Historical Indigenous Extinctions and Modern Deforestation in the Amazon

Bruno Barsanetti

FGV EPGE, Rio de Janeiro – RJ, 22250-145, Brazil

Alípio Ferreira

Economics Department, Southern Methodist University, Dallas – TX, 75206, USA

September, 2025

Abstract

We study the long-run determinants of Amazon deforestation by linking historical Indigenous extinctions to modern land use. Using the 1944 Ethnohistorical Map of Curt Nimuendajú and satellite-based deforestation data, we show that areas near pre-1944 extinctions experienced 26% more forest loss by 2010. The correlation is robust to extensive controls and alternative specifications. The association is stronger for extinctions attributed to violence or disease, which created demographic voids, but is absent for those attributed to assimilation. Our results are consistent with the hypothesis that historical Indigenous extinctions facilitated modern deforestation by reducing the deterrence of traditional populations against frontier occupation.

JEL Classification Codes: Q23, R11, J15.

Keywords: deforestation, Indigenous populations, history dependence.

[1]We appreciate the research assistance provided by Alan Gayger, Matheus Dutra, Lilian Kingston Freitas, and Alexandre Portugal. We thank Jennifer Alix-Garcia, Rafael Araújo, Mario Cannella, Jeffrey Lin, Ömer Özak, Marcos Salgado, and Katharine Sims for their comments. We are also grateful for seminar audiences at FGV EPGE, the 2023 SSUN Annual Conference, SMU, Insper, CPI-RJ, the 2023 LACEA-EHN Workshop on Historical Development, MAP, UFRJ, the 2023 LACEA-LAMES Meeting, UFPE, the 2023 Ridge Forum Workshop in Environmental Economics, the TrEnCE Working Group, UFRGS, the 14th Toulouse Conference on the Economics of Energy and Climate, the 5th Arthur Lewis Lab Conference, and the 2024 UEA North American Meeting. Bruno Barsanetti appreciates funding for this study by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. All errors are our own. Replication data and code are available at 10.5281/zenodo.17153102.

1 Introduction

Deforestation in tropical forests is a major global concern impacting greenhouse gas emissions (Mitchard 2018), biodiversity (Alroy 2017), and regional climates (Maeda et al. 2021). The Amazon, the world's largest tropical forest, has undergone rapid environmental change in recent decades due to human activities (Albert et al. 2023). While short-term drivers of Amazon deforestation, such as enforcement (Assunção, Gandour, and Rocha 2023), commodity prices (Assunção, Gandour, and Rocha 2015) or infrastructure (Pfaff et al. 2007; Araujo, Assunção, and Bragança 2023), have been extensively studied, the long-term determinants of land use change remain far less understood.

Large-scale Amazon deforestation began only in the 1960s with the advance of agricultural frontiers (Fearnside 2005). Yet modern-day deforestation occurs in a landscape whose social geography has been transformed over the last few centuries. As non-Indigenous settlement expanded into the Amazon, Indigenous people suffered profound social and demographic changes, becoming a minority of the population by the late 19th century (Moreira Neto 1988), while also frequently resisting occupation (Stokes et al. 2025). Many historically recorded Indigenous groups no longer exist as autonomous communities, often as a result of violent conflict, disease, or cultural assimilation (Cleary 2001). The ethnohistorical literature traditionally refers to these communities as "extinct" (Steward 1946; Ribeiro 1957; Nimuendajú 2017).

This paper investigates the association between historical extinctions of Indigenous peoples and modern deforestation in the Brazilian Amazon. Based on the 1944 Ethnohistorical Map of Curt Nimuendajú, we construct geo-referenced data on the historical location and the extinction status of Indigenous groups, which we compare with remotely sensed measures of deforestation. We find that modern, post-1960 deforestation is substantially higher near pre-1944 Indigenous extinctions. While this finding is correlational, it withstands an extensive set of robustness tests accounting for potential confounders and other sources of bias, suggesting a long-run connection between the historical events surrounding the extinctions and present-day land use. We provide evidence that the findings are strongest for extinctions attributed to disease or violence, which often created demographic voids, but absent for those attributed to assimilation, where traditional communities may have persisted. In line with qualitative evidence from history and anthropology (Cunha and De Almeida 2000; Stokes et al. 2025), these patterns suggest that historical Indigenous extinctions facilitated deforestation by removing the deterrence of traditional populations on frontier occupation.

We introduce the 1944 Ethnohistorical Map of Curt Nimuendajú to economics, constructing an ethnohistorical database which we use to quantitatively study the link between historical events and modern Amazon deforestation. The map records the locations of over two thousand South American Indigenous groups and whether each was extinct by 1944. We extend the data in it by systematically reviewing historical sources to classify extinctions into three main causes which reflect different forms of contact with non-Indigenous settlers and had distinct demographic consequences: assimilation, violence, and disease. We then overlay the ethnohistorical map with a remote-sensed map of accumulated deforestation in the Brazilian Amazon. To support the analysis, we also compile georeferenced data on land use, proximity to population centers, land tenure, and geographic fundamentals. We map these variables into a 4 km $\times$ 4 km grid.

As our main finding, we document a strong positive correlation between modern deforestation and proximity to the locations of historically extinct Indigenous groups. Specifically, we estimate a regression of accumulated deforestation by 2010 on (i) the number of Indigenous groups historically located in the neighborhood grid cell, and (ii) the number of historically extinct Indigenous groups near each cell. In our preferred specification, we use the historical locations of Indigenous groups recorded between 1800 and 1944, and include an extensive set of geographic controls. We find that accumulated deforestation is 68.3 hectares greater, on average, for each extinct group within a 20 km neighborhood of a grid cell. This coefficient is equivalent to 26% of the mean deforestation in the sample. By contrast, proximity to the historical locations of non-extinct groups is associated with a small, statistically insignificant decrease of 6.6 hectares.

While this result is correlational, it is robust to a wide range of robustness tests. We consider three potential sources of bias –reverse causality, omitted variables, and measurement error –, and argue that these sources of bias cannot entirely explain away our results. First, the time gap between the extinctions and modern deforestation rules out reverse causality: the 1944 Ethnohistorical Map records only extinctions occurring before that year, whereas large-scale deforestation began after the 1960s. Second, the results are stable to the inclusion of a broad set of geographic variables likely correlated with both historical Indigenous extinctions and modern deforestation, including relief, climate, agricultural suitability, malaria suitability, proximity to rivers and water bodies, the natural incidence of rubber trees, and distance to São Paulo. We further perform multiple complementary tests using additional historical data on pre-Columbian Indigenous presence and colonial settlements, and conduct sensitivity checks to assess the influence of unobserved variables. These exercises consistently suggest that our findings are not driven by extinct-group locations simply coinciding with areas more suitable for deforestation. Third, we exploit the detailed information in the Nimuendajú map to mitigate concerns about measurement error in the historical locations. Across all these exercises, the estimated association remains remarkably stable.

Besides deforestation, we find that historical extinctions are correlated with more roads, urban settlements, private land tenure, and government-sponsored land distribution projects. They also predict a smaller share of land under Indigenous tenure and substantially higher shares of non-ethnically Indigenous inhabitants and of migrants from outside the Brazilian Amazon. These results indicate a substantial difference in the institutional and demographic development of areas that historically had an extinct Indigenous group.

Why would extinction events that took place decades or centuries ago predict modern deforestation patterns? Drawing on previous research in history, anthropology, and geography, we investigate two hypotheses. The *loss of deterrence mechanism* posits that the disappearance of Indigenous and traditional populations lowered costs of later deforestation expansion by alleviating potential land conflicts with these populations. This hypothesis is supported by historical evidence (Stokes et al. 2025, Sellars and Alix-Garcia 2018) and underpins the view that the presence of traditional populations enhances forest conservation (Cunha and De Almeida 2000). This mechanism is more likely to play a role in extinctions related to diseases and violence, and not those related to assimilation, thus providing an opportunity to empirically test its plausibility. An alternative explanation is the *occupation persistence mechanism*, stating that extinctions occurred where non-Indigenous settlers established persistent settlements and land claims that facilitated later deforestation when agriculture became more profitable. This hypothesis is supported by a rich literature on the persistence of colonial and pre-colonial institutions, which we cite below. This mechanism can be tested by tracing historical settlement and land occupation patterns, investigating whether they date back to the historical extinctions.

We find a strong correlation between extinctions and deforestation for extinctions attributed to disease or violence, which typically created demographic voids, but a small and insignificant one for extinctions attributed to assimilation. Using city settlement dates and historical agricultural census data on private farm area, we find that nearby non-Indigenous occupation is typically recent and does not date back to the extinction events. Overall, these additional tests support the view that the continued presence of Indigenous or other traditional populations deters deforestation-related occupation, and that their demographic loss removed this deterrent effect. In contrast, the data do not support the occupation persistence mechanism.

This paper contributes to two important questions in the literature on tropical deforestation. The first is the role of Indigenous peoples in forest conservation, a long-running debated question in the social sciences (Raymond 2007). Many papers in economics have investigated the effects of protected Indigenous land rights on deforestation (BenYishay et al. 2017; Baragwanath and Bayi 2020; Sze et al. 2022). However, historical Indigenous experiences may impact forest conservation beyond the assignment of land rights to current Indigenous communities, including through their agency in deterring deforestation by non-Indigenous settlers. These papers have assessed the role of Indigenous presence together with a specific land-tenure policy that guarantees property rights to Indigenous communities. Our analysis differs by considering historical Indigenous presence and extinctions, which precede the assignment of legally protected Indigenous lands. The second question refers to the long-run roots of modern tropical deforestation. To our knowledge, this is the first paper quantitatively investigating such historical determinants.

We also contribute to a broader literature that documents the importance of history in explaining long-run development and economic geography (see Spolaore and Wacziarg 2013; Nunn 2020; Lin and Rauch 2022 for literature surveys). In particular, our paper contributes to a growing literature in economics studying the history of Indigenous groups in the Americas and their relationship with modern outcomes. Differently from the existing papers in this literature, we study how historical events related to Indigenous groups affected long term deforestation and land use. In contrast, the literature has documented that modern economic outcomes in the Americas relate to the historical locations and characteristics of Indigenous populations (Maloney and Valencia Caicedo 2016; Angeles and Elizalde 2017; Alix-Garcia and Sellars 2020; Elizalde 2020; Barsanetti 2021; Franco, Galiani, and Lavado 2021), and there is evidence of the long-term development effects of colonial institutions imposed on Indigenous peoples (Acemoglu, Johnson, and Robinson 2001; Dell 2010; Summerhill 2010; Naritomi, Soares, and Assunção 2012; Bruhn and Gallego 2012; Waldinger 2017; Valencia Caicedo 2019; Carpio and Guerrero 2021; Silva 2023). Other papers study the persistence of historical events or institutions on the modern socio-demographic conditions of Indigenous peoples (Agostini, Brown, and Roman 2010; Dippel 2014; Huaroto and Gallego 2023; Feir, Gillezeau, and Jones 2023; Feir, Gillezeau, and Jones 2024; Jordán, Jaimovich, and Heilmayr 2025).

The remainder of this paper is structured as follows. Section 2 provides historical background. Section 3 describes the data. Section 4 explains the empirical framework. Section 5 presents the main results and robustness tests. Section 6 investigates the mechanisms. Section 7 concludes.

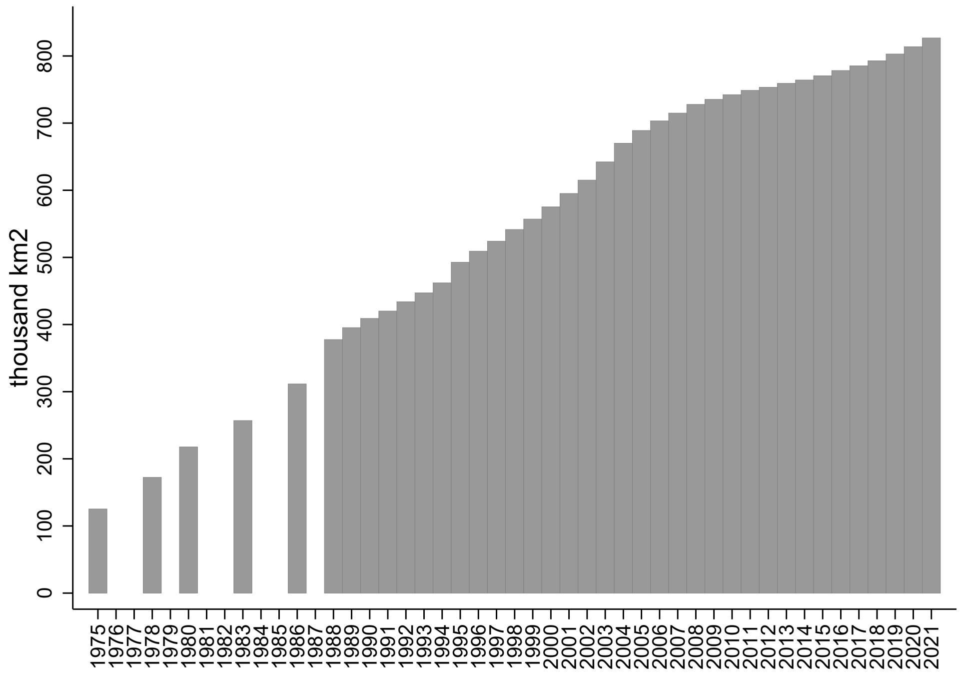
2 Historical Background

The human presence in the Amazon dates back at least 11,000 years, with agriculture emerging by at least 2,500 years ago (Roosevelt et al. 1996, Cleary 2001) and supporting the rise of complex societies that may have numbered 5 to 6 million people by 1500 (Denevan 2003, Denevan 2014). The region was peripheral to the Spanish and Portuguese empires and, compared with other parts of the Americas, European conquest was slower to take hold there (Cleary 2001). In the 16^{th} and 17^{th} centuries, colonial presence was primarily led by religious orders undertaking missionary activity across the Amazon lowlands (Souza 1994). By the mid-18^{th} century, Portugal intensified its colonization of the Amazon, promoting commodities such as cacao and expelling religious orders in an effort to modernize and secularize the regional economy (Souza 1994, Cleary 2001). The early 19^{th} century brought Brazilian independence, economic decline, and political unrest, particularly during the *Cabanagem* revolt (1835–1840), which killed an estimated 20% of the population (Souza 1994). In the late 19^{th} century, steamboats (Gregório 2009) and the rubber boom (Dean 2002) triggered a wave of settlement and economic growth, which began to fade in the 1910s due to competition from Asian plantations. There was a brief revival of the rubber industry during World War II, after which the region entered another period of stagnation.

For most of its history, from the European conquest until the 1960s, the main economic activities in the Amazon, such as rubber, nuts, and cacao, did not generate large-scale deforestation, and the region remained sparsely populated. In contrast, the period since the 1960s has been marked by large-scale forest loss, driven by the growth of modern agriculture. The Brazilian military dictatorship (1964–1985) pursued policies to occupy the Amazon for national security reasons (Souza 1994), including land distribution to newly arrived farmers, infrastructure projects, and subsidies for manufacturing, agriculture, and mining (Fearnside 2005). These policies spurred the expansion of agricultural frontiers, primarily for cattle pasture and soybeans, marking the onset of large-scale Amazon deforestation (Brondizio and Moran 2012; Pádua 2015). From 1975 to 2021, the share of deforested area in the Amazon forest rose from 3% to approximately 22%, implying that most deforestation is recent. Figure 1 displays the series of accumulated deforestation in the Brazilian Amazon, constructed based on satellite measurements from the *Instituto Nacional de Pesquisas Espaciais* (INPE); see Appendix A.1 for details.

This recent wave of deforestation took place in a region where Indigenous peoples had undergone pro-

Figure 1: Accumulated deforestation in the Brazilian Amazon since 1975



Note: Each bar shows the accumulated deforestation of primary forest up to the year on the horizontal axis for the Brazilian Amazon. Data from INPE; see Appendix A.1 for details on the construction of the series.

found social and demographic changes throughout much of the post-contact history. The Indigenous population declined sharply as contact with non-Indigenous populations brought frequent violent conflicts, the spread of lethal diseases, the imposition of forced labor, and cultural assimilation (Cleary 2001). Of hundreds of Indigenous groups historically documented, many no longer exist as autonomous communities. Following the usage in the ethnohistorical literature, we refer to these groups as "extinct."

Assimilation, violence, and disease were three common causes of extinctions. Extinctions through assimilation occurred when most members of a group joined non-Indigenous communities, with their descendants eventually adopting a new identity. These processes were often organized by religious missionaries or government authorities, who brought Indigenous groups to settlements, usually called *aldeamentos*, intended to assimilate them and often to extract their labor. Such efforts occurred from colonial times into the 20$^{\text{th}}$ century and led to the extinction of many Indigenous populations. For example, the Takonyapé had been in contact with Jesuit missionaries since the 1690s and were gradually resettled toward towns and missions over the next 170 years; the last Takonyapé *aldeamento* was established by the Pará government in 1859 (Steward 1946). Extinction by assimilation did not always involve demographic disappearance: part of the group might remain as a traditional population but no longer identify as Indigenous. Assimilated descendants, known as *tapuios* or *caboclos*, were the largest ethnic group in the Brazilian Amazon by 1800, although they were already a minority of the population by the end of the 19th century (Moreira Neto 1988; Cleary 2001).

Extinctions caused by violence or disease usually entailed the demographic disappearance of the group. Violent conflicts could result in the extermination of entire peoples. The Kanakateyé, for example, became extinct after their last village was destroyed in an enslavement raid in 1814 (Ribeiro 1841), while the Arupay were exterminated by a rival group in a war (Nimuendajú 1921). The violence of rubber tappers against Indigenous groups resulted in several extinctions, such as those of the Ararawa and of the Arikeme (Tastevin 1928, Rondon 1916). Epidemics of diseases such as smallpox and influenza were widespread and often caused catastrophic population losses (Cleary 2001). The Boanarí, a Karib people in the Middle Amazon, became extinct in the 19th century after repeated disease outbreaks following contact with Catholic missionaries (Souza 1875).

The 20th century saw a gradual acknowledgment of Indigenous rights and the development of explicit policies for their protection; see Mueller (2022) for an overview of Indigenous policy in Brazil. Beginning in the 1960s and accelerating in the 1980s and 1990s after democratization, the federal government recognized large areas of the Amazon as Indigenous lands. By 2022, these protected lands covered 23% of the Brazilian Amazon. Nonetheless, many Indigenous populations continue to face poverty, marginalization, and violence. Today, violence against Indigenous groups is still common in several contexts, including timber extraction (Chimeli and Soares 2017), mining (Pereira and Pucci 2022), and land disputes with farmers (Mueller 2022).

3 Data

3.1 Ethnohistorical Data

Data source. The central data source in our study is the 1944 Ethnohistorical Map by Curt Nimuendajú (1883-1945), a German-born ethnographer who migrated to Brazil in 1903, later worked for the Indigenous Protection Service (SPI), and wrote influential studies on several Brazilian Indigenous peoples (Grupioni 1998). Towards the end of his career, Nimuendajú produced ethnohistorical maps of Brazil and neighboring regions. The 1944 map used here was hand-drawn for the Brazilian National Museum, printed in 1981, and released in a digital edition in 2017 (Figure 2). The map records the location of over two thousand Indigenous groups based on 973 references, including anthropological studies, travel and scientific expedition reports, and historical documents.

Each map entry records an Indigenous group at a historical location and reports **(i)** the dates in which the group was recorded at that location, **(ii)** the group's linguistic family (color-coded), and **(iii)** whether Nimuendajú classified the group as "extinct" by 1944 (indicated by the calligraphic font). We interpret extinction as the disappearance of a socially distinct community, following the meaning used by Nimuendajú in his writings (e.g., his contributions to the Handbook of South American Indians Steward 1946). Hence, extinction in our database denotes not necessarily physical disappearance but also the dispersion of the group or its assimilation into other communities.

Data preparation. To construct the ethnohistorical data, we georeference the 2017 edition of the map. The map represents groups with line segments that indicate their territorial range. We georeference these segments in QGIS by recording their endpoints and approximating group locations as straight lines between these points. In the case of curved entries, we also input additional points to use in robustness tests. We then code the rich information Nimuendajú conveyed through color, calligraphic font, and text into a dataset with the group's name, its date of record, linguistic family, and the extinction status. Online Appendix A.2 contains additional details on the data preparation. The resulting database is deposited in a public data repository (Barsanetti and Ferreira 2024b).

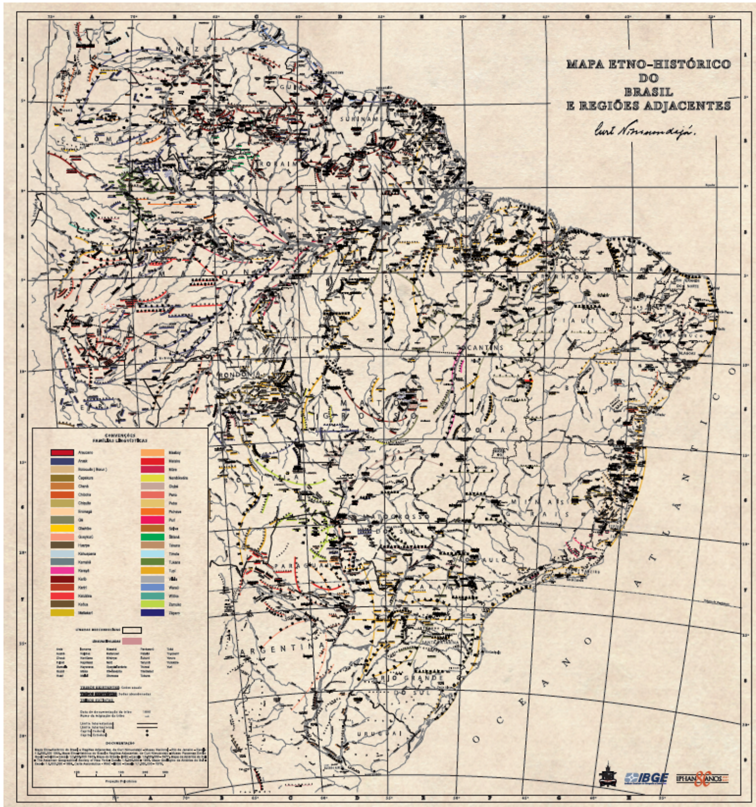
Sample restriction. A limitation of Nimuendajú's map is its reliance on historical sources. This implies that it does not survey the whole area equally, especially in colonial times (1500-1808) when the sources are sparser than in more recent periods (Moraes et al. 2021). For our study, we consider Indigenous groups recorded after 1800, a period in which the sources used by Nimuendajú were more numerous, covered a wider geographic area, and were more reliable than in earlier periods.[1] In robustness tests, we assess the sensitivity of our results to the inclusion of earlier entries.

Extinction circumstances. Beyond the information directly available from the map, we undertake a qualitative investigation of the circumstances that led to the extinction of Indigenous groups, a key infor-

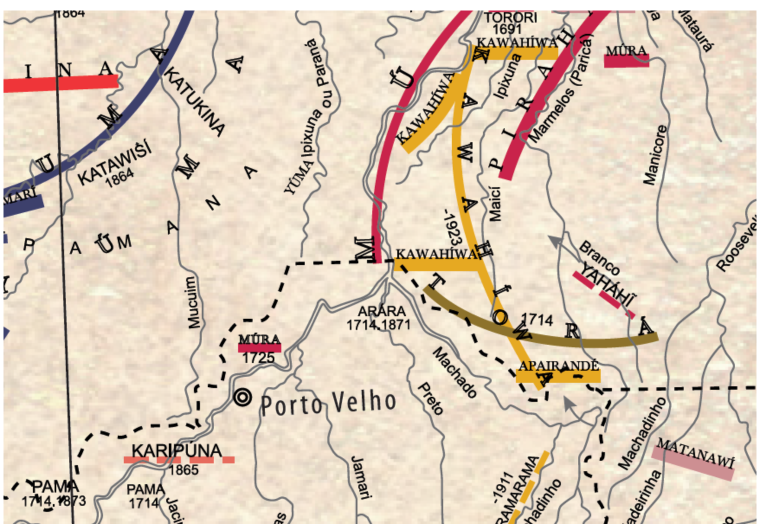
[1] For example, many 17$^{\text{th}}$ century entries are based on maps produced by the Jesuit priest Samuel Fritz in the 1690s, which list many Indigenous groups but contain almost no information about them. In contrast, the period after 1800 witnessed explorations by government officials and scientific expeditions which collected information about Indigenous groups. For example, the most cited source by Nimuendajú is a detailed description of Amazon peoples written by naturalist Carl Friedrich von Martius following his 1817–20 expedition.

Figure 2: Nimuendajú map: 2017 edition

(a) Full map



(b) Detail of the Madeira region



Note: The figure displays the digital version of the Ethnohistorical Map of Curt Nimuendajú published in 2017. This is the map we georeference and use to construct the data on Indigenous historical locations. Panel (a) displays the full map, while Panel (b) focuses on the Madeira River valley near the city of Porto Velho. The colors indicate the known language groups. The dashed lines indicate groups of known-language that were marked as extinct by Nimuendajú. Entries without color indicate groups with uncertain or unknown languages. The fonts also indicate the extinction status of the group; a finer writing denotes extinctions. The dates indicate the years associated with the historical record of the group.

mation that is not recorded in the map. This investigation allows us to test competing hypotheses about the mechanisms linking historical extinctions and modern deforestation. Specifically, we complement the ethnohistorical database by systematically reviewing the sources listed by Nimuendajú for each extinct group in the Brazilian Amazon. In addition, we consult the Handbook of South American Indians (Steward 1946) and the Ethnohistorical Dictionary of Colonial Amazon (Porro 2007). Based on the available textual evidence, we code the likely causes of extinction for each group.

These sources frequently report events that explain or suggest the causes of the groups' extinction, usually due to assimilation, violence, or diseases. We construct dummy variables for these three likely causes of extinction, allowing for overlaps when the sources mention more than one potential cause. We consider an extinction as likely due to assimilation when the sources mention that the group inhabited a non-Indigenous town or that there were attempts by missionaries or authorities to settle the group in a mission or government-established village. We consider an extinction as likely due to violence when the sources mention that the group experienced conflicts, such as war, violent deaths, and enslavement raids. We consider an extinction as likely due to disease when the sources mention epidemics or the death of many individuals due to illness.

These dummy variables are not mutually exclusive, and we assign more than one likely extinction cause for some groups. Note also that a value of zero should be understood as the absence of evidence for the particular extinction cause, but not necessarily the absence of its occurrence. These classifications are deposited in a public data repository, along with a dictionary containing a succinct description of the evidence supporting each classification (Barsanetti and Ferreira 2024a).

Language note. We use the term "extinction" following its widespread use by Nimuendajú and other historical sources consulted during data preparation. Although we recognize that the term may not adequately capture the complex and tragic processes behind the disappearance of distinct communities, we retain it due to its conventional usage in the ethnohistorical literature. Its use here carries no intent to diminish the human dimension of these events.

We also refer to each entry in the ethnohistorical map as an Indigenous "group." The choice of the word group, rather than "peoples" or "nations," is meant to remain agnostic about the cultural, social, and political organization of these communities. Nimuendajú used the term "tribe," which we avoid due to its potentially offensive connotations.

3.2 Other Data and Sample Construction

Data sources. We overlay the ethnohistorical data with high-resolution measures of deforestation, land use, economic activity, and geographical fundamentals. Our main outcome is accumulated deforestation by 2010, as measured by the PRODES program (INPE 2022). The PRODES data measure primary deforestation based on the interpretation of imagery from Landsat satellites of 1 arcsec $\times$ 1 arcsec (roughly 30 m $\times$ 30 m) resolution. We select 2010 as our benchmark year to allow us to merge the deforestation data with other spatial data relative to the Brazilian 2010 Census. However, we also consider deforestation in other years.[2]

Besides deforestation, we collect information on a wide range of other measures of present and past human geography: land use categories such as pastures or cropland, nightlights, proximity to cities and other settlements, roads, railroads, protected Indigenous lands, privately-owned farms, government-sponsored land distribution projects, self-reported ethnicity, and the share of migrants in the population. These variables are used as additional outcomes.

We also collect information on the following geographic characteristics: altitude, terrain ruggedness, water coverage, original natural non-forest vegetation, malaria suitability, climate (average temperature, precipitation, and relative humidity), proximity to main rivers, the potential yields of cassava, cacao, and soybeans, the presence of rubber trees (*Hevea brasiliensis*), caloric suitability (i.e., maximum calory production supported by local characteristics), and distance to São Paulo. These variables are used as controls.

Grid preparation. We carry out the empirical analysis in a grid with cell size of approximately 4 km $\times$ 4 km, covering the entire portion of the Amazon biome within Brazil. In what follows, we refer to our study area as the Brazilian Amazon. Each grid cell contains up to 18,225 pixels of the PRODES map and is a square of 0.036°in size. We calculate total deforestation within each grid cell by counting the deforestation pixels within it and multiplying it by pixel area. Aggregating the PRODES pixels reduces the computational burden of the estimation since it reduces the number of observations from billions of pixels to roughly 260 thousand grid cells. We aggregate most other spatial data at the grid cell. We calculate similar measures for an extensive list of other variables, which are used either as additional outcomes or controls in our analysis. Finally, we compute the nearest distance from the grid cells' centroids to the straight line formed by the two extreme points of the georeferenced location of Indigenous groups. We then calculate the number of

[2] PRODES is available in a harmonized corrected methodology starting in 2007. For this reason, we only consider maps starting in 2007 from this source.

Indigenous group entries and extinct group entries located within 20 km of each grid cell centroid. The count of extinct groups within this radius serves as our main explanatory variable.

Online Appendix A.3 provides full details on data sources and the construction of grid-level variables. Replication data and code is available at the DOI 10.5281/zenodo.17153102.

3.3 Summary Statistics

Our empirical analysis examines the association between historical Indigenous extinctions and modern deforestation in the Brazilian Amazon. Figure 3 presents a visual comparison of these variables. Panel (a) maps the locations of Indigenous groups recorded since 1800, distinguishing those marked as extinct by Nimuendajú with a square symbol, while Panel (b) shows the spatial distribution of accumulated deforestation by 2010. We focus on 599 groups whose centroids fall within the Brazilian Amazon and were recorded in or after 1800; of these, 130 are classified as extinct.[3] Based on our qualitative investigation, we classify 43 groups (33%) as likely extinct due to assimilation, 37 (28%) due to violence, and 14 (11%) due to disease. Figure 4 maps the spatial distribution of these extinction types.

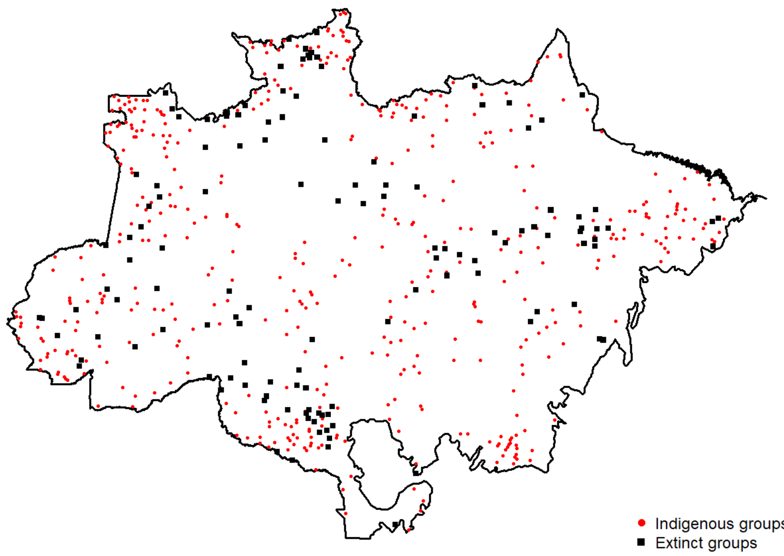
We report summary statistics on the Indigenous groups in Table 1. Column 1 displays statistics for all 599 entries, and Column 2 focuses on the 130 groups classified as extinct. Column 3 reports the differences between extinct and non-extinct groups. On average, the two groups do not differ significantly in their geographic coordinates or in the length of the line segments representing their spatial range. However, extinct group entries have fewer references listed by Nimuendajú and are less likely to have a known linguistic family than non-extinct ones. This likely reflects that the information available to Nimuendajú on extinct groups was more limited. Overall, he used an average of 10.3 sources per group and assigned a language family to 84% of the entries. Extinct entries are also more likely to be dated, probably because many non-extinct groups had their contemporary (1944) location marked; their recorded dates are, on average, slightly older.

Finally, we report summary statistics for the grid sample in Appendix Table B.1. On average, each grid cell has 0.59 Indigenous group entries and 0.12 extinct-group entries located within a 20 km radius of its centroid. In a typical grid cell of approximately 1,640 hectares, the mean accumulated deforestation by 2010 is 260.7 hectares. Land use data indicate that most of the deforested area has been converted to pasture,

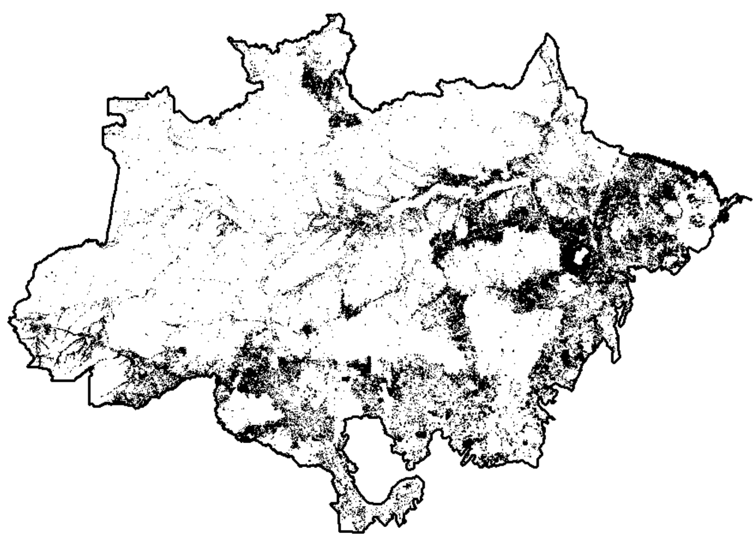
[3] These 599 groups represent 28% of all entries in the full Nimuendajú map, which covers a broader geography and historical period. Our grid-level regressions also incorporate groups with centroids outside the Brazilian Amazon whose mapped range intersects at least one grid cell in the study area.

Figure 3: Historical Indigenous group locations and deforestation in the Amazon

(a) Historical Indigenous group locations in the Nimuendajú map



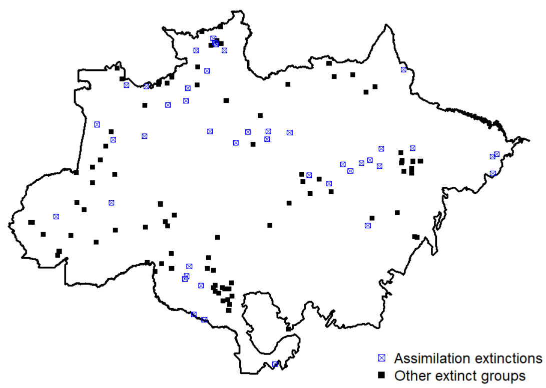
(b) Accumulated deforestation in 2010 (PRODES)



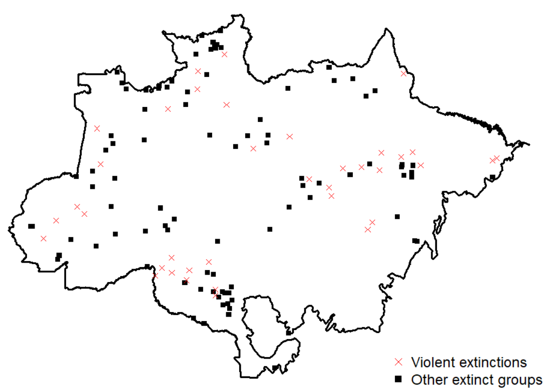
Note: The points in the map in Panel (a) indicate the midpoints of the Indigenous group entries in the Nimuendajú map for the Brazilian Amazon, which is the region defined by the black solid line, and that were recorded with dates since 1800. Black squares indicate groups assigned as extinct by Nimuendajú. Since the Nimuendajú map was produced in 1944, these extinctions occurred before 1944. Panel (b) displays, for the Brazilian Amazon, the area of accumulated primary forest clearing by 2010, as computed by the PRODES measurement program of the Brazilian Spatial Research Agency (INPE 2022).

Figure 4: Extinct groups by type of extinction

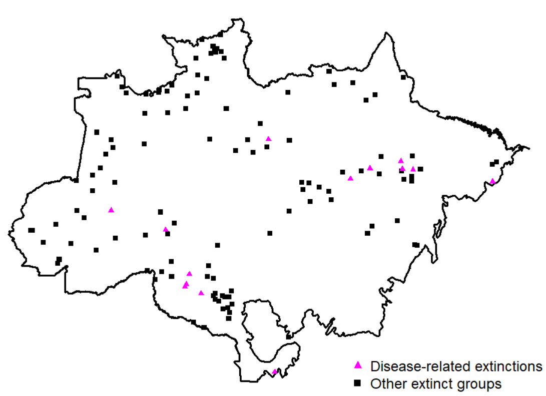
(a) Assimilation



(b) Violence



(c) Disease



Note: The maps display, for the Brazilian Amazon, the centroid of map entries that Nimuendajú indicated as extinct, categorized by the likely extinction causes (when the information was available in the sources).

Table 1: Summary statistics of Nimuendajú's map entries in the Brazilian Amazon

	All groups (1)	Extinct groups (2)	Mean difference extinct − non-ext. (3)
Extinct	0.22		
	(0.41)		
Latitude	-4.78	-4.67	0.14
	(5.06)	(4.85)	*0.767*
Longitude	-60.42	-60.60	-0.23
	(7.25)	(6.50)	*0.732*
Group entry range (km)	74.55	70.13	-5.64
	(64.50)	(58.63)	*0.345*
Number of sources	10.33	4.64	-7.27
	(12.60)	(4.22)	*0.000*
Known linguistic family	0.84	0.60	-0.31
	(0.37)	(0.49)	*0.000*
Dated entry	0.36	0.66	0.38
	(0.48)	(0.48)	*0.000*
Oldest year of record of the entry	1881.94	1875.36	-10.92
	(35.39)	(34.75)	*0.026*

Note: Each observation is a group entry in the Nimuendajú map. The table displays information only for the groups with centroids inside the Brazilian Amazon biome and that were recorded after 1800. There are 599 group entries, of which 130 refer to extinct groups. Column 1 displays the sample averages for all 599 group entries, while column 2 displays the sample averages for the 130 extinct group entries. Standard errors are shown in parentheses below the sample averages. Column 3 displays the difference in the means between extinct and non-extinct groups, with p-values in italics reported below. The statistics on the years is conditional on the entry being dated.

which covers an average of 206.8 hectares per cell. The Brazilian Amazon is sparsely populated, as reflected by low nightlight luminosity and the small area occupied by urban settlements.

4 Empirical Framework

Estimation. To investigate the association between historical Indigenous extinctions and modern Amazon deforestation, we estimate the following equation using ordinary least squares:

$$y_i = \beta\ ExtinctGroups_i + \eta\ Groups_i + X_i'\theta + u_i \tag{1}$$

Our primary dependent variable y_i is accumulated deforestation within each grid cell i, in hectares. We also consider various other dependent variables. Our main explanatory variable is $ExtinctGroups_i$, the number of extinct Indigenous groups within 20 km of the grid cell. This variable measures proximity to historical Indigenous extinctions. $Groups_i$ denotes the number of any historical Indigenous groups (extinct and non-extinct) within 20 km of the centroid of the grid cell, accounting for overall historical Indigenous presence. We include it as a control variable. Therefore, the coefficient on extinct groups represent the mean increase in deforestation for one additional extinct group in the cell's neighborhood, conditional on the number of recorded historical Indigenous groups (extinct or not) in the same neighborhood. In Online Appendix B.1, we show the robustness of the findings to alternative distance cutoffs, measures of proximity to historical Indigenous locations, year cutoffs when restricting the groups used in the analysis, and other specification choices.

We also include the set X_i of geographic controls: latitude, longitude, altitude, terrain ruggedness, water coverage, original non-forest coverage, malaria suitability, temperature, rainfall, relative humidity, proximity to main rivers, rubber tree range indicator, agro-ecological potential yields of soybeans, cassava and cacao, caloric suitability, and distance to São Paulo. These variables account for physical constraints to deforestation (e.g. ruggedness, water), remoteness (e.g. proximity to rivers, distance to São Paulo), land productivity (e.g. potential yields, rubber trees), and natural amenities (e.g. climate, malaria), reflecting natural conditions relevant for the historical occupation of the Amazon and for modern deforestation.

Inference. We account for spatial correlation in the errors by estimating standard errors that are robust to heteroskedasticity and cross-sectional correlation within clusters defined by grids of 1°of latitude by 1°of longitude. Because observations near the same latitude or longitude line can be geographically close yet assigned to different clusters, capturing correlation across such boundaries requires more than a single grid for clustering. We therefore use multi-way clustering (Cameron, Gelbach, and Miller 2011) with four interlaced 1°× 1°grids: the first grid aligned with 1°latitude and longitude lines, and three shifted horizontally, vertically, and diagonally by 0.5°. This design allows for correlation between the errors of any nearby observations, as they will belong to the same cluster in at least one grid. As an alternative, we compute spatial HAC standard errors using the method of Conley (1999); although computationally burdensome given the large sample size, they yield similar results (see Appendix Table B.12).

Interpretation of the estimates. Our coefficient of interest, β, captures the correlation between proximity to historically extinct Indigenous groups and deforestation. Since the regression controls for proximity to any Indigenous group in our data and for a rich set of geographic variables, this correlation is conditional on recorded historical Indigenous presence and geographic fundamentals. In practical terms, the estimates for β measure differences in average deforestation with each additional extinct group, holding constant the total number of Indigenous groups and the geographic characteristics of the cell. Due to the linear specification, β may be interpreted as the difference in deforestation between cells near one extinct group and geographically similar cells near one non-extinct group. To compare cells close to one extinct group to those without any nearby historical Indigenous group, one needs to compute $\beta + \eta$.[4]

We highlight that, because there is no clearly exogenous source of variation in extinctions, the estimates of β are correlational and should be interpreted with caution. We conduct a range of robustness checks to assess and mitigate potential sources of bias, and study potential mechanisms using additional historical information. Even if not causal, the estimates are informative about the historical events and processes associated with deforestation.

We consider three potential sources of bias that could invalidate causal interpretations of the estimates:

1. **Reverse causality.** This would occur if Indigenous group extinctions were a consequence of deforestation. The large time difference between the extinction events and deforestation rules out this concern. Given the date of the map, the extinction events necessarily predate the 1940s, whereas large-scale deforestation started after the 1960s. Therefore, it is not plausible that the extinctions of the Indigenous group in our data occurred due to deforestation.

2. **Omitted variables.** Neither the historical locations of Indigenous peoples nor their extinctions are random, and these locations may have characteristics that make them more or less likely to be deforested. To address this, we condition on a rich set of geographic fundamentals. We also apply sensitivity analysis methods to assess how strong selection on unobservables would need to be to explain our findings. Moreover, because we control for the total number of Indigenous groups near each grid cell ($Groups_i$), our design is robust to omitted variables that affect both extinct and non-extinct groups equally. In addition, we leverage the information from the Nimuendajú map to account for differences in Indigenous groups' characteristics.

3. **Measurement error.** Because Nimuendajú's map was based on historical sources, the location of extinct groups may be measured with error. We assess the potential impact of this issue using several

[4] In practice, estimates of η are usually close to zero, so the magnitude and sign of β broadly summarize the difference between cells near extinct groups and other locations.

robustness exercises, leveraging features of the ethnohistorical data such as the number and language of historical sources per group or whether we could confirm the extinctions in our reading of the sources. For instance, we run our regressions with subsamples of map entries for which we have greater confidence about their information, given the abundance of sources and our own assessment of what the sources said.

5 Results

We find a strong and consistent relationship between historical Indigenous extinctions and modern deforestation in the Brazilian Amazon. These patterns are robust to a long series of tests aimed at addressing concerns over selection in geographic or group characteristics, measurement errors in the historical data, or specific choices with respect to model specification and estimation. In contrast, there is no correlation between deforestation and proximity to non-extinct historical groups. This robust correlation between historical locations of extinct groups and modern deforestation is the central finding of the paper.

5.1 Main Results

Table 2 presents the estimates of Equation 1, where the outcome variable is accumulated deforestation by 2010, in hectares per cell. Column 1 presents the estimated results in a specification without any geographic controls. Column 2 presents the estimates in a parsimonious specification, which controls linearly for latitude and longitude. These geographic coordinates are strong predictors of deforestation, which is uneven across space, justifying their inclusion as controls. In column 3, we also control for all the additional geographic fundamentals computed explained earlier. This is our preferred specification, as it controls for an extensive list of geographical characteristics possibly correlated with historical Indigenous occupation, extinctions due to non-Indigenous contact, and suitability for modern agriculture. We always include the full set of geographic variables as controls in the results reported in later tables.[5]

The coefficient on $ExtinctGroups$, our main coefficient of interest, is positive in all columns, being statistically significant in both specifications with geographic controls (columns 2 and 3). In column 3, our preferred specification, one additional extinct group in a radius of 20 km of the grid cell is associated with

[5] Appendix Table B.2 shows the estimated coefficients on the controls.

Table 2: Association with hectares of accumulated deforestation by 2010

	(1)	(2)	(3)
Extinct groups	45.233	82.470**	74.740***
	(46.932)	(32.573)	(26.165)
Indigenous groups	-12.404	16.196	-6.617
	(22.254)	(19.072)	(13.488)
Controls	None	Coordinates	+ Fundamentals
Coeff. extinct vs. no group ($\beta + \eta$)	32.829	98.666	68.123
p-value $\beta + \eta = 0$	0.499	0.001	0.006
R-squared	0.001	0.261	0.436
Sample mean	260.722	260.722	260.722
Observations	259574	259574	259574

Note: The table displays estimates of the coefficients in Equation 1. The dependent variable is accumulated deforestation by 2010, in hectares, measured by PRODES. Column 1 reports estimates for a specification with no controls, column 2 for a parsimonious specification controlling linearly for the grid cell's coordinates (latitude and longitude), and column 3 for an extended specification that includes the following geographic fundamentals as controls: altitude, terrain ruggedness, water coverage, non-forest coverage, malaria suitability, temperature, rainfall, relative humidity, proximity to main rivers, rubber tree range indicator, agro-ecological potential yields of soybeans, cassava and cacao, caloric suitability, and distance to São Paulo. The table also shows the p-values for the test that the sum of the coefficients on "Extinct groups" (β) and "Indigenous groups" (η) are equal to zero. Standard errors in parentheses are multi-way clustered according to four interlaced grids of $1^\circ \times 1^\circ$. Statistical significance: * 10%, ** 5%, *** 1%.

additional 74.7 hectares of deforestation, significant at the 1% level. This increase is equivalent to 29% of the mean accumulated deforestation per cell in 2010, which was 260.7 hectares (16% of the cell's area).

The coefficients on $Groups$ are close to zero and statistically insignificant in all specifications. Therefore, proximity to historical locations of non-extinct Indigenous groups does not correlate with deforestation. This implies that the difference between cells near one historically extinct group and cells with no historical Indigenous presence, obtained by the sum of the two coefficients, is positive and similar to our main coefficient of interest. In our preferred specification, this difference is 68.1 hectares (26% of the mean deforestation), with areas near extinctions also featuring substantially more deforestation than areas with no recorded Indigenous presence.

In sum, our main finding is that areas near historical Indigenous extinctions have been 26–29% more deforested than other areas of the Amazon. We next show that this result is robust to a wide range of alternative specifications, measurements, and historical controls, and later explore the relationship between extinctions and other modern outcomes.

5.2 Robustness Tests

Selection on unobservables. Despite including a comprehensive list of geographic fundamentals, unobserved geographic variables could still influence both extinctions and deforestation, biasing the estimates. To assess these omitted variable bias concerns, we conduct a sensitivity analysis following Oster (2019). This exercise quantifies how strong the correlation of the regressor with unobservables would need to be for the true coefficient to be zero. The results imply that selection on unobservables would have to be 2.3 times as strong as selection on the list of included controls, which would be unlikely given the comprehensive set of included controls. Appendix B.1.1 reports the full exercise, along with a similar analysis following methods in Masten and Poirier (2022), both of which support the robustness of our main estimates.

Pre-1800 settlements as a confounder. A major concern is that historical settlement patterns may have influenced both extinctions and modern deforestation. Persistent effects of colonial settlements and institutions on economic development are well documented (e.g. Dell 2010; Naritomi, Soares, and Assunção 2012; Waldinger 2017; Valencia Caicedo 2019; Barsanetti 2021; Silva 2023). To address these concerns, we use georeferenced information on pre-Columbian Indigenous presence (before 1500) and colonial settlements in Brazil (1500-1808). Pre-Columbian Indigenous presence is measured by the location of Indigenous archaeological sites (from *Instituto do Patrimônio Histórico e Artístico Nacional*, IPHAN) and Indigenous earthworks (from Jacobs 2023; Walker et al. 2023). As for colonial settlements, we use the location of *aldeamentos* (religious missionary villages) from circa 1730 mapped by Bombardi (2014), historical (post-1500) archaeological sites, and municipality seats of cities settled before 1800.[6]

Appendix Table B.3 shows the results of the analysis for pre-Columbian sites. The extinctions are not correlated with proximity measures for archaeological sites or earthworks, and the inclusion of proximity to these sites as controls barely changes the coefficient on extinctions. Similar results are found for the analysis of proximity to colonial sites, shown in Appendix Table B.4. Neither *aldeamentos*, historical archaeological sites, nor colonial-origin cities feature a positive correlation with the extinctions: these correlations are either null or, for historical archaeological sites, negative and only marginally statistically significant. Moreover,

[6]For this exercise, we use the IBGE Localities data for 2010, define a city as a municipality seat, and identify the settlement year based on the texts about municipality histories in *IBGE Cidades*.

the positive association between extinctions and deforestation is robust to the inclusion of proximity to these colonial sites as controls.

Pre-1800 extinction events. As explained in Section 3, our main analysis uses the location of groups recorded after 1800 in the ethnohistorical map, so our findings refer to extinctions between 1800 and 1944. The map also contains information on earlier Indigenous group locations. Appendix Table B.5 reports estimates under alternative restrictions on the ethnohistorical entries. In Panel A, we vary the initial cutoff year, using all groups recorded after that year. A positive and statistically significant correlation between historical extinctions and modern deforestation holds for all cutoff years since 1700. The largest coefficient (91.2) arises when restricting to post-1850 groups; the coefficient declines and becomes statistically insignificant when pre-1700 groups are included (26.2 when including all groups).

The decline of the coefficient for the sample including pre-1800 groups likely reflects the lower quality of ethnohistorical data for earlier groups. Moraes et al. (2021) notes that coverage of the Nimuendajú map in the Amazon is limited for the initial colonial period, which is to be expected given that effective Portuguese colonization of the region began only in the 18^{th} century (Cleary 2001), suggesting greater measurement error in these entries. To assess whether information quality explains the loss of significance, Panel B repeats the analysis using only groups with a known linguistic family, varying the cutoff year. A known linguistic classification indicates that Nimuendajú had more reliable information about the group. In this subset, extinctions in all period samples remain positively and strongly correlated with deforestation, and the estimates are stable as we change the cutoff year. This pattern supports our interpretation that weaker correlations in the full sample stem mainly from low-quality historical records for earlier groups.

Cultural characteristics. Another selection concern is that extinction status may reflect unobservable group characteristics predictive of future deforestation, in which case the association with deforestation would reflect the long-run association with other characteristics of the historical Indigenous group inhabiting the area. For instance, cultural attributes of a group may determine both where it has historically lived (and the suitability for future land use) and whether it became extinct. To address this concern, we include information on language and record year in the Nimuendajú map as group-level controls in our specification. Appendix Table B.6 displays the results from a specification that is similar to Equation 1 but replaces the variable $Groups_i$ with variables that count the number of Indigenous groups recorded in the several historical periods (1800–1850, 1850–1900, and 1900–1944) and within each language family as classified by

Nimuendajú. Compared with our main estimates, controlling for these characteristics makes the coefficient size even larger (98.7 hectares in column 3), suggesting that cultural characteristics and period differences across groups do not explain our results.

Definitions of the proximity variables. In our main results, proximity to Indigenous group locations in the Nimuendajú map is measured as the number of groups within 20 km of each grid cell centroid. In the data preparation, we record the groups in the map as lines, with distance to the group computed to the closest point on the line. These choices could matter if the 20 km radius poorly approximates the historical extent of a group, or if curved entries are less accurately represented by straight segments. Appendix Table B.7 shows that our findings are robust to alternative definitions, including different ways of approximating curved entries, varying the distance cut-off, using binary indicators instead of counts, and measuring proximity by the minimum distance to a group. Detailed explanations of these robustness tests are in Appendix B.1.2.

Measurement errors. Despite the comprehensiveness of the sources used by Nimuendajú, the task of mapping historical Indigenous groups is subject to measurement error concerns. However, our results are robust to several tests that account for measurement errors: restricting the ethnohistorical sample to better-documented groups or registered also by non-Portuguese sources; controlling for non-Indigenous settlements at the time Nimuendajú prepared his map; using weights to address cartographic distortions; confirming whether the sources explicitly mention the extinction of the group to validate the extinction information. We provide a detailed rationale for these tests, along with the results tables, in Appendix B.1.3.

Inclusion of historical Indigenous presence as controls. In our main specification, we control for the total number of groups within 20 km of each cell, regardless of extinction status, so the coefficient on extinctions is conditional on overall historical Indigenous presence. Appendix Table B.11 reports robustness tests showing that our findings are not sensitive to this choice. Column 1 drops this control altogether, with coefficients remaining essentially unchanged. Column 2 uses a quadratic polynomial in the total number of groups, again yielding similar results, indicating that the estimates reflect proximity to extinct groups rather than areas with a higher overall concentration of historical groups. Column 3 restricts the control variable to groups not contemporary to the map, providing a closer comparison to the extinct groups, which by definition were not contemporary to the map, and the coefficients again remain similar.

Alternative estimation methods. We also perform exercises with alternative estimation methods, computing spatially-robust standard errors according to Conley (1999), clustering by modern municipality, and estimating a censored-variable Tobit model to account for the fact that deforestation is censored at the area of the grid cells. Detailed rationale and results are shown in Appendix B.1.4. Most of these tests only alter the main result on the margin, while the Tobit model doubles the size of the main estimate (see Table B.12).

5.3 Other Modern Outcomes

Settlements, modern infrastructure, and land tenure. We investigate how historical Indigenous extinctions are related to other geographic characteristics related to deforestation: proximity to economic activity, transportation infrastructure, and land tenure. We estimate Equation 1 replacing the dependent variable by (i) distance to cities and other non-Indigenous settlements (data from IBGE's 2010 Census), (ii) the length of roads and railroads in each grid cell (IBGE), and (iii) land tenure measures, namely: the presence in the grid cell of officially recognized Indigenous reserves (data from Funai), the area of private rural properties in the grid cell (data compiled by Imaflora), and the area of government-sponsored land distribution projects (data from INCRA).[7] All variables refer to modern outcomes circa 2010, except private rural properties, which are measured in 2020.[8]

The results for proximity to settlements and infrastructure are reported in Table 3. Columns 1 and 2 use as dependent variables indicators for being within 20 km of a city or a non-Indigenous settlement. In both cases, the coefficients are positive and statistically significant: each additional extinct group near the grid cell is associated with a 4.3 percentage points increase in the probability of being near a city, over a mean of 11.2%, and a 6.8 percentage points increase in the probability of being near a settlement, over a mean of 22.2%. Column 3 shows that an additional extinct group is significantly associated with 134 meters more automobile roads within the grid cell, consistent with the importance of roads for the expansion of agricultural frontiers in the region. In contrast, we find no significant association between extinctions and railroads (column 4).

The results for land tenure are reported in Table 4. Column 1 shows that each additional extinct group in the cell's neighborhood is associated with an 11 percentage points lower probability that the cell contains an officially recognized Indigenous land, while proximity to a non-extinct historical group is associated with

[7]Although comprising a small fraction of the land, government-sponsored land distribution was a historically important tool in the agricultural colonization of the Amazon (Kohlhepp 2015).

[8]The data on land ownership is built from recent administrative registries, not allowing us to map private rural land in 2010.

Table 3: Association with population centers and infrastructure

	Proximity to: city	Proximity to: settlement	Length of: roads	Length of: railroads
	(1)	(2)	(3)	(4)
Extinct groups	0.043***	0.068***	134.160***	5.318
	(0.017)	(0.024)	(48.882)	(7.112)
Indigenous groups	-0.005	-0.002	-17.042	-0.355
	(0.009)	(0.012)	(18.538)	(2.284)
R-squared	0.170	0.260	0.075	0.011
Sample mean	0.112	0.222	288.156	6.562
Observations	259574	259574	259574	259574

Note: The table displays estimates of the coefficients in Equation 1. In columns 1 and 2, the dependent variable is a dummy indicating if the grid cell is within 20 km of a settlement or city; the location of settlements and cities is from IBGE. In columns 3 and 4, the dependent variable is the extension of, respectively, roads (for automobiles) or railroads inside the grid cell. The road and railroad shapefiles are from IBGE's 2014 *Base Contínua ao Milionésimo*. The estimated equation includes latitude, longitude, and all geographic fundamentals from our preferred specification. Standard errors in parentheses are multi-way clustered according to four interlaced grids of $1° \times 1°$. Statistical significance: * 10%, ** 5%, *** 1%.

an 8.1 percentage points increase. Column 2 presents similar findings when using a dummy for being within 20 km of an Indigenous village.[9] These patterns are consistent with the designation process for protected Indigenous lands, which requires legal proof of historical occupation by the community. In contrast, cells present 59.0 more hectares of private rural properties per historically extinct group (column 3) and 0.4 more hectares of government-sponsored land distribution projects (column 4).

In sum, historical Indigenous extinctions are associated with proximity to cities and other population centers, more road infrastructure, fewer Indigenous land rights, and more private land ownership. Appendix Table B.14 displays estimates from "horse race" equations in which we augment Equation 1 with proximity to population centers, transportation, Indigenous land rights, and private land tenure as controls. The coefficient on extinct groups decreases but always remains large and statistically significant. The largest reduction occurs if we control for proximity to cities and settlements.

Hence, neither settlements, modern infrastructure, nor land tenure on their own entirely explain the correlation between extinctions and deforestation. However, the joint inclusion of all these variables in column 5 strongly reduces the coefficient on extinct Indigenous groups, which falls to 20.1 and becomes statistically insignificant.

[9]Proximity to Indigenous villages is an imperfect proxy for modern Indigenous presence, as many Indigenous individuals live outside villages, and the IBGE data only records villages with more than 20 inhabitants.

Table 4: Association with Indigenous lands and private land ownership

	Grid intersects Indig. reserve (1)	Proximity to Indig. village (2)	Private farm area (ha) (3)	Land projects area (ha) (4)
Extinct group	-0.110***	-0.037**	58.964**	0.351**
	(0.026)	(0.018)	(27.357)	(0.158)
Indigenous groups	0.081***	0.038***	-27.062	-0.278***
	(0.017)	(0.009)	(20.100)	(0.078)
R-squared	0.169	0.070	0.279	0.055
Sample mean	0.251	0.086	391.316	1.142
Observations	259574	259574	259574	259574

Note: The table displays estimates of the coefficients in Equation 1. In column 1, the dependent variable is a dummy equal to one if the grid cell overlaps with a homologated Indigenous land in 2010; these shapefiles are from Funai and the homologation dates are from *Terras Indígenas no Brasil*. In column 2, the dependent variable is a dummy indicating if the grid cell is within 20 km of an Indigenous village; the location of Indigenous villages is from the list of localities for the 2010 Census. In column 3, the dependent variable is the area of private farms in the grid cell, in hectares, obtained from the datasets SIGEF and CAR compiled by Imaflora for 2020. In column 4, the dependent variable is the area of government-sponsored land distribution projects in the grid cell, in hectares, obtained from INCRA data for 2010. The estimated equation includes latitude, longitude, and all geographic fundamentals from our preferred specification. Standard errors in parentheses are multi-way clustered according to four interlaced grids of $1° \times 1°$. Statistical significance: * 10%, ** 5%, *** 1%.

Modern demographics. Historical Indigenous extinctions likely influenced the ethnic composition of the population today. To evaluate this hypothesis, we use data on ethnicity from the 2010 Census at the level of census blocks (*setores censitários*).[10] There are almost 25 thousand inhabited census blocks in the Brazilian Amazon. We estimate Equation 1 at the census block level (instead of 4 km $\times$ 4 km grid cells). The dependent variables are the share of each ethnic group in the population of census block. The variables $ExtinctGroups$ and $Groups$ are the number of extinct and total historical groups within a 20 km distance from the census blocks centroids. We also calculate all geographic fundamentals used in our preferred specification at the census block level. Since census blocks have irregular shapes and areas, we weight each census block by area.

The results are shown in Table 5. Column 1 shows that one additional extinct Indigenous group near the census block is associated with an average reduction of 9.4 percentage points in the share of the population self-declaring as Indigenous. In contrast, the presence of non-extinct Indigenous groups is positively associated with the modern share of Indigenous population. These estimates confirm that areas with extinctions

[10] *Setores censitários* are similar in population size to the census blocks from the United States, which is why we refer to them as census blocks for simplicity.

have a lower Indigenous presence today. In columns 2 and 3, we show that the reduction in the Indigenous share occurs due to an increase in the shares of white (by 4.0 percentage points) and ethnically mixed residents (by 5.6 percentage points). Columns 4 and 5 report null correlations between extinct groups and the share of Blacks and Asians.

Table 5: Association with modern ethnic composition and migration

Share of:	Indigenous (1)	Whites (2)	Mixed (3)	Blacks (4)	Asians (5)	Migrants (6)
Extinct groups	-0.094***	0.040***	0.056**	-0.003	0.000	0.082***
	(0.029)	(0.012)	(0.022)	(0.005)	(0.001)	(0.028)
Indigenous groups	0.067***	-0.021**	-0.039***	-0.008**	0.000	-0.018
	(0.018)	(0.008)	(0.013)	(0.003)	(0.001)	(0.013)
R-squared	0.291	0.377	0.288	0.088	0.025	0.821
Sample mean	0.190	0.176	0.560	0.066	0.007	0.117
Observations	24790	24790	24790	24790	24790	501

Note: The table displays estimates of the coefficients in equation 1 at the Census block level (columns 1 to 5) and municipality level (column 6). For columns 1 to 5, the dependent variable is the share of the population of the ethnic group on the top; observations are non-zero population census blocks from the 2010 census. In column 6, the dependent variable is the share of migrants from outside the Amazon in the total population; we define an individual as a migrant if this individual was not born in one of the Amazonian states. These latter estimates are obtained in regressions at the municipality level, using municipalities whose seats are within the Brazilian Amazon, and the data stems from the extended questionnaire which was applied to a representative sample of the 2010 census. The estimated equation includes latitude, longitude, and all geographic fundamentals from our preferred specification. Observations (census blocks or municipalities) are weighted by area. The sample means reported at the bottom of the table are also weighted by area. Standard errors in parentheses are multi-way clustered according to four interlaced grids of $1^\circ \times 1^\circ$. Statistical significance: * 10%, ** 5%, *** 1%.

The recent expansion of agriculture in the Amazon has been accompanied by migration waves from other Brazilian states (Pellegrina and Sotelo 2024), raising the question of whether historical extinctions are linked to recent migration patterns. The strong correlation between extinctions and the white population is a high share of white migrants from other regions migrating into the Amazon.[11] Since migration data from the 2010 Census are available only at the municipality level, we estimate Equation 1 with the share of migrants from outside the Amazon region as the dependent variable, averaging $ExtinctGroups$, $Groups$, and the geographic fundamentals across grid cells within each municipality, and weighting by municipality area.[12] Column 6 of Table 5 shows a strong positive correlation: each extinct group is associated with an 8.2 percentage point higher share of migrants from outside the Amazon, which corresponds to 70% of the

[11] In 2010, 20% of white inhabitants of the Amazon states were born outside the region, compared with 9% of the non-white population.

[12] The sample consists of 501 municipalities with seats in the Brazilian Amazon.

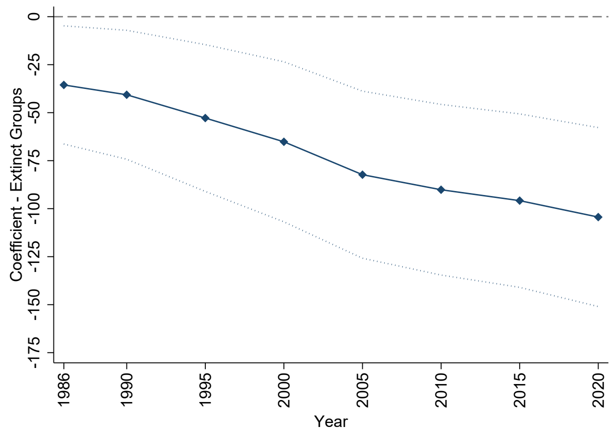
sample mean of 11.7%. Taken together, these results suggest that areas with historical Indigenous extinctions have fewer Indigenous and more white residents today, as well as a higher share of migrants from outside the Amazon, consistent with a more recent pattern of settlement.

Deforestation over time. Our main results refer to accumulated deforestation by 2010, but it is also relevant to understand when the correlation between extinctions and deforestation emerged. Amazon deforestation in Brazil accelerated in the second half of the 20$^{\text{th}}$ century, following the expansion of agriculture since the 1960s. To assess how the historical Indigenous extinctions correlate with deforestation in earlier periods, we use yearly primary forest cover maps from 1986 to 2020 from MapBiomas (Souza Jr et al. 2020), which are based on Landsat imagery, like PRODES. Primary forest cover is a mirror image of primary deforestation. We estimate Equation 1 using the area of primary forest in each year as the dependent variable. The estimates, reported in Figure 5, show that proximity to an extinct group is negatively and significantly correlated with less primary forest since the 1986 (conditional on historical Indigenous presence). The magnitude of the coefficient increases over time, from 33.9 hectares less primary forest in 1986 to 102.7 hectares less in 2020 per extinct group. These patterns suggest that, although deforestation near historical extinctions was already elevated by 1986, a substantial portion of the correlation in our main results reflects post-1986 forest loss in these areas.[13]

Modern land use. Understanding which land uses have replaced forests near historical extinctions helps to contextualize our main findings. We use MapBiomas maps to measure primary forest, farming (cropland plus pasture), cropland, pasture, and urban areas in each 4 km $\times$ 4 km grid cell. Nightlight intensity from DMSP–OLS serves as an alternative proxy for urbanization and economic activity. Appendix Table B.13 reports regressions using these variables as outcomes. Proximity to historically extinct Indigenous groups is associated with an average reduction of 88.6 hectares in primary forest cover and an average increase of 84.3 hectares in pasture, the most common replacement for forest. There is no significant association with cropland area, but a positive association with both urbanization and nightlight intensity. These results indicate that most of the deforestation in these areas reflects conversion to agriculture, with the expansion of urban settlements also contributing.

[13] Deforestation data from PRODES, our main source, are available under a harmonized methodology only from 2007. Appendix Figure B.1 shows that our main results are consistent with the MapBiomas results, with coefficients on extinctions increasing over time.

Figure 5: Association with primary forest cover over time, 1986–2020 (MapBiomas)



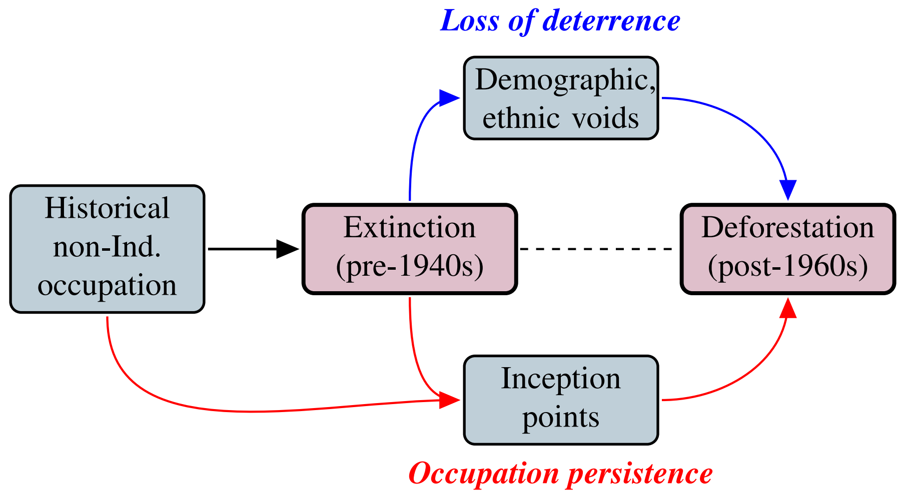
Note: The solid navy line and the markings display the estimates of the coefficient on $ExtinctGroups_i$ in Equation 1. The dependent variable is primary forest cover at the year indicated on the horizontal axis, in hectares, measured by MapBiomas. The equation also includes the variable $Groups_i$ and controls for latitude, longitude, and all geographic fundamentals used in the preferred specification. The dotted lines indicate the upper and lower bounds of the 90% confidence interval, based on multi-way clustered standard errors according to four interlaced grids of $1^\circ \times 1^\circ$. The dashed gray line indicates a horizontal line at zero.

6 Mechanisms

The previous section documents a positive correlation between historical Indigenous extinctions and accumulated deforestation, which also extends to modern land tenure, settlement, and demographic outcomes. Given the robustness of this association to omitted variables, specification, and measurement-error concerns, these findings suggest the presence of historical mechanisms shaping long-run land use. We investigate these mechanisms by incorporating historical data that allow us to characterize the extinction events and to trace settlement dynamics over time. In particular, there are two plausible mechanisms that can explain the data and are consistent with previous research in history, anthropology, and geography: loss of deterrence and occupation persistence. These mechanisms are illustrated in Figure 6 and explained below.

A first explanation for our findings is that Indigenous extinctions may have led to a *loss of deterrence* against deforestation-related activities due to a reduction in the presence of Indigenous or other traditional populations like *caboclos*. This deterrence often takes place due to the active resistance of these populations against external settlers, through protests, judicial disputes, or violent conflict. Stokes et al. (2025) provide historical evidence of how Indigenous peoples resisted the advance of non-Indigenous occupation in the Amazon, and Cunha and De Almeida (2000) describe how many traditional populations still counter

Figure 6: Mechanisms relating historical Indigenous extinctions and modern deforestation



Note: The directed acyclic graph displays two mechanisms leading to a positive association between Indigenous extinctions and modern deforestation (dashed line in the center). Connections in red refer to the occupation persistence mechanism and connections in blue refer to the loss of deterrence mechanism.

agricultural frontier expansion. In addition, continued Indigenous presence also strengthens their land rights over time, particularly after the formal recognition of Indigenous lands in the late 20th century, further blocking deforestation. The presence of traditional population has also sometimes deflected state-led development projects: after early colonization initiatives under the military dictatorship provoked conflicts with Indigenous and *caboclo* communities, the government revised land distribution policies in the 1980s to reduce such disputes (Kohlhepp 2015).

The alternative explanation for the link between historical extinctions and modern deforestation is *occupation persistence*. This mechanism is theoretically plausible because Indigenous extinctions were often the result of contact with non-Indigenous settlers (Cleary 2001), which frequently exploited, expelled, assimilated, or exterminated Indigenous communities. The extinctions observed in Nimuendajú's map can thus reflect the location of historical settlements and land claims, many of which would have persisted over time. The economic literature documents several cases of path dependence in settlement locations (Lin and Rauch 2022, Barsanetti 2021). From the 1960s onward, when economic conditions and state policies favored agricultural expansion, they would have become natural inception points for deforestation.

Both mechanisms have theoretical plausibility, are not mutually exclusive, and yield different testable implications. Loss of deterrence can explain our findings only to the extent that extinctions resulted in demographic and ethnic voids, reducing the presence of traditional populations. Therefore, we can investigate whether the coefficient on extinctions is driven by extinctions due to violence or disease and not those due to assimilation, in which the population changes its cultural identity but may remain in the area.[14]

Occupation persistence can explain our findings to the extent that the areas with historical extinctions also displayed historical non-Indigenous occupation. It thus predicts that areas near historical extinctions should also display early evidence of non-Indigenous settlement and land-tenure formation, potentially dating back to the extinction events. We can thus assess the importance of the occupation persistence hypothesis by examining data on historical occupation, to test whether such early occupation is systematically associated with extinctions as recorded in the Nimuendajú map.[15] We compile data on the dates of creation of cities in the Brazilian Amazon and historical land tenure data to perform this analysis.

6.1 Evidence on Loss of Deterrence

To assess the importance of loss of Indigenous deterrence in explaining the results, we estimate the correlation between historical extinctions and modern deforestation according to the evidence on the likely extinction causes. The analysis is performed by augmenting Equation 1 with three new terms, which account for the number of groups extinct by assimilation, violence, or disease within 20 km of the grid cell. Some extinctions are assigned to more than one of these variables, since the likely causes are not mutually exclusive, while some extinctions are not assigned to any of them, reflecting cases in which the historical sources are not informative about the extinction causes. Table 6 reports the results for this exercise, adding each type of extinction separately (columns 1–3) and then including all of them together (column 4). The coefficient on "extinct groups" captures the association between deforestation and proximity to extinct groups without mention of these causes, while the coefficients on "extinct groups (violence)," "extinct groups (assimilation)," and "extinct groups (disease)" measure differences in the association when the sources mention each cause relative to this baseline.

The coefficients in column 1 show that assimilation extinctions are not significantly correlated with modern deforestation. The coefficient on the number of groups extinct by assimilation is negative, large, and statistically significant, fully offsetting the positive coefficient on "extinct groups". In contrast, the

[14]The occupation persistence mechanism does not clearly predict such heterogeneity, as many assimilation episodes followed contact with non-Indigenous actors.

[15]In contrast, loss of deterrence is consistent with recent non-Indigenous settlement, which would be influenced by the lower presence of traditional populations.

Table 6: Association with hectares of accumulated deforestation by 2010 – Heterogeneity by extinct group information

	(1)	(2)	(3)	(4)
Extinct groups	98.575***	64.248**	60.835**	78.738***
	(29.466)	(26.585)	(25.809)	(28.466)
Indigenous groups	-7.393	-6.414	-6.614	-7.549
	(13.383)	(13.495)	(13.490)	(13.371)
Extinct groups (assimilation)	-69.500			-112.246***
	(43.596)			(38.878)
Extinct groups (violence)		43.016		67.590
		(53.975)		(49.231)
Extinct groups (disease)			158.323**	205.074**
			(72.003)	(83.675)
P-value (extinct + assimilation)	0.430			0.301
P-value (extinct + violence)		0.043		0.006
P-value (extinct + disease)			0.003	0.001
R-squared	0.437	0.436	0.437	0.439
Observations	259574	259574	259574	259574

Note: The table displays estimates of the coefficients in Equation 1, augmented with interaction terms for likely extinction causes. The dependent variable is accumulated deforestation by 2010, in hectares, measured by PRODES. The estimated equation also includes latitude, longitude, and the vector of geographic fundamentals from the preferred specification. Standard errors in parentheses are multi-way clustered according to four interlaced grids of $1^\circ \times 1^\circ$. The p-values indicate the statistical significance of the sum of "Extinct groups" and "Extinct groups (assimilation)," "Extinct groups" and "Extinct groups (violence)," or "Extinct groups" and "Extinct groups (disease)." Statistical significance: * 10%, ** 5%, *** 1%.

coefficients on the number of groups extinct by violence (column 2) or disease (column 3) are positive and large, although not statistically significant for the case of violence. These estimates remain qualitatively similar in column 4.[16]

In sum, the historical evidence on likely extinction causes supports the view that Indigenous extinctions influenced long-run deforestation through the loss of deterrence against new settlers and agricultural expansion. The results suggest that the correlation between historical extinctions and modern deforestation does not hold when groups were assimilated, while becoming stronger when there is evidence of violence or diseases. These findings underscore the importance of population removal in the long-term relationship between extinctions and deforestation, consistent with the loss of deterrence mechanism. While removing

[16]Note that the coefficient on "extinct groups" indicates a positive association between extinctions for which we could not assign an extinction cause and deforestation. It is plausible to conjecture that violence and disease are more under-reported than assimilation, which was less frequently seen as morally reprehensible and often organized by institutionalized actors like religious groups, potentially explaining this positive coefficient.

Indigenous presence through violence or diseases creates a demographic void that facilitates non-Indigenous land claims, ownership, and land use change, the presence of an assimilated population seems to preserve some deterrence.

6.2 Evidence on Historical Occupation

The occupation persistence mechanism postulates that historical extinctions are linked to persistent historical non-Indigenous settlements from where modern deforestation would later expand. If this mechanism explains our findings, we expect to find historical non-Indigenous settlements near the areas where the extinctions took place. To test this prediction, we use the settlement dates of Amazon cities and historical land tenure data.

We estimate Equation 1 using as outcome variable an indicator of whether the grid cell is within 20 km from the center of a city first settled in one of two periods: before or after 1944. To construct these variables, we combine the location of modern-day municipality seats with settlement dates drawn from the text describing each municipality's history in *IBGE Cidades*. We use the earliest available year in which the municipality seat was already a settlement; this year usually predates the incorporation date of the municipality. The 1944 cutoff reflects the maximum date for extinctions in the Nimuendajú map so that settlement events after 1944 necessarily occurred after any recorded Indigenous extinction in the map. The occupation persistence mechanisms predicts that the positive association between historical extinctions and modern cities, reported in column 1 of Table 3, should hold for the cities settled before 1944.

In a similar vein, we examine the correlation between Indigenous extinctions and the historical area of private farms in 1920 and 1940, as surveyed by the agricultural censuses of those years, which we digitized. Unlike the data on modern farms, which contain property polygons, the agricultural census data is only available at the municipality level, so we estimate Equation 1 at the (1920 or 1940) municipality level instead of the 4 km $\times$ 4 km grid used in most regressions. In these cases, the variables $ExtinctGroups$ and $Groups$ are the number of groups within a 20 km distance from the municipality seats. We also include the geographic coordinates of the municipality seats and the same set of geographic fundamentals as control variables.[17] We weight each observation by municipality area.

The results, shown in Table 7, do not support the occupation persistence mechanism as an explanation

[17] For each municipality seat, we assign the geographic fundamentals of its 4 km $\times$ 4 km grid cell.

for why Indigenous extinctions are related to long-term deforestation. Columns 1 and 2 of Table 7 show that the correlation between extinctions and proximity to cities only holds for cities settled after 1944, potentially multiple decades after the extinctions documented in the ethnohistorical data. With respect to private farm ownership in 1920 and 1940, the coefficients reported in columns 3 and 4 show that the extinctions do not correlate with historical land tenure patterns. Historical Indigenous extinction are not associated with pre-1944 settlements or farm ownership, suggesting that the correlation with these variables has emerged well after the extinction events recorded by Nimuendajú.

Table 7: Association with historical settlement and land tenure

	City by settlement date:		Share of mun. in farms:	
	≤ 1944	> 1944	1920	1940
	(1)	(2)	(3)	(4)
Extinct group	0.014	0.036***	-0.009	-0.038
	(0.014)	(0.011)	(0.028)	(0.034)
Indigenous groups	-0.003	-0.000	-0.023	-0.021
	(0.006)	(0.007)	(0.015)	(0.026)
R-squared	0.138	0.124	0.389	0.283
Sample mean	0.059	0.061	0.058	0.067
Observations	259574	259574	117	116

Note: The table displays of the coefficients in Equation 1, at the grid cell level (columns 1 and 2), the 1920 municipality level (column 3), or the 1940 municipality level (column 4). In columns 1 and 2, the dependent variables are dummies indicating if the grid cell is within 20 km of a city (in 2010) which was first settled in the historical period indicated on the top. Information on the settlement dates of each city is from *IBGE Cidades*. In columns 3 and 4, the dependent variables are the shares of the 1920 or 1940 municipality area that was inside a surveyed farm in the 1920 or 1940 Agricultural Census. The estimated equation includes latitude, longitude, and all geographic fundamentals from our preferred specification. In columns 1 and 2, the sample consists of the grid cells. In columns 3 and 4, the sample consists of all the municipalities from 1920 or 1940 with municipality seats within the Brazilian Amazon. Standard errors in parentheses are multi-way clustered according to four interlaced grids of $1° \times 1°$. Statistical significance: * 10%, ** 5%, *** 1%.

6.3 Discussion

The evidence in this section favors loss of deterrence over occupation persistence. By considering heterogeneity according to the likely extinction causes, we find that demographic and ethnic voids matter for the long-run association between the extinctions and deforestation. In contrast, using historical data on settlements and land tenure, we find no evidence of the inception points predicted by occupation persistence. This does not mean that occupation persistence is unimportant in general, only that it does not explain the association between historical Indigenous extinctions and modern deforestation. As noted in the literature, settlement path dependence plays a role in many other contexts.

The loss of deterrence explanation implies that areas that lost Indigenous presence following the extinctions were more likely to be occupied by future non-Indigenous settlers. This explanation is broadly consistent with findings from Section 5: areas with extinctions display a lower share of self-identified Indigenous inhabitants in 2010, a higher share of whites and migrants from outside the Amazon, and land tenure patterns characterized by more private farms and government-sponsored colonization projects but fewer Indigenous lands.

An important question about the loss of deterrence mechanism is whether our findings are mostly explained by the environmental protection provided by these land tenure differences, particularly through the assignment of Indigenous lands by the government or through state-led colonization projects, or whether the active resistance by traditional populations is essential for the patterns we document, driving the correlation even in the absence of such formal protections. As discussed in the "horse-race" exercise in Appendix Table B.14, land tenure variables explain little of the correlation between extinctions and modern deforestation, suggesting that official tenure choices do not fully account for our findings.

As additional evidence on this question, we use information on the creation dates of officially recognized Indigenous lands, collected from the encyclopedia *Terras Indígenas do Brasil*, and of government-sponsored land distribution projects, available from INCRA, to investigate when the correlation between extinctions and these variables first appeared. Appendix Figure B.2 shows that the negative correlation of extinctions with the official recognition of Indigenous territories appears only in the 2000s and, similarly, the positive correlation with land distribution projects emerges only in the 1990s. Hence, the correlation between historical Indigenous extinctions and these variables postdates the association with deforestation, which already exists in 1986 (see Figure 5). This timing indicates that formal tenure assignments occurred after ongoing deforestation patterns near extinction sites, which is consistent with the interpretation that later land tenure changes are not the primary driver of the observed association.

7 Conclusion

This paper investigates long-run determinants of tropical deforestation. We study the connection between the historical extinctions of Indigenous groups and modern Amazon deforestation, relying on the Ethnohistorical Map of Curt Nimuendajú, remote-sensed measures of deforestation, and other data sources. We find a strong positive correlation between past Indigenous extinctions and modern deforestation. Since our historical information on Indigenous extinctions predates 1944, whereas deforestation only took off in the 1970s, our analysis guarantees chronological precedence of the extinction events relative to the main outcome, thereby ruling out reverse causality. Our estimates are robust to a series of tests aimed at addressing a variety of concerns, such as selection in geographical or group characteristics or measurement error in the ethnohistorical data. Moreover, areas near historical extinctions have more cities, fewer Indigenous territories, more private land ownership, and a lower Indigenous composition today. Since extinctions are not exogenous, our findings are correlational and do not have a causal interpretation. However, the robustness of the association between extinctions and deforestation across the various empirical exercises suggests that there is a historical nexus between these events.

We investigate the mechanisms connecting historical extinctions and modern deforestation, evaluating two possibilities: the deterrent role of traditional populations against the development of modern agricultural activities and the persistence of historical non-Indigenous settlements. We confirm a prediction of loss of deterrence by showing that extinctions related to demographic voids, such as violence and diseases, are associated with deforestation, while assimilation extinctions are not. In contrast, we refute a direct prediction of occupation persistence by showing that the association between the extinctions and cities or land tenure are not observed in historical data and arose only recently. Hence, our analysis favors the mechanism of Indigenous deterrence over occupation persistence.

To the best of our knowledge, this paper is the first to explore econometrically how historical events have influenced modern tropical deforestation. Our results suggest that Indigenous extinction events have facilitated the expansion of agriculture into the Amazon rainforest since the 1970s, resulting in large-scale deforestation. A key takeaway is that, beyond direct economic factors like land quality, the long-run influence of historical events on land use is significant. Therefore, a comprehensive understanding of the drivers of deforestation should include the history of traditional populations. These historical determinants must also be considered in policy discussions, as they shape economic incentives and could impact the effectiveness of conservation policies. We hope that this paper will stimulate further research on the historical roots of tropical deforestation, opening new avenues for understanding this important environmental problem.

References

Acemoglu, Daron, Simon Johnson, and James A Robinson (2001). "The colonial origins of comparative development: An empirical investigation". *American Economic Review* 91.5, pp. 1369–1401.

Agostini, Claudio A, Philip H Brown, and Andrei C Roman (2010). "Poverty and inequality among ethnic groups in Chile". *World Development* 38.7, pp. 1036–1046.

Albert, James S, Ana C Carnaval, Suzette GA Flantua, Lúcia G Lohmann, Camila C Ribas, Douglas Riff, Juan D Carrillo, Ying Fan, Jorge JP Figueiredo, Juan M Guayasamin, et al. (2023). "Human impacts outpace natural processes in the Amazon". *Science* 379.6630, eabo5003.

Alix-Garcia, Jennifer and Emily A. Sellars (2020). "Locational fundamentals, trade, and the changing urban landscape of Mexico". *Journal of Urban Economics* 116, p. 103213.

Alroy, John (2017). "Effects of habitat disturbance on tropical forest biodiversity". *Proceedings of the National Academy of Sciences* 114.23, pp. 6056–6061.

Angeles, Luis and Aldo Elizalde (2017). "Pre-colonial institutions and socioeconomic development: the case of Latin America". *Journal of Development Economics* 124, pp. 22–40.

Araujo, Rafael, Juliano Assunção, and Arthur Amorim Bragança (2023). "The Effects of Transportation Infrastructure on Deforestation in the Amazon: A General Equilibrium Approach". *Policy Research Working Papers* 10415.

Assunção, Juliano, Clarissa Gandour, and Romero Rocha (2023). "DETERring deforestation in the Amazon: environmental monitoring and law enforcement". *American Economic Journal: Applied Economics* VOL. 15.2.

Assunção, Juliano, Clarissa Gandour, and Rudi Rocha (2015). "Deforestation slowdown in the Brazilian Amazon: prices or policies?" *Environment and Development Economics* 20.6, pp. 697–722.

Baragwanath, Kathryn and Ella Bayi (2020). "Collective property rights reduce deforestation in the Brazilian Amazon". *Proceedings of the National Academy of Sciences* 117.34, pp. 20495–20502.

Barsanetti, Bruno (2021). "Cities on pre-Columbian paths". *Journal of Urban Economics* 122, p. 103317.

Barsanetti, Bruno and Alipio Ferreira (Nov. 2024a). "Dataset of Circumstances around the Extinction of Historical Indigenous Groups in the Brazilian Amazon, Based on Nimuendajú's Ethnohistorical Map". DOI: 10.5281/zenodo.14186381. URL: https://doi.org/10.5281/zenodo.14186381.

— (2024b). *Georeferenced and cataloged dataset built from Nimuendajú's Ethnohistorical Map*. Version Version 2, July 2024. DOI: 10.5281/zenodo.12772630. URL: https://doi.org/10.5281/zenodo.12772630.

BenYishay, Ariel, Silke Heuser, Daniel Runfola, and Rachel Trichler (2017). "Indigenous land rights and deforestation: evidence from the Brazilian Amazon". *Journal of Environmental Economics and Management* 86, pp. 29–47.

Bombardi, Fernanda A. (2014). *Pelos interstícios do olhar do colonizador: descimentos de índios no Estado do Maranhão e Grão-Pará (1680-1750)*. Master's dissertation, Universidade de São Paulo.

Brondizio, Eduardo S and Emilio F Moran (2012). "Level-dependent deforestation trajectories in the Brazilian Amazon from 1970 to 2001". *Population and Environment* 34, pp. 69–85.

Bruhn, Miriam and Francisco A Gallego (2012). "Good, bad, and ugly colonial activities: do they matter for economic development?" *Review of economics and statistics* 94.2, pp. 433–461.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller (2011). "Robust inference with multiway clustering". *Journal of Business & Economic Statistics* 29.2, pp. 238–249.

Carpio, Miguel A. and María E. Guerrero (2021). "Did the colonial mita cause a population collapse? What current surnames reveal in Peru". *The Journal of Economic History* 81.4, pp. 1015–1051.

Chimeli, Ariaster B and Rodrigo R Soares (2017). "The use of violence in illegal markets: Evidence from mahogany trade in the Brazilian Amazon". *American Economic Journal: Applied Economics* 9.4, pp. 30–57.

Cleary, David (2001). "Towards an environmental history of the Amazon: from prehistory to the nineteenth century". *Latin American Research Review*, pp. 65–96.

Conley, Timothy G. (1999). "GMM estimation with cross sectional dependence". *Journal of econometrics* 92.1, pp. 1–45.

Cunha, Manuela Carneiro da and Mauro WB De Almeida (2000). "Indigenous people, traditional people, and conservation in the Amazon". *Daedalus* 129.2, pp. 315–338.

Dean, Warren (2002). *Brazil and the struggle for rubber: a study in environmental history*. Cambridge University Press.

Dell, Melissa (2010). "The persistent effects of Peru's mining mita". *Econometrica* 78.6, pp. 1863–1903.

Denevan, William M. (2003). "The native population of Amazonia in 1492 reconsidered". *Revista de Indias* 63.227, pp. 175–188.

— (2014). "Estimating Amazonian Indian numbers in 1492". *Journal of Latin American Geography*, pp. 207–221.

Dippel, Christian (2014). "Forced coexistence and economic development: evidence from Native American reservations". *Econometrica* 82.6, pp. 2131–2165.

Elizalde, Aldo (2020). "On the economic effects of indigenous institutions: evidence from Mexico". *Journal of Development Economics* 147, p. 102530.

Fearnside, Philip M. (2005). "Deforestation in Brazilian Amazonia: history, rates, and consequences". *Conservation biology* 19.3, pp. 680–688.

Feir, Donn L, Rob Gillezeau, and Maggie EC Jones (2023). "Institutional drift, property rights, and economic development: Evidence from historical treaties". *National Bureau of Economic Research Working Paper*.

— (2024). "The slaughter of the bison and reversal of fortunes on the Great Plains". *Review of Economic Studies* 91.3, pp. 1634–1670.

Franco, Ana P., Sebastian Galiani, and Pablo Lavado (2021). "Long-term effects of the Inca Road". *National Bureau of Economic Research Working Paper*.

Gregório, Vitor Marcos (2009). "O progresso a vapor: navegação e desenvolvimento na Amazônia do século XIX". *Nova Economia* 19, pp. 185–212.

Grupioni, Luís D.B. (1998). *Coleções e Expedições Vigiadas: Os Etnólogos no Conselho de Fiscalização das Expedições Artísticas e Científicas no Brasil*. São Paulo: Hucitec.

Huaroto, Cesar and Francisco A Gallego (2023). "The Legacy of the Spanish Conquista in the Andes: Mining Mita, Persistent Social Unrest, and Cultural Divergence". *Persistent Social Unrest, and Cultural Divergence*.

INPE (2022). "Metodologia utilizada nos sistemas PRODES e DETER - 2a edição (atualizada)." *Instituto Nacional de Pesquisas Espaciais, Brazilian National Institute for Spatial Research. Available at:* `http://urlib.net/ibi/8JMKD3MGP3W34T/47GAF6S`.

Jacobs, James Q. (2023). *Amazon Geoglyphs XLS*. Tech. rep. Accessed 10/26/2023. Available at: `http://jqjacobs.net/archaeology/geoglyph.html`.

Jordán, Felipe, D Jaimovich, and R Heilmayr (2025). "The impacts of restrictions to individual rights on Indigenous lands". *Working paper*.

Kohlhepp, Gerd (2015). "Tipos de Colonização Agrária Dirigida nas Florestas Brasileiras: Exemplos históricos". *Fronteira: Journal of Social, Technological and Environmental Science* 4.3, pp. 102–121.

Lin, Jeffrey and Ferdinand Rauch (2022). "What future for history dependence in spatial economics?" *Regional Science and Urban Economics* 94, p. 103628.

Maeda, Eduardo Eiji, Temesgen Alemayehu Abera, Mika Siljander, Luiz EOC Aragão, Yhasmin Mendes de Moura, and Janne Heiskanen (2021). "Large-scale commodity agriculture exacerbates the climatic impacts of Amazonian deforestation". *Proceedings of the National Academy of Sciences* 118.7, e2023787118.

Maloney, William F. and Felipe Valencia Caicedo (2016). "The persistence of (subnational) fortune". *The Economic Journal* 126.598, pp. 2363–2401.

Masten, Matthew A and Alexandre Poirier (2022). "The Effect of Omitted Variables on the Sign of Regression Coefficients". *arXiv preprint arXiv:2208.00552*.

Mitchard, Edward TA (2018). "The tropical forest carbon cycle and climate change". *Nature* 559.7715, pp. 527–534.

Moraes, Lana, Carlos Carvalho, Manoel Rendeiro, and Tiago Gil (2021). "Eliminando manchas brancas: um desmonte da cartografia indigenista de Curt Nimuendajú". *História da Historiografia: International Journal of Theory and History of Historiography* 14.37, pp. 17–61.

Moreira Neto, C.A. (1988). *Índios da Amazônia, de Maioria a Minoria: 1750-1850.* Petrópolis: Vozes.

Mueller, Bernardo (2022). "Property rights and violence in indigenous land in Brazil". *Land Use Policy* 116, p. 106063.

Naritomi, Joana, Rodrigo R Soares, and Juliano J Assunção (2012). "Institutional development and colonial heritage within Brazil". *The journal of economic history* 72.2, pp. 393–422.

Nimuendajú, Curt (1921). "Bruchstücke aus Religion und Überlieferung der Šipáia-Indianer. Beiträge zur Kenninis der Indianerstämme des Xingú-Gebietes, Zentralbrasilien." *Anthropos* H. 1./3, pp. 367–406.

— (2017). *Mapa Etno-Histórico do Brasil e Regiões Adjacentes*. Brasília: Instituto do Patrimônio Histórico e Artístico Nacional, Instituto Brasileiro de Geografia e Estatística.

Nunn, Nathan (2020). "The historical roots of economic development". *Science* 367.6485, eaaz9986.

Oster, Emily (2019). "Unobservable selection and coefficient stability: Theory and evidence". *Journal of Business & Economic Statistics* 37.2, pp. 187–204.

Pádua, José Augusto (2015). "A Mata Atlântica e a Floresta Amazônica na construção do território brasileiro: estabelecendo um marco de análise". *Revista de História Regional* 20.2.

Pellegrina, Heitor and Sebastian Sotelo (2024). "Migration, specialization, and trade: Evidence from the Brazilian March to the West". *Journal of Political Economy (forthcoming)*.

Pereira, Leila and Rafael Pucci (2022). *A Tale of Gold and Blood: the Consequences of Market Deregulation on Local Violence*. Insper.

Pfaff, Alexander, Juan Robalino, Robert Walker, Steven Aldrich, Marcellus Caldas, Eustaquio Reis, Stephen Perz, Claudio Bohrer, Eugenio Arima, William Laurance, et al. (2007). "Road investments, spatial spillovers, and deforestation in the Brazilian Amazon". *Journal of Regional Science* 47.1, pp. 109–123.

Porro, Antonio (2007). *Dicionário etno-històrico da Amazônia Colnial*. Instituto de Estudos Brasileiros, USP.

Raymond, Hames (2007). "The ecologically noble savage debate". *Annu. Rev. Anthropol.* 36.1, pp. 177–190.

Ribeiro, Darcy (1957). *Línguas e culturas indígenas do Brasil*. Rio de Janeiro: Centro Brasileiro de Pesquisas Educacionais.

Ribeiro, Francisco de Paula (1841). "Memoria sobre as nações gentias que presentemente habitam o Continente do Maranhão: analyse de algumas tribus mais conhecidas: ..." *Revista Trimensal de Historia e Geographia ou Jornal do Instituto Historico e Geographico Brasileiro*, n. 10, p. 184–197; n. 11, p. 297–322; n. 12, p. 442–456.

Rondon, Candido Mariano da Silva (1916). *Conferências realizadas nos dias 5, 7, e 9 de outubro de 1915 no Theatro Phenix do Rio de Janeiro*.

Roosevelt, Anna C, Marcondes Lima da Costa, C Lopes Machado, Mostafa Michab, Norbert Mercier, Hélène Valladas, James Feathers, William Barnett, M Imazio da Silveira, Anne Henderson, et al. (1996). "Paleoindian cave dwellers in the Amazon: the peopling of the Americas". *Science* 272.5260, pp. 373–384.

Sellars, Emily A and Jennifer Alix-Garcia (2018). “Labor scarcity, land tenure, and historical legacy: Evidence from Mexico”. *Journal of Development Economics* 135, pp. 504–516.

Silva, Vinicius Okada da (2023). “Jesuit missionaries in the colonial Amazon: Long-term effects on human capital”. *Working paper*.

Souza, Francisco Bernardino de (1875). *Pará e Amazonas*. Comissao do Madeira, Rio de Janeiro, Typographia Nacional.

Souza, Márcio (1994). *Breve História da Amazônia*. São Paulo: Marco Zero.

Souza Jr, Carlos M, Julia Z. Shimbo, Marcos R Rosa, Leandro L Parente, Ane A. Alencar, Bernardo FT Rudorff, Heinrich Hasenack, Marcelo Matsumoto, Laerte G. Ferreira, Pedro WM Souza-Filho, et al. (2020). “Reconstructing three decades of land use and land cover changes in brazilian biomes with landsat archive and earth engine”. *Remote Sensing* 12.17, p. 2735.

Spolaore, Enrico and Romain Wacziarg (2013). “How deep are the roots of economic development?” *Journal of Economic Literature* 51.2, pp. 325–69.

Steward, Julian Haynes (1946). *Handbook of South American Indians*. 143. US Government Printing Office.

Stokes, Freg J, Sandra Benites, Anita Ekman, Uraan Anderson Suruí, Laura Furquim, Ricarda Winkelmann, Jürgen Renn, and Patrick Roberts (2025). “Tropical deforestation and Indigenous resistance over the longue durée in South America”. *Journal of Global History*, pp. 1–22.

Summerhill, William (2010). “Colonial institutions, slavery, inequality, and development: Evidence from São Paulo, Brazil”. *Working paper*.

Sze, Jocelyne S, L Roman Carrasco, Dylan Childs, and David P Edwards (2022). “Reduced deforestation and degradation in Indigenous Lands pan-tropically”. *Nature Sustainability* 5.2, pp. 123–130.

Tastevin, Constantin (1928). “Le Riozinho da Liberdade”. *La Géographie* 49, p. 3.

Valencia Caicedo, Felipe (2019). “The mission: human capital transmission, economic persistence, and culture in South America”. *The Quarterly Journal of Economics* 134.1, pp. 507–556.

Waldinger, Maria (2017). “The long-run effects of missionary orders in Mexico”. *Journal of Development Economics* 127, pp. 355–378.

Walker, Robert S, Jeffrey R Ferguson, Angelica Olmeda, Marcus J Hamilton, Jim Elghammer, and Briggs Buchanan (2023). “Predicting the geographic distribution of ancient Amazonian archaeological sites with machine learning”. *PeerJ* 11, e15137.

Online Appendix

A Data Appendix

A.1 Descriptive Information on Historical Deforestation

The data on deforestation that we use to prepare Figure 1 is from the *Instituto Brasileiro de Pesquisas Espacias* (Brazilian National Institute for Spatial Research, INPE). It is based on satellite imagery in the framework of the PRODES project. This project was launched in the late 1980s with the aim of providing a reliable measurement of deforestation of primary forest in the Brazilian Amazon forest, and it is still active. Indeed, INPE's website (URL: http://www.obt.inpe.br/OBT/assuntos/programas/amazonia/prodes) provides information on yearly deforestation since 1988.

It is tricky to obtain reliable estimates of deforestation prior to 1988. Luckily, INPE had some attempts at estimating deforestation using early satellite imagery from the 1970s and early 1980s. INPE (1989) provides some statistics on deforestation starting in 1975, which we complement with information from INPE (1991). In particular, INPE (1991) explicitly mentions around 100 thousand square kilometers of deforestation prior to 1975 which is omitted in INPE (1989), and corrects upwards the deforestation in several states. We use these documents to draw the years before 1988 in Figure 1. The information mainly stems from INPE (1989), but we linearly correct the values according to corrections applied for each state in INPE (1991).

The conclusion based on the INPE series is that accumulated deforestation was indeed very low by 1975. Even accounting for the 100 thousand square kilometers of old deforestation, only about 3% of the primary forest cover had been deforested by 1975. Deforestation then picked up at a rate of 18 thousand square kilometers per year on average until the mid-2000s. Consequently, 85% of the deforestation of the Brazilian Amazon happened after 1975.

Since all these estimates are based on remote-sensing measures, and satellite images have been available only since the 1970s, it is even harder to estimate Amazon deforestation prior to 1975. A natural possibility is to examine historical agricultural censuses. Consider the 1940 Agricultural Census, which is roughly contemporary with the Nimuendajú map. It contains the information on the areas used for pastures and cropland; given the low urbanization of the region, most deforestation was likely occupied by these land uses. In 1940, they covered only 0.89% of the total area of the municipalities with seats within the Amazon biome. This number confirms that, at the time the Nimuendajú map was prepared, Amazon deforestation was indeed very low.

A.2 Ethnohistorical Data

Our main data contribution is the preparation of the information in the Ethnohistorical Map of Curt Nimuendajú for quantitative use. Our primary objective is to include as much of the rich information recorded by Nimuendajú following a transparent and manageable methodology which would be easily understood and replicable. Throughout this process, we sometimes had to take specific choices in how to prepare the data, and our decision has been to retain simple and transparent procedures. In the appendix below, we describe how we prepare the ethnohistorical data. Our ethnohistorical database is publicly available as Barsanetti and Ferreira (2024b).

Our work of converting the Nimuendajú map into useful spatial data is based on the high-resolution PDF file in the 2017 edition of the map. The map displays vertical and horizontal lines that comprise a $5° \times 5°$ grid. We use the intersection points of the lines to georeference the map. There are 67 such points. We georeference the map on QGIS Georeferencer Plugin, using a thin plate spline as the transformation type and a nearest-neighbor resampling. The resulting raster has the projection EPSG:4326 (WGS 84).

To identify the location of the map entries, we prepare a points shapefile on top of the Nimuendajú map raster. For each entry, we input two points, at each extreme point of the entry. If the entry is marked by a colorful line indicating the linguistic family of the group, we input the extreme points roughly at where the line ends. If there is no line (unknown language groups), we input the points at the extreme ends of the name.

We use a points shapefile instead of a line shapefile so the large information in the map could be imputed in a manageable way. We record the two extreme points to account for the range of the group and to facilitate reproducibility. Note this procedure entails some loss of information for the groups whose entries are curved lines, and could thus introduce some measurement error to our estimates. Most map entries are straight lines, so our procedure does not distort the distances for these cases. However, the shape of other entries remind a semi-circle or a S. To address concerns of measurement error with respect to them, we also manually add other points for the entries with non-straight shapes. For semi-circles, we add a third point along the curve, roughly equidistant from the extreme points, with which we can draw a triangle that approximates the curve. For S-shaped entries, we introduce two other points along the curves, roughly splitting the entry in three equal length parts which resemble a Z. We use these points to approximate the distances to the curves in a robustness test, showing that the approximation of the entry locations with straight lines does not impact our results.

On top of georeferencing and recording the map locations, our data set also includes the information that Nimuendajú recorded through text, color or font. For each map entry, we record the following information in the shapefile:

1. Group name, as written on the map.

2. All the dates which are written under the group name. We use this information to assign a year to dated entries. Sometimes Nimuendajú assigns more than one date for an entry. In this case, we assign the oldest year recorded to the entry. In a few cases, the date is not precise but indicates only the century; in this case, we assign the midpoint year in the century (e.g., 1850 for 19th century).

3. The calligraphic font used by Nimuendajú to write the group name.

4. The linguistic family, according to the color of the line and the legend at the left corner of the map.

5. A dummy variable indicating if the group is marked as extinct. This is done by examining whether the group name is written with a fine calligraphic font. In a few cases, the font in the 2017 map was confusing to understand, so in these cases we also examined a picture of the original 1944 map to ensure the correct assignment of this dummy. Moreover, for those entries which present a bar (i.e. feature a known linguistic family), the extinction information is also indicated by a dashed line pattern. To guarantee accuracy in the extinct information, we assign extinction status based on both the font and the line patterns; we then cross-check the resulting information.

Beyond the information directly available from the map, we digitize the index and the list of references in Nimuendajú (2017) and match each group to the respective references. This provides the information on how many sources refer to each group in the map. We manually assign a language to each reference, which allows us to redraw the ethnohistorical map according to the languages of the references.

A.3 Other Data

We complement our ethnohistorical data with a long list of other spatial databases, which we usually prepare as additional variables at the grid cell level.

Other outcome variables

Deforestation. To construct our main measure of deforestation, we use rasters from the *Projeto de Monitoramento do Desmatamento na Amazônia Legal por Satélite* (Satellite-based Legal Amazon Deforestation Satellite Monitoring Program, PRODES). The PRODES maps have been prepared every year since 1988 by the Brazilian National Institute for Spatial Research (*Instituto Nacional de Pesquisas Espaciais*, INPE). Consistent rasters of accumulated deforestation are publicly available since 2007 at Inpe's website, with a 30 m resolution. For each of these years, we compute the number of pixels within the grid cell that have been deforested and transform this count into area by considering that each pixel has 0.09 ha (900 sq. m). The PRODES rasters are available as "PRODES completo em formato matricial" at `https://terrabrasilis.dpi.inpe.br/downloads`.

Land use. We use maps of land use from MapBiomas (collection 8), produced using Landsat imagery at 1-arcsec resolution. MapBiomas maps classify each pixel according to several potential categories. We use primary forest data from the Deforestation and Secondary Vegetation maps; we use the maps for the years 1986, 1990, 1995, 2000, 2005, 2010, 2015, and 2020. From the Land Cover and Land Use map for 2010, we use the following variables: cropland, pasture, farming (cropland plus pasture, including pixels with a mosaic of both), and urban land. For each grid cell, we compute the number of pixels in each of these category and transform these counts into area according to the raster resolution. The MapBiomas rasters are available at the project website: `https://brasil.mapbiomas.org`.

Nightlights. We use the Defense Meteorological Program Operational Line-Scan System (DMSP-OLS) nighttime lights to measure the distribution of urban economic activity. We use the raster on average visible, stable lights, and cloud-free coverage for 2010, which we average within each grid cell. The nightlights raster is available at `https://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html`.

Population centers. We use the location of population settlements from the list of localities cataloged by the *Instituto Brasileiro de Geografia e Estatística* (Brazilian Institute of Geography and Statistics, IBGE) for the 2010 census in the *Cadastro de Localidades Selecionadas*. We consider three categories of locations.

The first category includes municipality seats (which we refer to as "cities"), which are the main urban center in the municipality areas. The second category includes any population center in the localities data that is not exclusively Indigenous, including municipality seats, municipal district seats (secondary urban centers), and rural villages; in the paper, we refer to these localities as settlements. The third category includes the Indigenous villages. We define a grid cell as close to each of these categories if its centroid is within 20 km of one. The georeferenced list of all these localities is available at https://www.ibge.gov.br/geociencias/organizacao-do-territorio/estrutura-territorial/27385-localidades.html?edicao=27386.

Road infrastructure. We use georeferenced data on Brazilian roads from version 4 of IBGE's *Base cartográfica contínua ao milionésimo*, a comprehensive dataset mapping Brazil. The road data is based on georeferenced information collected in 2007 and 2011. For each grid cell, we measure the length, in meters, of the roads crossing the cell. The shapefiles are available at https://www.ibge.gov.br/geociencias/cartas-e-mapas/bases-cartograficas-continuas/15759-brasil.html.

Railroad infrastructure. The data on railroads is also from version 4 of IBGE's *Base cartográfica contínua ao milionésimo*. The railroad data refers to the information available from the Ministry of Transportation in the *Banco de Informações de Transportes* for the period 2007–2009. For each grid cell, we measure the length, in meters, of the railroads crossing the cell. See the item on road infrastructure for where to access the shapefiles.

Indigenous lands. The location of protected Indigenous lands stems from shapefiles released by Funai, the federal institution responsible for Indigenous policy. By scrapping information on each Indigenous land from the website *Terras Indígenas no Brasil*, we obtain the dates of the decrees that created these reserves so we can restrict our analysis to the ones created up to 2010. More precisely, we use the decrees that "homologated" each land. Homologated lands are the ones receiving full legal protection from the Brazilian state. We then use this information to assign, to each grid cell, a dummy variable indicating whether it is intersected by some Indigenous land that was homologated up to 2010. For the results presented in Section 6, we use different cutoff years with respect to the homologation dates. The Funai shapefiles on Indigenous lands are available for dowload at https://www.gov.br/funai/pt-br/atuacao/terras-indigenas/geoprocessamento-e-mapas. The information on each Indigenous land is available from https://terrasindigenas.org.br.

Private land ownership. To measure the location of private farms, we rely on the Agricultural Atlas, a consolidated map of land tenure prepared by the NGO Imaflora drawing from multiple official registries. This atlas refers to 2020; since the CAR is a recent registry, we cannot measure private farm locations in 2010, the year we use for most other variables. The location of private farms is based on an official public registry called *Sistema de Gestão Fundiária* (Land Tenure System, SIGEF), complemented by a recent self-declaratory dataset, the *Cadastro Ambiental Rural* (Rural Environmental Registry, CAR), managed by the Brazilian Ministry of the Environment. For detailed information on these data, see Freitas et al. (2018). Based on these polygons, we measure the area of private farms within each grid cell, in hectares. These data is available for download as shapefiles at `https://atlasagropecuario.imaflora.org/`.

Land distribution projects. Shapefiles of land settlement projects are from the website *Acervo Fundiário* held by the *Instituto Nacional de Colonização e Reforma Agrária* (National Institute of Colonization and Land Reform, INCRA). We use the shapefile "Projetos de Assentamento Total". We use the creation dates available as an attribute of the shapefile to know whether each project existed by 2010. For the results presented in Section 6, we use different cutoff years with respect to the creation dates. The shapefiles are available for download at `http://acervofundiario.incra.gov.br`.

Ethnic composition. To obtain information on the ethnic composition of the population across space, we use the results of the 2010 Demographic Census which were reported at the *setor censitário* (census block) level. We match these data with the shapefile of *setores censitários* from IBGE. We further measure all geographic controls at the *setor censitário* level. The ethnic classification of the population is self-reported into five categories: white, black, ethnically mixed (*pardo*), Asian, and Indigenous. The shapefile of *setor censitários* is available for download at `https://www.ibge.gov.br/geociencias/organizacao-do-territorio/malhas-territoriais/26565-malhas-de-setores-censitarios-divisoes-intramunicipais.html`. The census data at the *setor censitário* level is available at `https://ftp.ibge.gov.br/Censos/Censo_Demografico_2010/Resultados_do_Universo/Agregados_por_Setores_Censitarios`.

Migration. As information about the migration history of the population inhabiting the Brazilian Amazon, we use information on the place of birth. This information is available from the microdata of the 2010 Demographic Census. In each municipality, a representative sample of the municipal population answered an extended questionnaire which recorded the country and state of birth. We refer to an individual as a migrant from outside the Amazon if they were born outside one of the states which contain the Brazilian Amazon (Acre, Amapá, Amazonas, Maranhão, Mato Grosso, Pará, Rondônia, Roraima, Tocantins). The census microdata is available for download at https://ftp.ibge.gov.br/Censos/Censo_Demografico_2010/Resultados_Gerais_da_Amostra/Microdados.

Municipality settlement dates. For each municipality in the Amazon region (as of 2010), we identify their settlement dates by scrapping the municipality histories available from the website *IBGE Cidades*. We then split the municipalities into two categories: first settled up to 1944 (the year of the Nimuendajú map), and first settled after 1944. For use as a control variable in a robustness test, we also consider municipalities settled before 1800. *IBGE Cidades* can be accessed at https://cidades.ibge.gov.br.

1920 Agricultural Census. To obtain information on the extent of farm area in 1920, we manually digitize the information in Table I from the agricultural section of the 1920 Brazilian Census. This table contains both the farm area and the total municipality areas. The .pdf file with the agricultural information from the 1920 Census is available for download at https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=26452.

1940 Agricultural Census. To obtain information on the extent of farm area in 1940, we manually digitize the information in Table 29 from the agricultural section of the 1940 Brazilian Census. We also digitize the municipality areas reported in the census. The .pdf files with the 1940 Agricultural Census is available for download at https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?id=765&view=detalhes.

Control variables

Relief. Elevation data is from the CGIAR-SRTM dataset at a 30-arcsec resolution. For each grid cell in our sample, we compute the average altitude and terrain ruggedness index (TRI) in meters. The elevation data is available for download at https://bigdata.cgiar.org/srtm-90m-digital-elevation-database.

Water coverage. We compute the water coverage of each grid cell, as measured by PRODES. As for deforestation, we measure water coverage in hectares. See the item on deforestation for where to access the data.

Non-forest coverage. We compute the original non-forest formation coverage of each grid cell, as measured by PRODES. As for deforestation, we measure non-forest coverage in hectares. See the item on deforestation for where to access the data.

Malaria suitability. We measure the climatic suitability of malaria with information from the Malaria Atlas, averaging the suitability indexes for *Plasmodium vivax* and *P. falciparum*; see Gething et al. (2011) for details on these indexes. We average the resulting index within each grid cell. The rasters are available for download at `https://data.malariaatlas.org`.

Climate variables. To measure climatic fundamentals, we use average temperatures, precipitation, and relative humidity from the ten-minute climatology dataset of the Climatic Research Unit (CRU); see New et al. (2002). We match each grid cell with the closest point in the CRU data, and assign it the information for that point. The CRU data is available for download at `https://crudata.uea.ac.uk/cru/data/hrg/tmc`.

Rivers. We identify the main rivers of the Amazon Basin based on shapefiles and maps from the *Agência Nacional de Águas* (National Water Agency, ANA). The main rivers in our sample are (in alphabetical order, as written in the shapefile) the: Abacaxi, Amazonas, Anamu, Araguaia, Arinos, Aripuanã, Branco, Canumã, Culuene, Cuniuá, Envira, Guaporé, Içá, Iriri, Iriri Novo, Jamanxim, Japurá, Jari, Jatapu, Javari, Juína, Juruá, Juruena, Jutaí, Madeira, Mamoré, Negro, Parima, Purus, Roosevelt, São Manuel, São Manoel, Solimões, Sucunduri, Tapajós, Tapauá, Tarauacá, Teles Pires, Tocantins, Trombetas, Uanamu, Uatumã, Uaupés, Uraricoera, and Xingu; these are the rivers highlighted in the map from ANA (ANA 2015). We use the river shapefile to assign each grid cell a dummy which indicates if it is within the 10 km buffer around these main rivers. The geopackage with the original river shapefiles is available at `https://metadados.snirh.gov.br/geonetwork/srv/api/records/0f57c8a0-6a0f-4283-8ce3-114ba904b9fe`.

Rubber tree range. We use the range of *Hevea brasiliensis*, the most productive and economically relevant species of rubber trees, mapped by the *Plano Nacional da Borracha* (National Rubber Plan); see Comércio (1971). We georeference the map on the range of *H. brasiliensis* (Figure 3 of the document), create a polygon with the range, and assign a dummy equal to one if the grid cell intersects the species range. The .pdf file with the map we georeference is available for download at `https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=247667`.

Crop suitability. We use the rain-fed agro-ecological attainable yields of cassava and cacao under low-input use and of soybeans under high-input use from FAO's Global Agro-Ecological Zones project (FAO-

GAEZ v.4). We average each of these attainable yields within each grid cell. The rasters are available at https://gaez.fao.org/pages/data-viewer-theme-4.

Distance to São Paulo. We use the distance between each grid cell and the city of São Paulo as a measure of remoteness and distance to the frontier. We calculate these distances using the Haversine formula, using the coordinates (23.5558 S, 46.6396 W) for São Paulo.

Caloric suitability. We use the caloric suitability index of Galor and Özak 2015, Galor and Özak 2016 and Galor, Özak, and Sarid 2016, which measures how many calories can be produced in a hectare. We average their index within each grid cell. The rasters with caloric suitability are available for download from one of the authors at https://ozak.github.io/Caloric-Suitability-Index.

1940 Population Census. To obtain a measure of proximity to non-Indigenous population at the time of the Nimuendajú map, we use the shapefiles released by IBGE on the 1940 municipality seats; see IBGE (2011). We also digitize the column "População de Fato" from Table 48 of the 1940 Brazilian Census. We digitize these values for all municipalities that belong to the states that contain the Brazilian Amazon. For each grid cell, we compute: **(i)** the distance to the closest 1940 municipality seat, and **(ii)** a market potential variable, which sums the municipality populations divided by the distances to their municipality seats. The .pdf files with the census information are available for download at https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?id=765&view=detalhes.

Archaeological sites. The location of archaeological sites is from the official list of Brazilian archaeological sites (*Cadastro Nacional de Sítios Arqueológicos*, CNSA) held by the *Instituto do Patrimônio Histórico e Artístico Nacional* (IPHAN) and publicly available as a points shapefile. We consider any site recorded as "Pré colonial" or "Pré contato" in this database as pre-colonial. We consider any site recorded as "Histórico" or "Pós contato" in this database as historical (that is, after the European arrival). The shapefiles are available for download at http://portal.iphan.gov.br/pagina/detalhes/1701/

Pre-Columbian earthworks. The precise georeferenced location of 1269 geoglyphs and mound villages is from a publicly available compilation made by Jacobs (2023). These locations are based on many references in the literature and most were checked on satellite imagery by Walker et al. 2023. The georeferenced list is available in Excel format at https://jqjacobs.net/archaeology/geoglyph.html.

Historical aldeamentos. The list of 66 religious *aldeamentos* circa 1730 in the Portuguese Amazon is from Bombardi (2014). We obtain their location by georeferencing Map 6 in page 40 of her thesis, which is available at https://www.teses.usp.br/teses/disponiveis/8/8138/tde-08012015-164954/en.php.

A.4 Data and Code Availability

Our georeferenced ethnohistorical database is deposited in a public repository under the DOI 10.5281/zenodo.8338508. Our classification of likely extinction causes, including a dictionary explaining it, is publicly available under the DOI 10.5281/zenodo.14186380. All other data used are also public, with their sources listed above. The data files and the R and Stata codes used in the analysis are available as part of a replication package deposited in a public repository under the DOI 10.5281/zenodo.17153102.

B Additional Empirical Results

B.1 Additional Discussion of Robustness Tests

B.1.1 Selection in Omitted Variables

Selection in geographic variables is one of our main concerns with respect to bias in our estimates. In our main analysis, we carefully address this issue by including a long list of geographic variables as controls. Despite considering a comprehensive list of fundamentals based on the history of the region and of known determinants of deforestation, there could still be other relevant variables that determine both deforestation and extinctions which are missing from these controls. Some of these factors may even be non-observable. Our failure to include them would bias the estimates for the association between deforestation and extinctions. To investigate the severity of this problem, we use the method from Oster (2019), which estimates the minimum degree of selection in the omitted variables, relative to selection in the included variables, so that our positive coefficients would be consistent with a null association between extinctions and deforestation if all the omitted variables were also included. Following Masten and Poirier (2022), we refer to the estimates for this relative degree of selection in omitted variables as the explain-away breakdown point. As we show below, this breakdown point is too high for the selection in omitted variables to be a plausible threat, since it requires a degree of selection in omitted variables that is much greater than the selection in included controls.

For this exercise, we make the assumption that including the omitted fundamentals in the regression would lead to a model that fully explains the variation in the dependent variable (that is, with an R-squared equal to 1). This is the most conservative assumption for this method; see Oster (2019) for details. The estimates indicate that, for selection in omitted fundamentals to fully explain away our findings, the degree of selection in omitted variables would have to be at least 2.341 times larger than the degree of selection in the long list of geographical fundamentals included as control variables (altitude, terrain ruggedness, water coverage, non-forest coverage, malaria suitability, temperature, rainfall, relative humidity, proximity to main rivers, Hevea brasiliensis range indicator, agro-ecological potential yields of soybeans, cassava and cacao, caloric suitability, and linear distance to São Paulo). Since our analysis already includes all the most relevant variables that could bias our estimates if omitted, this test indicates that it is unlikely that our findings could be explained by geographic variables we fail to include as controls.

The value of 2.341 refers to the smallest ratio of selection on omitted variables relative to selection on included controls for which the data would be consistent with a zero association between Indigenous extinctions and deforestation. Masten and Poirier (2022) refer to this ratio as the explain away breakdown point. Masten and Poirier (2022) also show that sometimes a lower ratio allows the data to be consistent with a non-zero coefficient of opposite sign; they refer to this alternative ratio as the sign-change breakdown point. If we impose no magnitude restriction on the negative coefficient that could be consistent with the data, then the sign-change breakdown point is 1; note that 1 is the highest possible absolute value for the sign-change breakdown point according to Theorem 4 in Masten and Poirier (2022). However, if we impose a magnitude restriction of 1,640 ha (the size of the grid cell) on the absolute value of the coefficient, then the sign change breakdown point becomes equal to the explain away breakdown point, 2.341.

Moreover, the method by Oster (2019) can be inverted to produce bounds on the association between extinctions and deforestation. Instead of estimating the explain away breakdown point, we can assume an interval for the degree of selection in omitted controls relative to selection in included geographical controls and derive bounds on the coefficient β. We calculate these bounds under the assumption that the degree of selection in the omitted controls is not stronger, in absolute terms, than the degree of selection in the included geographic controls. By doing so, we find a lower bound of 46.827 and an upper bound of 98.346 to the coefficient on extinctions.

B.1.2 Alternative Measures of Proximity to Groups

Our main proximity measures count the number of Indigenous group locations in the Nimuendajú map within 20 km of each grid cell centroid. We construct these counts both for extinct groups, which is our main variable of interest, and for all groups, which we use as a control. The distance from a centroid to a map entry is the minimum distance to the straight line connecting the two extreme points of the entry. Table B.7 shows that our results are robust to alternative definitions of proximity.

First, column 1 shows that the results hold when distance is measured to the midpoint of the entry rather than to the line segment. This approach removes any influence of the entry's range, which may reflect the group's actual extent but could also arise from uncertainty, long names, or vcartographic style choices by Nimuendajú.

Second, we consider a different way to approximate the shape of each map entry. For tractability and simplicity, our main results measure distance from the grid cell centroid to the straight line connecting the two extreme points of the entry. Most entries are straight lines, so this approach introduces no distortion. A small number of entries, however, have curved shapes, either semicircles or, less frequently, S shapes. To approximate these shapes more closely, we add extra points along the curve: one point at the midpoint for semicircles, and two intermediate points for S shapes. We then connect these points to form a polyline figure that approximate the shape of the map entry and compute distances to this figure. Column 2 shows that using this alternative method produces estimates that are very similar to our main results, indicating that accounting for the curvature in the original map does not affect our findings.

Third, we vary the cutoff radius used to count nearby groups. Columns 3 to 6 report coefficients for cutoffs ranging from 5 to 40 km. The association between proximity to extinct groups and deforestation remains positive and declines monotonically as the cutoff increases, consistent with the idea that precise group locations are important predictors of deforestation.

Fourth, we replace counts with dummies indicating whether any group lies within 20 km of a cell. This divides the sample into three regions. The first region contains the 11% of cells within 20 km of an extinct group. The second contains the 31% of cells within 20 km of a non-extinct group only. The third contains the remaining 59% of cells that are farther than 20 km from any Indigenous group. Column 6 shows that, conditional on controls, deforestation is on average 81.6 ha higher in the first region than in the second, and 75.0 ha higher than in the third.

Finally, we use the distances, in kilometers, to the nearest extinct group and the nearest group. The results, reported in column 7, confirm the same qualitative patterns.

B.1.3 Measurement error

Measurement error is an important concern when using historical data sources such as the Nimuendajú map, as it may bias the coefficient estimates. Measurement error may arise due to the quality of the information in the map, distortions in the group locations, or the selective recording of the Indigenous groups and extinctions by the sources consulted by Nimuendajú. We present five robustness tests below to address these measurement errors concerns.

First, note that Nimuendajú relied on almost a thousand sources to place the groups on the map and to assign the extinction status. Hence, the map quality depends on the sources' quality and quantity, both of which are unevenly distributed across groups. In column 1 of Table B.8, we use the number of sources assigned to each entry to address concerns relative to the amount of information used by Nimuendajú. We restrict the Indigenous groups only to those cited by at least two references. By doing so, our analysis is based only on better documented (in a quantitative sense) map entries. The coefficients are robust to this restriction, with proximity to each well-documented extinct Indigenous group associated with additional 72.7 ha of accumulated deforestation by 2010.

Second, map distortions in the production and georeferencing of Nimuendajú's ethnohistorical map are another potential source of measurement error. To address this concern, we exploit the map markings, vertical and horizontal lines forming a $5° \times 5°$ grid, which we use to georeference the map. By construction, there are no distortions near the marks, and the map distortions plausibly increase as a grid cell is farther away from these markings. Hence, we mitigate this measurement error concern if we weight the observations by the inverse of their distance to the map markings. The estimates, shown in column 2 of Table B.8, are again robust to this procedure.

A third measurement error concern is that, since the map is based mainly on European or Euroamerican sources, group locations in the Nimuendajú map may be influenced by the location of non-Indigenous settlements at the time Nimuendajú worked on the map. These settlements may also correlate with modern deforestation, especially if later agricultural expansion followed from pre-existing non-Indigenous settlements. As evidence that our findings are not due to this concern, we consider two variables related to population distribution as measured by the 1940 Census, which is roughly contemporary to the Nimuendajú map (from 1944). First, we use a dummy of proximity to a municipality seat in 1940. Second, we use the population tables in the census to construct a measure of market potential, which averages the population of surrounding towns by the inverse of their distance. In Table B.9, we show that these two variables do not feature strong correlations with Indigenous extinctions and that our findings are robust to their inclusion as control variables.

A related non-classical measurement error concern is that the information about the Indigenous groups reflects locations visited by Brazilian sources, which may be related to the occupation of the Brazilian Amazon by non-Indigenous settlers. Fortunately, Nimuendajú used a variety of references which were written in many different languages and published in different countries. His map does not rely exclusively on Brazilian sources. We manually classify the languages of the sources used by Nimuendajú and find that, although Portuguese is the most common language, it accounts for only 39.6% of the references. Sources in German (17.1% of the references), Spanish (14.9%), French (14.9%) and English (8.6%) were also extensively used by Nimuendajú. There are also a few references in Italian, Swedish, Dutch and Latin. To better account for this concern, we show that our findings are robust to the restriction of the measures of proximity to Indigenous groups to those cited by at least one reference in a language other than Portuguese; see column 3 of Table B.8. By doing so, the reflect only proximity to groups recorded by non-Brazilian sources, in which case the concern above is less relevant.

Finally, measurement error could also occur in the classification of the Indigenous group as extinct by Nimuendajú. Nimuendajú did not detail the exact procedure he used to assign extinction status to the map entries. Based on our own reading of the sources, such assignment could have occurred in two ways. In many cases, the references used by Nimuendajú are quite specific about the extinction, sometimes even containing a historical account of it. For these groups, the extinction assignment is *positive*, that is, based on an explicit reference to the extinction of the groups. In many other cases, in contrast, the references used by Nimuendajú do not explicitly indicate that the group was extinct. Nimuendajú may have assigned them an extinct status based on his own field knowledge or on the lack of more recent sources which indicate a group with similar name and characteristics in the region. In particular, this may have been the case for earlier group entries, which were often scarcely recorded in the historical sources and sometimes under unusual or vague names. For these groups, the extinction assignment is *negative*, that is, based on an absence of modern references to these groups. Such difference in the extinction assignment procedure may result in measurement errors, as the extinction information for positively assigned extinct groups is plausibly more reliable than for negatively assigned ones.

To address this concern, we use our systematic reading of the historical sources, explained and listed in detail in Barsanetti and Ferreira (2024a), and we show that our findings are indeed driven by proximity to positively assigned extinct groups. Based on our own reading of the sources, we identify whether we could confirm the extinction in the sources. Our reading of the sources confirms the extinction of roughly half (48%) of the extinct groups used in our baseline analysis. In Table B.10, we estimate coefficients for proximity to confirmed and unconfirmed extinct groups. In column 1, we include only the proximity to confirmed groups, while in column 2 we include proximity to both types of extinct groups. The show that our findings are mainly due to extinctions which are well documented in the historical sources: the coefficient on confirmed extinct groups (102.9 ha) is twice as large as the coefficient on unconfirmed extinct groups (47.7), although the latter is still large and statistically significant at the 10% level.

B.1.4 Alternative Estimation and Inference

Our findings are robust to alternative estimation methods, as we show in Table B.12.

We first consider alternative approaches to inference, that is, to the estimation of the standard errors of the equation coefficient. For our baseline , we account for heteroskedasticity and spatial correlation by multi-way clustering according to the four interlaced grids of $1^\circ \times 1^\circ$. In column 1, we estimate robust standard errors using instead the approach proposed by Conley (1999); although this method is computationally demanding given our large sample and each equation takes many hours to be estimated, the standard error remain similar. In column 2, we show that our findings are robust to clustering standard errors by municipality, the administrative boundaries that define local governments and within which deforestation may be correlated due to local economic shocks.

We also consider an alternative estimation method for the coefficients of Equation 1. By construction, our dependent variable, the area of accumulated deforestation within a grid cell, has values bounded between 0 and 16,040 ha (the size of the grid cell). The linear model we use may lead to predicted values that fall outside these bounds. In fact, for 16.8% of our grid cells, our linear model predicts negative values of accumulated deforestation. A natural question is whether our findings hold when estimated in a model that does not allow for out-of-bounds predicted values. In column 3 of Table B.12, we show that our findings hold when we estimate a tobit model in which the dependent variable has these bounds. Note that the estimated coefficient on extinct Indigenous groups, 149.7, is larger than the ordinary least squares coefficients. This difference may be explained by the fact that, relative to the ordinary least squares, the tobit model assigns more weight to variations in deforestation and extinct groups across grid cells for which the geographic fundamentals predict higher deforestation. These are likely to be the regions where the historical extinctions may feature a stronger association with modern deforestation.

B.2 Appendix Tables and Figures

Table B.1: Summary statistics of grid sample

	Mean (1)	Std. Dev. (2)
Distance to closest map entry (km)	33.79	(29.68)
Distance to closest extinct group map entry (km)	76.09	(48.46)
Map entries within 20 km	0.59	(0.85)
Extinct group map entries within 20 km	0.12	(0.38)
Cumulative deforestation area (ha) in 2010	260.72	(487.26)
Additional deforestation area (ha), 2010-2020	26.15	(97.67)
Primary forest area (ha), 2010	1348.68	(476.24)
Farming area (ha), 2010	218.45	(408.02)
Cropland area (ha) 2010	10.78	(82.71)
Pasture area (ha), 2010	206.75	(395.36)
Urban area (ha), 2010	1.43	(25.87)
Nighlight intensity, 2010	0.20	(2.11)
Distance to nearest city (km), 2010	83.75	(64.73)
Distance to nearest settlement (km), 2010	61.17	(54.01)
Road extension (m), 2011	288.16	(1114.10)
Railroad extension (m), 2009	6.56	(163.32)
Protected indigenous land, 2010	0.25	(0.43)
Distance to nearest indigenous village (km), 2010	99.03	(68.34)
Privately owned farms (ha), 2020	391.32	(603.52)
Land reform projects (ha), 2010	1.14	(3.80)
Latitude	-5.03	(4.34)
Longitude	-59.22	(6.76)
Altitude (m)	173.59	(129.18)
Terrain ruggedness index (m)	30.44	(37.60)
Water (ha)	38.24	(172.34)
Non-forest coverage (ha)	114.02	(357.60)
Malaria index	0.76	(0.06)
Average temperature (C)	26.12	(0.70)
Average monthly rainfall (mm)	189.24	(29.42)
Average relative humidity (%)	83.51	(3.81)
Close (up to 10 km) of a main river	0.17	(0.37)
Hevea br. range	0.57	(0.50)
Attainable yield: soybeans (kg DM)	1525.77	(594.90)
Attainable yield: cassava (kg DM)	904.66	(533.22)
Attainable yield: cacao (kg DM)	1172.53	(375.27)
Caloric suitability index (cals)	8412.21	(1288.00)
Distance to São Paulo (km)	2545.67	(541.43)
Observations	259,574	

Note: Each observation is a 4 km $\times$ 4 km grid cell within the Brazilian Amazon. The area of each grid cell is approximately 1,640 ha. There are 259,574 grid cells in the sample.

Table B.2: Full estimates of the association with deforestation

	(1)	(2)
Extinct groups	82.470**	74.740***
	(32.573)	(26.165)
Indigenous groups	16.196	-6.617
	(19.072)	(13.488)
Longitude	30.725***	75.857***
	(3.607)	(10.539)
Latitude	-36.279***	-128.334***
	(4.456)	(17.028)
Altitude (m)		-0.578**
		(0.231)
Terrain ruggedness index (m)		0.878***
		(0.297)
Water (ha)		-0.228***
		(0.047)
Non-forest coverage (ha)		-0.341***
		(0.039)
Malaria index		-1230.574**
		(503.869)
Average temperature (C)		28.057
		(29.313)
Average monthly rainfall (mm)		0.049
		(0.771)
Average relative humidity (%)		22.711***
		(6.131)
Close (up to 10 km) of a main river		-2.433
		(20.168)
Hevea br. range		-230.554***
		(37.420)
Attainable yield: soybeans (kg DM)		0.187***
		(0.044)
Attainable yield: cassava (kg DM)		0.056*
		(0.033)
Attainable yield: cacao (kg DM)		-0.151**
		(0.066)
Caloric suitability index (cals)		0.064***
		(0.024)
Distance to São Paulo (km)		1.060***
		(0.178)
R-squared	0.261	0.436
Observations	259574	259574

Note: The table displays estimates of the coefficients in Equation 1. The dependent variable is accumulated deforestation by 2010, in hectares, measured by PRODES. Column 1 reports estimates for a parsimonious specification controlling linearly for the grid cell's coordinates (latitude and longitude), and column 2 for an extended specification the full set of geographic fundamentals as controls. Standard errors in parentheses are multi-way clustered according to four interlaced grids of $1^\circ \times 1^\circ$. Statistical significance: * 10%, ** 5%, *** 1%.

Table B.3: Robustness to pre-Columbian sites

	Arch. site (1)	Earthwork (2)	Deforestation (3)
Extinct groups	-0.000	-0.020	79.140***
	(0.019)	(0.012)	(25.029)
Indigenous groups	-0.003	0.005	-7.106
	(0.007)	(0.007)	(12.761)
Proximity to pre-Columbian arch. site			186.097***
			(28.508)
Proximity to earthwork			215.603***
			(51.419)
R-squared	0.085	0.090	0.454
Dependent variable mean	0.088	0.030	260.722
Observations	259574	259574	259574

Note: The table displays estimates of the coefficients in Equation 1. The dependent variable is accumulated deforestation by 2010, in hectares, measured by PRODES. The dependent variable is: in column 1, a dummy indicating that a pre-colonial archaeological is within 20 km of the grid cell; in column 2, a dummy indicating that an earthwork (geoglyph or mound village) is within 20 km of the grid cell; in column 3, accumulated deforestation by 2010, in hectares, measured by PRODES. The estimated equation includes latitude, longitude, and all geographic fundamentals from our preferred specification. Standard errors in parentheses are multi-way clustered according to four interlaced grids of $1^\circ \times 1^\circ$. Statistical significance: * 10%, ** 5%, *** 1%.

Table B.4: Robustness to colonial sites

	Aldeamento (1)	Arch. site (2)	Settlement (3)	Deforestation (4)
Extinct groups	-0.005	-0.009*	0.000	75.082***
	(0.004)	(0.005)	(0.008)	(26.170)
Indigenous groups	-0.003	-0.003	-0.001	-6.368
	(0.002)	(0.004)	(0.004)	(13.443)
Proximity to religious *aldeamentos*				-27.364
				(45.155)
Proximity to historic arch. site				56.465
				(51.308)
Proximity to pre-1800 settlements				126.445**
				(55.586)
R-squared	0.051	0.060	0.096	0.438
Dependent variable mean	0.017	0.019	0.026	260.722
Observations	259574	259574	259574	259574

Note: The table displays estimates of the coefficients in Equation 1. The dependent variable is: in column 1, a dummy indicating if any of the religious *aldeamentos* circa 1730 from Bombardi 2014 is within 20 km of the grid cell; in column 2, a dummy indicating that a historic archaeological is within 20 km of the grid cell; in column 3, a dummy indicating that a municipality seat of a municipality which was originally settled before 1800 is within 20 km of the grid cell; in column 4, accumulated deforestation by 2010, in hectares, measured by PRODES. The estimated equation includes latitude, longitude, and all geographic fundamentals from our preferred specification. Standard errors in parentheses are multi-way clustered according to four interlaced grids of $1^\circ \times 1^\circ$. Statistical significance: * 10%, ** 5%, *** 1%.

Table B.5: Robustness to alternative sample period cutoffs to map entries

	All dates	Entries dated since: 1600	1650	1700	1750	1800	1850	1900
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: all map entries								
Extinct groups	26.153	28.215	31.502	42.410*	58.577**	74.740***	91.227***	90.175**
	(18.277)	(18.044)	(19.377)	(22.282)	(23.485)	(26.165)	(30.013)	(35.778)
Indigenous groups	-6.435	-6.157	-5.583	-6.596	-5.929	-6.617	-11.990	-8.522
	(12.976)	(12.967)	(12.984)	(12.961)	(13.224)	(13.488)	(13.921)	(13.503)
R-squared	0.434	0.434	0.434	0.434	0.435	0.436	0.436	0.436
Observations	259574	259574	259574	259574	259574	259574	259574	259574
Panel B: entries with a known linguistic family								
Extinct groups	70.003**	73.467***	73.664**	85.559***	91.860***	94.066***	103.691***	105.799**
	(27.255)	(26.748)	(29.548)	(31.949)	(32.740)	(33.316)	(38.407)	(45.438)
Indigenous groups	-6.385	-6.101	-5.660	-6.581	-6.276	-6.141	-11.337	-7.403
	(13.149)	(13.137)	(13.216)	(13.154)	(13.456)	(13.642)	(14.096)	(13.316)
R-squared	0.435	0.436	0.435	0.436	0.436	0.436	0.436	0.436
Observations	259574	259574	259574	259574	259574	259574	259574	259574

Note: The table displays estimates of the coefficients in Equation 1. The dependent variable is accumulated deforestation by 2010, in hectares, measured by PRODES. The estimated equation includes latitude, longitude, and all geographic fundamentals from our preferred specification. The variables $ExtinctGroups_i$ and $Groups_i$ include, respectively, all extinct and all entries in the Nimuendajú map which were dated after the year shown on the top of the column. Standard errors in parentheses are multi-way clustered according to four interlaced grids of $1^{\circ} \times 1^{\circ}$. Statistical significance: * 10%, ** 5%, *** 1%.

Table B.6: Robustness: accounting for Indigenous group characteristics

	(1)	(2)	(3)
Extinct groups	76.182***	80.912***	98.682***
	(27.971)	(22.957)	(26.435)
R-squared	0.436	0.444	0.445
Observations	259574	259574	259574
Year of record?	Y		Y
Linguistic family?		Y	Y

Note: The table displays estimates of the coefficients in Equation 1. The dependent variable is accumulated deforestation by 2010, in hectares, measured by PRODES. The estimated equation includes latitude, longitude, and all geographic fundamentals from our preferred specification. Column 1 includes the count of groups within 20 km of the grid cell by historical period (from 1800 to 1850, from 1850 to 1900, from 1900 to 1944, and undated); column 2 includes the count of groups in each linguistic family recorded by Nimuendajú within 20 km of the grid cell; column 3 includes all the controls above. The variable ”Groups” is not included in the table due to multicollinearity. Standard errors in parentheses are multi-way clustered according to four interlaced grids of $1^\circ \times 1^\circ$. Statistical significance: * 10%, ** 5%, *** 1%.

Table B.7: Robustness to alternative proxies for proximity to map entries

	Dist. to midpoint (1)	Dist. to approx. shape (2)	Count of groups within: 5 km (3)	10 km (4)	30 km (5)	40 km (6)	Binary var. (7)	Distance (km) (8)
Extinct groups	82.528**	71.685***	94.223***	85.938***	63.920***	54.688***	81.335***	-0.729**
	(34.757)	(25.970)	(34.431)	(31.786)	(21.761)	(18.741)	(29.260)	(0.310)
Indigenous groups	-7.262	-5.293	-9.370	-6.896	-4.241	-2.664	-6.669	-0.372
	(17.898)	(13.588)	(16.584)	(15.914)	(11.462)	(9.910)	(18.094)	(0.396)
R-squared	0.434	0.436	0.434	0.435	0.437	0.438	0.435	0.450
Observations	259574	259574	259574	259574	259574	259574	259574	212404

Note: The table displays estimates of the coefficients in Equation 1. The dependent variable is accumulated deforestation by 2010, in hectares, measured by PRODES. The estimated equation includes latitude, longitude, and all geographic fundamentals from our preferred specification. The measures of proximity to Indigenous group entries in the Nimuendajú map vary by column. In column 1, we use the count of groups whose midpoints are within 20 km of the grid cell. In column 2, we use distances to the approximate of the entries. In columns 3 to 6, we count groups that are within different distance cutoffs, from 5 km to 40 km. In column 7, we use binary variables that indicate if any group or any extinct group is within 20 km of the grid cell. In column 8, we use the distances (in kilometers) to the nearest group and to the nearest extinct group. Standard errors in parentheses are multi-way clustered according to four interlaced grids of $1^\circ \times 1^\circ$. Statistical significance: * 10%, ** 5%, *** 1%.

Table B.8: Robustness to measurement error in the Nimuendajú map

	Entries with more than 2 references (1)	Weight cells by inverse distance to map markings (2)	Entries with non-Portuguese references (3)
Extinct groups	72.677**	65.881**	63.344**
	(30.001)	(29.972)	(26.068)
Indigenous groups	-6.455	1.719	-4.907
	(14.393)	(13.515)	(14.324)
R-squared	0.435	0.447	0.435
Observations	259574	259574	259574

Note: The table displays estimates of the coefficients in Equation 1. The dependent variable is accumulated deforestation by 2010, in hectares, measured by PRODES. The estimated equation includes latitude, longitude, and all geographic fundamentals from our preferred specification. In column 1, the variables $ExtinctGroups_i$ and $Groups_i$ count only map entries of groups with 2 or more citations in the accompanying dictionary. In column 2, we weight the observations according to the inverse of the distance to the $5° \times 5°$ map markings that we used to georeference Nimuendajú's map. In column 3, the variables $ExtinctGroups_i$ and $Groups_i$ count only map entries of groups for which there is at least one reference in a language other than Portuguese. Standard errors in parentheses are multi-way clustered according to four interlaced grids of $1° \times 1°$. Statistical significance: * 10%, ** 5%, *** 1%.

Table B.9: Robustness to proximity to non-Indigenous population in 1940

	Proximity to city (1)	Log market potential (2)	Deforestation, 2010 (3)
Extinct groups	0.003	-0.009	77.768***
	(0.008)	(0.009)	(26.644)
Indigenous groups	-0.004	-0.005	-4.640
	(0.004)	(0.005)	(13.691)
Proximity to 1940 city			40.834
			(55.838)
Log market potential in 1940			335.061***
			(118.109)
R-squared	0.113	0.905	0.444
Observations	259574	259574	259574

Note: The table displays estimates of the coefficients in Equation 1. In column 1, the dependent variable is a dummy indicating that the grid cell is within 20 km of a 1940 municipality seat. In column 2, the dependent variable is the logarithm of market potential in 1940. In column 3, the dependent variable is accumulated deforestation by 2010, in hectares, measured by PRODES. The estimated equation includes latitude, longitude, and all geographic fundamentals from our preferred specification. Standard errors in parentheses are multi-way clustered according to four interlaced grids of $1° \times 1°$. Statistical significance: * 10%, ** 5%, *** 1%.

Table B.10: Robustness: consider whether the historical sources confirm the extinction

	(1)	(2)
Extinct groups (confirmed)	98.741**	102.867**
	(40.787)	(41.045)
Extinct groups (unconfirmed)		47.721*
		(26.403)
Indigenous groups	-0.529	-6.180
	(12.995)	(13.441)
R-squared	0.436	0.436
Observations	259574	259574

Note: The table displays of the coefficients in Equation 1. The dependent variable is accumulated deforestation by 2010, in hectares, measured by PRODES. The estimated equation includes latitude, longitude, and all geographic fundamentals from our preferred specification. In column 1, the variable $ExtinctGroups_i$ consider only extinct groups for which our reading of the historical sources confirm the extinction of the group. In column 2, we include proximity to both extinct Indigenous group about which we confirm and about which we do not confirm the extinction in the sources. Standard errors in parentheses are multi-way clustered according to four interlaced grids of $1° \times 1°$. Statistical significance: * 10%, ** 5%, *** 1%.

Table B.11: Robustness to the use of other groups as a control variable

	Only extinct group entries (1)	Quadratic term (2)	Past (pre-1944) location entries (3)
Extinct groups	68.465***	74.473***	73.647**
	(24.674)	(26.239)	(30.685)
Indigenous groups		0.384	-5.121
		(18.922)	(18.401)
Indigenous groups (squared)		-2.706	
		(7.187)	
R-squared	0.436	0.436	0.436
Observations	259574	259574	259574

Note: The table displays of the coefficients in Equation 1. The dependent variable is accumulated deforestation by 2010, in hectares, measured by PRODES. The estimated equation includes latitude, longitude, and all geographic fundamentals from our preferred specification. In column 1, we do not include the variable $Groups_i$, only the count of extinct Indigenous groups within 20 km of the grid cell. In column3, we include both the $Groups_i$ and $Groups_i^2$. In column 3, the variable $Groups_i$ counts only map entries that refer to past group locations, as opposed to Indigenous group locations that were contemporary to Nimuendajú when he prepared the map (circa 1944). Standard errors in parentheses are multi-way clustered according to four interlaced grids of $1° \times 1°$. Statistical significance: * 10%, ** 5%, *** 1%.

Table B.12: Robustness to alternative estimation or inference

	Alternative inference		Tobit
	Conley (1999)	Cluster by mun.	model
	(1)	(2)	(3)
Extinct groups	74.740***	74.740***	149.728***
	(26.474)	(19.200)	(44.478)
Indigenous groups	-6.640	-6.617	-20.536
	(13.668)	(10.314)	(27.551)
Observations	259574	259574	259574

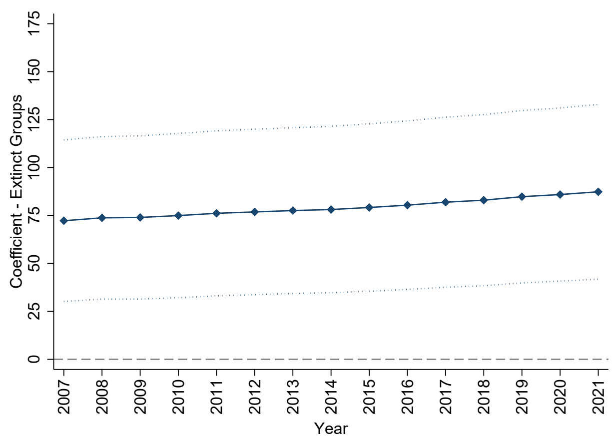
Note: The table displays of the coefficients in Equation 1. The dependent variable is accumulated deforestation by 2010, in hectares, measured by PRODES. The estimated equation includes latitude, longitude, and all geographic fundamentals from our preferred specification. Columns 1 and 2 display ordinary least square ; column 3 displays from a tobit model. In column 1, the standard errors in parentheses are robust to heteroskedasticity and spatial correlation as in Conley 1999, using an uniform kernel with distance cutoff of 100 km. In column 2, the standard errors in parentheses are robust to heteroskedasticity and cross-sectional correlation within clusters defined by municipality. In column 3, standard errors in parentheses are multi-way clustered according to four interlaced grids of $1^\circ \times 1^\circ$. Statistical significance: * 10%, ** 5%, *** 1%.

Table B.13: Association with land use

	Primary forest area (ha)	Farming area (ha)	Cropland area (ha)	Pasture area (ha)	Urban area (ha)	Light intensity
	(1)	(2)	(3)	(4)	(5)	(6)
Extinct groups	-88.584***	84.294***	1.575	82.629***	0.993**	0.150***
	(27.068)	(24.266)	(2.698)	(24.129)	(0.409)	(0.052)
Indigenous groups	9.608	-9.148	-3.064	-5.735	-0.494**	-0.060**
	(13.788)	(13.127)	(2.142)	(12.534)	(0.245)	(0.025)
R-squared	0.449	0.431	0.135	0.405	0.007	0.020
Sample mean	1348.675	218.450	10.782	206.746	1.433	0.201
Observations	259574	259574	259574	259574	259574	259574

Note: The table displays estimates of the coefficients in Equation 1. From columns 1 to 5, the dependent variable is the land use coverage indicated at the top, in hectares, for 2010; these data are from MapBiomas. In column 6, the dependent variable is the average nightlight intensity of the grid cell, from DMSP-OLS. The estimated equation includes latitude, longitude, and all geographic fundamentals from our preferred specification. Standard errors in parentheses are multi-way clustered according to four interlaced grids of $1^\circ \times 1^\circ$. Statistical significance: * 10%, ** 5%, *** 1%.

Figure B.1: Association with deforestation over time, 2007–2021 (PRODES)



Note: The solid navy line and the markings display the estimates of the coefficient on $ExtinctGroups_i$ in Equation 1. The dependent variable is accumulated deforestation at the year indicated on the horizontal axis, in hectares, measured by PRODES. The equation also includes the variable $Groups_i$ and controls for latitude, longitude, and all geographic fundamentals used in the preferred specification. The dotted lines indicate the upper and lower bounds of the 90% confidence interval, based on multi-way clustered standard errors according to four interlaced grids of $1^{\circ} \times 1^{\circ}$. The dashed gray line indicates a horizontal line at zero.

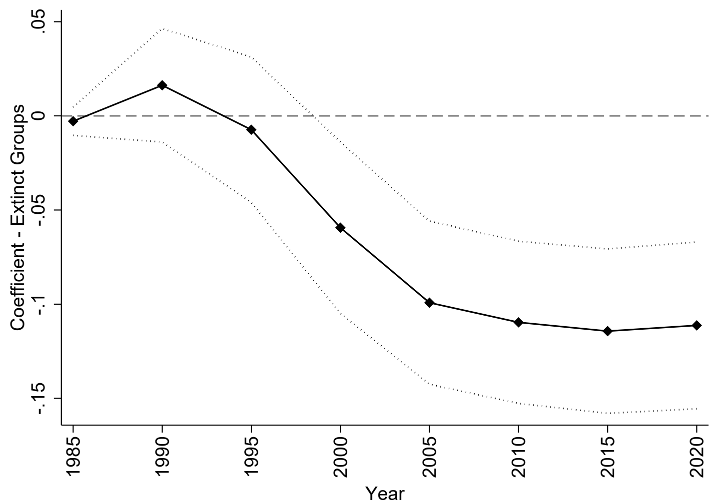
Table B.14: Horse race with settlements, infrastructure and land tenure

	(1)	(2)	(3)	(4)	(5)
Extinct groups	44.237**	59.288***	55.366**	55.121**	20.055
	(19.129)	(21.142)	(24.810)	(22.755)	(14.942)
Indigenous groups	-5.105	-4.662	7.269	4.862	8.764
	(10.351)	(11.823)	(12.428)	(10.669)	(7.729)
Proximity to settlement	318.334***				236.440***
	(25.706)				(20.327)
Proximity to city	200.366***				168.868***
	(27.516)				(22.305)
Road extension		0.114***			0.080***
		(0.007)			(0.005)
Railroad extension		0.026			0.013
		(0.022)			(0.017)
Indigenous land dummy			-192.523***		-64.305***
			(30.090)		(17.283)
Proximity to indigenous village			46.935**		16.136
			(22.092)		(15.177)
Area of private farms				0.207***	0.144***
				(0.024)	(0.017)
Area of land projects				21.180***	15.380***
				(3.260)	(2.442)
Observations	259574	259574	259574	259574	259574
R-squared	0.537	0.499	0.459	0.495	0.608

Note: The table displays estimates of the coefficients in Equation 1. The dependent variable is accumulated deforestation by 2010, in hectares, measured by PRODES. The estimated equation includes latitude, longitude, and all geographic fundamentals from our preferred specification. Additional controls are distances to settlements and cities (column 1), road extension and main road extension (column 2), Indigenous reserves and distance to Indigenous villages (column 3), and the area within privately-owned farms (column 4); all these additional controls are jointly included in column 5. Standard errors in parentheses are multi-way clustered according to four interlaced grids of $1° \times 1°$. Statistical significance: * 10%, ** 5%, *** 1%.

Figure B.2: Association with Indigenous lands or land projects by year

(a) Indigenous lands



(b) Land distribution projects

Note: The solid line and the markings display the of the coefficient on $ExtinctGroups_i$ in Equation 1. The dependent variable is a dummy equal to one if the grid cell overlaps with a homologated Indigenous land by the year on the horizontal axis (to panel) or the area of the grid cell covered by land redistribution projects created up to the year on the horizontal axis (bottom panel). The estimated equation includes latitude, longitude, and all geographic fundamentals from our preferred specification.The dotted lines indicate the upper and lower bounds of the 90% confidence interval, based on multi-way clustered standard errors according to four interlaced grids of $1^\circ \times 1^\circ$. The dashed gray line indicates a horizontal line at zero. The coefficients differ from those in columns 1 and 4 of Table 4 by considering the location of, respectively, Indigenous lands and land distribution projects not only by 2010, but also by other years.

References

ANA (2015). *Conjuntura dos recursos hídricos no Brasil: regiões hidrográficas brasileiras.* Agência Nacional de Águas.

Barsanetti, Bruno and Alipio Ferreira (Nov. 2024a). "Dataset of Circumstances around the Extinction of Historical Indigenous Groups in the Brazilian Amazon, Based on Nimuendajú's Ethnohistorical Map". DOI: 10.5281/zenodo.14186381. URL: https://doi.org/10.5281/zenodo.14186381.

— (2024b). *Georeferenced and cataloged dataset built from Nimuendajú's Ethnohistorical Map*. Version Version 2, July 2024. DOI: 10.5281/zenodo.12772630. URL: https://doi.org/10.5281/zenodo.12772630.

Bombardi, Fernanda A. (2014). *Pelos interstícios do olhar do colonizador: descimentos de índios no Estado do Maranhão e Grão-Pará (1680-1750)*. Master's dissertation, Universidade de São Paulo.

Comércio, Ministério da Indústria e do (1971). *Plano Nacional da Borracha. Anexo VII. O Gênero Hevea: Descrição das Espécies e Distribuição Geográfica.* Brasília MDIC.

Conley, Timothy G. (1999). "GMM estimation with cross sectional dependence". *Journal of econometrics* 92.1, pp. 1–45.

Freitas, Flávio L. M. de, Vinícius Guidotti, Gerd Sparovek, and Caio Hamamura (2018). "Malha fundiária do Brasil, v.1812." *Atlas - A Geografia da Agropecuária Brasileira; IMAFLORA: Piracicaba, Brazil*.

Galor, Oded and Ömer Özak (2015). "Land productivity and economic development: Caloric suitability vs. agricultural suitability". *Brown University Working paper*.

— (2016). "The agricultural origins of time preference". *American Economic Review* 106.10, pp. 3064–3103.

Galor, Oded, Ömer Özak, and Assaf Sarid (2016). "Geographical origins and economic consequences of language structures". *Brown University Working paper*.

Gething, Peter W., Thomas P. Van Boeckel, David L. Smith, Carlos A. Guerra, Anand P. Patil, Robert W. Snow, and Simon I. Hay (2011). “Modelling the global constraints of temperature on transmission of Plasmodium falciparum and P. vivax”. *Parasites & Vectors* 4.1, pp. 1–11.

IBGE, Diretoria de Geociências (2011). *Evolução da divisão territorial do Brasil 1872-2010*.

Jacobs, James Q. (2023). *Amazon Geoglyphs XLS*. Tech. rep. Accessed 10/26/2023. Available at: http://jqjacobs.net/archaeology/geoglyph.html.

Masten, Matthew A and Alexandre Poirier (2022). “The Effect of Omitted Variables on the Sign of Regression Coefficients”. *arXiv preprint arXiv:2208.00552*.

New, Mark, David Lister, Mike Hulme, and Ian Makin (2002). “A high-resolution data set of surface climate over global land areas”. *Climate research* 21.1, pp. 1–25.

Nimuendajú, Curt (2017). *Mapa Etno-Histórico do Brasil e Regiões Adjacentes*. Brasília: Instituto do Patrimônio Histórico e Artístico Nacional, Instituto Brasileiro de Geografia e Estatística.

Oster, Emily (2019). “Unobservable selection and coefficient stability: Theory and evidence”. *Journal of Business & Economic Statistics* 37.2, pp. 187–204.

Walker, Robert S, Jeffrey R Ferguson, Angelica Olmeda, Marcus J Hamilton, Jim Elghammer, and Briggs Buchanan (2023). “Predicting the geographic distribution of ancient Amazonian archaeological sites with machine learning”. *PeerJ* 11, e15137.