$\pi_d^{so} = \left(\frac{\tau_d^o p^{so}}{p_d^s}\right)^{1-\sigma}. \tag{B.7}$$

Composite goods price indices in region r

$$p_r^A \equiv \left(\sum_{k=1}^{K} (p_r^{Ak})^{1-\theta}\right)^{\frac{1}{1-\theta}}, \quad p_r = (p_r^A)^{\phi} (p_r^{NA})^{1-\phi} \tag{B.8}$$

Optimal level of deforestation

$$T_r^D = \min\left\{ (Z_r^D)^{\frac{1}{1-\delta}} \left(\frac{\delta q_r}{p_r}\right)^{\frac{\delta}{1-\delta}}, T_r^N \right\} \tag{B.9}$$

Labour shares and labour income

$$s_r^A = Pr\left[z_{ir}^{NA}/z_{ir}^A \leq w_r^A/w_r^{NA_r}\right] \tag{B.10}$$

$$\overline{y}_r^{LA} = w_r^A \mathbb{E}\left[z_{ir}^A \middle| z_{ir}^{NA}/z_{ir}^A \leq w_r^A/w_r^{NA_r}\right] \tag{B.11}$$

$$\overline{y}_r^{LNA} = w_r^{NA} \mathbb{E}\left[z_{ir}^{NA} \middle| z_{ir}^{NA}/z_{ir}^A > w_r^A/w_r^{NA_r}\right] \tag{B.12}$$

Local expenditure

$$E_r = \overline{y}_r^L L_r + v_r T_r^A - p_r I_r^D, \quad \overline{y}_r^L = s_r^A \overline{y}_r^{LA} + (1-s_r^A) \overline{y}_r^{LNA} \tag{B.13}$$

Optimal consumption share in agriculture given final goods prices and local consumption

$$\vartheta_r^A = \phi + \nu\Big(\frac{p_r^A}{p_r^{NA}}\Big)\Big(\frac{E_r}{L_r}\Big)^{-\eta} p_r^{\eta}\lambda_r. \tag{B.14}$$

where $\lambda_r \equiv \frac{\mathbb{E}\left((y_{ir})^{1-\eta}\right)}{\left(\mathbb{E}(y_{ir})\right)^{1-\eta}}$ Thus the overall expenditure in agriculture, including for investment, equals

$$X_r^A = \vartheta_r^A E_d + \phi p_r I_r^D, \tag{B.15}$$

and the expenditure in agricultural good k equals

$$X_r^{Ak} = X_r^A\Big(\frac{p_r^{Ak}}{p_r^A}\Big)^{1-\theta}, \tag{B.16}$$

Goods market clearing

$$\begin{cases} \lambda(s_r^A)^{-\frac{1}{\chi}} s_r^{Ak} L_r w_r^A & = \sum_{d=1}^R (1-\alpha_k)\pi_d^{Akr} X_d^{Ak} \quad \forall k \\ \lambda(1-s_r^A)^{\frac{\chi-1}{\chi}} L_r w_r^{NA} & = \sum_{d=1}^R \pi_d^{NAr} X_d^{NA} \end{cases} \tag{B.17}$$

Which could be rewritten as the following two equations instead:

$$\begin{cases} \overline{y}_r & = \frac{1}{L_r}\sum_{d=1}^R \left[\sum_{k=1}^K (1-\alpha_k)\pi_d^{Akr} X_d^{Ak} + \pi_d^{NAr} X_d^{NA}\right] \\ v_r & = \frac{1}{T_r^A}\sum_{k=1}^K \sum_{d=1}^R \alpha_k \pi_d^{Akr} X_d^{Ak} \end{cases} \tag{B.18}$$

Set numeraire

$$(p^A)^{\phi}(p^{NA})^{1-\phi} = 1 \tag{B.19}$$

where

$$p^A \equiv \Big(\sum_{k=1}^K (p^{Ak})^{1-\theta}\Big)^{\frac{1}{1-\theta}}.$$

and

$$p^s \equiv \Big(\sum_{r=1}^R (p^{sr})^{1-\sigma}\Big)^{\frac{1}{1-\sigma}} \quad \forall s \in \{A \times \{1,\ldots,K\}, NA\}.$$

Migration equilibrium

$$L_d = \sum_{r=1}^R \rho_{rd} L_{0r}. \tag{B.20}$$

where

$$\rho_{od} = \frac{(V(e_d,p_d)\mu_{od}B_d)^{\epsilon}}{\sum_{r=1}^R (V(e_r,p_r)\mu_{or}B_r)^{\epsilon}}. \tag{B.21}$$

B.3 Model inversion

Given all production and preference parameters, the distribution of population, initial land endowments, and K+3 vectors of observable endogenous quantities

$$\{s_r^A, T_r^D, p^{NAr}Y_r^{NA}, \{T_r^{Ak}, p^{Akr}Y_r^{Ak}\}_k\}_r,$$

below I formulate the system of equations that need to be solved in order to find the regional TFPs and amenities.

The Cobb-Douglas form of agricultural income of a each commodity k in region r yields

$$p^{Akr}Y_r^{Ak} = \frac{v_r T_r^{Ak}}{\alpha_k} \tag{B.22}$$

Adding up over all k,

$$v_r = \frac{1}{T_r^A}\sum_k \alpha_k p^{Akr}Y_r^{Ak}. \tag{B.23}$$

GDP accounting

$$\sum_k p^{Akr}Y_r^{Ak} + p_r^{NA}Y_r^{NA} = L_r\overline{y}_r^L + v_r T_r^A \tag{B.24}$$

so

$$\overline{y}_r^L = \frac{1}{L_r}\Big(\sum_k p^{Akr}Y_r^{Ak} + p_r^{NA}Y_r^{NA} - v_r T_r^A) \tag{B.25}$$

Having the regional prices at origin (start with a guess), we can get goods prices at destination

$$p_d^s = \Big(\sum_{r=1}^{R}(\tau_d^r p^{sr})^{1-\sigma}\Big)^{\frac{1}{1-\sigma}} \tag{B.26}$$

$$p_r^A = \Big(\sum_{k=1}^{K}(p^{Akr})^{1-\theta}\Big)^{\frac{1}{1-\theta}} \tag{B.27}$$

$$p_r = (p_r^A)^{\phi}(p_r^{NA})^{1-\phi} \tag{B.28}$$

Land prices

$$q_r = \frac{1}{1-\beta(1-\rho)}v_r \tag{B.29}$$

Deforestation investment

$$p_r I_r^D = \delta q_r T_r^D \tag{B.30}$$

Local household expenditure

$$E_r = v_r T_r^A + \overline{y}_r^L L_r - p_r I_r^D \tag{B.31}$$

Local expenditure shares in agriculture

$$\vartheta_r^A = \phi + \nu \left(\frac{p_r^A}{p_r^{NA}} \right) \left(\frac{E_r}{p_r L_r} \right)^{-\eta} \lambda_r \tag{B.32}$$

where $\lambda_r \equiv \frac{\mathbb{E}\left((y_{ir})^{1-\eta}\right)}{\left(\mathbb{E}(y_{ir})\right)^{1-\eta}}$ Regional expenditures in each sector

$$X_d^A \equiv \vartheta_d^A E_d + \phi p_d I_d^D, \quad X_d^{NA} \equiv (1 - \vartheta_d^A) E_d + (1 - \phi) p_d I_d^D$$

$$X_d^{Ak} = X_d^A \Big(\frac{p_d^{Ak}}{p_d^A} \Big)^{1-\theta}$$

Rewrite trade shares as

$$\pi_d^{so} = (\tau_d^o p^{so})^{1-\sigma} (p_d^s)^{\sigma - 1}$$

in order to solve for updated origin prices (for contraction mapping) in the market clearing equations (B.18)

$$T_r^{Ak} v_r = \sum_{d=1}^{R} \alpha_k (\tau_d^r p^{Akr})^{1-\sigma} (p_d^{Ak})^{\sigma-1} X_d^{Ak} \tag{B.33}$$

$$= (p^{Akr})^{1-\sigma} \sum_{d=1}^{R} \alpha_k (\tau_d^r)^{1-\sigma} (p_d^{Ak})^{\sigma-1} X_d^{Ak}. \tag{B.34}$$

Solving for p^{Akr},

$$p^{Akr} = \Big(\frac{T_r^{Ak} v_r}{\alpha_k} \Big)^{\frac{1}{1-\sigma}} \Big(\sum_{d=1}^{R} (\tau_d^r)^{1-\sigma} (p_d^{Ak})^{\sigma-1} X_d^{Ak} \Big)^{\frac{1}{\sigma-1}}. \tag{B.35}$$

Similarly, for the non-agricultural sector,

$$p^{NAr} = \big(\overline{y}_r^L (1 - s_r^A) L_r \big)^{\frac{1}{1-\sigma}} \Big(\sum_{d=1}^{R} (\tau_d^r)^{1-\sigma} (p_d^{NA})^{\sigma-1} X_d^{NA} \Big)^{\frac{1}{\sigma-1}}. \tag{B.36}$$

To back out the Z_r^s, first solve for w_r^A and w_r^{NA},

$$w_r^A = \lambda^{-1} \overline{y}_r^L (s_r^A)^{1/\chi}, \quad w_r^{NA} = \lambda^{-1} \overline{y}_r^L (1 - s_r^A)^{1/\chi} \tag{B.37}$$

Then

$$Z_r^{Ak} = \frac{1}{p_r^{Ak}} \Big(\frac{w_r^A}{1-\alpha_k}\Big)^{1-\alpha_k} \Big(\frac{v_r^A}{\alpha_k}\Big)^{\alpha_k} \tag{B.38}$$

$$Z_r^{NA} = \frac{1}{p_r^{NA}} w_r^{NA} \tag{B.39}$$

Aggregate market clearing. Because of equation (B.17) and the fact that the trade shares from all origins add up to 1 for each destination, it follows that

$$\sum_{r=1}^{R} L_r(1-s_r^A)y_r^{LNA} = \sum_{r=1}^{R}\sum_{d=1}^{R} \pi_d^{NAr} X_d^{NA} = \sum_{d=1}^{R} X_d^{NA} \tag{B.40}$$

$$= \sum_{d=1}^{R} \Big[(1-\phi)p_d Y_d - \nu E_d^{1-\eta}\Big(\frac{p_d^A}{p_d^{NA}}\Big)(p_d L_d)^{\eta} \frac{\mathbb{E}\big((y_{id})^{1-\eta}\big)}{\big(\mathbb{E}(y_{id})\big)^{1-\eta}}\Big] \tag{B.41}$$

And if we decompose region-sector consumer price indices p_d^s as the product of a Brazil-wide sectoral price index $p^s \equiv \Big(\sum_{o=1}^{R}(p^{so})^{1-\sigma}\Big)^{\frac{1}{1-\sigma}}$ and a region specific component $\tilde{p}_d^s \equiv p_d^s/p^s$, then we can solve for a Brazil-wide relative price index of agriculture:

$$\frac{p^A}{p^{NA}} = \Bigg(\Big[\sum_{d=1}^{R}(1-\phi)p_d Y_d - \sum_{r=1}^{R} L_r(1-s_r^A)y_r^{LNA}\Big]\Big/\Big[\sum_{d=1}^{R} \nu E_d^{1-\eta}\Big(\frac{\tilde{p}_d^A}{\tilde{p}_d^{NA}}\Big)(p_d L_d)^{\eta}\frac{\mathbb{E}\big((y_{id})^{1-\eta}\big)}{\big(\mathbb{E}(y_{id})\big)^{1-\eta}}\Big]\Bigg) \tag{B.42}$$