Human and Nature:
Economies of Density and Conservation in the Amazon Rainforest*

Shunsuke Tsuda, *Brown University*

Yoshito Takasaki, *The University of Tokyo*

Mari Tanaka, *Hitotsubashi University*

Please check HERE for the latest version.

June 14, 2023

Abstract

Conserving tropical forests impacts the standard of living of local populations. Moreover, human adaptation through sectoral or spatial reallocation of economic activity may undermine conservation policy goals. To derive policies that balance human and ecological well-being, this paper estimates a multi-sector spatial model that formalizes human-nature interactions using high-resolution georeferenced data from roadless river basins in the Peruvian Amazon. Identification comes from plausibly exogenous variation in the structure of river networks. We find that the agglomeration externality in agricultural production outweighs dispersion forces in access to land, implying that higher concentration leads to higher productivity with less deforestation per farmer. We also find a strong congestion externality with spatial spillovers in natural resource extraction. The estimated agglomeration externality, primarily driven by economies of scale in transport technology and agricultural intensification, generates large welfare and forest cover gains but leads to natural resource depletion through general equilibrium effects. Counterfactuals demonstrate that combining well-targeted place-based protection policies and transport infrastructure can simultaneously achieve higher welfare, lower deforestation, and less natural resource depletion.

Keywords: Tropical Forests, Agriculture, Natural resource extraction, Conservation, Externalities, Spatial models

*Tsuda is deeply grateful to Andrew Foster and Matt Turner for continuous guidance and support, and to Dan Björkegren, Jesse Shapiro, and Peter Hull for detailed discussions and feedback. We thank Dave Donaldson, Mateusz Filipski, John Friedman, Allan Hsiao, Jean Imbs, Shoya Ishimaru, Yoshi Konishi, Bingjing Li, Ellen McCullough, Yuhei Miyauchi, Kazuki Motohashi, Kaivan Munshi, Takeshi Murooka, Tomoya Mori, Mike Neubauer, Jordan Norris, Matt Pecenco, Heitor Pellegrina, Mark Rosenzweig, Veronica Salazar-Restrepo, Xiaodong Zhu, and audiences at Brown University, Hitotsubashi University, JADE-CEPR-TIME Development Economics Conference, Kyoto Institute of Economic Research, LSE Environment Day, Occasional Workshop in Environmental and Resource Economics (UCSB), Osaka University, UEA Meeting (Bocconi), University of Essex, University of Georgia, University of Hong Kong, Young Urban Economist Workshop (World Bank), and Y-RISE Conference for helpful comments. We acknowledge the PARLAP project collaborators, Oliver Coomes and Christian Abizaid, for their invaluable inputs. Nicholas Cadieux and Tristan Grupp provided excellent research assistance. Margaret Kalacska generously assisted with remote sensing. Tsuda thanks Margarita Gafaro for providing community-level identifiers of the Peruvian Agricultural Census. Tsuda acknowledges financial support from the Interdisciplinary Opportunity Dissertation Completion Fellowship at Brown University, a research grant from the Murata Science Foundation, and research support from CREPE. Takasaki acknowledges grants from the Japan Society for the Promotion of Science (23243045; 26245032; 18H05312; 18KK0042; 20K20332). Tanaka acknowledges grants from the Japan Society for the Promotion of Science (18KK0042). All remaining errors are ours.

Contacts: Shunsuke_Tsuda@brown.edu (https://shunsuketsuda.com/), takasaki@e.u-tokyo.ac.jp, mari.tanaka@r.hit-u.ac.jp.

1 Introduction

Conservation of tropical forests is a key element of global efforts to slow climate change and preserve biodiversity (Malhi et al. 2008; Pimm et al. 2014). Meanwhile, many populations in rainforest areas have low living standards and rely on forest and environmental income sources (Angelsen et al. 2014). Policymakers face the inherent trade-off between conserving the rainforest and improving the welfare of local populations. For example, increasing the cost of forest clearing may reduce deforestation, but it may also reduce agricultural income and thus undermine human welfare. Moreover, human adaptation through sectoral or spatial reallocation of economic activity may undermine the goals of conservation. For example, if the cost of forest clearing is increased, then local populations may shift from agriculture to other extractive activities such as fishing and hunting, which could lead to biodiversity loss; if a place-based protection policy prohibits resource extraction in one area, then local populations may move from that area to others to extract resources, and the impact on overall resource depletion is unclear. How can we design a policy that resolves these trade-offs?

To answer this question, it is necessary to consider general equilibrium forces in human-nature interactions across space. Several recent studies in economics investigate the effectiveness of policy interventions such as environmental conditional cash transfers, regulation and taxes, and protected areas (e.g., Assunçao et al. 2022; Hsiao 2022; Jayachandran et al. 2017; Sims and Alix-Garcia 2017; Souza-Rodrigues 2019), as well as infrastructure projects (e.g., Asher et al. 2020; Madhok 2022) on deforestation and biodiversity outcomes. However, most prior studies have not considered the general equilibrium contexts. Moreover, many previous studies of tropical forests focus on the contexts of extensive commercial investments (e.g., logging, soybeans and cattle-ranching in the Brazilian Amazon, and palm oil plantations in Indonesia; see Balboni et al. 2022 for a comprehensive review). It is not clear the degree to which findings from these contexts can be generalized to less developed areas where deforestation and other resource extractions are primarily practiced by local populations for their livelihood.

In this paper, we build a multi-sector quantitative spatial general equilibrium model that features the concentration and dispersion forces of the population in tropical forests. We estimate the model using fine-resolution georeferenced data from four roadless river basins in the Peruvian Amazon. We identify externality parameters that explain how population density affects productivity in rural sectors by exploiting plausibly exogenous variation in the structure of river networks. We then implement counterfactual simulations to evaluate policy impacts on human welfare, deforestation, and natural resource depletion. These experiments demonstrate that well-targeted place-based protection policies and transport infrastructure are complementary to improving both human welfare and ecological conservation.

The Peruvian Amazon is an ideal setting to study fundamental human-nature interactions for two reasons. First, most of the population engages in traditional ways of life in remote areas without modern technology and large-scale external investments. These features enable us to attribute resource extraction to small-scale farmers and hunter-gatherers and thus focus on density externalities that they cause.[1] Second, river networks almost solely constitute the transportation routes in this region. This feature allows us to identify key structural parameters by exploiting plausibly exogenous variation in the structure of river networks. In particular, our dataset covers major river basins of the Peruvian Amazon (Figure 1) and consists of grid cell-level geographical and remote sensing data, original community-level census from rural communities, and national censuses.

Using this dataset, we start with three stylized facts that motivate the theoretical model. First, we observe both spatial concentration and dispersion of populations and community locations. This fact suggests the presence of both strong agglomeration and congestion forces of economic activities in rainforests. Second, while the land footprint around a community increases with settlement size, *per capita* land footprint decreases. This pattern suggests that as the community population increases, the cost of access to farmland through forest clearing might become higher because of congestion forces around the settlement. Moreover, this negative relationship is convex, which suggests that a mean-preserving increase in the variance of settlement size may increase total deforestation if this convexity is strong enough. Third, while agriculture is widely observed in both concentrated and dispersed areas, natural resource extraction tends to occur more in locations surrounded by areas with lower population densities. These observations suggest that the spatial extent over which density externalities operate might vary across sectors.

Motivated by the stylized facts, we build a quantitative spatial model that highlights the rainforest population's trade-off: natural resource and land endowments are more accessible in sparse areas owing to weak congestion, while dense areas have higher market access and agglomeration benefits. Our model includes two rural sectors and an urban sector. The rural sectors include an agricultural sector (crop production), which is associated with deforestation, and a natural resource extraction sector (fishing, hunting, and forest products), which is linked to biological resource depletion. The model analyzes trade of the products of the three sectors based on comparative advantages across multiple locations (grid cells) in a general equilibrium framework within a river basin. We explicitly incorporate agglomeration and congestion externalities for productivities in the two rural sectors. The model primarily focuses on the rural spatial structure and incorporates the rural-urban linkage to capture rural remoteness. The balance between the concentration and dispersion forces, including the agglomeration and congestion externalities, determines the spatial distribution of economic activities in equilibrium.

We estimate the model in three steps. First, we estimate trade cost and demand parameters without solving the model. Second, we calibrate the productivities that rationalize the observed sectoral populations in all locations by inverting the model, given the parameters obtained

[1] This empirical setting is also important in that it is closely related to recent concerns and trends of deforestation in the Amazon. Kalamandeen et al. (2018) report that small-scale deforestation has recently increased throughout the Amazon (in Brazil and other countries), raising concerns about the role of small-scale farmers in forest conversion.

in the first step. The calibrated productivities contain both productivity fundamentals and endogenous terms caused by the density externalities. Third, we use a generalized method of moments (GMM) procedure to estimate to estimate the density externalities, using the inverted productivity composites as data and exploiting plausibly exogenous variation in the structure of river networks. We construct a measure of "river network access" (RNA), which captures the distance-weighted sum of access to other locations via the river network, as an instrumental variable for each location's population. The logic behind the identification is that the variation in RNA, as a market potential shifter, affects productivity only through its effect on employment and thus through externalities that arise, rather than through productivity fundamentals.[2]

The estimation reveals three types of sector-specific density externalities. First, we find a strong congestion externality in forest clearing for access to agricultural lands. Second, we find a strong agglomeration externality in agricultural production. More importantly, this agglomeration externality outweighs the congestion externality in access to land, implying that higher concentration leads to higher productivity with less deforestation per farmer. Third, we also find a strong congestion externality with spatial spillovers in natural resource extraction.

The estimated density externalities are quantitatively important. For example, the agglomeration externality contributes to substantially improving human welfare and reducing deforestation, at the expense of other natural resource endowments. Without the agglomeration externality, human welfare decreases by about 7–13% and deforestation increases by about 18–56% depending on the basin. On the other hand, without the agglomeration externality, the natural resource depletion also decreases by about 1–3% through general equilibrium effects in the basin. These impacts are primarily due to the sectoral reallocation of workers. Since agricultural productivity is diminished, the economy needs more agricultural employment to satisfy the population's demand for agricultural goods, given that consumption demands across the agricultural and natural resource goods are complementary.

We investigate mechanisms behind the agglomeration externality, focusing on the transaction environment and agricultural intensification. Forest clearing and natural resource extraction are rivalrous, so it is straightforward that they face a congestion externality. In contrast, the agglomeration externality in agriculture may not be immediately obvious, in part because agglomeration has typically been discussed in urban settings. Using the original community-level census information, we provide supportive evidence that agglomeration is primarily driven by economies of scale in trade costs. The empirical results are consistent with the view that transport modes, transport costs, and transaction costs are endogenous and are affected by the community's population. This mechanism is also consistent with the model, since the original model is isomorphic to an alternative model in which a higher origin population leads to lower

[2]The immediate concern may be that being located near a river means high RNA but also high productivity. However, this concern is mitigated because RNA also captures distant river geometries. The independence assumption is that, after controlling for a rich set of geographical conditions in the own location, unobservable productivity fundamentals are not correlated with the variation in RNA that can be generated by exogenous river shapes in locations far away from the own location. We also consider another instrument motivated by history. Section 5 provides a more detailed discussion supporting the identifying assumptions.

iceberg trade costs. In addition, using the producer-level agricultural census, we argue that economies of scale in accessing inputs and technologies in the cropping and marketing process are also behind the agglomeration externality, but to a lesser degree.

We assess two types of counterfactual experiments—transport infrastructure investments and protection policies—with the aim of finding a "win-win" policy that simultaneously achieves higher welfare, lower deforestation for agriculture, and less natural resource depletion. The transport infrastructure investments aim to reduce high trade costs, which primarily stem from the asymmetry of transport costs due to river orientations and the slow speed of river boats. We evaluate transportation investments such as high-quality boats or dredging rivers as proposed in the Amazon Waterway project (outlined in Lu 2019). Protection policies aim to answer the following question: In the presence of density externalities, is it beneficial to concentrate the ecological footprint in fewer spots rather than to have many small communities?

The counterfactuals show that well-targeted river infrastructure investments and place-based protection policies are complementary to improving human welfare and ecological conservation. In particular, the composite intervention that combines the transport infrastructure investments that connect hinterlands to the central area of a basin with protecting the rural frontier simultaneously achieves the win-win outcome.

Figure 2 presents a simplified illustration of this policy outcome that depicts the deforestation impact, abstracting from the sectoral reallocation of workers. The left-hand maps represent a stylized river basin with a fixed number of agricultural populations where each brown area inside the boundaries is the deforested area by each farmer. The right-hand diagrams represent the relationship between the agricultural population and the deforested area per farmer in each rural community, which is negative and convex, consistent with the stylized fact introduced above. First, compare panels (A) and (B). As hinterlands become more integrated by the improvement in river infrastructure (the light blue line) in panel (B), we observe spatial reallocation of farmers from dense areas toward remote areas compared with the benchmark equilibrium in panel (A). In the remote areas, populations enjoy more agglomeration benefits and deforestation *per farmer* decreases. Total deforestation has reduced, since this forest gain in remote areas outweighs the increase in deforestation per farmer in the central area. Next, compare panels (B) and (C). Protecting the rural frontier in panel (C) generates a more compact basin for human settlements, and the population concentrates in a smaller set of communities. Total deforestation decreases further, given the congestion in forest clearing. This protection policy also reduces natural resource depletion by increasing the congestion force in resource extractions, which depends on surrounding populations, within the more compact area.

Obviously, this figure abstracts from any other general equilibrium effects. Actual simulations in the four river basins reveal that this composite intervention increases welfare by about 1–2.1%, decreases deforestation by about 1–7%, and decreases natural resource depletion by about 0.5–2.4%. However, neither of these two types of interventions achieves this joint outcome alone. Section 7 describes the results and mechanisms in more detail.

Spatial targeting matters in both types of interventions. With respect to the transport infrastructure, where the improvement takes place within the spatial structure of river networks determines whether total deforestation increases or decreases. In contrast to the previous case, improving the transport infrastructure only in densely populated areas increases total deforestation in the basin. This comparison implies that policy interventions that spread the agglomeration externalities more evenly across the basin are desirable in terms of reducing deforestation. That is, policies that result in more moderate-sized but dispersed settlements are preferable to policies that result in a separation in the basin with highly populated areas of agricultural intensification and very small communities with low agricultural productivity in the hinterland. With respect to the protection policies, a comparison between different protected areas, each directly treating the same number of populations, illustrates the policymaker's trade-off in mitigating different types of environmental costs. For example, a resettlement policy that targets the smallest communities reduces deforestation more than protecting the frontier, but it increases natural resource depletion. Moreover, we also discuss unintended consequences of a sector-specific protection policy as well as pathways to external validity.

Related literature. This paper contributes to three strands of literature. First, it adds to a large body of literature on the trade-off between economic development and environmental goals (see Jayachandran 2021; Jayachandran 2022 for review). Specifically, this paper contributes to the debate on the relationship between agriculture and deforestation (e.g., Abman and Carney 2020; Abman et al. 2020; Angelsen 1999; Angelsen 2010; Carreira et al. 2022; Foster et al. 2002; Heß et al. 2021; Szerman et al. 2022). Empirical evidence on whether higher agricultural productivity causes more deforestation is mixed. We consider a new channel, the agglomeration benefits, through which higher concentration leads to higher productivity and lower deforestation per farmer. Improving agricultural productivity through our channel can reduce deforestation without the strong conditions that other recent papers have used to draw the same conclusions. In particular, agriculture is the most land-intensive sector in our model (unlike Szerman et al. 2022), and we are not imposing any assumption about factor market constraints (unlike Abman et al. 2020). More importantly, we argue that total deforestation as an aggregate outcome in the economy depends on how the agglomeration benefits are spatially spread. This lesson suggests the importance of analyzing the relationship between agriculture and deforestation in a general equilibrium framework. Furthermore, we incorporate distinct types of environmental costs—deforestation and other natural resource extractions such as fishing and hunting—that have often been studied in isolation, in a unified framework. Our counterfactual experiments relate to the policy discussions on preserving rainforests, including protected areas, taxes, and infrastructure (e.g., Alix-Garcia et al. 2013; Alix-Garcia et al. 2015; Araujo et al. 2020; Assunçao et al. 2022; Madhok 2022; Naughton-Treves et al. 2011; Sims and Alix-Garcia 2017; Souza-Rodrigues 2019). We contribute to them by demonstrating the complementarity between protection policies and transport infrastructure investments, which have often been investigated separately, in reducing the different types of environmental costs. In addition, this paper joins the discussion on commons (e.g., Dasgupta and Mäler 1995; Hardin 1968; Ostrom 1990; Ryan and Sudarshan 2022). To our knowledge, this paper is the first to investigate spatial competition over common pool resources in rural areas using a quantitative spatial model.

Second, this paper is connected to the literature of economic geography and quantitative spatial models that investigate spatial distribution of economic activities (e.g., Ahlfeldt et al. 2015; Allen et al. 2020; Fajgelbaum and Redding 2022; Fujita et al. 1999; Nagy 2020; Redding 2016). In particular, this paper contributes to environmental considerations in economic geography. A growing number of studies are examining the impact of a path of climate change that is taken as given on economic activities and welfare using spatial models at various levels, including country (e.g., Balboni 2019; Rudik et al. 2021), regional and continental (e.g., Conte 2022; Jedwab et al. 2022), and global (e.g., Costinot et al. 2016; Cruz and Rossi-Hansberg 2021; Nath 2022). This paper complements this literature by analyzing new endogenous environmental outcomes in rainforests. Moreover, this paper also contributes to the literature on intra-country agricultural trade in developing countries (e.g., Bergquist et al. 2022; Pellegrina 2022; Rivera-Padilla 2020; Sayre 2022; Sotelo 2020). This paper is distinct from the previous literature in that it applies a quantitative spatial model in a more spatially granular setting in rural areas. In addition, it incorporates extraction of natural resource products, which has contrasting characteristics to agriculture, and features sector-specific density externalities.

Third, this paper enriches our understanding of agglomeration economies (see Ahlfeldt and Pietrostefani 2019; Duranton and Puga 2004; Duranton and Puga 2020 for comprehensive review). While agglomeration has typically been discussed in urban settings in the literature, this paper is the first to uncover the presence of agglomeration externality in rainforests. We present the mechanism behind the agglomeration using rich community- and household-level census data, which amplifies this contribution. In this respect, this paper is also related to previous research that study economies of density in agriculture (Boserup 1965; Caunedo et al. 2020; Holmes and Lee 2012; Salehi-Isfahani 1993; Stryker 1976) and contributes to them by deriving general equilibrium implications for human welfare and environmental costs.

Roadmap. The remainder of this paper proceeds as follows. Section 2 introduces the empirical setting and data sources. Section 3 presents stylized facts. Section 4 describes the quantitative spatial model. Section 5 calibrates and estimates parameters of the model. Section 6 investigates mechanisms behind the agglomeration externality. Section 7 provides counterfactual experiments. Section 8 concludes the paper and presents directions for future research.

2 Empirical Setting and Data

The Peruvian Amazon, our study area, is an ideal setting to study fundamental human-nature interactions in rainforests for two reasons. First, most of the population engages in traditional ways of life in remote areas without modern technology and large-scale external investments. This feature allows us to attribute resource extraction to small-scale farmers and hunter-gatherers and focus on the density externalities that they cause. Second, river networks almost solely constitute the transportation routes in this region. We take advantage of this feature to identify key structural parameters by exploiting the exogenously given structure of river networks.

This empirical setting is important in that it is also closely related to recent concerns and trends of deforestation in the Amazon. In recent years, small-scale deforestation has increased throughout the Amazon (in Brazil and other countries), raising concerns about the role of small-scale farmers in forest conversion. Kalamandeen et al. (2018) find that the number of small forest loss patches (< 1 ha) increased by more than 30% between 2001–2007 and 2008–2014 in the entire Amazon, while the number of large ones (> 50 ha) declined significantly. In the context of the Peruvian Amazon, government documents tend to blame smallholder farmers who practice shifting cultivation for most small-scale deforestation (Ravikumar et al. 2017). Potapov et al. (2014) estimate that the forest cover lost between 2000 and 2010 in Peru is equivalent to 2.44% of 78.6 million ha of the Peruvian humid tropical forest biome area and that the majority of this forest loss (92%) was due to clearing. Moreover, recent hotspots of forest loss have been shifting from the Brazilian Amazon to Peru and Bolivia (Kalamandeen et al. 2018). This fact also suggests the importance of research on rainforest conservation outside the Brazilian Amazon, where less is known.

Our study area consists of two administrative departments, Loreto and Ucayali, that cover an area of 471,199 km^2 or about 85% of the Peruvian Amazon. Iquitos and Pucallpa are the capitals of these departments, respectively. Figure 1 shows our study area. The area covered by our data consists of four major river basins of the Peruvian Amazon (Napo-Amazon, Pastaza, Lower Ucayali, and Upper Ucayali), which together encompass a vast area of 117,680 km^2, more than double the total area of Costa Rica and close to that of Malawi. We will consider a general equilibrium in the model in each of these major river basins.

The main data and their sources are presented in the following sections. Appendix B provides more detail about the data and introduces additional data.

2.1 Grid Cell-Level Geographic Information

Our primary units of empirical analysis are 1 km $\times$ 1 km grid cells covering the four basins. We construct:

Distance matrix. We select all the grid cells within 5km from river lines. We consider rivers of river orders 1–5 where higher river orders mean more splits from the central river. We then implement Dijkstra shortest path calculations between all pairs of the selected grid cells. To calculate the shortest path, we take into account total upstream and downstream distances by river travel as well as distance by land travel.[3]

Forest cover. We classify Landsat satellite imageries from 1985, 2001, and 2015 as forest, non-forest, and masked (cloud and water), using CLASlite (Asner et al. 2009),[4] and we aggregate them into grid cell-level information. The forest or non-forest area within a grid cell is measured by the number of 30 m $\times$ 30 m pixels classified as forest or non-forest. From the multi-period information, we also construct variables of forest loss, forest recovery, and forest disturbance. Forest loss and recovery are measured between 1985 and 2015. The area of forest loss is the area within a grid cell that changed from forest to soil. The area of forest recovery is the area within a grid cell that changed from soil to forest. Forest disturbance is measured between 2001 and 2015. Forest disturbance includes any events that disturbed primary forests, such as the appearance of secondary forests. Secondary forests are forests that re-grow during the fallow phase of the shifting cultivation system after initial clearing of primary forests and cropping.

Our empirical analyses use many other cell-level geographic data and variables, which we list in Appendix B.

2.2 Peruvian Amazon Rural Livelihoods and Poverty Project

The Peruvian Amazon Rural Livelihoods and Poverty (PARLAP) project[5] collects key community- and household-level data from the population living in the four major river basins in the Peruvian Amazon. All the surveyed communities are geo-referenced and can be matched with other publicly available census data by unique community-level identifiers. The PARLAP data consist of the following types of data.

Community Census (CC). CC is a basin-wide census that collects community-level information from almost all communities in the four river basins (919 communities in total). The field survey team conducted the data collection over the course of 19 months during 2012–2014. In each community, the field team sought out the local authorities and conducted a focus group interview. The CC data include information on community history (foundation, relocation, past shocks), population size, infrastructure (such as transport modes and communication methods), public services, across-community formal and informal networks, commodity prices, initial and current economic activities, and initial and current natural resource endowments. We rely on these data for various purposes throughout this paper. Appendix B describes the data in more detail. Henceforth we refer to these data as the "CC data."

[3] Cadieux et al. (2020) provide a detailed explanation of the Python algorithm used for these calculations.

[4] See Supplementary Information of Coomes et al. (2021) for the detailed procedure of processing satellite images. We are also working on constructing the forest cover data from the Google Earth Engine.

[5] Detailed information about the project can be found at: `https://parlap.geog.mcgill.ca/`. We gratefully acknowledge two PARLAP field teams that conducted the survey work, often under challenging conditions, in Loreto (Carlos Rengifo Upiachihua, Iris Anelís Arevalo Piña, Judiht del Castillo Macedo, Jacob Gonzales Bardales, Kathicsa Naydu Mendoza Montalvan, Norith Paredes Salas, and Inelza Zumbilla Ajón) and Ucayali (Luis Angel Collado Panduro, Claudio Sinuri Lomas, Santiago Nunta, Diego Fernando Dávila Gomez, Eduardo Carlos Perea Tuesta, and Segundo Jorge Vázquez Flores). This study would not have been possible without their tireless efforts and steadfast dedication to the project, or without the support of community authorities and from participating households throughout the study region.

Community/Household Survey (CS/HS). The PARLAP also surveyed 235 communities from among the 919 communities in the census. This subset of communities was selected by stratified random sampling based on ethnicity, location, initial resource endowments, history, and public services. From these 235 surveyed communities, about 4,000 households were surveyed.[6] These survey data were collected between 2014 and 2016. The community survey obtains information on history, school and education, health care, institutions, and natural resource conservation. The household survey contains detailed information on forest clearing, land use and agricultural production, and natural resource extraction. We use these data for some descriptive analyses. Henceforth we refer to these data as the "CS data" and the "HS data."

Unless otherwise noted, we aggregate variables from these data into the grid cells by matching each community with a corresponding grid cell.

2.3 National Censuses

We use the following censuses to complement the dataset.

Peru Population and Housing Census. The Institute Nacional de Estadistica e Informatica (INEI) in Peru conducted population censuses in 1981, 1993, 2007, and 2017. The census generally contains information on total population size, total number of households, and basic demographic characteristics in the entire country, with variations in the specific content across census years. We primarily use the data from 2007 and 2017.[7] We aggregate variables from these data into the grid cells. Populations in the urban centers (and a few rural communities) not covered by the CC data are extracted from the 2017 census to construct comprehensive population data for the general equilibrium units of the spatial model. Henceforth we call these data the "INEI population census."

Peruvian Agricultural Census (CENAGRO). The INEI collected these data in 1994 and 2012, conducting direct interviews with agricultural producers throughout Peru. We use the data from 2012. These data contain comprehensive farm-level and plot-level information about land use, land tenure, crop choices, irrigation, marketing, livestock practices, credit access, infrastructure and machinery, labor, input use, and demographic characteristics of the producers' household members. Since a community-level unique identifier is available for each agricultural producer's location, we can match these data from Loreto and Ucayali departments in the Peruvian Amazon with the others. We primarily use these data to investigate the mechanisms underlying the density externalities in agriculture implied by the structural model.

[6] If the surveyed community had more than 20 household, then we surveyed the 20 households selected by the stratified random sampling based on wealth status. Otherwise, we surveyed all households in the community.

[7] In order to construct the long-term panel data, we are currently constructing the data in 1993 by manually matching its community names with the 2007 and 2017 data.

3 Stylized Facts

This section presents stylized facts that motivate the model and counterfactual policy interventions.

3.1 Spatial Distribution of Communities and Populations

Fact 1A: Spatial concentration and dispersion of populations and communities.

Figure 3 indicates the concentration and dispersion of populations. The circle dots in the figure represent the centroids of 1 km $\times$ 1 km populated grid cells in rural locations ($n = 900$). The square symbol in each basin represents the urban center. As shown in the figure legends, the circle dots represent quantiles of population sizes whereby each of the five ranges contains a nearly equal number of locations. It is apparent from the legend that, in all four basins, more than 80% of rural locations have populations smaller than 450. Moreover, about 40% of rural locations are very small, with populations of 150 or less. In contrast, fewer than 50 rural locations (out of 900) have populations larger than 1,000.

Figure A.1 and Figure A.2 illustrate the concentration and dispersion of locations community locations. These maps show the establishment of communities in the Napo and Upper Ucayali basins over decades. Every two decades, newly founded communities appear in both sparse and concentrated areas.

Implication. This fact suggests the presence of both strong agglomeration and congestion forces of economic activities in rainforests.

3.2 Human Settlements and Forest Cover

It is not feasible to identify the direct cause of deforestation from information on forest loss obtained from satellite images. It might nevertheless be useful to investigate the relationship between human settlements and forest cover. We implicitly presume that the deforestation observed around a rural community is primarily caused by the population that settles in the community. Before stating the stylized facts that connect to the model, we first report the overall relevance of the local population for the forest cover in the four basins.

Figure 4 visualizes the significant overlap between community locations, populations, and deforestation in the Upper Ucayali basin. To give a sense of the overall magnitude of this relationship, Table 1 presents summary statistics on forest loss from 1985 to 2015 (in hectares) for each grid cell category according to the presence of settlements. Panel (A) reports summary statistics in all grid cells in the four basins. According to the first row, the average deforested area in cells with a census community (10.434 ha) is significantly higher than in cells without a census community (2.842 ha). However, the areas where people clear forest might not be exactly the same as their residential locations reflected in the census community location. Therefore, other rows in this panel present summary statistics on deforestation in grid cells within 2 km and 5 km of a census community. Importantly, the column of sum in the last two rows indicates that more than 50% of total deforestation in the four basins is observed in areas within 5km from the census communities. Panel (A) reports summary statistics in grid cells within 2km of the census communities in the four basins, given that most communities are located along the river lines. Among these selected grid cells, the last two rows indicates that almost 60% of total deforestation in the four basins is observed in areas within 5km of the census communities.

Figure 5 shows the cumulative distributions of the percentage of deforested area in each grid cell for cells with and without census communities. This figure implies the same takeaway as in the previous table. The top-left diagram implies that cells with census communities have much higher deforestation. The other diagrams imply that the difference in the distribution of deforested area decreases as the buffer size increases. Moreover, these patterns are robust to the period of measuring deforestation. Figure A.3 shows the same relationship, but with deforestation measured between 2001 and 2015.

These patterns stress the importance of small-scale farmers as a cause of deforestation in the study area. This takeaway is indeed consistent with the findings from the existing literature that we introduced at the beginning of section 2. Meanwhile, the positive relationship between human settlements and deforestation is straightforward in theory. We next present a more important fact that motivates the model and policy interventions.

Fact 2A: The relationship between population and per capita land footprint is negative and convex.

Figure 6 plots populations and per capita forest loss areas (between 1985 and 2015) or per capita non-forest areas (in 2015) within 2 km buffers from the 2007 INEI population census communities in the Napo-Amazon basin. The total population in the buffer surrounding a community is measured by summing populations from communities whose centroids are inside the buffer. Note that, if we assume that Amazon river basins were fully covered by rainforests at the very beginning (i.e., before any human settlements), then non-forest area could also be a reliable measure of total net deforestation associated with human settlements until today. With the use of either way of measuring the deforestation, the figure is consistent with a negative relationship between population and per capita land footprint.

Furthermore, the negative relationship between population and per capita land footprint is convex as the lowess lines illustrate. Figure A.4 and Figure A.5 show this relationship for different buffer sizes (1 km, 2 km, 5 km) and for all four basins. These figures illustrate that the convex relationship is robust to different buffer sizes and different basins.

This relationship is correlational, but suppose for simplicity that we take this relationship as structural as a thought experiment. Then, this relationship suggests the following two takeaways toward the modeling, estimation, and policy simulations. First, even without a land market, a congestion force may exist in forest clearing depending on the population size in a residential location. Second, and more importantly, if this convexity is strong enough, a mean-preserving increase in the variance of settlement size might increase total deforestation.[8]

Recall Figure 2 that indeed illustrates the second point. We ignore the left-hand maps now and focus on the right-hand diagrams in scenarios (A) and (B), which depicts the convex relationship. We consider these two scenarios for an economy with five communities and a fixed total population of 16. In scenario (A), four communities each have a population of two and a deforested area of 10 hectares per capita. Another community has a population of eight and a per capita deforested area of 3 hectares. In this scenario, the total deforestation in the economy is 104 hectares. In scenario (B), four communities each have a population of three and a deforested area of 6 hectares per capita. Another community has a population of four and a per capita deforested area of 5 hectares. In this scenario, the total deforestation in the economy is 92 hectares and is reduced from the previous scenario.

These cross-sectional relationships are important even though there are dynamic processes of deforestation and reforestation in the real world. Farmers in our study area (and in many tropical forest areas in the world) practice shifting cultivation with the following swidden-fallow cycle. First, farmers clear primary (old-growth) forests and burn the vegetation to obtain land plots, and then they plant crops. Second, when plots become no longer productive, plots are left in fallow and the secondary forest regrows. After several years of fallow, farmers clear such secondary forests again and the cycle is repeated. Although the deforested locations around the community are moving over time due to this cycle, at any given moment the stock of forest fallow and the total deforested area around the community remain relatively constant (Coomes et al. 2021). Therefore, the cross-sectional relationship between the settlement size and deforestation has a more significant variation, which motivates our static theoretical model.

Implication. To summarize, these facts imply the following two takeaways. The first take-away is the importance of modeling the congestion externality in forest clearing in a land-intensive sector and estimating its magnitude. The second takeaway is that policy interventions that lead to a change in the variance of settlement size even with the same total population might have an important ecological consequence.

3.3 Spatial Distribution of Agriculture and Natural Resource Extraction

We finally focus on sectoral difference in the spatial distribution of economic activities. The two main types of activities in our study area are agriculture and natural resource extractions. Agricultural sector includes food crops (e.g., manioc, farina, plantain) and cash crops (e.g., rice, maize, beans, sugar cane, vegetables, fruits). Natural resource extractions include wildlife extractions and forest products. Wildlife extractions include fishing and game meat (hunting).

[8] Figure A.6 shows a clean log linear form of the relationship between population and per capita deforestation with different buffer sizes. This relationship motivates the decision of the functional form of the congestion externality in the model. Appendix C.1 investigates the relationship between the Market Access and forest cover changes (forest disturbance, forest loss, and forest recovery) implied by a simple one sector model and it reports consistent facts.

Forest products consist of non-timber forest products (NTFP) and timber.[9]

Exploiting the HS data for detailed household-level activity choices, we find the following two facts.

Fact 3A: Agriculture is widely observed in both concentrated and dispersed areas.

Fact 3B: Natural resource extraction is observed more in areas with lower surrounding population densities.

Figure 7 and Figures A.7–A.9 illustrate Facts 3A and 3B. These figures present cumulative distributions of households that do and do not engage in each activity in terms of the surrounding population density of their settlements with different buffer sizes.[10] The population density in a smaller buffer of a community becomes close to the population density of the community itself. A smaller population density in a large buffer of a community implies more sparseness of communities and less competition over natural resources. Figure A.7 shows that distributions of households with and without any activity in terms of the population density within 1km from the community are close. In contrast, Figure 7 shows that distributions of households with and without each natural resource activity in terms of the population density within 10km from the community become distinct, while those with and without agricultural activities (food crops and cash crops) remain close. Comparing those with other buffer sizes as well ($x =$2 km, 5km), we can summarize that spatial distributions of natural resource extractions become more distinct from agricultural activities as the buffer size enlarges.

Implication. These observations suggest that the spatial extent over which density externalities operate is heterogeneous across sectors. This takeaway motivates the model that incorporates multiple sectors with distinct types of density externalities.

4 Quantitative Spatial Model of Rainforest Communities

We construct a multi-sector spatial model that features the concentration and dispersion forces of the population in tropical forests. The model considers a general equilibrium in a river basin in the Amazon rainforest and analyzes trade across communities based on comparative advantage. The model is built on Michaels et al. (2011) and incorporates novel sector-specific density externalities and a rural-urban linkage in an environment of missing land market. While the model incorporates the rural-urban linkage to capture rural remoteness, it focuses primarily on the rural spatial structure and abstracts from the urban form.

[9]Timber logging also involves deforestation. For the deforestation outcome in counterfactuals, however, we focus on deforestation for agriculture. Moreover, timber logging constitutes the smallest share in the natural resource extraction sector and we abstract from its deforestation impact.

[10]The total population in the xkm buffer surrounding a community is measured by summing populations from census communities whose centroid are inside the buffer.

The model has three purposes. First, the model rationalizes the stylized facts presented in the previous section. Second, we model rainforest populations' fundamental trade-offs between richer resource endowments in sparse areas and higher market access and agglomeration benefits in dense areas. Third, counterfactual experiments based on the model quantify two types of trade-offs from policy planners' perspective. The first trade-off is between rainforest conservation and local populations' welfare. The second type of trade-off is between different aspects of conservation, such as between deforestation and biological resource depletion.

4.1 Geography and Population

We consider a general equilibrium in each river basin b. Each basin consists of a set of rural locations ($\mathcal{R}_b$) and one urban center (u_b), indexed by o (origin) and d (destination), where $o, d \in \mathcal{I}_b = \mathcal{R}_b \bigcup u_b$.[11] There are three sectors—agricultural production (Ag) and natural resource extraction (Nr) in the rural locations and an urban sector (M) in the urban center. Locations exogenously differ in productivity fundamentals (e.g., soil quality and access to water) in each sector and distance to other locations given the river networks. The total population in each basin is $\overline{N}_b$ and fixed, while we assume the free mobility of the population within the basin.[12] The total land area (including areas covered by rainforests) in each basin is $\overline{L}_b$. The notation of basin b is omitted for simplicity of exposition hereafter in this section.

Each worker in o obtains wage w_o by inelastically supplying one unit of labor and chooses how much of each variety (indexed by j) of the rural sectoral goods and of the urban good, all of which are produced in the basin. In particular, a worker solves the following problem:

$$\begin{aligned} &\max_{\{c_{o,Ag}(j), c_{o,Nr}(j), C_{o,M}\}} \left[\alpha_{Ag} C_{o,Ag}^{\frac{\bar{\sigma}-1}{\bar{\sigma}}} + \alpha_{Nr} C_{o,Nr}^{\frac{\bar{\sigma}-1}{\bar{\sigma}}} + \alpha_M C_{o,M}^{\frac{\bar{\sigma}-1}{\bar{\sigma}}}\right]^{\frac{\bar{\sigma}}{\bar{\sigma}-1}} \\ \text{s.t.} \quad &\int_0^{n_{Ag}} p_{o,Ag}(j) c_{o,Ag}(j) dj + \int_0^{n_{Nr}} p_{o,Nr}(j) c_{o,Nr}(j) dj + P_{o,M} C_{o,M} = w_o \end{aligned}$$

where $C_{o,Ag} \equiv \left[\int_0^{n_{Ag}} c_{o,Ag}(j)^{\frac{\sigma-1}{\sigma}} dj\right]^{\frac{\sigma}{\sigma-1}}$ and $C_{o,Nr} \equiv \left[\int_0^{n_{Nr}} c_{o,Nr}(j)^{\frac{\sigma-1}{\sigma}} dj\right]^{\frac{\sigma}{\sigma-1}}$ are continuums of varieties of the agricultural and natural resource goods. $C_{o,M}$ represents consumption of the single urban good and $P_{o,M}$ is its price in o. σ represents the elasticity of substitution across varieties within each sector and $\bar{\sigma}$ represents the elasticity of substitution across sectors. Solving this problem yields the following indirect utility:

$$V_o = \frac{w_o}{\left[\sum_{K=Ag,Nr,M} \alpha_{o,K}^{\bar{\sigma}} P_{o,K}^{1-\bar{\sigma}}\right]^{\frac{1}{1-\bar{\sigma}}}}$$

[11] The urban centers in the four basins are Iquitos in the Napo-Amazon basin, San Lorenzo in the Pastaza basin, Orellana in the Lower Ucayali basin, and Pucallpa in the Upper Ucayali basin, respectively.

[12] Since indigenous and folk populations tend to have been living within a particular region for a long time since their ancestors, we do not consider population inflows to a river basin or outflows from a river basin. Indeed, the CC data show that only 1.4% of the rural communities are colonist settlements. At the same time, within the narrow geographical scale of a river basin, we observe migration and relocating community locations in response to economic opportunities, which would be consistent with the assumption of labor mobility inside a river basin.

expenditure shares across varieties within each rural sector in each location:

$$\tilde{\alpha}_{o,K}(j) \;=\; \frac{P_{o,K}^{\sigma} p_{o,K}(j)^{(1-\sigma)}}{\sum_{K'=Ag,Nr,M} P_{o,K'}} \quad K = Ag, Nr \tag{1}$$

and expenditure shares across sectors in each location:

$$\tilde{\alpha}_{o,K} \;=\; \frac{\alpha_K^{\bar{\sigma}} P_{o,K}^{(1-\bar{\sigma})}}{\sum_{K'=Ag,Nr,M} \alpha_{K'}^{\bar{\sigma}} P_{o,K'}^{1-\bar{\sigma}}} \quad K = Ag, Nr, M \tag{2}$$

where $P_{o,Ag} \equiv \left[\int_0^{n_{Ag}} p_{o,Ag}(j)^{1-\sigma} dj\right]^{\frac{1}{1-\sigma}}$ and $P_{o,Nr} \equiv \left[\int_0^{n_{Nr}} p_{o,Nr}(j)^{1-\sigma} dj\right]^{\frac{1}{1-\sigma}}$ are sectoral price indices.

4.2 Production with Density Externalities

Agricultural production with congestion and agglomeration externalities

Agricultural production consists of two steps. The first step is to access land for cultivation by clearing forest. There is no land market. We thus define the production function of "land access" for cropping variety j as follows:

$$L_o(j) = A_{o,L} N_{o,Ag}^{-\mu_L} \cdot N_{o,L}(j), \quad o \in \mathcal{R} \tag{3}$$

where $L_o(j)$ is the amount of land cleared for cropping variety j of the agricultural products and $N_{o,L}(j)$ represents employment used to access this land.

The first factor of the production function $(A_{o,L} N_{o,Ag}^{-\mu_L})$ represents the productivity composite which combines productivity fundamentals $(A_{o,L})$ and a congestion force. We assume the common productivity fundamentals $A_{o,L}$ of land access for cropping different varieties in each location, given that mixed cropping is commonly observed in the shifting cultivation system in tropical forests. The congestion force in forest clearing depends on total agricultural employment in the location $(N_{o,Ag})$ and is governed by parameter μ_L. A positive value for this parameter indicates that as the community population increases, the accessibility of land available to each person is reduced.[13]

The second step is to produce agricultural goods given the accessed land. We define the

[13] Farmers clear forests to obtain agricultural land only nearby their residential locations in the absence of the land market and property rights and with a high monitoring cost. As the community population increases, the cost of access to farmland through forest clearing would become higher. In the absence of a land market, community members negotiate with each other to distribute the land they each use as farmland. The larger the community population, the higher the cost of negotiating and monitoring areas for forest clearing. For these reasons, we hypothesize that $\mu_L > 0$, which is also consistent with Fact 2A.

production function of variety j of agricultural goods as follows:

$$Q_{o,Ag}(j) \quad = \quad z_{o,Ag}(j) N_{o,Ag}^{\mu_{Ag}} \cdot N_{o,C}(j)^{\gamma} L_o(j)^{(1-\gamma)}, \quad o \in \mathcal{R} \tag{4}$$

where $N_{o,C}(j)$ represents employment for cropping variety j on the cleared land. That is, the total agricultural employment in o is $N_{o,Ag} = N_{o,L} + N_{o,C} = \int_0^{n_{Ag}} (N_{o,L}(j) + N_{o,C}(j))dj$.

The first factor $(z_{o,Ag}(j) N_{o,Ag}^{\mu_{Ag}})$ represents the productivity. A stochastic factor in the productivity, $z_{o,Ag}(j)$, follows the Fréchet distribution such that $F_{Ag}(z) = \exp(-A_{o,Ag} z^{-\theta})$, where $A_{o,Ag}$ represents the absolute advantage and θ represents the comparative advantage. Lower θ corresponds to higher dispersion of productivity. Productivity in this second step again depends on the total agricultural employment in the location. Parameter μ_{Ag} governs the agglomeration force in agricultural production and marketing. A positive value for this parameter indicates that as the community agricultural population increases, agricultural productivity increases.[14]

Natural resource extraction with congestion externality across space

We define the production function of extracting variety j of natural resources as follows:

$$Q_{o,Nr}(j) \quad = \quad z_{o,Nr}(j) \Big[\sum_{d \in \mathcal{R}} D_{od}^{-\nu} N_{d,Nr} \Big]^{-\mu_{Nr}} \cdot N_{o,Nr}(j), \quad o \in \mathcal{R} \tag{5}$$

where $N_{o,Nr}(j)$ represents employment for extracting variety j of natural resource. Note that the amount of $Q_{o,Nr}(j)$ is natural resource extracted by individuals residing at o from surrounding areas by search and travel (i.e., not extracted only from o). The total employment for natural resource extraction in o is $N_{o,Nr} = \int_0^{n_{Nr}} N_{o,Nr}(j)dj$. In contrast to agricultural production, labor is the only factor for natural resource extraction. In rural locations $o \in \mathcal{R}$ the total population is thus $N_o = N_{o,Ag} + N_{o,Nr}$.

The first factor $(z_{o,Nr}(j) \Big[\sum_{d \in \mathcal{R}} D_{od}^{-\nu} N_{d,Nr} \Big]^{-\mu_{Nr}})$ represents the productivity. A stochastic factor in the productivity, $z_{o,Nr}(j)$, also follows the Fréchet distribution such that $F_{Nr}(z) = \exp(-A_{o,Nr} z^{-\theta})$, where $A_{o,Nr}$ represents the absolute advantage and θ is the same comparative advantage parameter defined above. Productivity depends on employment of natural resource extraction not only in the own location but also in surrounding locations. Parameter μ_{Nr} governs the congestion force with spatial spillover from the surrounding population in the common pool natural resource extraction. Variable D_{od} represents the river-equivalent distance between cells o and d along the shortest path. Parameter ν governs the spatial decay in access to natural resources. The implicit assumption underlying this specification is that people travel longer distances for natural resource extraction than for agriculture.[15]

[14]We hypothesize that $\mu_{Ag} > 0$, but the empirical analysis for estimating this parameter does not restrict it to be positive. There are several possible mechanisms under which the agglomeration externality may exist. Section 6 provides a detailed discussion.

[15]We hypothesize that $\mu_{Nr} > 0$ and $\nu > 0$, which are consistent with Fact 3B, but the empirical analysis for

Urban good production at the urban center

The urban good contains miscellaneous non-food (manufacturing and service) items, but for simplicity we express it as a single good and define its production function as:

$$Q_{u,M} \;=\; A_{u,M} \cdot N_{u,M} \tag{6}$$

where $N_{u,M}$ $(= N_u)$ is employment in the urban center and $A_{u,M}$ is exogenous productivity.

4.3 Prices and Trade

Let $p_{od,K}(j)$ be the price of product j in sector K produced in o (origin) to be purchased in d (destination). We define the iceberg trade cost of sector K goods, denoted by $\tau_{od,K}$, such that it satisfies the following relationships: for $K \in \{Ag, Nr\}$, $p_{od,K}(j) = \tau_{od,K} p_{oo,K}(j)$, $\tau_{oo,K} = 1$, $\tau_{od,K} > 1$ for $o \neq d$, and $\tau_{od,K} < \tau_{oi,K}\tau_{id,K}$ for $o \neq i \neq d$, where $o \in \mathcal{R}$ and $i, d \in \mathcal{I}$. For the urban good, $P_{d,M} = p_{ud,M} = \tau_{ud,M} p_{uu,M} = P_{u,M}$, where $\tau_{ud,M} > 1$ for $d \in \mathcal{R}$ and $\tau_{uu,M} = 1$.

Trade cost between a pair of locations may be asymmetric. That is, while the above specification of iceberg trade cost is standard in the literature, $\tau_{od,K} = \tau_{do,K}$ does not necessarily hold because of river orientations. The trade literature often assumes symmetric trade costs, but the asymmetry would matter in particular environments such as transports along a river or a high-slope road and during periods before modern transport technologies were adopted. For example, Chen et al. (2022) discuss that exogenously determined asymmetric trade costs play an important role in predicting city locations in ancient Greece where the major transport mode was sailing.[16]

We define D_{od} as the downstream-river-equivalent kilometers along the lowest-cost route from o to d and assume that:

$$D_{od} = D_{od,down} + \lambda_{up} D_{od,up} + \lambda_{land} D_{od,land} \tag{7}$$

where $D_{od,down}$, $D_{od,up}$, and $D_{od,land}$ are the distances (in kilometers) of going downstream and upstream on the river and the distance of land travel. λ_{up} and λ_{land} are the parameters capturing the downstream-river-equivalent distance per kilometer of upstream and land travels (i.e., relative travel costs in terms of downstream-river travel). We then parameterize the iceberg trade cost as:

$$\tau_{od,K} = D_{od}^{\delta_K} \tag{8}$$

estimating these parameters does not restrict them to be positive. In addition, forest clearing for agriculture may also affect biological resource endowments and this sector's productivity. However, we are not incorporating this across-sector externality. Section 5.3.2 describes the reasons for this model choice in detail.

[16] Trade costs could also be endogenous. See, for example, Brancaccio et al. (2020) and section 6 for more detailed discussion.

where δ_K is the elasticity of trade cost of sector K goods with respect to the effective distance.

We assume perfect competition among producers of all the sectors. Under perfect competition, prices are equalized to marginal costs as follows. For the agricultural goods, $p_{oo,Ag}(j) = \frac{w_o}{z_{o,Ag}(j)A_{o,L}^{(1-\gamma)}N_{o,Ag}^{\tilde{\mu}_{Ag}}\kappa_1}$, where $\tilde{\mu}_{Ag} \equiv \mu_{Ag} - \mu_L(1-\gamma)$ represents the composite of density externalities in agriculture and κ_1 contains constant terms.[17] For the natural resource goods, $p_{oo,Nr}(j) = \frac{w_o}{z_{o,Nr}(j)[\sum_{d\in\mathcal{R}} D_{od}^{-\nu}N_{d,Nr}]^{-\mu_{Nr}}}$. For the urban good, $P_{u,M} = \frac{w_o}{A_{o,M}}$.

Following Eaton and Kortum (2002), we can derive each destination's expenditure share on goods shipped from each origin. For the rural goods $K = Ag, Nr$, the expenditure share across locations within a sector is expressed as:[18]

$$\pi_{od,K} \;=\; \frac{\tilde{A}_{o,K}(w_o\tau_{od,K})^{-\theta}}{\sum_{o'\in\mathcal{R}}\tilde{A}_{o',K}(w_{o'}\tau_{o'd,K})^{-\theta}}, \quad o\in\mathcal{R},\, d\in\mathcal{I} \tag{9}$$

where $\tilde{A}_{o,K}$ represents the productivity composite which contains both exogenous and endogenous components of productivity:

$$\begin{aligned}\tilde{A}_{o,Ag} &\equiv A_{o,Ag}A_{o,L}^{(1-\gamma)\theta}N_{o,Ag}^{\tilde{\mu}_{Ag}\theta}\kappa_1^{\theta} \\ \tilde{A}_{o,Nr} &\equiv A_{o,Nr}[\sum_{d\in\mathcal{R}} D_{od}^{-\nu}N_{d,Nr}]^{-\mu_{Nr}\theta}\end{aligned} \tag{10}$$

The urban good is only purchased from the urban center u in all locations.

4.4 Spatial Equilibrium

The spatial equilibrium is defined as the following conditions.

1. The labor market for agricultural production clears in all the rural locations:

$$w_oN_{o,Ag} \;=\; \sum_{d\in\mathcal{I}}\pi_{od,Ag}\tilde{\alpha}_{d,Ag}w_dN_d \quad \forall\, o\in\mathcal{R} \tag{11}$$

2. The labor market for natural resource extraction clears in all the rural locations:

$$w_oN_{o,Nr} \;=\; \sum_{d\in\mathcal{I}}\pi_{od,Nr}\tilde{\alpha}_{d,Nr}w_dN_d \quad \forall\, o\in\mathcal{R} \tag{12}$$

3. The labor market for the urban good clears in the urban center:

$$w_uN_{u,M} \;=\; w_uN_u = \sum_{d\in\mathcal{I}}\tilde{\alpha}_{d,M}w_dN_d \tag{13}$$

[17]Specifically, $\kappa_1 = \gamma^{\gamma}(1-\gamma)^{(1-\gamma)}$.

[18]= Pr (d buys a good from o) = Pr (o sells a good at the lowest price to d) = Fraction of goods that d buys from o

4. The overall labor market clears:

$$\overline{N} = \sum_{o\in\mathcal{R}} N_o + N_u = \sum_{o\in\mathcal{R}} \sum_{K\in\{Ag,Nr\}} N_{o,K} + N_{u,M} \tag{14}$$

5. Utility is equalized across populated locations due to free labor mobility:

$$V_o = \frac{w_o}{\left[\sum_{K=Ag,Nr,M} P_{o,K}^{1-\bar{\sigma}}\right]^{\frac{1}{1-\bar{\sigma}}}} = \overline{U} \quad \forall\, o \in \tilde{\mathcal{I}} \tag{15}$$

   where $\tilde{\mathcal{I}}$ is the set of all locations that have positive populations.

6. The total deforested area does not exceed the available land area: $\sum_{o\in\mathcal{R}} L_o \leq \bar{L}$

Solving the model involves solving for endogenous employment shares and wages, given the geography, productivity fundamentals, parameters, and the total population in the economy.

5 Estimating the Model

Table 2 summarizes the parameters of the model. We estimate these parameters in the following sequential steps and then quantify the effects of density externalities.

5.1 Obtaining Parameters without Solving the Model (Step 1)

Trade cost parameters

We first calibrate the elasticity of trade cost with respect to the effective downstream-river-equivalent distance for each sector ($\{\delta_K\}$), using price observations in the populated grid cells and the downstream river distance. The CC data collect prices of representative products of agricultural and natural resource sectors in the census communities (see Appendix B.1 for details). We compile the grid cell-level price data by computing average prices of goods in each cell. We then obtain price ratios of these products between all pairs of grid cells that have non-missing price observations. We calibrate δ_K by minimizing the squared sum of $(p_{od,K}^{Max} - D_{od,down}^{\delta_K})$ where $p_{od,K}^{Max} \equiv \max_j\{p_{d,K}(j)/p_{o,K}(j)\}$ is the maximum price ratio across all the varieties in each sector among location pairs that satisfy $D_{od,up} = D_{od,land} = 0$. Although the price ratio should be constant in j under the model, we need to take the maximum price ratio among the observed prices for the following two reasons. First, this approach mitigates the concern that observed price gaps of each product might underestimate trade costs, given our limited price information[19] and that price ratios between two locations are bounded above by trade costs

[19]The price data have the following limitations. First, we do not have multiple periods of price observations. Second, we have a smaller number of products compared with the actual potential number of varieties that people in our study area are producing. Third, we do not observe origins and destinations of these varieties.

between them (Eaton and Kortum 2002). Second, we assume that $\tau_{od,K}$ is the cost of transport by the most widely available transport mode ("peque-peque," shown in the top-right picture of Figure A.10). However, even if the location pair consists of actual origin and destination, the observed price ratio may reflect transports by other boat types that are superior (e.g., the bottom two pictures of Figure A.10), in which case the trade cost is underestimated. We obtain $\hat{\delta}_{Ag} = 0.178$, $\hat{\delta}_{Nr} = 0.137$, and $\hat{\delta}_M = 0.098$. These values lie in a standard range of values from the related literature. Values of trade costs that are higher than our estimates nevertheless lead to qualitatively robust subsequent results and just amplify our core findings.

We next calibrate the effective distance parameters in terms of the downstream distance on the river (λ_{up}, λ_{land}). We obtain $\hat{\lambda}_{up} = 1.282$, using records of travel time by peque-peque for several routes in the Peruvian Amazon.[20] We obtain $\hat{\lambda}_{land} = 36.767$, using records of transportation costs by land travels. Appendix C.2 provides the detail procedure to estimate these parameters.

Having obtained the lengths of river (downstream and upstream) and land travels in the shortest paths and the relative distance parameters, we compute the effective downstream-river-equivalent distance for all pairs of grid cells. To solve the spatial equilibrium in a later stage, we set $\tau_{od,K} = \hat{D}_{od}^{\hat{\delta}_K}$, where $\hat{D}_{od}$ is calculated by (22) using $\hat{\lambda}_{up}$ and $\hat{\lambda}_{land}$.

Demand parameters

We estimate elasticity of substitution between varieties within each sector (σ) and between sectoral composite products ($\bar{\sigma}$), using household-level information on expenditures and unit values (interpreted as buying prices) from the Peru National Household Survey. Appendix C.2 describes the data and provides the detailed procedure for estimating these parameters. We obtain $\hat{\sigma} = 2.401$ (substitutes since $\hat{\sigma} > 1$) and $\hat{\bar{\sigma}} = 0.752$ (complements since $\hat{\bar{\sigma}} < 1$).[21]

Other parameters

We assume that the labor cost share in the agricultural production is $\hat{\gamma} = 0.6$, which lies in a standard range of values from the related literature. For example, Sotelo (2020) reports that the average labor cost share in crop production in Peru is 0.55. We assume that the productivity dispersion (the shape parameter) in the rural sectors is $\hat{\theta} = 7.8$, drawn from the study by Donaldson (2018) on agricultural trade in the context of developing economies.[22]

[20]The downstream-river-equivalent distance measure specified by (22) implicitly assumes the constant slope everywhere. Moreover, going upstream may need more fuel and thus upstream costs may be higher than accounted for by travel time. Using a value of upstream cost higher than the estimate amplifies our core story. Future work will construct a more sophisticated measure of the effective distance that takes river orientations with slopes in each travel route and fuel costs into account.

[21]Under alternative specifications (i.e., alternative controls, instruments, and methods for selecting the tuning parameter of LASSO), estimates of σ range around [2.4, 3.3] and those of $\bar{\sigma}$ range around [0.63, 0.8]. Subsequent results with different values of (σ, $\bar{\sigma}$) are qualitatively robust.

[22]Donaldson (2018) estimates this parameter by a gravity equation using data on quantities of agricultural commodities traded internally in India during 1882-1920. His empirical setting has a similarity to ours because

5.2 Model Inversion to Recover Productivity Composites (Step 2)

Given the parameters obtained in the previous step, we invert the model to obtain productivity composites of all the sectors and wages ($\{\tilde{A}_{o,Ag}, o \in \tilde{\mathcal{R}}\}$, $\{\tilde{A}_{o,Nr}, o \in \tilde{\mathcal{R}}\}$, $A_{u,M}$, $\{w_o, o \in \tilde{\mathcal{I}}\}$) that rationalize the observable data (sectoral employment share and total population in each location) as a spatial equilibrium. In each basin, we use $2|\tilde{\mathcal{R}}| + 1 + |\tilde{\mathcal{I}}|$ $(= 3|\tilde{\mathcal{I}}| - 1)$ equations with the observables to solve for $2|\tilde{\mathcal{R}}| + 1 + |\tilde{\mathcal{I}}|$ unknowns, where $\tilde{\mathcal{R}}$ is the set of rural locations that have positive populations. We observe sectoral employments in grid cells of rural locations and total populations in the urban centers.[23] The $2|\tilde{\mathcal{R}}| + 1 + |\tilde{\mathcal{I}}|$ equations involve sectoral labor market clearing in all locations and utility equalization across space, expressed by (11), (12), (13), and (15). The $2|\tilde{\mathcal{R}}| + 1 + |\tilde{\mathcal{I}}|$ unknowns are the productivity composites and wages.

The model inversion reduces to a nested fixed point problem. The algorithm is described in Appendix C.3. By the construction of the inversion problem, the model perfectly fits the data of sectoral employment.[24] Proving the uniqueness of the solution to this inversion problem follows Michaels et al. (2011).

5.3 GMM Estimation of Density Externality Parameters (Step 3)

Using productivity composites obtained in the previous step as data and exploiting exogenous river shape, we employ the GMM to estimate parameters governing the density parameters. We first estimate the parameter governing density externalities in agriculture using both grid cell-level model-driven data and community-level information of forest cover. We then estimate congestion externality in natural resource extraction.

Table 3 summarizes the estimation results of these parameters. Below we describe each step.

5.3.1 Density externalities in agriculture

Net agglomeration externality in the productivity composite in agriculture

The inverted productivity composite of agricultural production is $\tilde{A}_{o,Ag} \equiv A_{o,Ag} A_{o,L}^{(1-\gamma)\theta} N_{o,Ag}^{\tilde{\mu}_{Ag}\theta} \kappa_1^{\theta}$. Taking the logarithm yields the following linear empirical specification:

$$\ln \tilde{A}_{o,Ag} = \tilde{\mu}_{Ag} \theta \ln N_{o,Ag} + X_o' \beta + \phi_B + \epsilon_{o,Ag} \tag{16}$$

agriculture constituted a significant share in India's economy during this period. Moreover, this parameter value is in the standard range of estimates from developing countries in the literature and is higher than estimates from developed countries. Given that our study area is narrower than the usual context (e.g., a single country), comparative advantage may play a smaller role and the geography exerts a stronger force than in other settings in the literature. However, larger values of θ will amplify the main results and will not change the core story of this paper.

[23] See Appendix B.1 for the construction of the sectoral employment shares.

[24] The calibrated productivity composites are correlated with non-targeted data and moments. For example, panel (A) of Table A.1 reports a significant positive correlation between the reported number of species found around a community (from the CC data) and the calibrated productivity of the natural resource sector.

where $\ln N_{o,Ag}$ is the logarithm of agricultural employment at location o, X_o is a vector of geographical controls that can be regarded as exogenous productivity fundamentals (including the constant)[25], ϕ_B represents basin fixed effect, and $\epsilon_{o,Ag}$ includes unobservable factors in $A_{o,Ag}$ and $A_{o,L}$. $\ln N_{o,Ag}$ is likely to be endogenous in this econometric specification, since the productivity fundamentals ($A_{o,Ag}$ and $A_{o,L}$) affect the agricultural population in the theoretical model. That is, agricultural labor demand would be higher at locations that have higher productivity fundamentals, but unobservable elements captured in $\epsilon_{o,Ag}$ might remain. To estimate the coefficient of interest, $\tilde{\mu}_{Ag}\theta$, we instrument $\ln N_{o,Ag}$ by $\ln RNA_o$. RNA_o is the "river network access" measure at o defined by:

$$RNA_o = \sum_{d \in RC} \tau_{od}^{\theta} \tag{17}$$

where RC is a set of all river cells in the basin (cells that contain a river) whether or not they have positive populations.[26] Column (1) of Table A.2 presents the first-stage regression result and that RNA significantly predicts the current agricultural population.

In addition to the river structure being pre-determined and unaffected by human settlements, we need the following identifying assumptions. The independence assumption is that, after controlling for own-location characteristics, productivity fundamentals are uncorrelated with accessibility to other locations. The corresponding moment condition is $\mathbb{E}[\epsilon_{o,Ag} \ln RNA_o] = 0$. The exclusion restriction is that market opportunity (due to the accessibility to other locations) affects productivity only through its effect on employment and thus through externalities that arise.

Figure 8 provides intuition behind these identifying assumptions. This figure represents three similar areas inside a river basin. We focus on the relationship between agricultural population and productivity by comparing the three cells with different colors (pink, red, and brown) that are located next to the bottom-left cells. We assume that these three cells have the same observable productivity fundamentals (e.g., soil types, distance to the nearest river point, lake size, elevation, flood risk) that affect labor demand. The only observable difference between these cells in this figure is RNA, that is, the weighted accessibility to other cells that face rivers (whether or not these cells have positive populations). The middle and right maps have one more river cell than the left map, each in a different location. Therefore, RNA is higher in the red and brown cells than that in the pink cell. Comparing the red cell in the middle map with the brown cell in the right map, RNA is higher in the brown cell because the additional cell next to the top-right is located closer to the brown cell. This variation in RNA, stemming from the river shape in nature, is regarded as a shifter of market potential. That is, RNA affects the population size because of the trade mechanism. There are still unobservable productivity

[25] The geographical controls include a dummy of high river orders (4 and 5), distance to the urban center, distance to the river, squared distance to the river, interaction terms of these two variables with a river cell dummy, elevation, river confluences, flood vulnerability, soil conditions, and water shares (main and non-main channels). We provide definitions of these variables in Appendix B.3.

[26] For RNA, we use the sector-independent symmetric proximity measure $\tau_{od} = D_{od}^{\delta}$ where D_{od} is the travel distance without taking the river asymmetry into account and δ is obtained from price gaps pooling all the sectors.

shifters remaining in the own location. Rivers around a location and water level fluctuations may also affect unobservable soil quality directly. However, there would be no theoretically plausible reason to believe that, given the same observable geographical conditions in the own location, the unobservable productivity shifters are associated with the variation in RNA that can be generated by exogenous river shapes in locations far away from the own location. The comparison between the three panels illustrates this point.

Holding observable productivity fundamentals fixed in this stylized example corresponds to controlling for observable geographic characteristics in the actual estimation. Panel (A) of Table A.3 reports that, given a subset of geographical controls,[27] RNA is uncorrelated with potential productivity shifters: the share of water other than main channel rivers, river confluences, flood vulnerability, and soil conditions unrelated to rivers. These non-correlations imply the plausibility of assuming that RNA is also uncorrelated with other unobservable productivity shifters given the rich set of full geographical controls, although there is no formal way to prove this claim.

Furthermore, although the theoretical model is static, the agglomeration effect in the empirical setting could contain the effect of past history of human settlement in a particular location (e.g., accumulation of knowledge). Even if predetermined productivity fundamentals are the same, locations with longer histories may attract larger populations today through agglomeration spillovers over time. That is, if early settlements are not correlated with productivity fundamentals, then they would affect current productivity only through their effects on current populations caused by the agglomeration spillovers over time.

Motivated by this conjecture and supported by the subsequent arguments, we also construct, as an instrumental variable, a dummy for whether the community was established in its current location by 1940. Importantly, the primary reason for early settlement was the opportunity to obtain natural resource products. The rubber boom, which began in the late 19th century but collapsed around 1940, had a significant impact on initial settlements (Barham et al. 1996; Coomes 1995). Moreover, even after the collapse of the wild rubber economy, the primary activity of many communities was initially natural resource extraction and then has shifted towards agriculture over time (Coomes et al. 2016). Therefore, it would be plausible to assume that the locations of communities established before 1940 were determined primarily by natural resource endowments, not by advantages in agricultural productivity.[28]

The data also support the validity of using this instrumental variable. Column (2) of Table A.2 presents the first-stage regression result and that the early settlement significantly predicts

[27]The geographical characteristics controlled in this panel includes elevation and river characteristics including distance to the river, share of main channel rivers, and floodplain soil share. The share of main channel rivers is controlled as it could be theoretically correlated with RNA: RNA may be higher where the share of main channel river water is higher, possibly because such larger river lines tend to be located in the central area of a basin. The floodplain soil share is controlled because floodplain soils are formed along whitewater rivers, which may also be mechanically correlated with RNA.

[28]For this reason, it is not valid to use this instrument for estimating the density externality in natural resource extraction in section 5.3.2.

the current agricultural population. Moreover, panel (B) of Table A.3 reports that, given a subset of geographical controls,[29] the early settlement location is uncorrelated with potential agricultural productivity shifters: the share of water (both main channel rivers and non-main channels), flood vulnerability, and all of the three soil quality variables. This randomness implies the plausibility of assuming that the early settlement location is also uncorrelated with other unobservable productivity shifters given the rich set of full geographical controls.

Table A.4 reports the IV estimation results and confirms the presence of the agglomeration externality in the agricultural sector. Columns (2) and (3) report that the point estimate of $\tilde{\mu}_{Ag}\theta$ is positive and statistically significant with either RNA or the historical community existence as an IV. The point estimate is larger with the historical IV. This difference would be because the latter estimate contains more of the agglomeration spillover from the history of human settlements. In our preferred specification, we use both IVs given that these two IVs exploit distinct variations that influence contemporary populations. The right two columns in Table A.2 indicate that these two instruments are uncorrelated both unconditionally and conditionally on the geographic controls. This combination also achieves a significant first stage fit (column (3) of Table A.2). In this preferred specification (column (4) of Table A.4), the point estimate $\tilde{\mu}_{Ag}\theta$ is 0.501 at the 1% level of statistical significance. The J-statistic of the overidentifying restrictions J test fails to reject the null hypothesis that the instruments are valid (p-value = 0.648). Dividing the IV estimate by $\hat{\theta}$, we obtain the reported point estimate $\hat{\tilde{\mu}}_{Ag} = 0.064$ (standard error = 0.010) in Table 3. This result confirms the presence of agglomeration externality in the agricultural sector *on net*. That is, this result implies that the agglomeration externality in agricultural production outweighs the congestion externality in land access. Table A.4 also shows that the OLS estimate is larger than the IV estimate in any specification. This observation is consistent with the view that agricultural employment is correlated with agricultural productivity unobservable by econometricians and thus the OLS estimate has an upward bias.

The result is robust to different sample selections and different sets of geographic controls. In particular, this result is not driven by a small number of locations with very large populations or by a specific set of geographical control variables. Columns (5) through (8) report the estimation results by restricting the sample locations to only those with small populations (e.g., $< 1,000$). While the point estimates are slightly smaller, the statistical significance of the agglomeration externality remains the same. Moreover, Table A.5 presents that the significance (both statistically and economically) of the agglomeration externality is stable across different combinations of the geographical control variables.

[29]The geographical characteristics controlled for in this panel includes elevation and basic river characteristics such as river proximity and river confluences. River confluences are controlled because we expect that such locations may have higher accessibility to various destinations and thus attract people for reasons unrelated to agriculture.

Decomposition into congestion externality in forest clearing and agglomeration externality in agricultural production

Recall from section 4.3 that $\tilde{\mu}_{Ag} \equiv \mu_{Ag} - (1-\gamma)\mu_L$. That is, the overall agglomeration in agriculture based on the estimation in the previous step consists of both congestion externality in forest clearing and agglomeration externality in agricultural production.

To decompose the net density externality into these two components, we first estimate the parameter governing congestion externality in forest clearing, using community-level (not grid cell-level) information. We use community-level information because it is ideal to have the information of deforestation and associated population that caused it. It is not feasible to identify the population that caused each deforestation area with the grid cell-level data. Note also that in the estimation of density externality in forest clearing, we do not use any information from solving the model. This is another reason why we do not have to stick with using the grid cell-level data. In particular, we use information of the community-level land footprint after clearing the primary forest within each voronoi polygon around the settlement, constructed by Coomes et al. (2021) and shown in Figure A.11. This community-level land footprint contains all patches of agricultural fields and secondary forests detected in satellite images.

The land access function (3) and $N_{o,L} = (1-\gamma)N_{o,Ag}$ at the optimal input choice yield the following empirical specification:

$$\ln \frac{L_o}{N_{o,Ag}} = -\mu_L \ln N_{o,Ag} + X_o'\beta + \phi_B + \epsilon_{o,L} \tag{18}$$

To estimate μ_L, we follow the same identification strategy as above. The same identifying assumption holds because the residual in (18) is also contained in the residual in (16). Therefore, we use exactly the same set of IVs and geographical controls. Table A.6 reports the results. We obtain the point estimate $\hat{\mu}_L = 0.522$ (standard error = 0.094).

From $\hat{\mu}_L$, $\hat{\tilde{\mu}}_{Ag}$, and the relationship $\tilde{\mu}_{Ag} \equiv \mu_{Ag} - (1-\gamma)\mu_L$, we back out the point estimate of the agglomeration externality in agricultural production (given the land cleared for agriculture) $\hat{\mu}_{Ag} = 0.273$ reported in Table 3.

5.3.2 Congestion externality in natural resource extraction with Spatial Spillovers

The inverted productivity composite of natural resource extraction is $\tilde{A}_{o,Nr} \equiv A_{o,Nr}[\sum_d D_{od}^{-\nu} N_{d,Nr}]^{-\mu_{Nr}\theta}$. By taking the logarithm, we can express the residual variation in $\ln A_{o,Nr}$ (productivity fundamentals) as follows:

$$\epsilon_{o,Nr} = \ln \tilde{A}_{o,Nr} + \mu_{Nr}\theta \ln\Big(\sum_d D_{od}^{-\nu} N_{d,Nr}\Big) - X_o'\beta - \phi_B \tag{19}$$

Likewise, employment of natural resource extraction in the own and surrounding locations is likely to be endogenous. We thus estimate the congestion externality parameter (μ_{Nr}) and its spatial decay parameter (ν) by non-linear GMM with the instruments plausibly satisfying the following moment conditions:

$$\mathbb{E}[\epsilon_{o,Nr} \ln RNA_o] = 0 \quad \text{and} \quad \mathbb{E}[\epsilon_{o,Nr} \ln(\sum_{d|D_{o,d} \leq x} RNA_d)] = 0, \; x \in \mathcal{X} \tag{20}$$

Panel (B) of Table 3 reports the results of the nonlinear GMM estimation. We select our instruments with $\mathcal{X} = \{2, 5, 10, 25, 50, 75, 100\}$. The point estimates and standard errors are based on the two-step estimation using the optimal weight matrix. We construct the optimal weight matrix using the first-step estimates of the parameters that were estimated using the identical matrix as a weight matrix. We get the point estimate of the congestion externality parameter $\hat{\mu}_{Nr} = 0.335$ (standard error = 0.042) with the spatial decay parameter $\hat{\nu} = 0.593$ (standard error = 0.075). Given the number of instruments, these parameters are over-identified. We report the result of the overidentifying restrictions J test. The resulting J-statistic fails to reject the null hypothesis that the instruments are valid (p-value = 0.821).[30]

To understand the natural resource competition realistically—that is, to understand the actual distance where spatial spillovers of congestion externality in natural resource extraction are strong—we first approximate the productivity composite by the following:

$$\tilde{A}_{o,Nr} \equiv A_{o,Nr} N_{o,Nr}^{-\mu_{Nr}\theta} \prod_{x \in \mathcal{X}} \Big[\sum_{d|D_{o,d} \leq x} N_{d,Nr} \Big]^{-\mu_{Nr,x}\theta}$$

and employ a linear specification to estimate μ_{Nr} and $\{\mu_{Nr,x}\}$ with instruments $\ln RNA_o$ and $\ln \sum_{d|D_{o,d} \leq x} RNA_d$ for $x \in \mathcal{X}$. Table A.7 reports results with different $\mathcal{X}$s, but they are subsets of $\mathcal{X}$ used for the main non-linear estimation.

There are two noteworthy findings. First, the negative effect of a population engaging in natural resource extraction in the own location becomes weaker both economically and significantly compared with the effects of surrounding populations once we also consider them with wider distance ranges. This empirical pattern is consistent with the presence of congestion externality with spatial spillovers. Interestingly, once we control for surrounding populations with comprehensive distance ranges, the coefficient sign of the own population even becomes positive. This observation is consistent with the view that there is also an agglomeration benefit as we saw in the agricultural sector even though the congestion force outweighs on net. Note also that the OLS estimate of the coefficient of own population is larger than the IV estimate. This observation suggests that employment of natural resource extraction is correlated with its unobservable productivity and thus the OLS estimate has an upward bias.

Second, as we increase the number of distances in $\mathcal{X}$, the size of point estimate $\hat{\mu}_{Nr,x_{max}}$ gradually decreases. All of their signs remain the same up to $x_{max} = 100$, but this regular pattern diminishes with $x = 150$. This empirical pattern implies that the strength of congestion

[30] The point estimates from the first and second steps of the GMM are very close. Moreover, we are also working on the iterative GMM estimation (Hansen and Lee 2021) and the result will be available soon.

externality has spatial decay. Moreover, this observation justifies our choice of instruments in the main non-linear estimation, exploiting the variation from surrounding populations up to $x = 100$.

There is a concern that deforestation also negatively affects biological resource endowments (Barlow et al. 2016; Giam 2017). That is, there may also exist the across-sector externality—the effect of clearing forests for agriculture on the productivity of the natural resource extraction sector.

However, in the context of our study, we are focusing on the first-order direct effects of the population engaging in the natural resource sector on natural resource depletion and not incorporating the across-sector externality. We have three comments on this decision. First, the main reason for this decision is that the spatial extent of these sectors' activities is distinct in our study area. Deforestation for agricultural land is mostly distributed along the rivers. The mean, median, and maximum land footprint depths among the rural communities in our study area are about 1 km, 0.85 km, and 5.5 km, respectively. The land footprint depth represents the distance from the river to the furthest inland point in the community-level land footprint within a voronoi polygon around the settlement detected in satellite images (Figure A.11). In contrast, the spatial extent of some natural resource extractions (hunting wild animals and collecting forest products) is much broader, including deep inland areas away from the river where forests are not cleared at all. In addition, fishing is obviously conducted on the river without forests. Second, natural resource endowments are not significantly correlated with the community-level land footprint. Panel (B) of Table A.1 reports the correlation between the community-level land footprint and the number of species found around the community (reported in the CC data). Column (1) reports that the number of total species pooling all types of extractive activities is not significantly correlated with the land footprint. Columns (3)-(5) imply that the land footprint is not significantly correlated with resource endowments for any inland extractive activities (timber; NTFP; game). The significant negative correlation between the land footprint and the number of fish species (column (2)) may be reflecting that higher fish endowments attracts more workers for fishing relative to agriculture, which results in the smaller deforestation for agriculture, given that deforestation would not have direct effects on the fish stock. Panel (C) of Table A.1 reports that the correlation between the land footprint depth and natural resource endowments is also insignificant (except for timber). Third, although the data support our model choice, it is not possible to formally prove the absence of the across-sector externality. This model choice does not affect the inversion problem, but affects outcomes in counterfactual policy simulations. We investigate policies that reduce total deforestation in a river basin in a later section. Therefore, we can interpret these policies' welfare effects as lower bounds because the reduction in deforestation may generate additional gains of natural resource endowments that our model is not accounting for.

5.4 Quantifying the Effects of Density Externalities

We investigate the quantitative importance of the estimated density externalities in determining human welfare and environmental costs. The primary outcome variables at each basin level are welfare of the population, total deforestation, and natural resource depletion. The welfare measures is based on the real wage that is equalized across locations in the spatial equilibrium.[31] We define natural resource depletion in each basin as the total expected amount natural resources extracted in the spatial equilibrium, denoted by Q (Nr). We also report the counterfactual outcomes of total sectoral employment in the basin, denoted by N (Ag), N (Nr), and N (Urban).

First, we focus on the agglomeration externality in agriculture. Given the total population in each basin and the calibrated productivity fundamentals, we solve the model by shutting down the agglomeration externality ($\mu_{Ag} = 0$). Note that we maintain the estimated congestion externalities in forest clearing and natural resource extractions.

Panel (A) of Table 4 indicates that the agglomeration externality has large effects. Without the agglomeration externality, deforestation increases by about 18–56%, but total agricultural production decreases by about 29–35% from the benchmark equilibrium. The human welfare decreases by about 7–13% without the agglomeration. However, the natural resource depletion also decreases by about 1–3% without the agglomeration externality. That is, another trade-off between two types of environmental costs remains—deforestation versus biological natural resource depletion.

Figure 9 and Figure A.12 show spatial distributions of relative values of key outcome variables caused by this counterfactual experiment (in terms of those in the benchmark spatial equilibrium) in the Upper Ucayali and Napo-Amazon river basins.[32] As the upper middle maps indicate, agricultural productivity decreases in all locations with the lack of agglomeration force, since the productivity gains from concentration decline. According to the upper left maps, there is a significant spatial reallocation of the agricultural population: it has increased without the agglomeration externality in more than 80% of rural locations. This spatial reallocation is from concentration into dispersion because the congestion externality in forest clearing dominates. A sectoral reallocation of workers also occurs: the overall agricultural employment increases (by 6.6% and 2.9%) and the overall employment for natural resource decreases (by 1.8% and 4.4%). This sectoral reallocation reflects general equilibrium effects. Since the agricultural productivity has declined, the economy needs more agricultural employment to satisfy

[31]The real wage is equivalent to the indirect utility based on the specified utility function. Note also that the value of indirect utility varies with monotone transformations of the utility function (that result in the same observed endogenous variables). Therefore, while the real wage is uniquely determined, it is possible to construct other welfare measures. Nevertheless, the sign of the welfare effect and the order of its magnitude in different counterfactuals are uniquely identified, although the magnitude of the welfare value varies depending on how the utility function is defined.

[32]To save space, we present maps from the two basins (Upper Ucayali and Napo-Amazon, which have the two largest urban centers) out of the four basins throughout the counterfactual experiments. Maps from the other two basins are available upon request.

the demand for agricultural goods by the population, given that consumption demands across the agricultural and natural resource goods are complementary.

Next, we focus on the congestion externality in natural resource extraction. Unlike the agglomeration externality, shutting down or reducing the congestion externality is impractical, given the finite nature of natural resource endowments. Therefore, we instead solve the model by increasing the congestion externality in natural resource extractions. In particular, we increase the congestion externality parameter (μ_{Nr}) by 25% from its estimated value. Note that we maintain the estimated agglomeration externality in agriculture and congestion externality in forest clearing as well as the estimated spatial decay parameter in natural resource extractions.

Panel (B) of Table 4 indicates that the increased congestion externality also has large effects. With the increased congestion externality, total natural resource depletion decreases by about 40–50% and human welfare decreases by about 11–30%. A sectoral reallocation of workers also occurs: the overall agricultural employment decreases (by about 2–9%), the overall employment for natural resource increases (by about 9–16%), and the overall employment in the urban center decreases (by about 4–7%). This sectoral reallocation reflects general equilibrium effects and is explained in the same way as in the case of shutting down the agglomeration externality above.

6 Mechanisms Underlying the Agglomeration Externality

We investigate mechanisms behind the density externalities that we found in the previous section. Interpretation of the congestion externality in forest clearing or natural resource extractions is straightforward given its rivalrous nature. However, interpreting the agglomeration externality in agriculture is not straightforward, and thus we examine its microfoundation. This is also new to the literature because agglomeration has typically been discussed in urban settings.

6.1 Economies of Scale in Transport and Transaction Costs

One hypothesis is that the cost of transporting products from a community decreases with increases in the community's production and export of the products. This mechanism could be the case if large-scale commercial river boats (e.g., "lancha," shown in the bottom-right picture of Figure A.10) are more likely to stop by communities that export large amounts of their products. The mechanism could also be the case if populations in larger communities cooperate to invest in fast motor boats. Another simple possibility could be that the average transport cost charged by an intermediary agent decreases with the amount of transported products.[33] In any of the ways, the actual transport costs could be endogenous depending on the exporting community's population. However, the model takes river transport costs as

[33] Foster and Gafaro (2017) also report a consistent finding that the economies of scale in transport technology allow small-scale farmers to shift from subsistence to market participation.

given such that they depend only on distance and river orientations based on the most widely available transport mode (peque-peque, shown in the top-right picture of Figure A.10).

Nevertheless, the present model with agglomeration externality in agricultural productivity is indeed isomorphic to a model with endogenous transport costs as long as the endogeneity stems from the population of an exporting community; that is, the transport cost depends on an origin-specific endogenous factor. To see this point, consider the following model. Suppose that there is no agglomeration externality in agricultural productivity such that we replace (4) with the following agricultural production function:

$$Q_{o,Ag}(j) = z_{o,Ag}(j) \cdot N_{o,C}(j)^{\gamma} L_o(j)^{(1-\gamma)}, \quad o \in \mathcal{R}$$

while keeping the same land access function as (3). Suppose also that the trade cost of agricultural products is characterized as:

$$\tilde{\tau}_{od,Ag} = N_{o,Ag}^{-\mu_{Ag}} \tau_{od,Ag} \tag{21}$$

where $\tau_{od,Ag}$ is the exogenous term of trade cost defined by (8) and $\mu_{Ag} > 0$ such that the trade cost decreases with the agricultural employment. These assumptions lead to the same expenditure share across locations within the agricultural sector as (9).[34] This model thus leads to the same set of spatial equilibrium conditions. Therefore, the inverted productivity composites could contain the endogenous term of trade costs.

The CC data allow us to test this hypothesis. Panel (A) of Table 5 reports the results of estimating the scale effect on the availability of different transport modes in a community. First of all, according to column (8), the mean value of peque-peque availability is close to one (0.972) and much higher than the availability of the other three transport modes. This observation justifies our decision to calculate the asymmetric trade cost based on peque-peque, the most widely available mode. Next, according to columns (2), a 1% increase in the agricultural employment increases the probability of lancha being available in the community by 0.14% at the 1% level of statistical significance. As the primary purpose of lancha is to carry products, this result is consistent with our hypothesis.

Panel (B) of Table 5 shows the scale effect on the frequency of these transport modes available. According to column (2), the mean number of days per week when lancha is available is 3.429. Lancha is a large commercial boat and is not generally owned by each small-scale community. We thus interpret this value in a way that lancha passes a community about half a week on average. This column reports that a higher agricultural employment significantly

[34] To see this isomorphism,

$$\pi_{od,Ag} = \frac{A_{o,Ag} N_{o,Ag}^{-(1-\gamma)\mu_{Ag}\theta} \kappa_1^{\theta} (w_o \tilde{\tau}_{od,Ag})^{-\theta}}{\sum_{o' \in \mathcal{R}} A_{o',Ag} N_{o',Ag}^{-(1-\gamma)\mu_{Ag}\theta} \kappa_1^{\theta} (w_{o'} \tilde{\tau}_{o'd,Ag})^{-\theta}} = \frac{\tilde{A}_{o,Ag} (w_o \tau_{od,Ag})^{-\theta}}{\sum_{o' \in \mathcal{R}} \tilde{A}_{o',Ag} (w_{o'} \tau_{o'd,Ag})^{-\theta}}, \quad o \in \mathcal{R},\, d \in \mathcal{I}$$

where $\tilde{A}_{o,Ag}$ is same as (10).

increases the number of days when lancha is available to the community. This result is also consistent with our hypothesis of the scale economy in transport costs.

A related mechanism is that other transaction costs decrease as population size increases, which is also consistently explained by the specification of (21). According to column (6) of panel (A) of the same table, the probability of "rapido" (a commercial speedboat shown in the bottom-left picture of Figure A.10) being available in the community also increases by 0.06% at the 10% level of statistical significance. In contrast to lancha, the primary purpose of rapido is to move people around. Therefore, this result does not seem to be closely relevant to the scale economy in transporting products. This result is instead consistent with the scale economy in transaction costs through various possible mechanisms. For example, in communities accessible by such a speedboat, contracts with intermediaries of commodity trading may be facilitated.

Table A.8 adds a more direct argument. According to this table, probabilities that a river trader is available to the community, the community population is contracted to selling a product, or there are contractors living in the community are significantly higher with a higher community population.[35] Although we do not directly observe activities of these intermediaries, a straightforward interpretation is that their presence decreases transaction costs for selling community's products.[36]

6.2 Economies of Scale in Agricultural Intensification

Another hypothesis is that there are economies of scale in accessing agricultural inputs. The 2012 CENAGRO allows us to test this hypothesis. Table 6 reports the scale effect on the household-level use of various agricultural infrastructure, technology, and inputs into land and crops in the communities in our study area.

Modern technology is limited in the Peruvian Amazon. We pick all the related variables available in CENAGRO that have non-missing information. All the variables are dummies. The mean value is less than 0.1 for all the variables (i.e., less than 10% of agricultural producers use them) with one exception for boat/canoe/speedboat having about 0.6. Moreover, the mean value is less than 0.01 for 13 items (out of 24).

In this limited environment of available agricultural inputs, we nevertheless observe the scale effect on some technologies and inputs. For example, the effects on some direct inputs into land and crops (fertilizers, insecticides, herbicides, and fungicides) as well as on their complementary equipment (sprayers) have higher point estimates than others at the 1% or 5% level of statistical significance. We also see a significant scale effect on the use of crop processing technology (grain mill). Since grain mills facilitate processing raw grains into marketable products, this result is broadly consistent with the previous finding of the economies of scale in improving trade environment.

[35]Table A.9 reports product-specific contracts.

[36]Investigating the incidence between traders and community population (e.g., Atkin and Donaldson 2015; Bergquist and Dinerstein 2020) is not feasible with our data. Future work is warranted in this research area.

The CC data from PARLAP collect detailed information on the form of seed acquisition. The people in the community are asked whether they obtain their seeds or other planting materials from others in the community, from people in other communities, or from a city. They are also asked about the principle form of acquisition if they obtain seeds from other places. Table A.10 reports the results. The point estimates reported in panel (A) imply that as the community population increases, people in the community are more likely to obtain crop seeds from others in the community and less likely to obtain them from a city or from other communities,[37] although the statistically significance is not strong. According to panel (B) of the same table, the form of transaction (market or non-market transaction) when people obtain seeds from outside the community does not significantly vary in the community population.[38] To summarize, we find suggestive evidence of the economies of scale in accessing agricultural inputs. Given the very low rate of adopting modernized agricultural technologies and inputs, however, the economies of scale in trading products is revealed to be the primary mechanism underlying agglomeration in the quantitative sense.

7 Counterfactual Experiments

Counterfactual experiments investigate environmental protection policies as well as improving infrastructure and technology to design the "win-win policy" that increases both human and ecological well-being. The set of primary outcome variables at each basin level is same as in section 5.4. We define a win-win policy as the policy that achieves the following three outcomes: human welfare is increased, total deforestation is decreased, and total natural resource depletion is decreased.

Note that a counterfactual experiment requires solving the model and obtaining population distributions in a new spatial equilibrium. In the presence of externalities, it is not feasible to prove the uniqueness of the equilibrium. To address this issue, throughout the counterfactual experiments we use the same simulation algorithm, described in Appendix C.3, with the aim of reaching the spatial equilibrium closest to the benchmark equilibrium given the set of populated locations. We have also tried to solve the model with various combinations of initial guesses to search for multiple equilibria, but have not found any equilibria other than those presented in the following results.

7.1 Protection Policies

In this section, given the fixed number of total population and estimated density externalities, we ask the following questions: is it beneficial to concentrate the ecological footprint in fewer spots rather than to have many small communities? If so, how do different ways of setting

[37] Table A.11 reports crop-specific results and that this effect largely stems from maize and plantain.

[38] Market transactions include purchasing and borrowing. Non-market transactions include exchanging, donating, and giving.

protected areas and encouraging resettlement from the protected locations affect welfare and outcomes? To approach these questions, we implement the following protection policy experiments. We simulate the model to derive endogenous outcomes given the additional condition imposed by each experiment. Note that the model does not incorporate the cost of associated resettlements (or a subsidy to compensate for them) in response to policy interventions. Therefore, implementing a cost-benefit analysis of a single policy intervention is beyond the scope of this paper. To interpret and compare policy outcomes meaningfully, we thus conduct multiple policy experiments, each directly treating an equal size of population. We design each policy such that it directly treats 2.5% of rural populations in the benchmark equilibrium in each basin for resettlement.

(A) Protecting the rural frontier

We first present another fact about the spatial distribution of community locations.

Fact 1B: The formation of new rural communities occurs not only by the rural frontier expanding but also by the interior of the rural frontier filling in

We define the rural frontier in each river basin as the rural settlement that takes the maximum distance from the urban center (represented by a square in each map of Figure 3). Beyond the rural frontier, the land is covered by forests and no human settlements are assumed. Figure A.1 and Figure A.2 show the establishment of communities in the Napo-Amazon and Upper Ucayali basins over decades. These maps illustrate that the rural frontier expanded during most of the two-decade intervals. This process is consistent with the explanation provided by the classical model of dynamic city formation (Fujita and Krugman 1995; Fujita et al. 1999), which argues that the agricultural frontier expands as the population grows. In contrast to the assumption in the classical model that areas inside the agricultural frontier are filled with farmers and farmlands, we also observe the formation of new communities as the interior of the rural frontier fills in during most periods. Since our model is static, modeling this complex dynamic process of rural community formation is left for future research. Nevertheless, this fact motivates the question of what happens to the spatial distribution of economic activity and environmental costs when some protection policy controls the expansion of the rural frontier.

We consider a place-based protection policy of controlling the expansion of rural frontier. This simulation chooses rural locations to be treated in order, starting with those farthest from the urban center, until the treated population reaches 2.5% of the total rural population in each basin.

Panel (A) of Table 7 reports the results. In the Upper Ucayali basin, total deforestation has decreased by 2% with a very small decrease in welfare (0.2%). Similar results follow in the other three basins. Intuition behind these results is as follows. In the presence of congestion externality in clearing forest, total deforestation decreases by reducing the number of populated locations while fixing the total population size. The capacity of agricultural production is affected, depending both on the decreased amount of land inputs and the increased productivity gain due to the agglomeration externality in agricultural production. The minimal welfare effect implies that these two forces almost cancel each other out.

This policy has also reduced natural resource depletion (in three of the four basins). In the Upper Ucayali basin, total natural resource depletion has decreased by 0.8%. Intuition is as follows. With the protected areas targeting the rural frontier, the overall scope of natural resource extraction activities is narrowed. This policy increased the overall population density in the basin as it relocated the population closer to the urban center. This reallocation means that the surrounding populations increase in most of the populated areas. Therefore, the overall productivity of the natural resource sector would be negatively affected because of the congestion externality with spatial spillovers. As a result, the total outputs of the natural resource products decrease (i.e., natural resource depletion is reduced), which is an ecological benefit.

Figure 10 and Figure A.13 show spatial distributions of relative values of key outcome variables caused by this counterfactual experiment (in terms of those in the benchmark spatial equilibrium) where red **x** marks indicate the treated locations whose residents resettled to other locations. We will discuss these maps later when we compare the counterfactual outcomes between different protection policies.

(B) Not allowing new community formation

We next implement a simple experiment of not allowing new community formation. In this experiment, we simply evaluate economic and environmental outcomes with the same total population but without new community formation. This experiment chooses rural locations to be treated in reverse order of establishment, starting from the most recently established communities, until the treated population reaches 2.5% of the total population in each basin. We then solve the model with the rest of the communities established earlier given the total population.

Panel (B) of Table 7 reports the results. In the Upper Ucayali basin, total deforestation decreases by 3% with a small decrease in welfare (0.1%). Similar results follow in the other three basins. The welfare outcome is almost the same as in the previous experiment. This experiment has also reduced deforestation with a similar mechanism to the previous experiment. Compared with the previous experiment that protects the rural frontier, however, the forest cover gain is larger. After presenting the next experiment, this difference in the forest cover gain will be discussed.

(C) Minimum population threshold to form a community

We also consider a place-based experiment of specifically targeting small communities. This experiment chooses rural locations to be treated in order, starting with the location with the smallest population size, until the treated population reaches 2.5% of the total population in each basin. We then solve the model with the rest of the locations, keeping the total population in each basin fixed.

Panel (C) of Table 7 reports the results. The welfare outcome is almost the same as in the previous experiment. However, targeting residents of the smallest communities for resettlement leads to a higher gain in preserving forest cover. In all four basins, this policy has the largest negative deforestation impact among the three experiments. This policy has reduced deforestation by 7.3% in the Upper Ucayali basin and 13.1% in the Napo basin. The right five columns in (B) and (C) indicate that the outcomes of sectoral reallocation of the population and sectoral outputs are almost the same as in the previous experiment (except for the Lower Ucayali basin). Therefore, the forest cover benefit relative to the previous experiment arises without sacrificing other objectives. This benefit primarily stems from the agglomeration and congestion externalities. This experiment reduced communities that had both low congestion in accessing land and low agglomeration benefit for agricultural production and caused the population to enjoy more agglomeration externality after resettlement. The resulting increase in productivity due to the agglomeration externality compensated for the increased congestion for land access. Figure 11 and Figure A.14 show spatial distributions of relative values of key outcome variables caused by this counterfactual experiment (in terms of those in the benchmark spatial equilibrium) where red x marks indicate the treated locations whose residents resettled to other locations.

A comparison between policies (A) and (C) illustrates that policymakers face a trade-off in mitigating different types of environmental costs: deforestation and depletion of other natural resources. The protection policy that targets the smallest communities (C) reduces deforestation more than the frontier protection (A), but increases natural resource depletion. A comparison between the bottom-middle maps of Figure 10 and Figure 11 indicates the primary source of these contrasted outcomes from the two experiments (A) and (C). These two maps report that the productivity of the natural resource sector (relative to its original values) becomes smaller than 0.96 for more than 80% of the remaining locations by protecting the frontier (A), while it becomes larger than 0.97 for all of the remaining locations by this experiment (C).

We of course understand the feasibility and ethical concerns of resettling populations in the smallest communities that already exist. This experiment nevertheless generates an important policy implication. This experiment can indirectly suggest the importance of regulating new community formations to minimize deforestation. For example, one policy idea might be to regulate the formation of new communities with a threshold of minimum settlement size. Indeed, the deforestation impact in the previous experiment (B) of not allowing new community formation is smaller than in this experiment (C) but larger than in the experiment (A) of protecting the frontier (in three of the four basins). This fact reflects that newly formed communities includes more small-scale communities than those locating near the frontier.

7.2 Improvement of River Transport Infrastructure

We consider improving the transport infrastructure that aims to reduce high trade costs in the environment of Amazon river networks. The high trade costs primarily stem from the asymmetry of transport costs due to river orientations, the seasonality of transport costs due to water level fluctuations, and the slow speed of river boats. Recall (22) where we defined the downstream-river-equivalent kilometers along the lowest-cost route. With the improved transport infrastructure, the downstream-river-equivalent distance is expressed as follows:

$$D_{od} = D_{od,initial,down} + \lambda_{up} D_{od,initial,up} + \lambda_{land} D_{od,land} + \lambda_{upgraded} D_{od,upgraded} \tag{22}$$

where $D_{od,upgraded}$ corresponds to the upgraded part of the river transport network. $D_{od,initial,down}$ and $D_{od,initial,up}$ correspond to the portions of the river transport network that are not upgraded.

Specifically, we consider the quality improvement of boats and river dredging as the improvement of river infrastructure. In the model, the infrastructure improvement has two elements. The first element is that there is no asymmetric cost in the upgraded part. The second element is that the travel cost of the upgraded part decreases. To capture this effect, we set $\lambda_{upgraded} = 0.8$. These two elements are indeed consistent with the trajectory of transportation infrastructure development in the Peruvian Amazon over time.[39] River dredging is a complementary devise for reducing transport costs as it enables larger ships to travel.[40]

Table 8 reports the results of a variety of experiments. Reducing the trade costs increases human welfare in all the cases; however, the effects on environmental costs are heterogeneous across different ways of improving the transport infrastructure.

Infrastructure investments that make the areas enjoying agglomeration externalities more even across the basin are preferable in terms of reducing deforestation

We first focus on the agricultural sector and the deforestation impact. Figure 12 shows locations of transport infrastructure improvement along river lines according to river orders. Higher river orders mean more splits from the central river.

Panel (A) of Table 8 reports the results of improving the transport infrastructure in a way that hinterlands are connected to the central area of a basin (by targeting river lines with river

[39] For example, canoes, row and sail boats, and steamers were the primary means of river transportation during 1860-1940. Using historical records on travel time (e.g., Herndon and Gibbon 1853; Paz Soldan 1877; Schurz et al. 1925) and the same approach described in Appendix C.2, we estimate that the value of the upstream cost parameter (λ_{up}) is 2.37 in this historical period. This value is close to double the estimated value $\hat{\lambda}_{up} = 1.282$ for the modern period. That is, the cost of the asymmetry of river orientations is decreasing over time. Also evident is the increase in speed over time. Meanwhile, the exact speed increase of a next generation boat is ambiguous. Therefore, $\lambda_{upgraded}$ is set arbitrarily. Smaller values of this parameter amplify the reported results of the counterfactuals.

[40] The Amazon Waterway is an ongoing project (but started after our data collection) that dredges rivers. See, for example, Abizaid et al. (2022) and Lu (2019) for more details. We are currently working on incorporating a potential negative externality of dredging rivers on fish stock, which affects the productivity of the natural resource sector.

order 2). In all basins except Lower Ucayali, total deforestation is decreased by this experiment. Figure 13 and Figure A.15 show spatial distributions of the key outcome variables caused by this counterfactual experiment. As hinterlands become more integrated in the trade network, we observe spatial reallocation of farmers toward remote areas compared with the benchmark equilibrium (the top-left maps). Farmers in the remote areas enjoy more agglomeration benefits. The top-middle maps show agricultural productivity gains in remote areas. Given the congestion externality in forest clearing, deforestation *per farmer* decreases in remote areas. The reduction of total deforestation in the basin means that this forest gain in remote areas outweighs the forest loss in denser areas.

Panel (B) of Table 8 reports the results of improving the transport infrastructure only in densely populated areas (by targeting river lines with river order 1 in the case of Napo basin). In contrast to the previous case, total deforestation increases. Figure 14 shows spatial distributions of the key outcome variables caused by this counterfactual experiment. The agricultural population is more concentrated in the central area of the basin (the top-left map). Due to the density externalities, agricultural productivity increases and deforestation per farmer decreases in the central areas. On the flip side, this policy also generates much smaller communities with much higher deforestation per farmer in hinterlands. The increase in total deforestation in the basin means that the latter effect dominates. From these two experiments, we learn that the *direction* of deforestation impact depends on where in the spatial structure of river networks the improvement takes place.

Additional experiments help to elucidate the importance of place-based targeting of infrastructure investments to mitigate environmental costs. Panel (C) of Table 8 reports the results of improving the transport infrastructure in a way that combines both of the previous two experiments (i.e., targeting river lines with both river orders 1 and 2 in the Napo basin). The welfare increase is higher than the previous two experiments simply because this experiment treats more river lines. The deforestation impact is between those of the previous experiments, but much closer to the second experiment targeting dense areas. Panel (D) of Table 8 reports the results of just reducing the asymmetry in all river lines. Total deforestation also increases by uniformly improving the transport infrastructure across the basin.

To summarize, policy interventions that make the areas enjoying agglomeration externalities more even across the basin are preferable in terms of reducing deforestation. In other words, policies causing more even distribution of middle-sized communities across the basin are more preferable than those causing a separation in the basin between highly concentrated locations with agricultural intensification and very small communities in hinterlands with low productivity agriculture. This implication is also consistent with the finding in the previous section that the protection policy targeting the smallest communities has reduced deforestation most.

A sector-specific protection policy leads to an unintended consequence

We next focus on the natural resource extraction. Let us first focus on the experiment of targeting river lines with river order 2 (panel (A) of Table 8; Figure 13; Figure A.15). As hinterlands become more integrated in the trade network, we observe spatial reallocation of this activity toward remote areas from the benchmark equilibrium (the bottom-left maps). The degree of this spatial reallocation is lower than that of agricultural population according to the comparison between the two maps on the left. This difference is due to the different extent of density externalities between agriculture and natural resource extraction. Owing to the congestion externality with spatial spillovers, productivity of the natural resource sector increases in remote areas and decreases in the central dense areas (bottom-middle maps). Although these two opposing forces exist, the net depletion of natural resources increases in remote areas connected by the improved transport infrastructure. The impact on total natural depletion in the basin is ambiguous. It increases by 0.3% in Upper Ucayali, but it decreases in the other three basins.

To combat the natural resource depletion, a sector-specific protection policy such as banning hunting in specific location might be considered. However, the feasibility of such a policy is unrealistic and even more difficult than resettlement policies. It would be more difficult to monitor people's activities than whether areas are inhabited. Therefore, as a more feasible policy, we consider checkposts to monitor traded goods in the improved transport infrastructure. That is, we consider a protection policy of using the improved transport infrastructure only for transporting agricultural and urban goods, but not for transporting natural resource products. The trade cost of natural resource products is thus unchanged in this experiment.

Panel (E) of Table 8 reports the results of this experiment. In contrast to the policy objective of protecting natural resources, it leads to an unintended consequence of increasing extractive activities. In all four basins, the resulting natural resource depletion is higher in this policy (panel E) than in the policy without the sector-specific checkposts (panel A). This unintended consequence is primary driven by the sectoral reallocation of populations through general equilibrium effects. The overall relative productivity in the natural resource sector (in terms of the amount of output that can be delivered to destinations through trade) is diminished because of the sector-specific restriction in trade. The economy thus needs more employment in this sector to meet the consumption demand by the population, which increases the total depletion of natural resources.

7.3 Toward External Validity and the Win-Win Policy

Conceptually, the set of counterfactual experiments derives an implication toward external validity. Counterfactual exercises based on a class of general equilibrium models often derive a conclusion from a single economy (e.g., from one country/region based on economic geography models or from one city based on a quantitative urban model that studies internal city structure). In the context of our study, this practice corresponds to analyzing counterfactuals only from one of the four basins. However, the conclusions that can be drawn from this practice may not be as conclusive as one might expect. Panel (A) of Table 8 illustrates this point. One might conclude from the Napo basin (or Pastaza and Lower Ucayali) that the improvement of transport infrastructure targeting river order 2 is a win-win policy because it increases welfare and reduces deforestation and natural resource depletion. In contrast, the natural resource depletion is increased in the Upper Ucayali basin even with the same model, the same parameter values, and a similar study area (compared with other rainforest areas outside the Peruvian Amazon). This example illustrates the caveat of deriving conclusions from a single economy.

Therefore, this study states conclusions if and only if primary outcomes from all four basins are qualitatively same. Although it is still not perfect to argue the external validity, this way would improve the credibility of policy implications compared with those derived from a single economy. To achieve the win-win outcome, we consider a composite intervention that combines the protection policy with the improvement of transport infrastructure. Table 9 reports results of the composite counterfactual experiment that contains the following two components. The first component sets protected areas that control the expansion of rural frontier by directly targeting 2.5% of the total rural population in each basin. The second component improves the transport infrastructure in a way that it connects hinterlands to the central area of the basin (by targeting river lines with river order 2). Each component is exactly the same as in a previous experiment. According to the table, this policy is win-win because it increases welfare by about 1–2.1%, decreases deforestation by about 1–7%, and decreases natural resource depletion by about 0.5–2.4%. These outcomes imply that, while limiting the expansion of rural frontier, connecting the given rural frontier with a central area by eco-friendly transport infrastructure is preferable to simultaneously achieve multiple objectives that are seemingly trade-offs.

8 Conclusion

This paper investigates the concentration and dispersion forces of economic activity to balance human and ecological well-being in tropical forests. We formalize these forces in a multi-sector spatial model, which we estimate with detailed georeferenced data from major river basins in the Peruvian Amazon and plausibly exogenous variation in Amazon river networks. Counterfactual simulations demonstrate that, in the presence of density externalities, combining transport infrastructure investments and protection policies can simultaneously achieve higher welfare, lower deforestation, and less natural resource depletion. This paper also emphasizes that such a win-win outcome of welfare and ecosystem conservation can be achieved subject to careful spatial targeting of both types of policy interventions.

Several limitations remain. First, the counterfactual spatial distribution of human settlements is studied in a static sense and determined from the set of populated locations in the benchmark equilibrium. Investigating the dynamic process and consequences of community formation, resource depletion, policy responses, and the entry of external investments is an important agenda. Second, we are not directly quantifying a collective value for the preservation of a traditional way of life. In other words, we are not directly capturing the utility cost of spatial reallocation of the population caused by the counterfactual policy interventions, although some policies treat the equal number of populations for a meaningful comparison between outcomes. Moreover, it may be important in studying traditional societies to incorporate local populations' values of their living with others of common ancestry, which could also potentially incite certain political movements such as indigenous rights protections. Incorporating these factors is beyond the scope of the current model. We nevertheless believe that this paper has provided a benchmark framework upon which future research can be based to address these agendas.

References

Abizaid, Christian, Oliver T Coomes, and Yoshito Takasaki, "Lifeways and currents of change in the Peruvian Amazon: A 1000 km boat journey down the Ucayali River," *Focus on Geography*, 2022, *65.*

Abman, Ryan and Conor Carney, "Agricultural productivity and deforestation: Evidence from input subsidies and ethnic favoritism in Malawi," *Journal of Environmental Economics and Management*, 2020, *103*, 102342.

_ , Teevrat Garg, Yao Pan, and Saurabh Singhal, "Agriculture and deforestation," *Available at SSRN 3692682*, 2020.

Ahlfeldt, Gabriel M and Elisabetta Pietrostefani, "The economic effects of density: A synthesis," *Journal of Urban Economics*, 2019, *111*, 93–107.

_ , Stephen J Redding, Daniel M Sturm, and Nikolaus Wolf, "The economics of density: Evidence from the Berlin Wall," *Econometrica*, 2015, *83* (6), 2127–2189.

Alix-Garcia, Jennifer, Craig McIntosh, Katharine RE Sims, and Jarrod R Welch, "The ecological footprint of poverty alleviation: evidence from Mexico's Oportunidades program," *Review of Economics and Statistics*, 2013, *95* (2), 417–435.

Alix-Garcia, Jennifer M, Katharine RE Sims, and Patricia Yañez-Pagans, "Only one tree from each seed? Environmental effectiveness and poverty alleviation in Mexico's Payments for Ecosystem Services Program," *American Economic Journal: Economic Policy*, 2015, *7* (4), 1–40.

Allen, Treb, Costas Arkolakis, and Yuta Takahashi, "Universal gravity," *Journal of Political Economy*, 2020, *128* (2), 393–433.

Angelsen, Arild, "Agricultural expansion and deforestation: modelling the impact of population, market forces and property rights," *Journal of development economics*, 1999, *58* (1), 185–218.

_ , "Policies for reduced deforestation and their impact on agricultural production," *Proceedings of the national Academy of Sciences*, 2010, *107* (46), 19639–19644.

_ , Pamela Jagger, Ronnie Babigumira, Brian Belcher, Nicholas J Hogarth, Simone Bauch, Jan Börner, Carsten Smith-Hall, and Sven Wunder, "Environmental income and rural livelihoods: a global-comparative analysis," *World development*, 2014, *64*, S12–S28.

Araujo, Rafael, Francisco Costa, and Marcelo Sant'Anna, "Efficient Forestation in the Brazilian Amazon: Evidence from a Dynamic Model," 2020.

Asher, Sam, Teevrat Garg, and Paul Novosad, "The ecological impact of transportation infrastructure," *The Economic Journal*, 2020, *130* (629), 1173–1199.

Asner, Gregory Paul, David E Knapp, Aravindh Balaji, and Guayana Páez-Acosta, "Automated mapping of tropical deforestation and forest degradation: CLASlite," *Journal of Applied Remote Sensing*, 2009, *3* (1), 033543.

Assunçao, Juliano, Robert McMillan, Joshua Murphy, and Eduardo Souza-Rodrigues, "Optimal Environmental Targeting in the Amazon Rainforest," 2022.

Atkin, David and Dave Donaldson, "Who's Getting Globalized? The Size and Implications of Intra-national Trade Costs," Technical Report, National Bureau of Economic Research 2015.

Balboni, Clare, Aaron Berman, Robin Burgess, and Benjamin A Olken, "The Economics of Tropical Deforestation," 2022.

Balboni, Clare Alexandra, "In harm's way? infrastructure investments and the persistence of coastal cities." PhD dissertation, The London School of Economics and Political Science (LSE) 2019.

Barham, Bradford L, Oliver T Coomes et al., *Prosperity's promise: the Amazon rubber boom and distorted economic development.*, Westview Press, Inc., 1996.

Barlow, Jos, Gareth D Lennox, Joice Ferreira, Erika Berenguer, Alexander C Lees, Ralph Mac Nally, James R Thomson, Silvio Frosini de Barros Ferraz, Julio Louzada, Victor Hugo Fonseca Oliveira et al., "Anthropogenic disturbance in tropical forests can double biodiversity loss from deforestation," *Nature*, 2016, *535* (7610), 144–147.

Bergquist, Lauren, Benjamin Faber, Thibault Fally, Matthias Hoelzlein, Edward Miguel, and Andres Rodriguez-Clare, "Scaling Agricultural Policy Interventions: Theory and Evidence from Uganda," 2022.

Bergquist, Lauren Falcao and Michael Dinerstein, "Competition and entry in agricultural markets: Experimental evidence from Kenya," *American Economic Review*, 2020, *110* (12), 3705–47.

Boserup, Ester, *The conditions of agricultural growth: The economics of agrarian change under population pressure*, Routledge, 1965.

Brancaccio, Giulia, Myrto Kalouptsidi, and Theodore Papageorgiou, "Geography, transportation, and endogenous trade costs," *Econometrica*, 2020, *88* (2), 657–691.

Cadieux, Nicolas, Margaret Kalacska, Oliver T Coomes, Mari Tanaka, and Yoshito Takasaki, "A Python Algorithm for Shortest-Path River Network Distance Calculations Considering River Flow Direction," *Data*, 2020, *5* (1), 8.

Carreira, Igor, Francisco Costa, and Joao Paulo Pessoa, "The Deforestation Effects of Trade and Agricultural Productivity in Brazil," 2022.

Caunedo, Julieta, Namrata Kala, , and Haimeng Zhang, "Economies of Density and Congestion in Capital Rental Markets," 2020.

Chen, Yuxian, Yannis M Ioannides, and Ferdinand Rauch, "Asymmetric Trading Costs and Ancient Greek Cities," 2022.

Conte, Bruno, "Climate change and migration: the case of africa," *Unpublished manuscript*, 2022.

Coomes, Oliver T, "A century of rain forest use in western Amazonia: lessons for extraction-based conservation of tropical forest resources," *Forest and Conservation History*, 1995, *39* (3), 108–120.

_, Margaret Kalacska, Yoshito Takasaki, Christian Abizaid, and Tristan Grupp, "Smallholder agriculture results in stable forest cover in riverine Amazonia," *Environmental Research Letters*, 2021.

_, Yoshito Takasaki, Christian Abizaid, and J Pablo Arroyo-Mora, "Environmental and market determinants of economic orientation among rain forest communities: Evidence from a large-scale survey in western Amazonia," *Ecological Economics*, 2016, *129*, 260–271.

Costinot, Arnaud, Dave Donaldson, and Cory Smith, "Evolving comparative advantage and the impact of climate change in agricultural markets: Evidence from 1.7 million fields around the world," *Journal of Political Economy*, 2016, *124* (1), 205–248.

Cruz, José Luis and Esteban Rossi-Hansberg, "The economic geography of global warming," *University of Chicago, Becker Friedman Institute for Economics Working Paper*, 2021, (2021-130).

Dasgupta, Partha and Karl-Göran Mäler, "Poverty, institutions, and the environmental resource-base," *Handbook of development economics*, 1995, *3*, 2371–2463.

Donaldson, Dave, "Railroads of the Raj: Estimating the impact of transportation infrastructure," *American Economic Review*, 2018, *108* (4-5), 899–934.

Duranton, Gilles and Diego Puga, "Micro-foundations of urban agglomeration economies," in "Handbook of regional and urban economics," Vol. 4, Elsevier, 2004, pp. 2063–2117.

_ and _ , "The economics of urban density," *Journal of economic perspectives*, 2020, *34* (3), 3–26.

Eaton, Jonathan and Samuel Kortum, "Technology, geography, and trade," *Econometrica*, 2002, *70* (5), 1741–1779.

Fajgelbaum, Pablo and Stephen J Redding, "Trade, Structural Transformation, and Development: Evidence from Argentina 1869–1914," *Journal of Political Economy*, 2022, *130* (5), 1249–1318.

Foster, Andrew D and Margarita Gafaro, "Subsistence Agriculture with Economies of Scale in Transport Technology," 2017.

_ , Mark R Rosenzweig, and Jere R Behrman, "Population Growth, Income Growth and Deforestation: Management of Village Common Land in India," 2002.

Fujita, Masahisa and Paul Krugman, "When is the economy monocentric?: von Thünen and Chamberlin unified," *Regional science and urban Economics*, 1995, *25* (4), 505–528.

_ , Paul R Krugman, and Anthony Venables, *The spatial economy: Cities, regions, and international trade*, MIT press, 1999.

Giam, Xingli, "Global biodiversity loss from tropical deforestation," *Proceedings of the National Academy of Sciences*, 2017, *114* (23), 5775–5777.

Hansen, Bruce E and Seojeong Lee, "Inference for iterated GMM under misspecification," *Econometrica*, 2021, *89* (3), 1419–1447.

Hardin, Garrett, "The Tragedy of the Commons," *Science*, 1968, *162*, 1243–1248.

Herndon, William Lewis and Lardner Gibbon, *Exploration of the Valley of the Amazon: Made Under Direction of the Navy Department*, Vol. 1, Robert Armstrong, public printer, 1853.

Heß, Simon, Dany Jaimovich, and Matthias Schündeln, "Environmental effects of development programs: Experimental evidence from West African dryland forests," *Journal of Development Economics*, 2021, *153*, 102737.

Holmes, Thomas J and Sanghoon Lee, "Economies of density versus natural advantage: Crop choice on the back forty," *Review of Economics and Statistics*, 2012, *94* (1), 1–19.

Hsiao, Allan, "Coordination and Commitment in International Climate Action: Evidence from Palm Oil," 2022.

Jayachandran, Seema, "How economic development influences the environment," 2021.

_ , "The Inherent Trade-Off Between the Environmental and Anti-Poverty Goals of Payments for Ecosystem Services," Technical Report, National Bureau of Economic Research 2022.

— **, Joost De Laat, Eric F Lambin, Charlotte Y Stanton, Robin Audy, and Nancy E Thomas**, "Cash for carbon: A randomized trial of payments for ecosystem services to reduce deforestation," *Science*, 2017, *357* (6348), 267–273.

Jedwab, Remi, Federico Haslop, Roman David Zarate, and Rodriguez-Castelan Carlos, "The Real Effects of Climate Change in the Poorest Countries: Evidence from the Permanent Shrinking of Lake Chad," 2022.

Kalamandeen, Michelle, Emanuel Gloor, Edward Mitchard, Duncan Quincey, Guy Ziv, Dominick Spracklen, Benedict Spracklen, Marcos Adami, Luiz EOC Aragão, and David Galbraith, "Pervasive rise of small-scale deforestation in Amazonia," *Scientific reports*, 2018, *8* (1), 1–10.

Lu, Mercedes, "Amazon waterway and fluvial transportation: Lessons from environmental planning and evaluation," 2019.

Madhok, Raahil, "Infrastructure, Institutions, and the Conservation of Biodiversity in India," 2022.

Malhi, Yadvinder, J Timmons Roberts, Richard A Betts, Timothy J Killeen, Wenhong Li, and Carlos A Nobre, "Climate change, deforestation, and the fate of the Amazon," *science*, 2008, *319* (5860), 169–172.

Michaels, Guy, Ferdinand Rauch, and Stephen J Redding, "An Eaton and Kortum (2002) Model oF Urbanization and Structural TransFormation," 2011.

Nagy, Dávid Krisztián, "Hinterlands, city formation and growth: Evidence from the US westward expansion," 2020.

Nath, Ishan, "Climate Change, The Food Problem, and the Challenge of Adaptation through Sectoral Reallocation," Technical Report 2022.

Naughton-Treves, Lisa, Jennifer Alix-Garcia, and Colin A Chapman, "Lessons about parks and poverty from a decade of forest loss and economic growth around Kibale National Park, Uganda," *Proceedings of the National Academy of Sciences*, 2011, *108* (34), 13919–13924.

Ostrom, Elinor, *Governing the commons: The evolution of institutions for collective action*, Cambridge university press, 1990.

Paz Soldan, Mariano Felipe, *Diccionario geográfico del Perú* 1877.

Pellegrina, Heitor S, "Trade, productivity, and the spatial organization of agriculture: Evidence from Brazil," *Journal of Development Economics*, 2022, *156*, 102816.

Pimm, Stuart L, Clinton N Jenkins, Robin Abell, Thomas M Brooks, John L Gittleman, Lucas N Joppa, Peter H Raven, Callum M Roberts, and Joseph O Sexton, "The biodiversity of species and their rates of extinction, distribution, and protection," *science*, 2014, *344* (6187), 1246752.

Potapov, PV, J Dempewolf, Y Talero, MC Hansen, SV Stehman, C Vargas, EJ Rojas, D Castillo, E Mendoza, A Calderón et al., "National satellite-based humid tropical forest change assessment in Peru in support of REDD+ implementation," *Environmental Research Letters*, 2014, *9* (12), 124012.

Ravikumar, Ashwin, Robin R Sears, Peter Cronkleton, Mary Menton, and Matías Pérez-Ojeda del Arco, "Is small-scale agriculture really the main driver of deforestation in the Peruvian Amazon? Moving beyond the prevailing narrative," *Conservation Letters*, 2017, *10* (2), 170–177.

Redding, Stephen J, "Goods trade, factor mobility and welfare," *Journal of International Economics*, 2016, *101*, 148–167.

Rivera-Padilla, Alberto, "Crop choice, trade costs, and agricultural productivity," *Journal of Development Economics*, 2020, *146*, 102517.

Rudik, Ivan, Gary Lyn, Weiliang Tan, and Ariel Ortiz-Bobea, "The Economic Effects of Climate Change in Dynamic Spatial Equilibrium," 2021.

Ryan, Nicholas and Anant Sudarshan, "Rationing the commons," *Journal of Political Economy*, 2022, *130* (1), 210–257.

Salehi-Isfahani, Djavad, "Population pressure, intensification of agriculture, and rural-urban migration," *Journal of development economics*, 1993, *40* (2), 371–384.

Sayre, James, "Farm to Firm: Clustering and Returns to Scale in Agricultural Value Chains," 2022.

Schurz, William Lytle, OD Hargis, Courtland Brenneman Manifold, and Curtis Fletcher Marbut, *Rubber Production in the Amazon Valley: By William L. Schurz, Commercial Attaché, and OD Hargis, Special Agent, of the Department of Commerce, and CF Marbut, Chief, Division of Soil Survey, and CB Manifold, Soil Surveyor, of the Department of Agriculture...* number 23, US Government Printing Office, 1925.

Sims, Katharine RE and Jennifer M Alix-Garcia, "Parks versus PES: Evaluating direct and incentive-based land conservation in Mexico," *Journal of Environmental Economics and Management*, 2017, *86*, 8–28.

Sotelo, Sebastian, "Domestic trade frictions and agriculture," *Journal of Political Economy*, 2020, *128* (7), 2690–2738.

Souza-Rodrigues, Eduardo, "Deforestation in the Amazon: A unified framework for estimation and policy analysis," *The Review of Economic Studies*, 2019, *86* (6), 2713–2744.

Stryker, J Dirck, "Population density, agricultural technique, and land utilization in a village economy," *The American Economic Review*, 1976, *66* (3), 347–358.

Szerman, Dmitri, Juliano Assunção, Molly Lipscomb, and Ahmed Mushfiq Mobarak, "Agricultural Productivity and Deforestation: Evidence from Brazil," 2022.

Figures

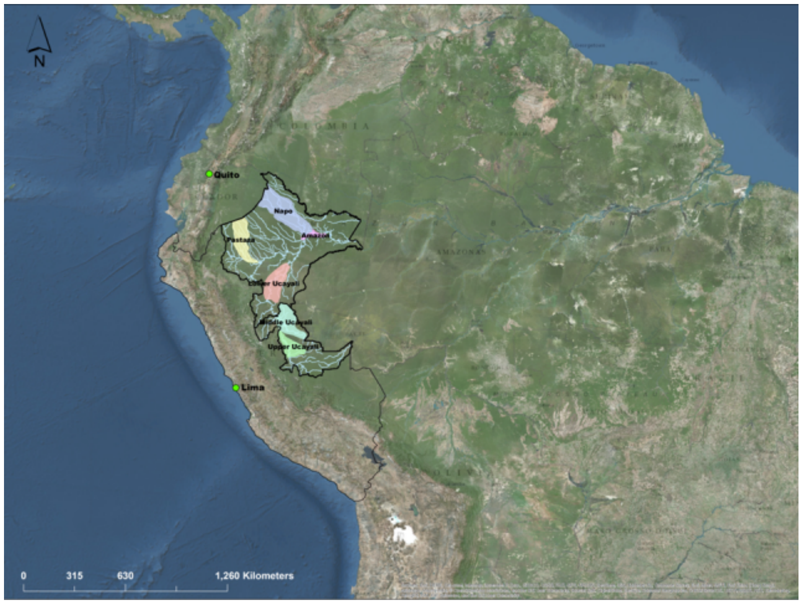
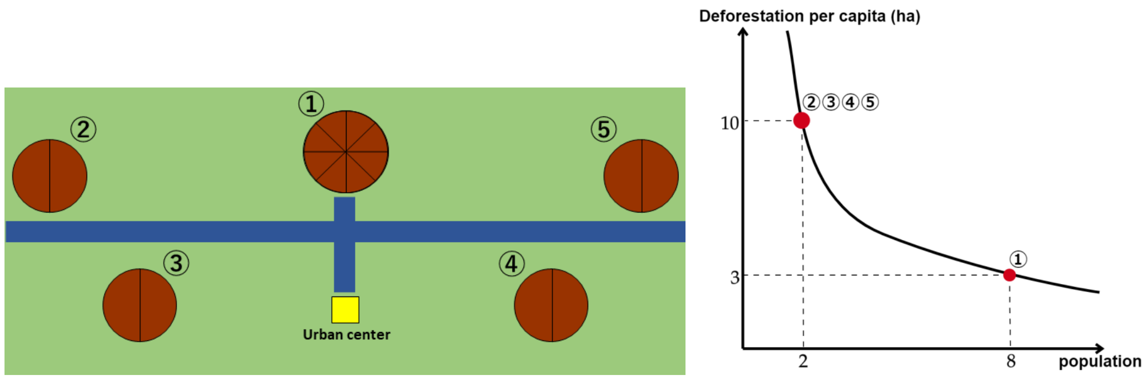
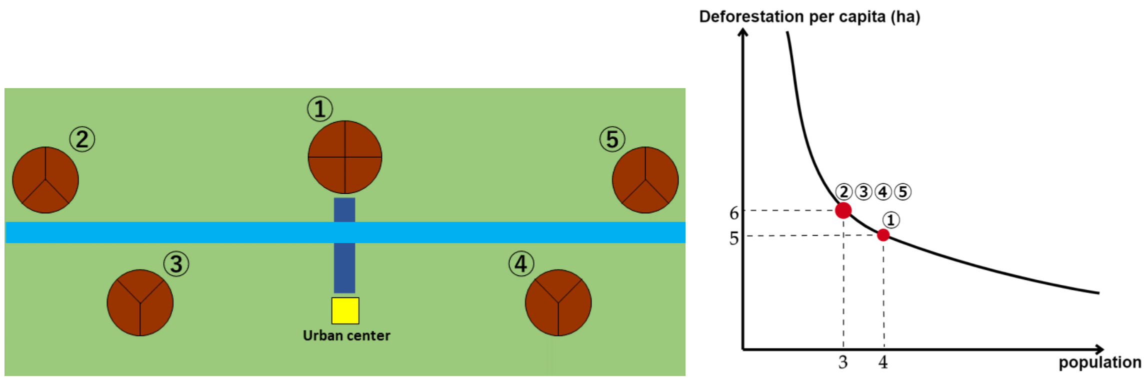


Figure 1: Study Area

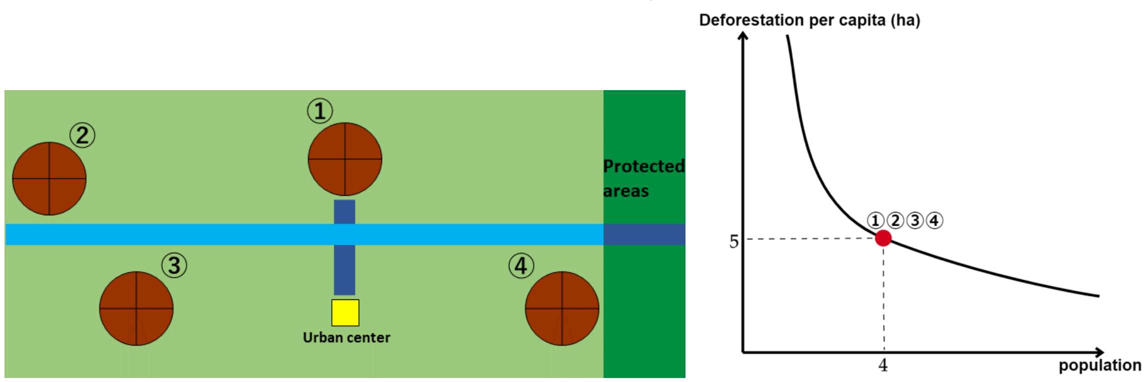
Notes: This map shows the Peruvian Amazon. Two regions inside the thick boundaries represent Loreto and Ucayali departments. Our study area includes four major river basins: the Napo-Amazon basin (combining Napo and Amazon basins in the map), the Pastaza basin, the Lower Ucayali basin, and the Upper Ucayali basin (combining Middle Ucayali and Upper Ucayali basins in the map).



(A) Benchmark (Deforestation = 104 ha)



(B) Transport infrastructure investments that integrate hinterlands (Deforestation = 92 ha)



(C) Transport infrastructure + Protecting the rural frontier (Deforestation = 80 ha)

Figure 2: Intuitive Illustration of the Deforestation Impacts of Counterfactual Policies

Notes: These figures illustrate the deforestation outcome in a simple river basin with a fixed number of total population (= 16) under three different scenarios. For simplicity, these figures focus on the agricultural sector and abstract from sectoral reallocation of the population through general equilibrium effects. There are 16 agricultural populations. In the left maps, brown areas represent deforested areas for agricultural land around each community. Each area inside borders represents the deforested area by each farmer. A dark blue line represents rivers. A light blue line represents rivers with upgraded transport infrastructure. The right diagrams plot the agricultural population in each community and the size of deforested area per farmer. ①–⑤ represent community identifiers and correspond between the left maps and the right diagrams.

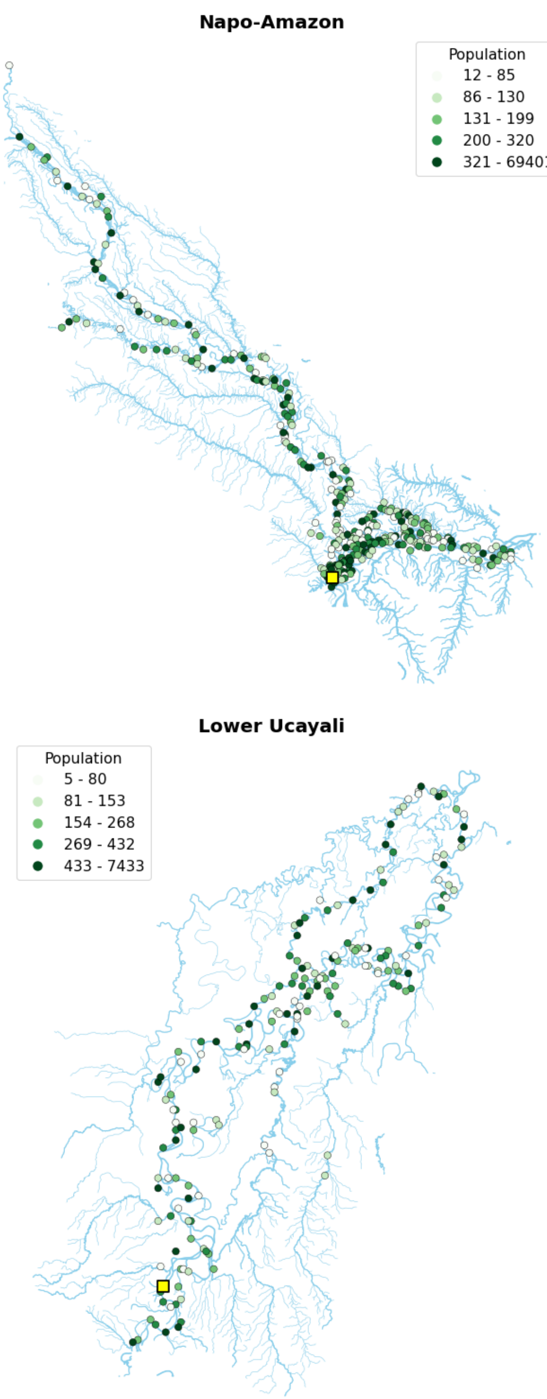


Figure 3: Population in Rural Locations

Notes: Each map shows population in rural locations in each basin of the study area. The legend is based on quantiles. Yellow squares represent the urban centers.

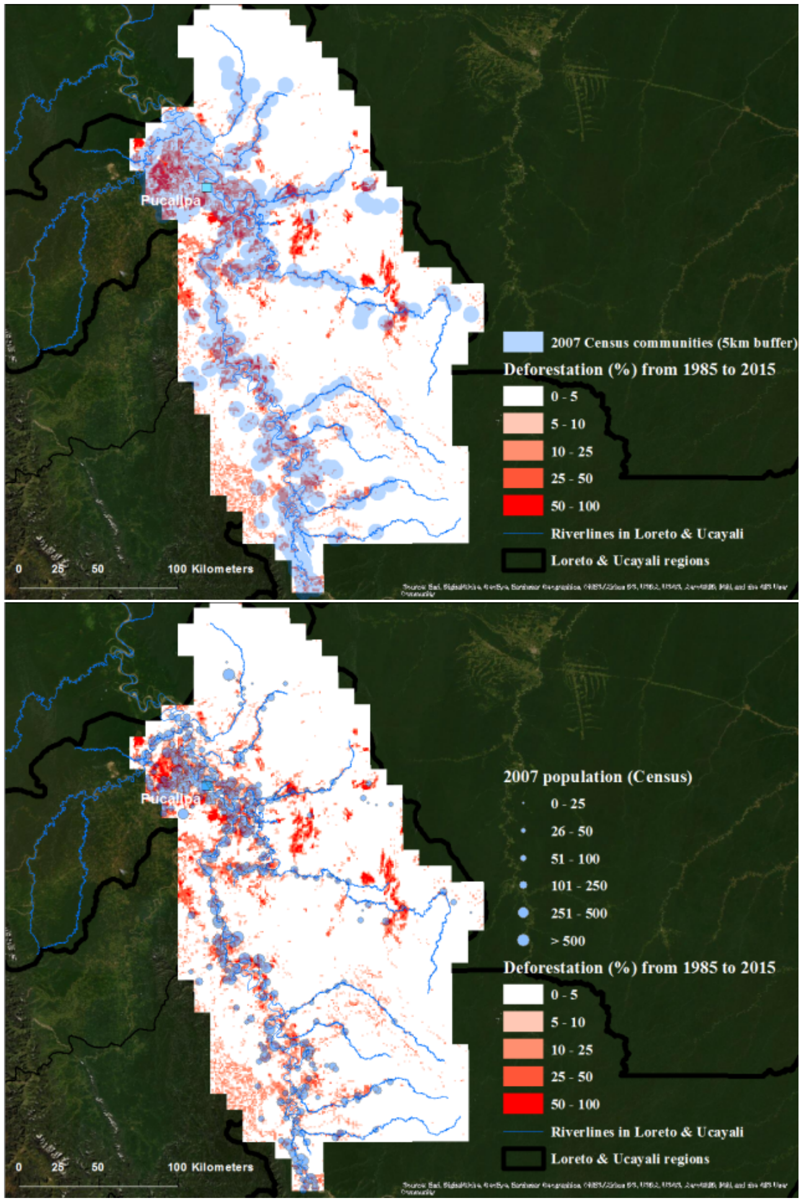


Figure 4: Communities, Populations, and Deforestation

Notes: The maps show the Upper-Ucayali basin. The forest loss measure represents a percentage of the deforested area inside each grid cell.

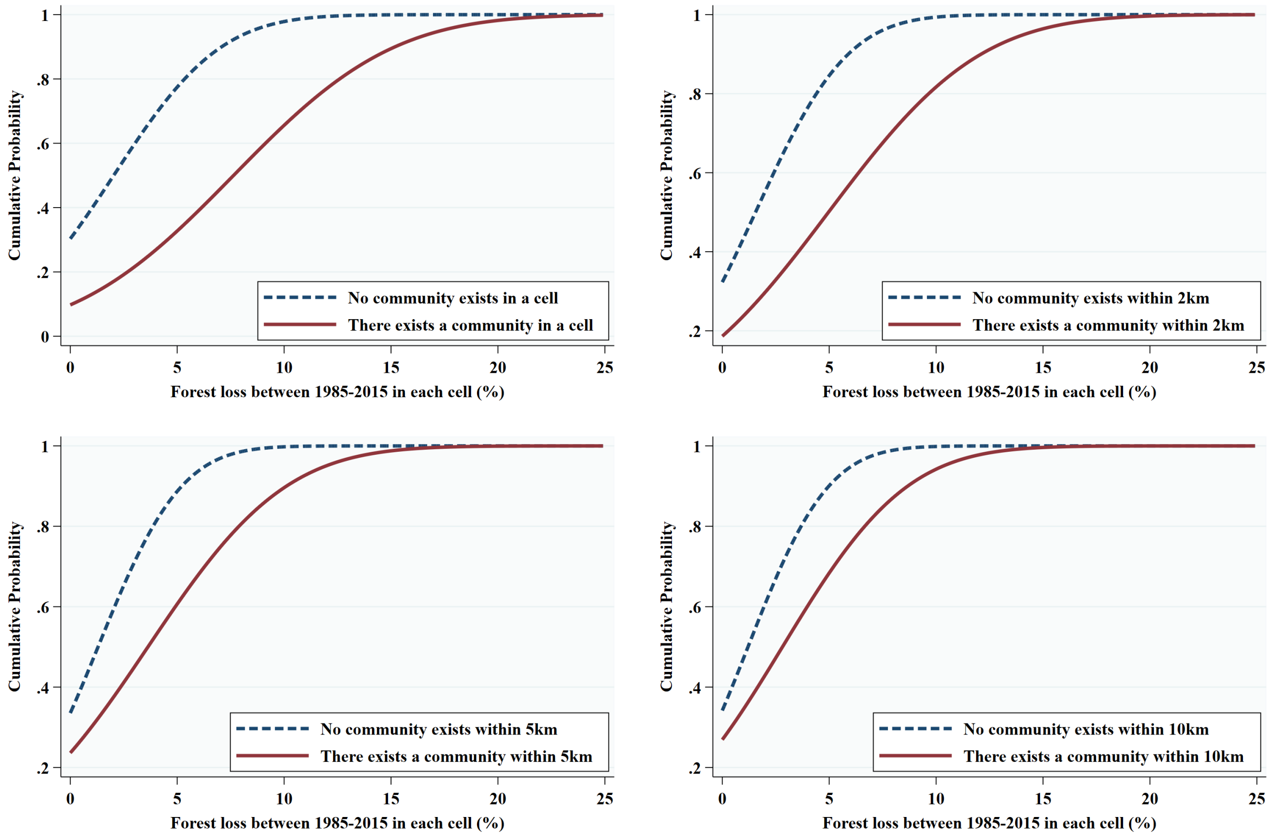


Figure 5: Communities and Deforestation (1985-2015)

Notes: The unit of analysis is a 1km × 1km grid cell in the four river basins. The forest loss measure represents a percentage of the deforested area inside each grid cell. The information of community locations is from the Peru Population and Housing Census in 2007.

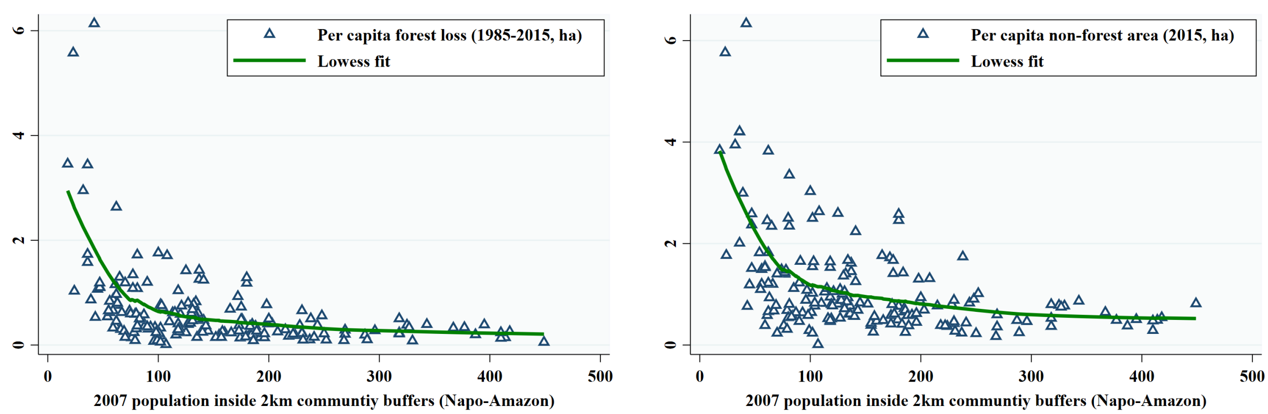


Figure 6: Populations and Per Capita Land Footprint

Notes: The population information is from the Peru Population and Housing Census in 2007. The total population in the 2km buffer surrounding a community is measured by summing populations from communities whose centroid are inside the buffer.

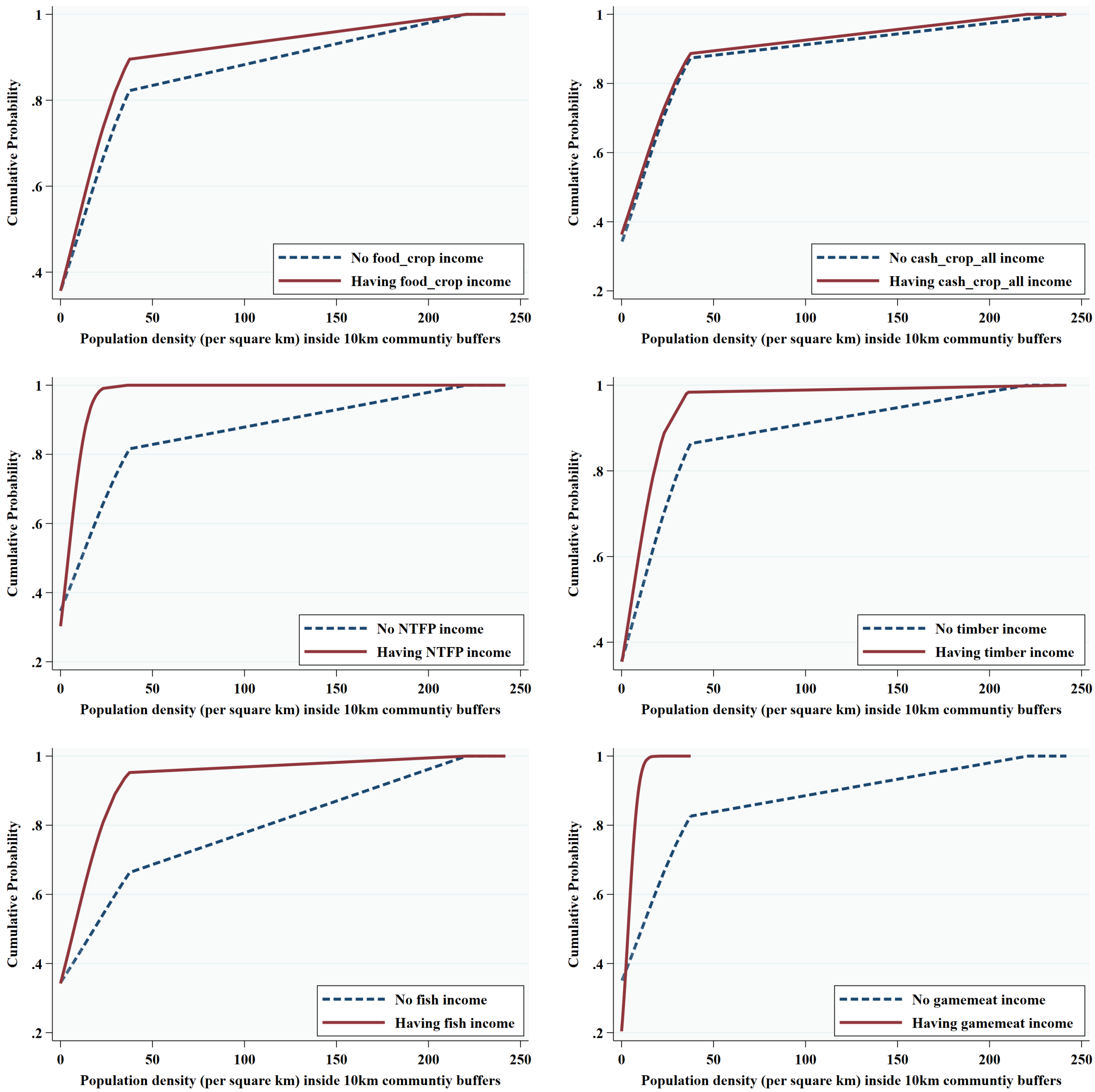


Figure 7: Population Density (10km buffer) and Distribution of Activities (Household-level)—Agriculture vs. Forest Products vs. Wildlife Extractions

Notes: The population information is from the PARLAP community census. We divide the total population in the $x$km buffer by the area of buffer ($x^2\pi$) to calculate the population density in the $x$km buffer. The total population in the $x$km buffer surrounding a community is measured by summing populations from communities whose centroid are inside the buffer.

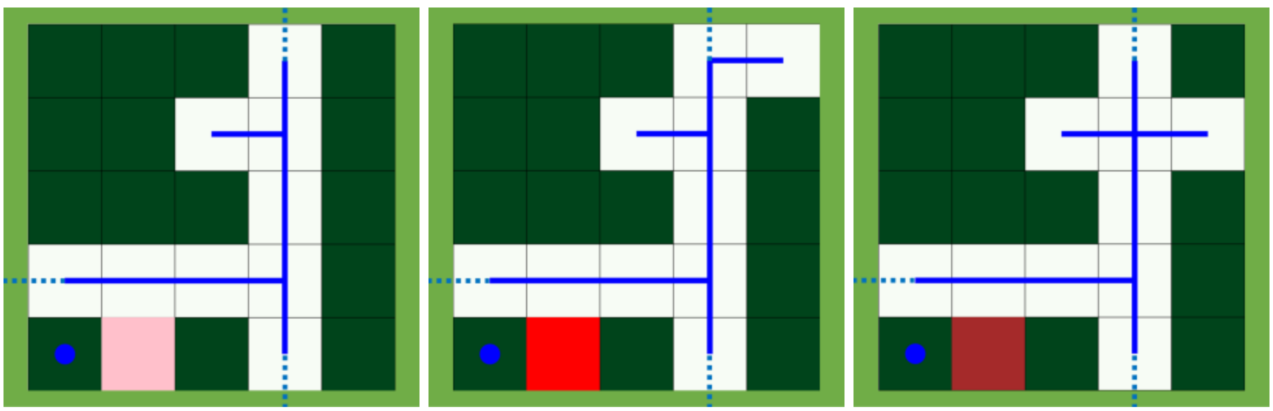


Figure 8: Intuition of Identifying the Density Externalities

Notes: This figure shows three similar areas inside a river basin and illustrates a comparison of the River Network Access (RNA) measures between pink (in the left), red (in the middle), and brown (in the right) cells as a plausibly exogenous variation to identify the density externalities. The blue lines indicate rivers and the white cells indicate river cells (cells that face rivers). RNA captures the weighted sum of the accessibility to other river cells (the white cells in this figure) whether or not these cells have positive populations. The middle and right maps have one more river cell than the left map, each in a different location. Therefore, RNA is higher in the red and brown cells than that in the pink cell. Comparing the red cell in the middle map with the brown cell in the right map, RNA is higher in the brown cell because the additional cell next to the top-right is located closer to the brown cell.

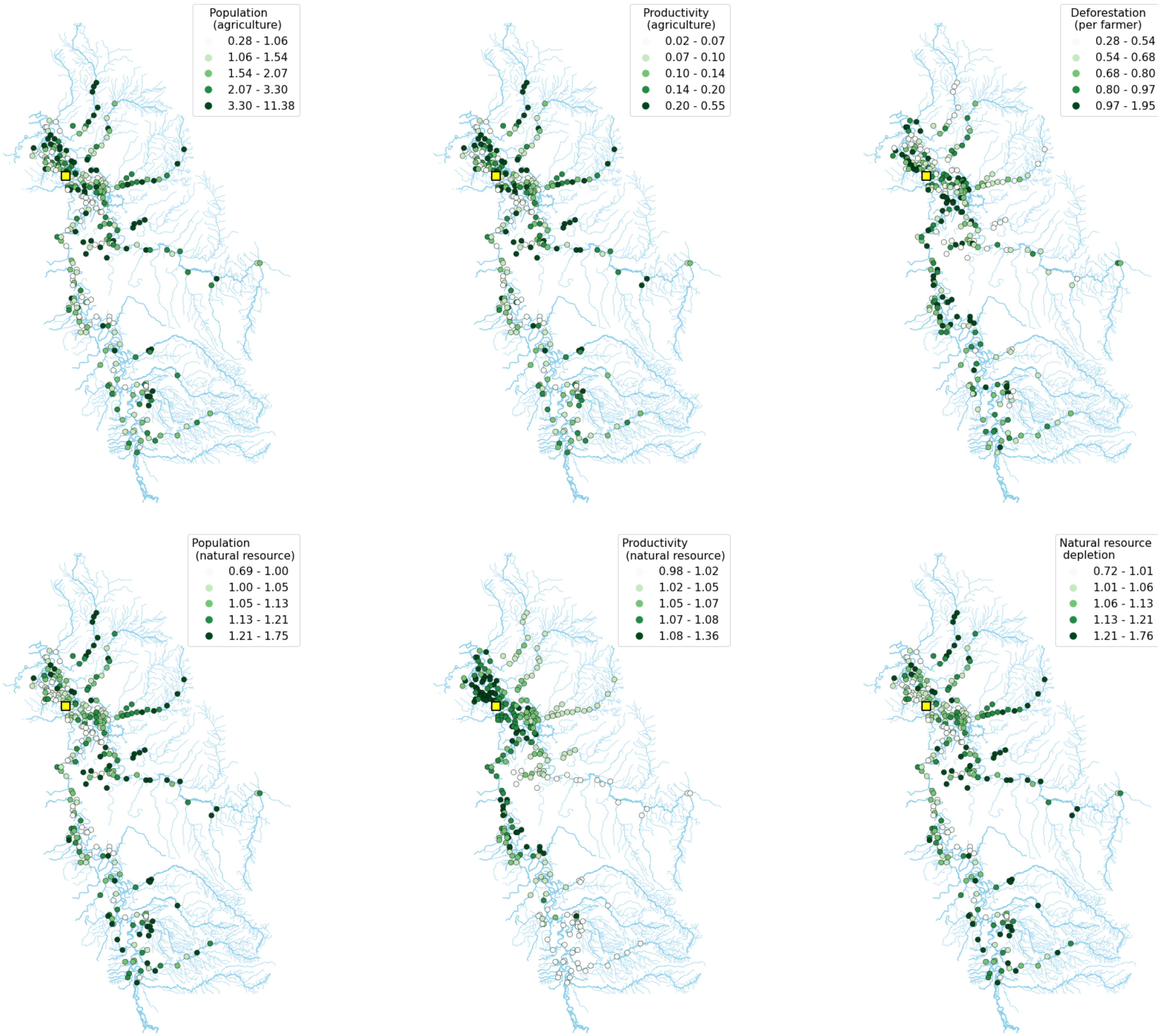


Figure 9: Counterfactual Outcomes without the Agglomeration Externality (Upper Ucayali)

Notes: These maps present the counterfactual outcomes of shutting down the agglomeration externality in agriculture. Values shown in the legend of the circle dots are relative values in the counterfactual scenarios in terms of those in the benchmark spatial equilibrium in rural locations. The legend is based on quantiles. The yellow square represents the urban center.

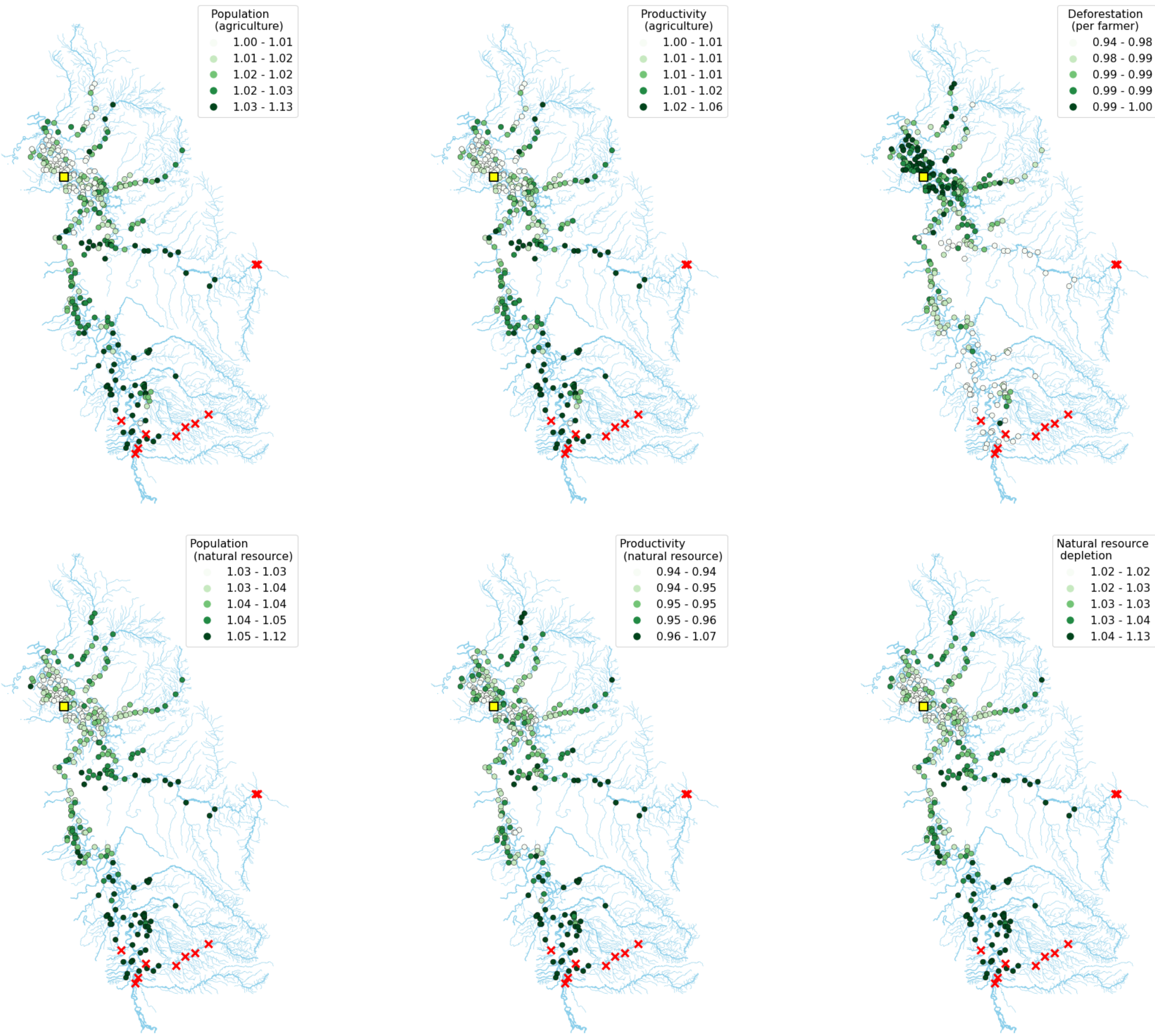


Figure 10: Counterfactual Outcomes of Protecting the Rural Frontier (Upper Ucayali)

Notes: These maps present the counterfactual outcomes of a place-based protection policy of controlling the expansion of rural frontier. This experiment chooses rural locations to be treated in order, starting with those farthest from the urban center, until the treated population reaches 2.5% of the total rural population in each basin. The red **x** marks indicate the treated locations. Values shown in the legend of the circle dots are relative values in the counterfactual scenarios in terms of those in the benchmark spatial equilibrium in rural locations. The legend is based on quantiles. The yellow square represents the urban center.

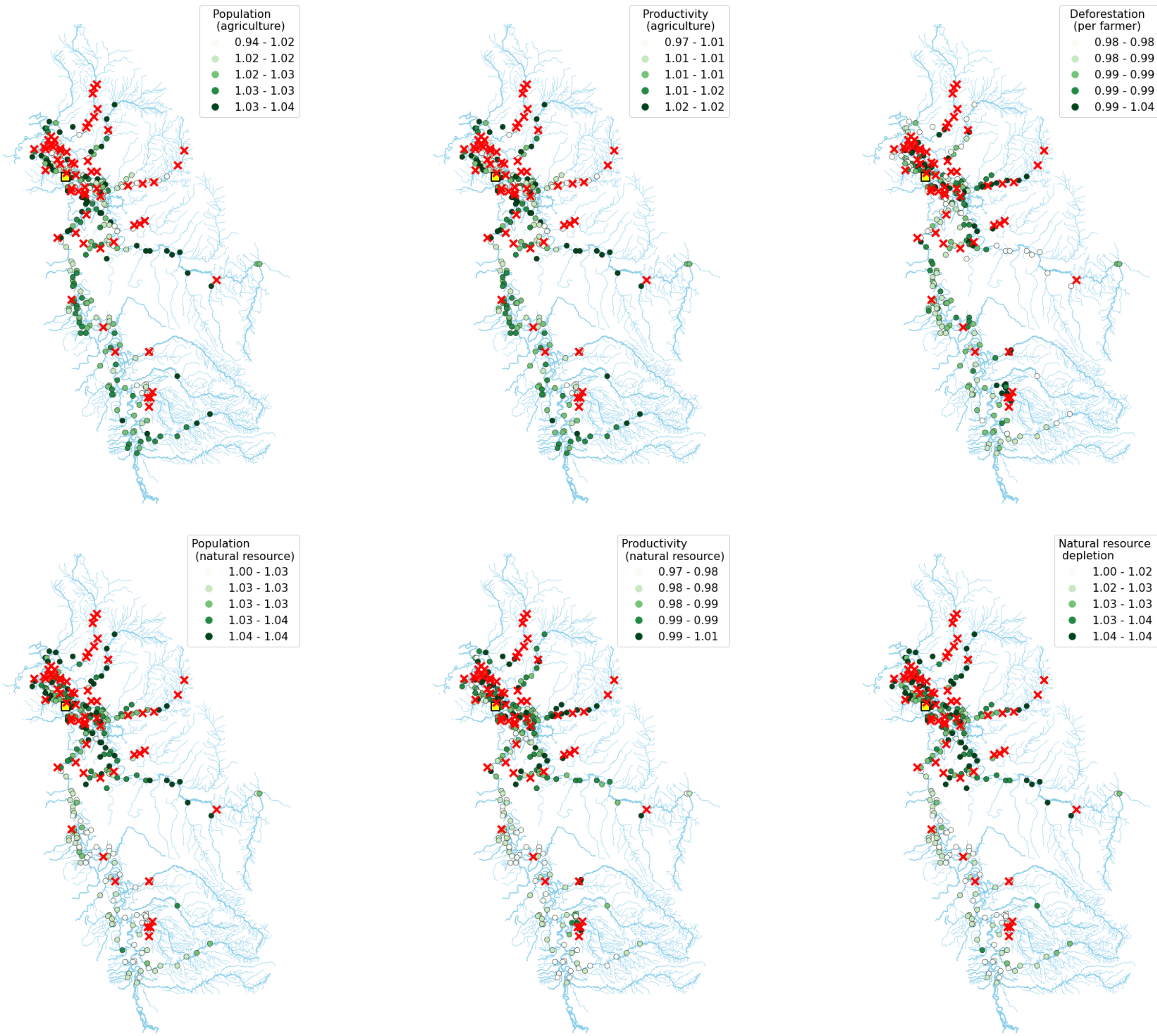


Figure 11: Counterfactual Outcomes without Small Communities (Upper Ucayali)

Notes: These maps present the counterfactual outcomes of a place-based protection policy that targets the smallest communities. This experiment chooses rural locations to be treated in order, starting with the location with the smallest population size, until the treated population reaches 2.5% of the total population in each basin. The red **x** marks indicate the treated locations. Values shown in the legend of the circle dots are relative values in the counterfactual scenarios in terms of those in the benchmark spatial equilibrium in rural locations. The legend is based on quantiles. The yellow square represents the urban center.

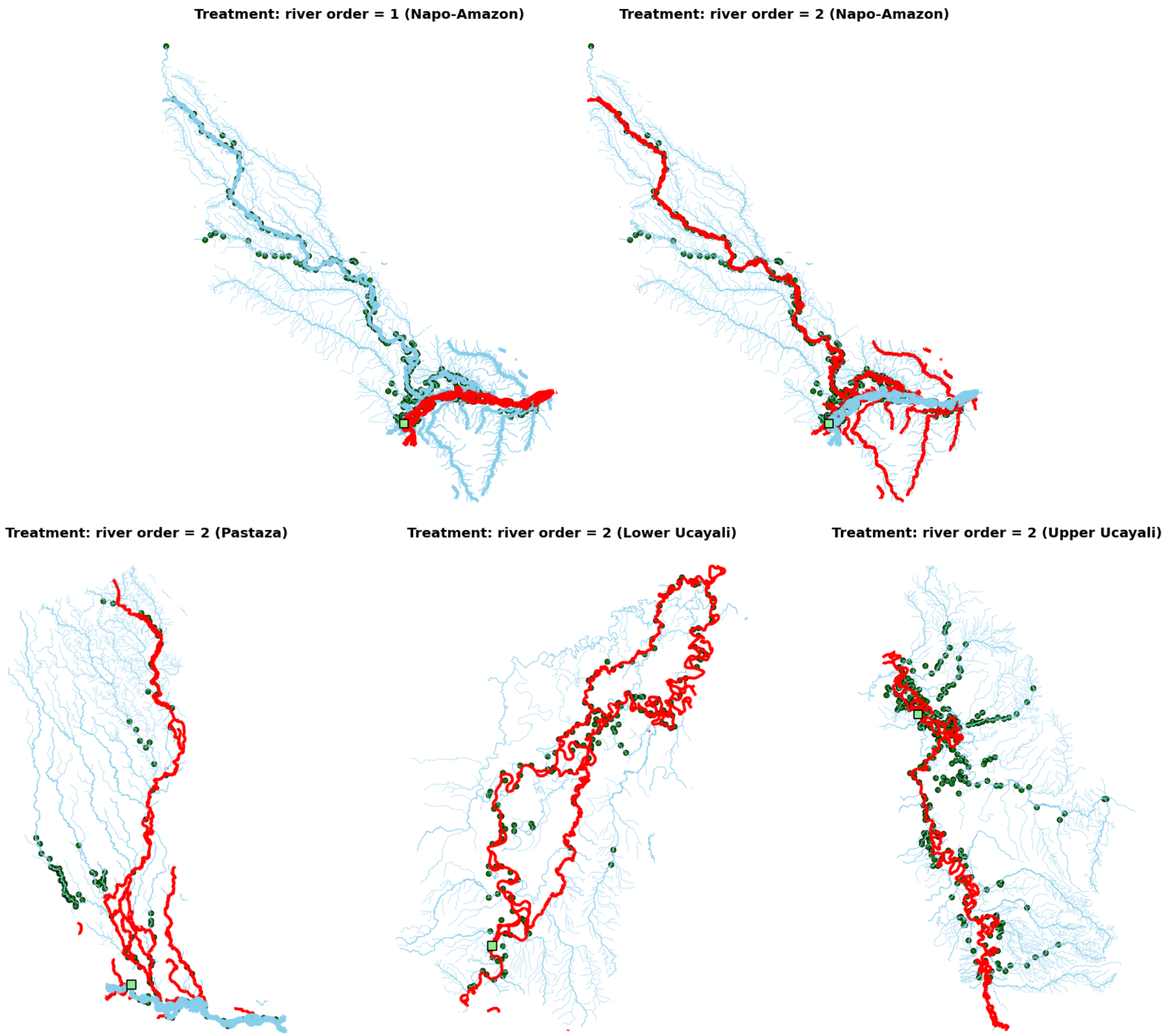


Figure 12: River Orders and Counterfactual Experiments of Improving Transport Infrastructure

Notes: The light blue lines indicate the river lines without the improvement of transport infrastructure. The red lines indicate the river lines where the transport infrastructure is upgraded. The yellow square represents the urban center.

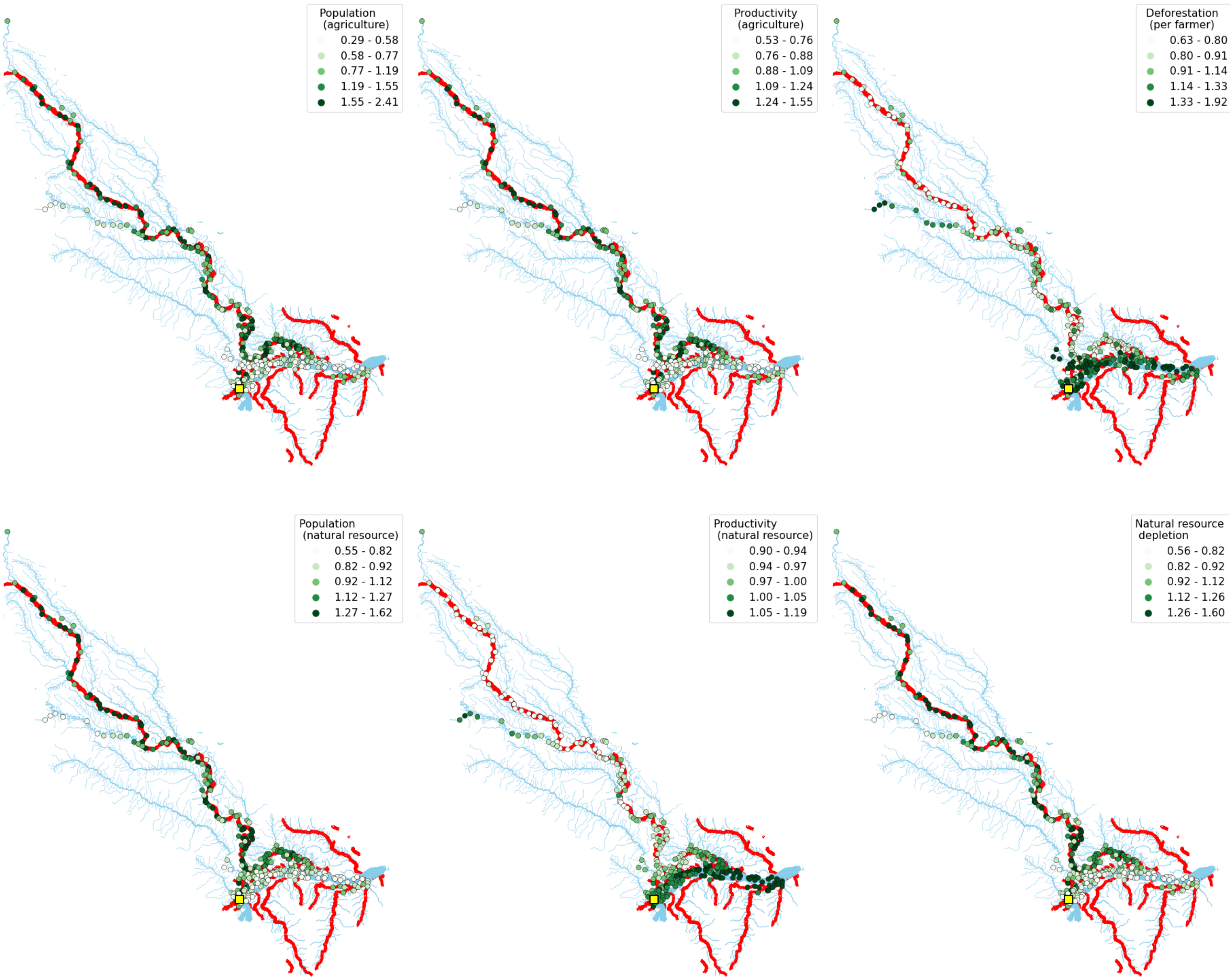


Figure 13: Counterfactual Outcomes with Improved Transport Infrastructure along River Order 2 (Napo-Amazon)

Notes: These maps present the counterfactual outcomes of improving the transport infrastructure in a way that hinterlands are connected to the central area of the basin. The light blue lines indicate the river lines without the improvement of transport infrastructure. The red lines indicate the river lines where the transport infrastructure is upgraded. Values shown in the legend of the circle dots are relative values in the counterfactual scenarios in terms of those in the benchmark spatial equilibrium in rural locations. The legend is based on quantiles. The yellow square represents the urban center.

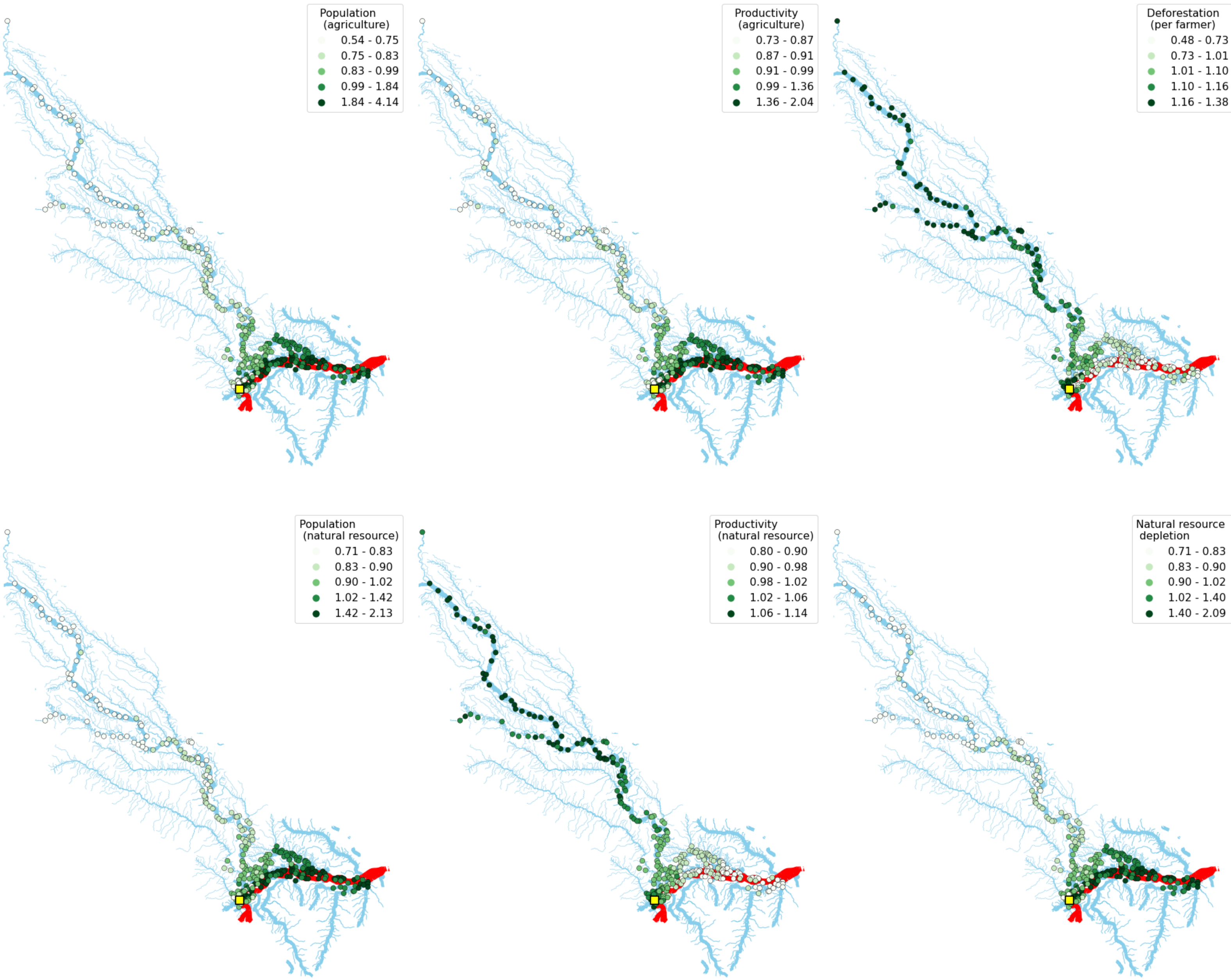


Figure 14: Counterfactual Outcomes with Improved Transport Infrastructure along River Order 1 (Napo-Amazon)

Notes: These maps present the counterfactual outcomes of concentrating the transport infrastructure investments in the central area of the basin. The light blue lines indicate the river lines without the improvement of transport infrastructure. The red lines indicate the river lines where the transport infrastructure is upgraded. Values shown in the legend of the circle dots are relative values in the counterfactual scenarios in terms of those in the benchmark spatial equilibrium in rural locations. The legend is based on quantiles. The yellow square represents the urban center.

Tables

Table 1: Human Settlements and Deforestation

	Sum	N	Mean	Std. Dev.	Min	Max
(A) All grid cells in the four basins						
1.1. No community exists in a cell	424160.7	149259	2.842	7.601	0	101.97
1.2. There exists a community in a cell	13606.38	1304	10.434	10.8	0	96.93
2.1. No community exists within 2km	302599.5	131962	2.293	6.883	0	101.97
2.2. There exists a community within 2km	135167.6	18601	7.267	10.865	0	101.79
3.1. No community exists within 5km	205061.6	106244	1.93	6.217	0	101.97
3.2. There exists a community within 5km	232705.5	44319	5.251	9.963	0	101.79
(B) Grid cells within 2km from a river line						
1.1. No community exists in a cell	301283.9	102075	2.952	7.533	0	101.97
1.2. There exists a community in a cell	12243.42	1232	9.938	10.007	0	88.11
2.1. No community exists within 2km	197397	86501	2.282	6.772	0	101.97
2.2. There exists a community within 2km	116130.3	16806	6.91	10.078	0	101.79
3.1. No community exists within 5km	126734.8	67652	1.877	6.134	0	101.97
3.2. There exists a community within 5km	186792.6	35786	5.22	9.417	0	101.79

Notes: This table presents summary statistics on forest loss from 1985 to 2015 for each grid cell category according to the presence of settlements. The unit of analysis is a 1km $\times$ 1km grid cell in the four river basins. The unit of measurement for forest loss measure is the hectare. The information of community locations is from the Peru Population and Housing Census in 2007.

Table 2: Summary of Model Parameters

Parameter	Description	Estimation strategy
δ_{Ag}, δ_{Nr}, δ_M	Elasticity of trade cost	Commodity prices from the CC
λ_{up}, λ_{land}	Relative distance in terms of downstream river	Travel time and transport costs survey
α_{Ag}, α_{Nr}, α_M	Utility weights	Normalized to = 1
σ	Within-sector elasticity of substitution	Expenditure information from ENAHO
$\bar{\sigma}$	Across-sector elasticity of substitution	Expenditure information from ENAHO
γ	Labor share in agricultural production function	From the literature (e.g., Sotelo 2020)
θ	Trade elasticity	From the literature (Donaldson 2018)
μ_L	Congestion in forest clearing	Linear IV using the community-level data
μ_{Ag}	Agglomeration in agricultural production	Model inversion and linear IV
μ_{Nr}	Congestion in natural resource extraction	Model inversion and non-linear GMM
ν	Spatial decay in natural resource access	Model inversion and non-linear GMM

Notes: See section 4 for the model and section 5 for the structural estimation.

Table 3: Density Externalities in Agriculture and Natural Resource Extraction

Parameter	Point estimate	Standard errror	Description
(A) Agriculture			
$\tilde{\mu}_{Ag}$	0.064	0.010	$= \mu_{Ag} - (1-\gamma)\mu_L$
	J test p-value = 0.648		
μ_L	0.522	0.094	Congestion in forest clearing
μ_{Ag}	0.273		Agglomeration in agricultural production
(B) Natural resource extraction			
μ_{Nr}	0.335	0.042	Congestion in natural resource extraction
ν	0.593	0.075	Spatial decay of congestion externality
	J test p-value = 0.821		

Notes: Estimates of density externalities in agriculture (panel A) are based on the linear specification using $\ln RNA_o$ and the initial community existence in 1940 as instruments. Estimates of parameters governing congestion externality in natural resource extraction (panel B) are based on the non-linear GMM using $\ln RNA_o$ and $\{\ln \sum_{d|D_{o,d} \leq x} RNA_d\}$ for $x \in \mathcal{X} = \{2, 5, 10, 25, 50, 75, 100\}$ as instruments.

Table 4: Quantifying the Effects of Density Externalities

Basin	Welfare	Deforestation	Q (Ag)	Q (Nr)	N (Ag)	N (Nr)	N (Urban)
(A) Without the Agglomeration Externality in Agriculture ($\mu_{Ag} \to 0$)							
Napo	0.888	1.555	0.654	0.991	1.029	0.956	1.011
Pastaza	0.911	1.241	0.688	0.992	1.044	0.974	0.988
LowerUcayali	0.873	1.214	0.718	0.968	1.051	0.959	0.987
UpperUcayali	0.929	1.177	0.709	0.987	1.066	0.982	0.985
(B) Higher Congestion Externality in Natural Resource Extractions ($\mu_{Nr} \uparrow$ by 25%)							
Napo	0.770	1.131	0.927	0.497	0.925	1.138	0.955
Pastaza	0.757	1.035	0.909	0.585	0.930	1.087	0.938
LowerUcayali	0.704	1.015	0.901	0.536	0.914	1.088	0.926
UpperUcayali	0.889	0.996	0.980	0.586	0.976	1.156	0.975

Notes: Values shown in the table are relative values in the counterfactual scenarios in terms of those in the benchmark spatial equilibrium.

Table 5: Community Population and Availability of Transport Modes

(A)	Availability of Transport Modes in a Community							
	Lancha		Colectivo		Rapido		Peque-peque	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
$\log(N_{o,Ag})$	0.0469***	0.144***	0.0478***	0.0280	0.0522***	0.0566*	-0.00528	0.00418
	(0.0111)	(0.0430)	(0.0115)	(0.0383)	(0.0108)	(0.0292)	(0.00576)	(0.0156)
Mean (Dep. var.)	0.492	0.492	0.386	0.386	0.110	0.110	0.972	0.972
SD (Dep. var.)	0.500	0.500	0.487	0.487	0.314	0.314	0.164	0.164
First stage F-stat		24.845		24.845		24.845		24.845
Observations	906	906	906	906	906	906	906	906
(B)	Frequency of Transport Modes Passing a Community per Week							
	Lancha		Colectivo		Rapido		Peque-peque	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
$\log(N_{o,Ag})$	0.162	1.055**	0.213**	0.336	0.0439	-0.287	0.225	0.518
	(0.118)	(0.468)	(0.0958)	(0.305)	(0.283)	(0.574)	(0.177)	(0.677)
Mean (Dep. var.)	3.429	3.429	5.154	5.154	6.164	6.164	5.835	5.835
SD (Dep. var.)	2.468	2.468	2.368	2.368	1.772	1.772	2.019	2.019
First stage F-stat		7.351		10.260		3.758		3.233
Observations	330	330	276	276	61	61	144	144
Basin FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Robust standard errors in parentheses. The unit of anaylysis is a community in the PARLAP Community Census (CC) in 2014. We use $\log(RNA_o)$ and the initial community existence in 1940 as instruments for $\log(N_{o,Ag})$. Geographical controls include a dummy of high river orders (4 and 5), distance to the urban center, distance to the river, squared distance to the river, interaction terms of these two variables with a river cell dummy, elevation, river confluence, flood vulnerability, geology measures, and open water access measures for a grid cell where each census community belongs.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Community Population and Modern Technology Use

(A) Basic infrastructure	(1)	(2)	(3)	(4)	(5)	(6)
	Irrigation	Certified seed	Crops have been certified organic	Electricity for agricultural work	Animals for agricultural work	Tractors for agricultural work
$\log(N_{o,Ag})$	-0.00329*	-0.000857	0.0000692	-0.000688	0.00315	0.000476
	(0.00180)	(0.00430)	(0.000584)	(0.000863)	(0.00206)	(0.000811)
Mean (Dep. var.)	0.013	0.064	0.001	0.003	0.010	0.002
SD (Dep. var.)	0.112	0.245	0.037	0.054	0.098	0.044
First stage F-stat	1649.082	1649.082	1649.082	1649.082	1649.082	1649.082
Observations	25827	25827	25827	25827	25827	25827
(B) Inputs into land and crops	(1)	(2)	(3)	(4)	(5)	(6)
	Guano/manure/ compost	Chemical fertilizers	Insecticides	Herbicides	Fungicides	Biologic control
$\log(N_{o,Ag})$	0.000807	0.00265**	0.0228***	0.0314***	0.0118***	-0.00239
	(0.00111)	(0.00115)	(0.00353)	(0.00371)	(0.00219)	(0.00239)
Mean (Dep. var.)	0.005	0.004	0.040	0.051	0.012	0.020
SD (Dep. var.)	0.069	0.063	0.197	0.221	0.111	0.140
First stage F-stat	1649.082	1649.082	1649.082	1649.082	1649.082	1649.082
Observations	25827	25827	25827	25827	25827	25827
(C) Animal, electrical, or mechanical energy	(1)	(2)	(3)	(4)	(5)	(6)
	Iron plow of animal traction	Wooden plow of animal traction	Harvester	Foot plow	Motorized sprayer	Manual sprayer
$\log(N_{o,Ag})$	-0.000796	-0.000223	-0.000229	-0.000806	0.00197**	0.0214***
	(0.000523)	(0.000311)	(0.000282)	(0.000556)	(0.000815)	(0.00401)
Mean (Dep. var.)	0.001	0.000	0.001	0.001	0.002	0.062
SD (Dep. var.)	0.035	0.022	0.025	0.035	0.043	0.241
First stage F-stat	1649.082	1649.082	1649.082	1649.082	1649.082	1649.082
Observations	25827	25827	25827	25827	25827	25827
(D) Electrical or mechanical energy	(1)	(2)	(3)	(4)	(5)	(6)
	Grain mill	Grass chopper	Thresher	Electric generator	Wheel tractor	Boat/canue/ speedboat
$\log(N_{o,Ag})$	0.00696***	0.000462	-0.00102	-0.0103***	0.000932	-0.0187**
	(0.00194)	(0.000448)	(0.000674)	(0.00323)	(0.000590)	(0.00746)
Mean (Dep. var.)	0.013	0.001	0.004	0.036	0.001	0.618
SD (Dep. var.)	0.111	0.025	0.061	0.186	0.030	0.486
First stage F-stat	1649.082	1649.082	1649.082	1649.082	1649.082	1649.082
Observations	25827	25827	25827	25827	25827	25827
Basin FE	Yes	Yes	Yes	Yes	Yes	Yes
Geographic controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Robust standard errors in parentheses. The unit of anaylysis is a household in the 2012 Peruvian Agricultural Census. We use $\log(RNA_o)$ and the initial community existence in 1940 as instruments for $\log(N_{o,Ag})$. Geographical controls include a dummy of high river orders (4 and 5), distance to the urban center, distance to the river, squared distance to the river, interaction terms of these two variables with a river cell dummy, elevation, river confluence, flood vulnerability, geology measures, and open water access measures for a grid cell where each census community belongs.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Counterfactual Experiments of Protection Policies

Basin	Welfare	Deforestation	Q (Ag)	Q (Nr)	N (Ag)	N (Nr)	N (Urban)
(A) Protecting the rural frontier							
Napo	0.998	0.948	0.999	0.997	1.002	1.003	0.996
Pastaza	0.999	0.975	0.999	1.001	1.001	1.000	0.998
LowerUcayali	0.997	0.967	0.999	0.997	0.999	1.001	0.996
UpperUcayali	0.998	0.980	1.000	0.992	1.001	1.004	0.999
(B) Not allowing new community formation							
Napo	0.998	0.915	1.008	1.002	1.002	1.002	0.997
Pastaza	0.997	0.951	1.004	1.000	0.999	1.003	0.997
LowerUcayali	0.997	0.979	0.996	0.994	0.998	1.002	0.998
UpperUcayali	0.999	0.970	1.004	1.001	1.001	1.002	0.999
(C) Not allowing for small communities							
Napo	0.997	0.869	1.009	1.002	1.002	1.003	0.996
Pastaza	0.996	0.936	1.003	1.002	0.999	1.003	0.996
LowerUcayali	0.998	0.900	1.008	1.003	1.001	1.000	0.994
UpperUcayali	0.999	0.927	1.009	1.005	1.001	1.003	0.999

Notes: Values shown in the table are relative values in the counterfactual scenarios in terms of those in the benchmark spatial equilibrium. Each policy directly treats 2.5% of rural populations in the benchmark equilibrium in each basin for resettlement.

Table 8: Counterfactual Experiments of Improving Transport Infrastructure

Basin	Welfare	Deforestation	Q (Ag)	Q (Nr)	N (Ag)	N (Nr)	N (Urban)
(A) Transport infrastructure improved ($\lambda_{upgraded} = 0.8$) along river order = 2							
Napo	1.018	0.989	0.992	0.997	0.996	0.998	1.004
Pastaza	1.013	0.976	1.028	0.995	1.003	0.999	0.998
LowerUcayali	1.023	1.014	1.022	0.978	1.000	0.995	1.023
UpperUcayali	1.012	0.989	1.024	1.003	1.002	1.005	0.998
(B) Transport infrastructure improved ($\lambda_{upgraded} = 0.8$) along river order = 1							
Napo	1.010	1.060	0.974	0.997	0.994	0.997	1.007
(C) Transport infrastructure improved ($\lambda_{upgraded} = 0.8$) along river order = 1, 2							
Napo	1.029	1.051	0.969	0.995	0.991	0.995	1.011
(D) Symmetric trade cost ($\lambda_{up} = 1$)							
Napo	1.011	1.023	0.985	0.997	0.996	0.997	1.005
Pastaza	1.008	1.006	1.002	0.999	1.000	1.000	1.000
LowerUcayali	1.009	1.035	1.003	0.992	0.999	0.998	1.018
UpperUcayali	1.005	1.011	1.005	1.004	1.003	1.006	0.998
(E) Transport infrastructure improved ($\lambda_{upgraded} = 0.8$) along river order = 2 with checkposts that prohibit transporting natural resource goods							
Napo	1.012	0.968	0.998	1.002	1.000	1.002	0.999
Pastaza	1.007	0.979	1.021	0.999	1.000	1.002	0.996
LowerUcayali	1.010	1.020	1.011	0.991	0.996	1.002	1.007
UpperUcayali	1.009	0.988	1.026	1.007	1.002	1.009	0.998

Notes: Values shown in the table are relative values in the counterfactual scenarios in terms of those in the benchmark spatial equilibrium.

Table 9: Counterfactual Outcomes of a Composite Intervention

Basin	Welfare	Deforestation	Q (Ag)	Q (Nr)	N (Ag)	N (Nr)	N (Urban)
Protecting the rural frontier &							
Transport infrastructure improved ($\lambda_{upgraded} = 0.8$) along river order = 2							
Napo	1.016	0.933	0.990	0.994	0.999	1.001	1.011
Pastaza	1.010	0.953	1.024	0.997	1.004	0.999	0.996
LowerUcayali	1.021	0.990	1.019	0.976	1.000	0.996	1.021
UpperUcayali	1.010	0.969	1.024	0.995	1.003	1.009	0.997

Notes: Values shown in the table are relative values in the counterfactual scenarios in terms of those in the benchmark spatial equilibrium. The frontier protection directly treats 2.5% of rural populations in the benchmark equilibrium in each basin for resettlement.

— Appendix —

"Human and Nature: Economies of Density and Conservation in the Amazon Rainforest"

Shunsuke Tsuda, Yoshito Takasaki, Mari Tanaka

A Additional Figures and Tables

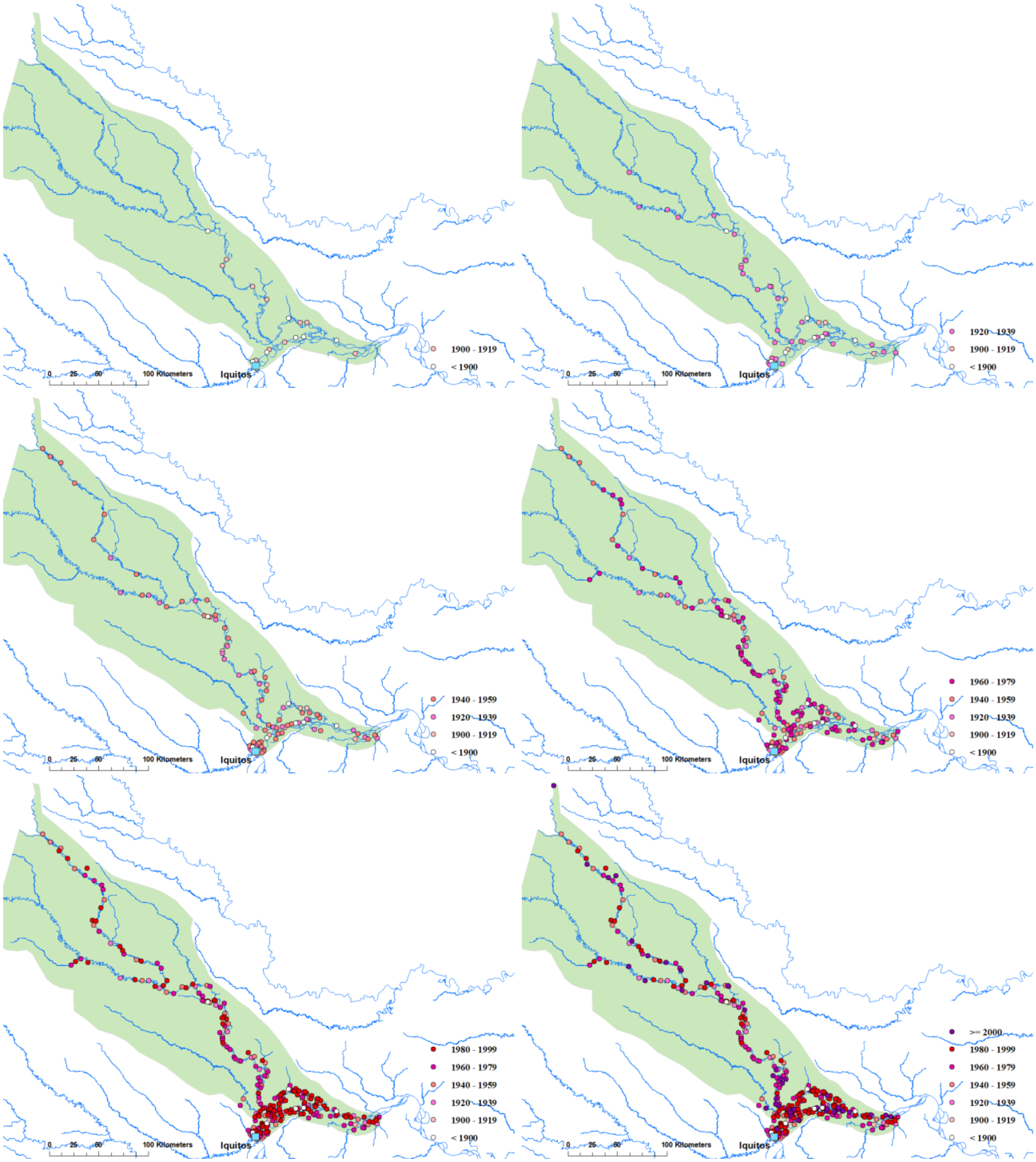


Figure A.1: Establishment of Rainforest Communities over Decades in the Napo-Amazon Basin

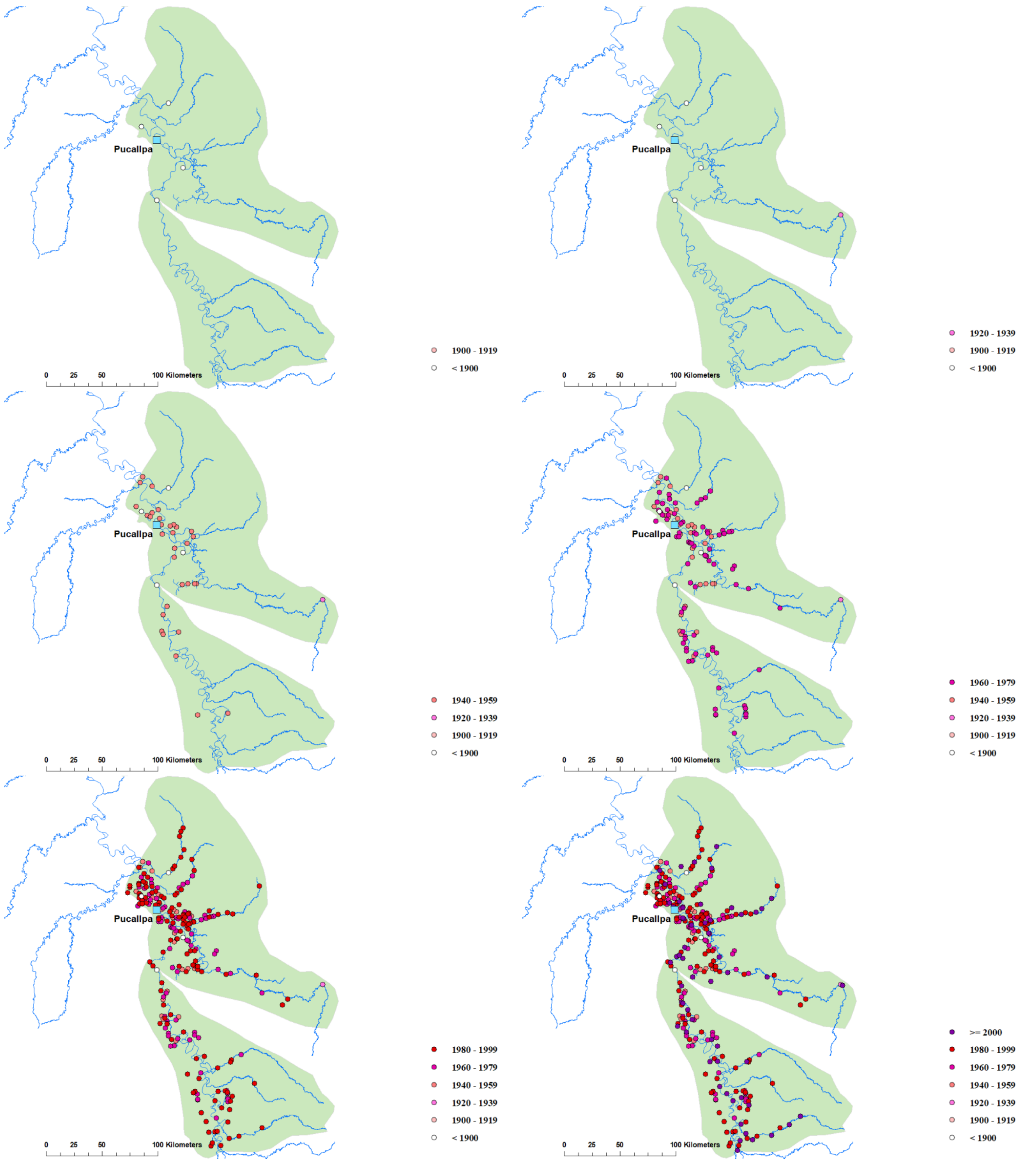


Figure A.2: Establishment of Rainforest Communities over Decades in the Upper Ucayali Basin

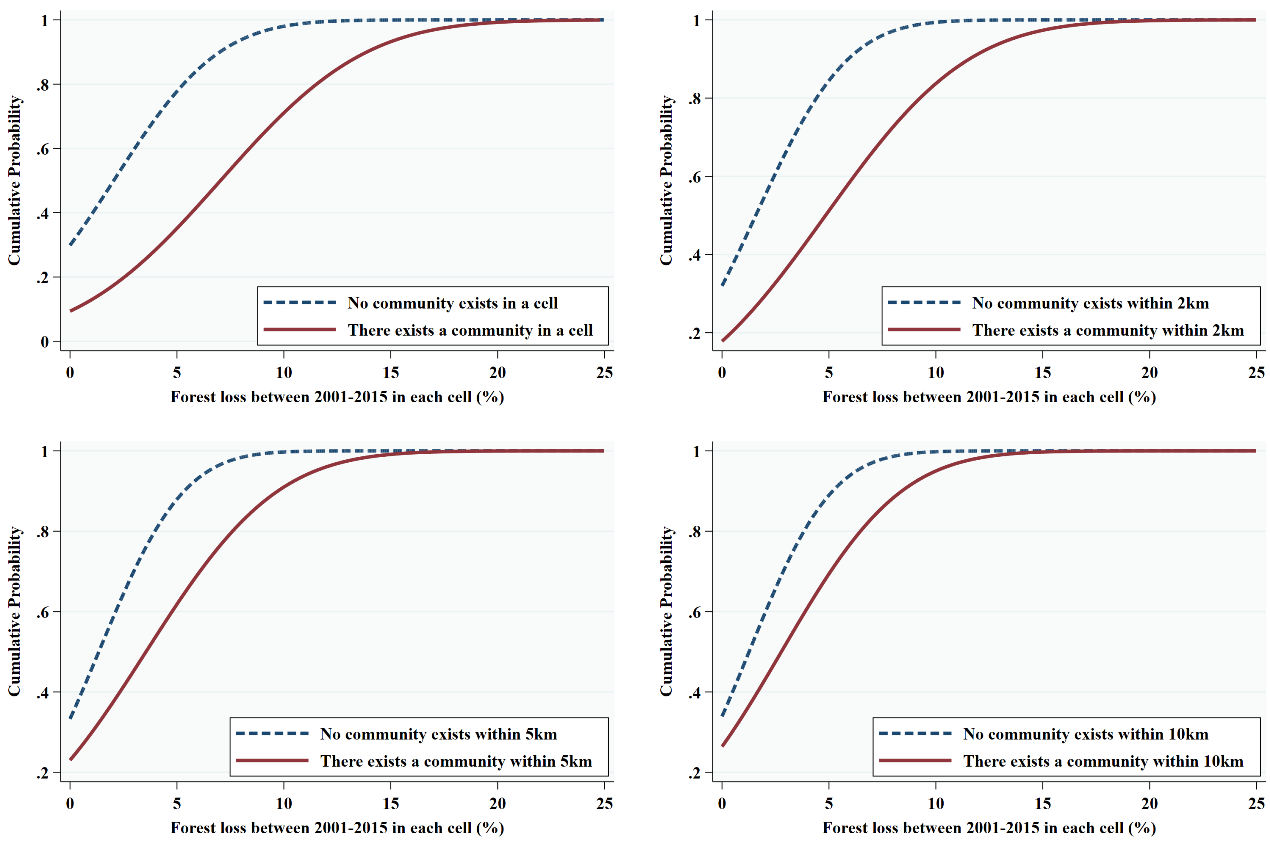


Figure A.3: Communities and Deforestation (2001-2015)

Notes: The unit of analysis is a 1km $\times$ 1km grid cell in the four river basins. The forest loss measure represents a percentage of the deforested area inside each grid cell. The information of community locations is from the Peru Population and Housing Census in 2007.

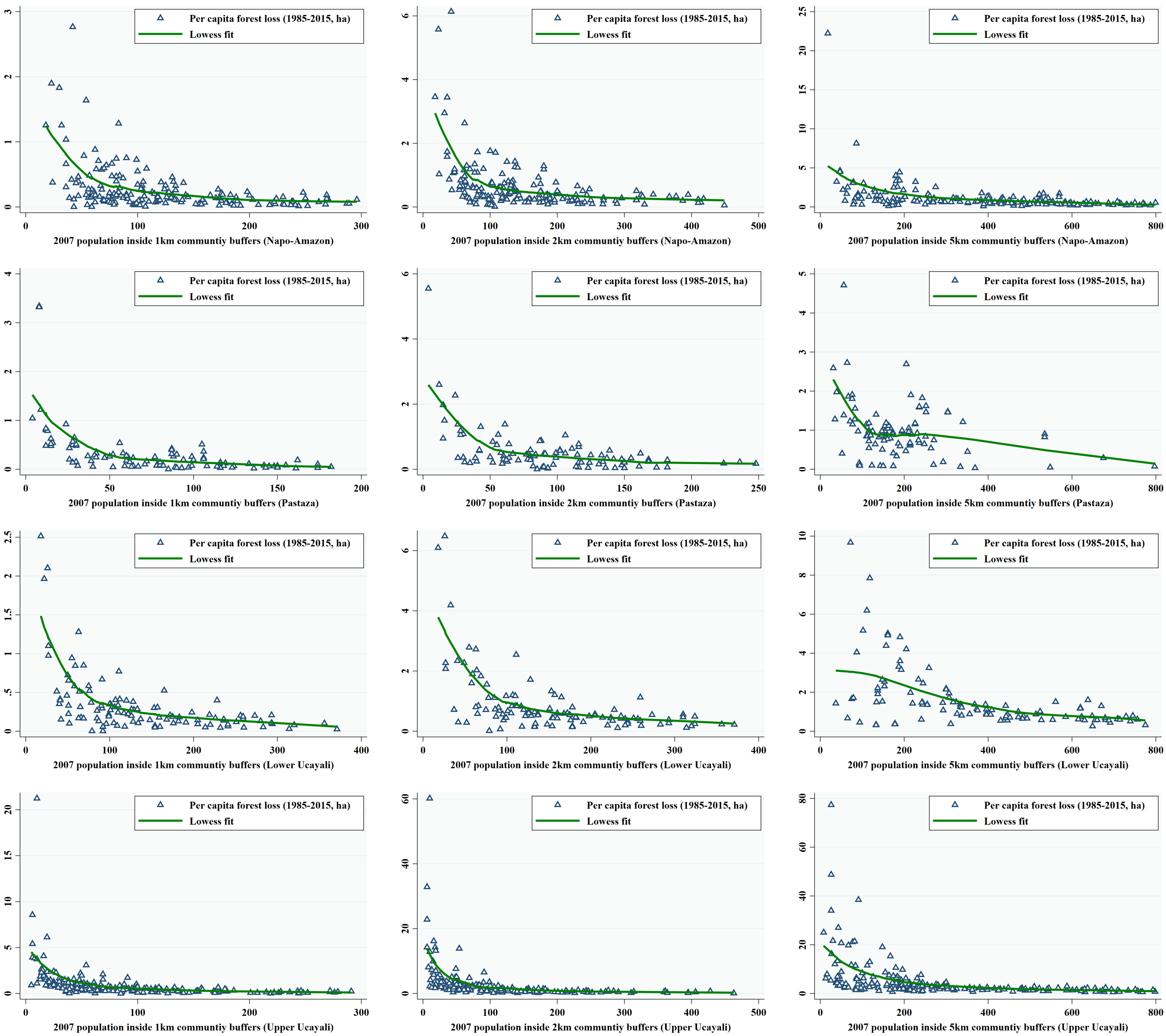


Figure A.4: Populations and Per Capita Deforestation with Different Buffers

Notes: The population information is from the Peru Population and Housing Census in 2007. The total population in the $x$km buffer surrounding a community is measured by summing populations from communities whose centroid are inside the buffer.

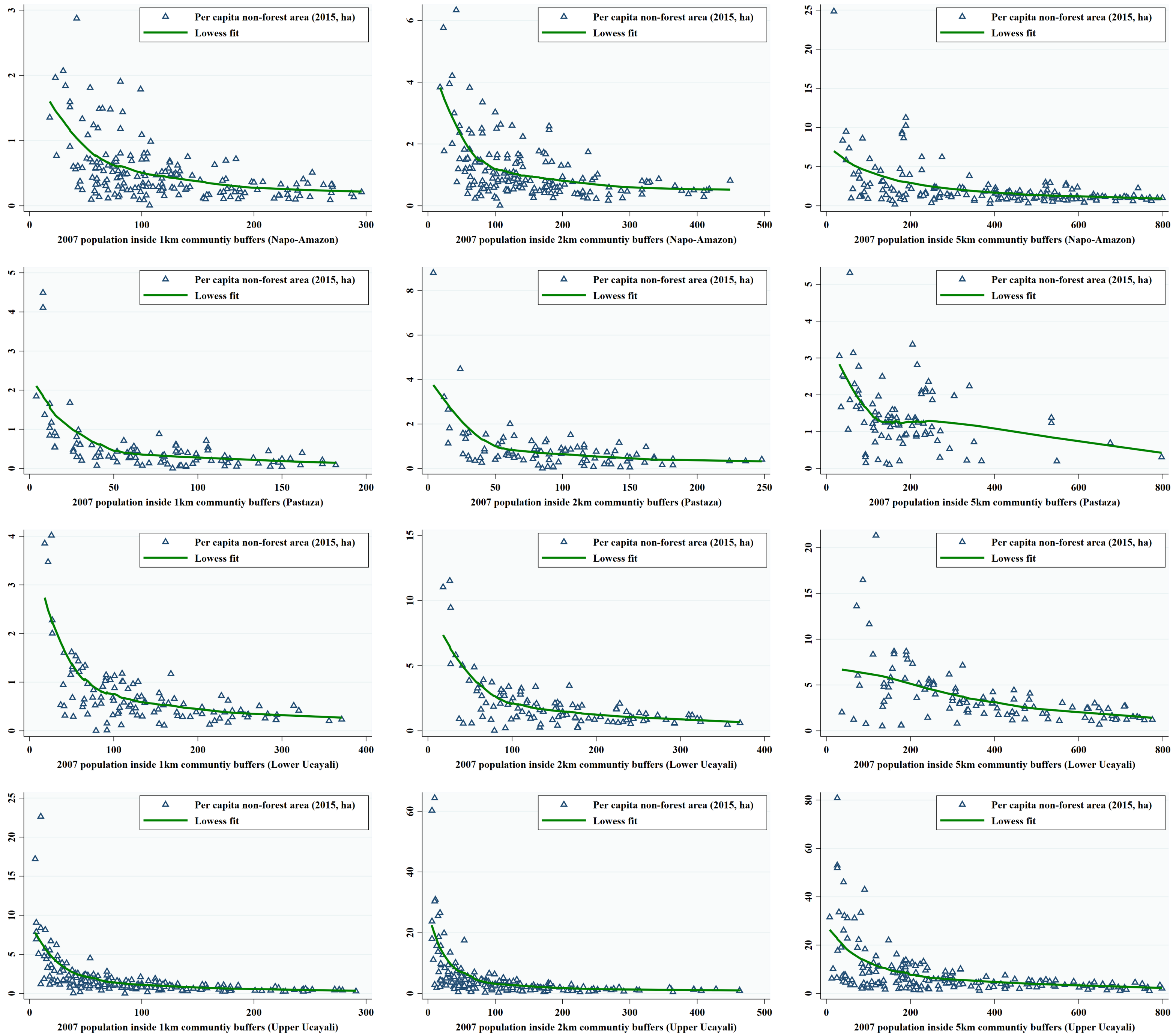


Figure A.5: Populations and Per Capita Non-Forest Area with Different Buffers

Notes: The population information is from the Peru Population and Housing Census in 2007. The total population in the *x*km buffer surrounding a community is measured by summing populations from communities whose centroid are inside the buffer.

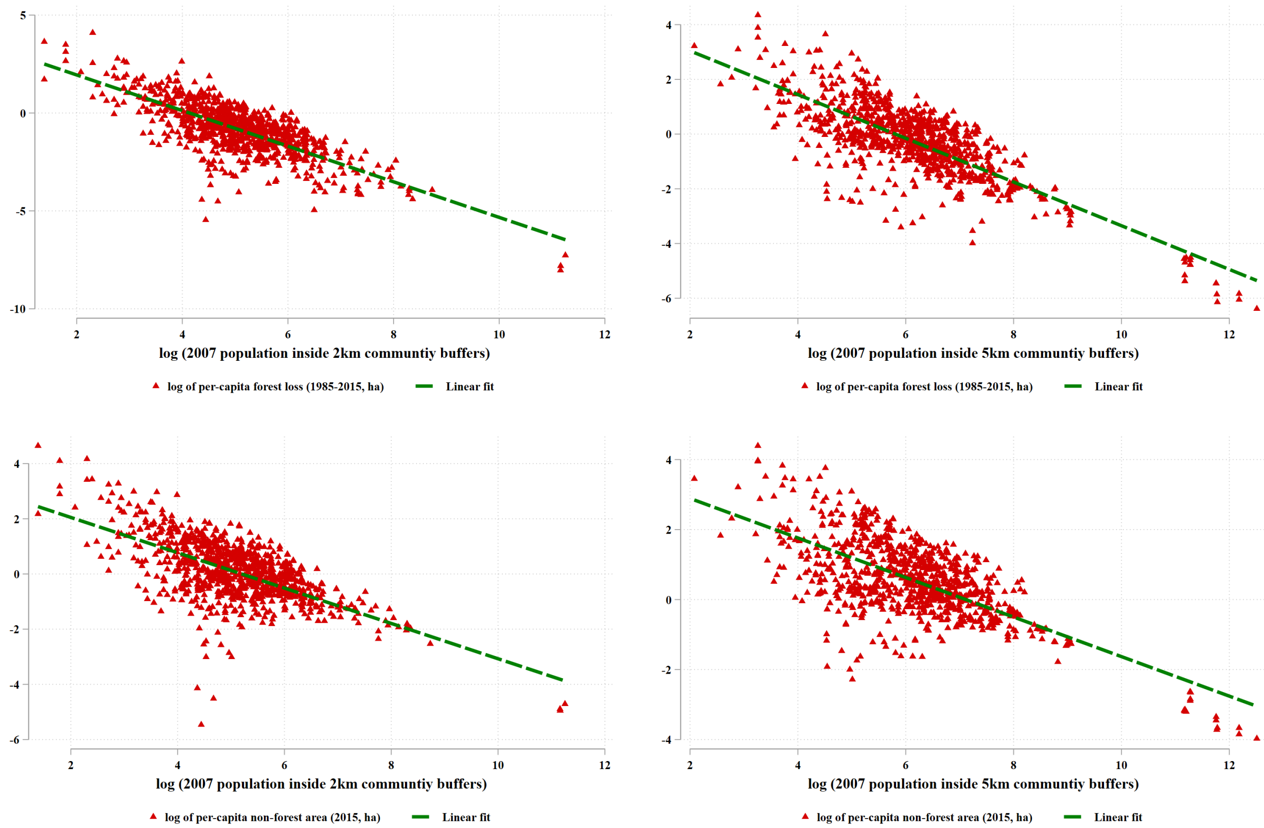


Figure A.6: Populations and Per Capita Deforestation

Notes: The population information is from the Peru Population and Housing Census in 2007. The total population in the 2 km and 5 km buffers surrounding a community is measured by summing populations from communities whose centroid are inside the buffers.

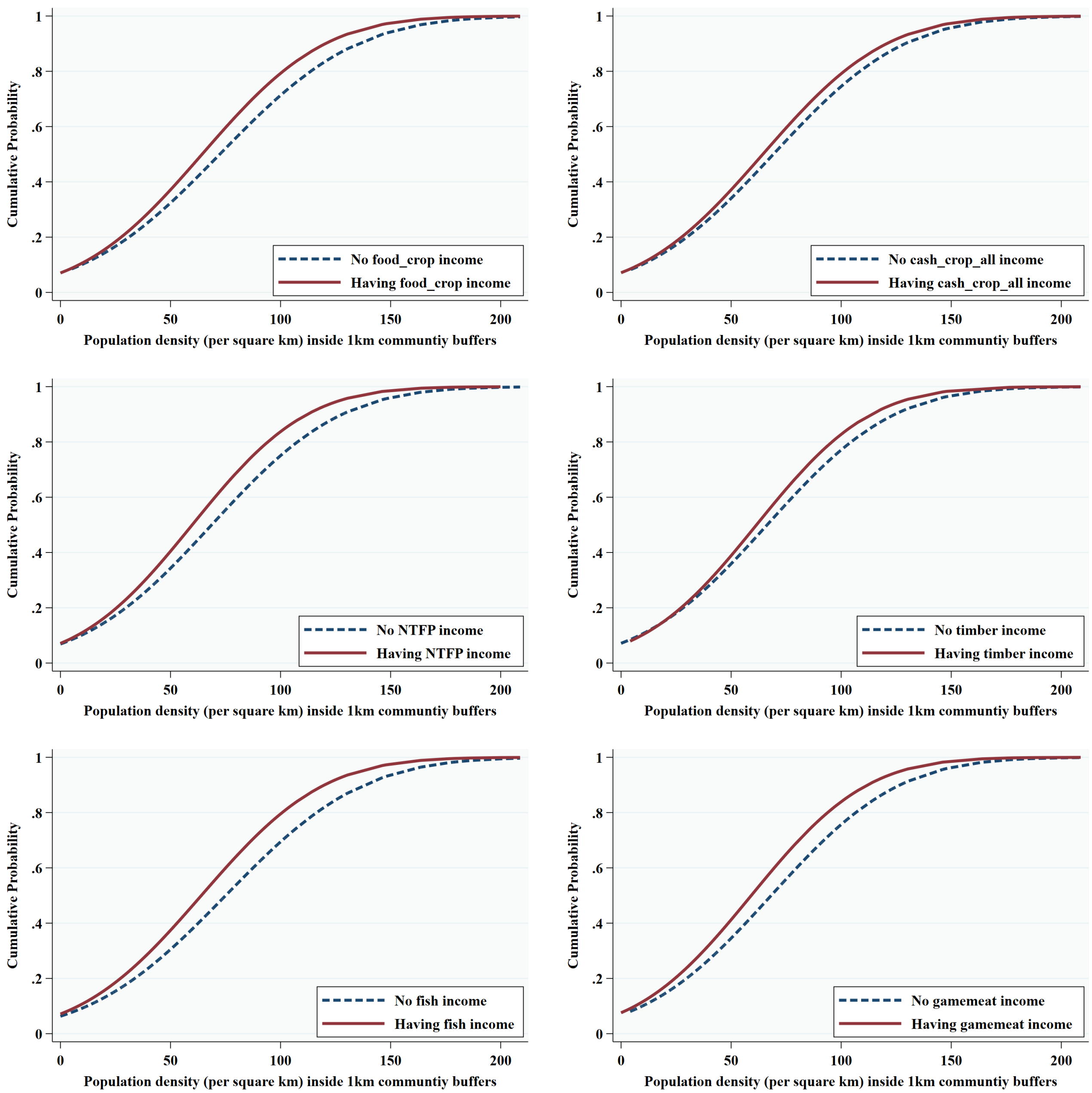


Figure A.7: Population Density (1km buffer) and Distribution of Activities (Household-level)—Agriculture vs. Forest Products vs. Wild Animal Extractions

Notes: The population information is from the PARLAP community census. We divide the total population in the $x$km buffer by the area of buffer ($x^2\pi$) to calculate the population density in the $x$km buffer. The total population in the $x$km buffer surrounding a community is measured by summing populations from communities whose centroid are inside the buffer.

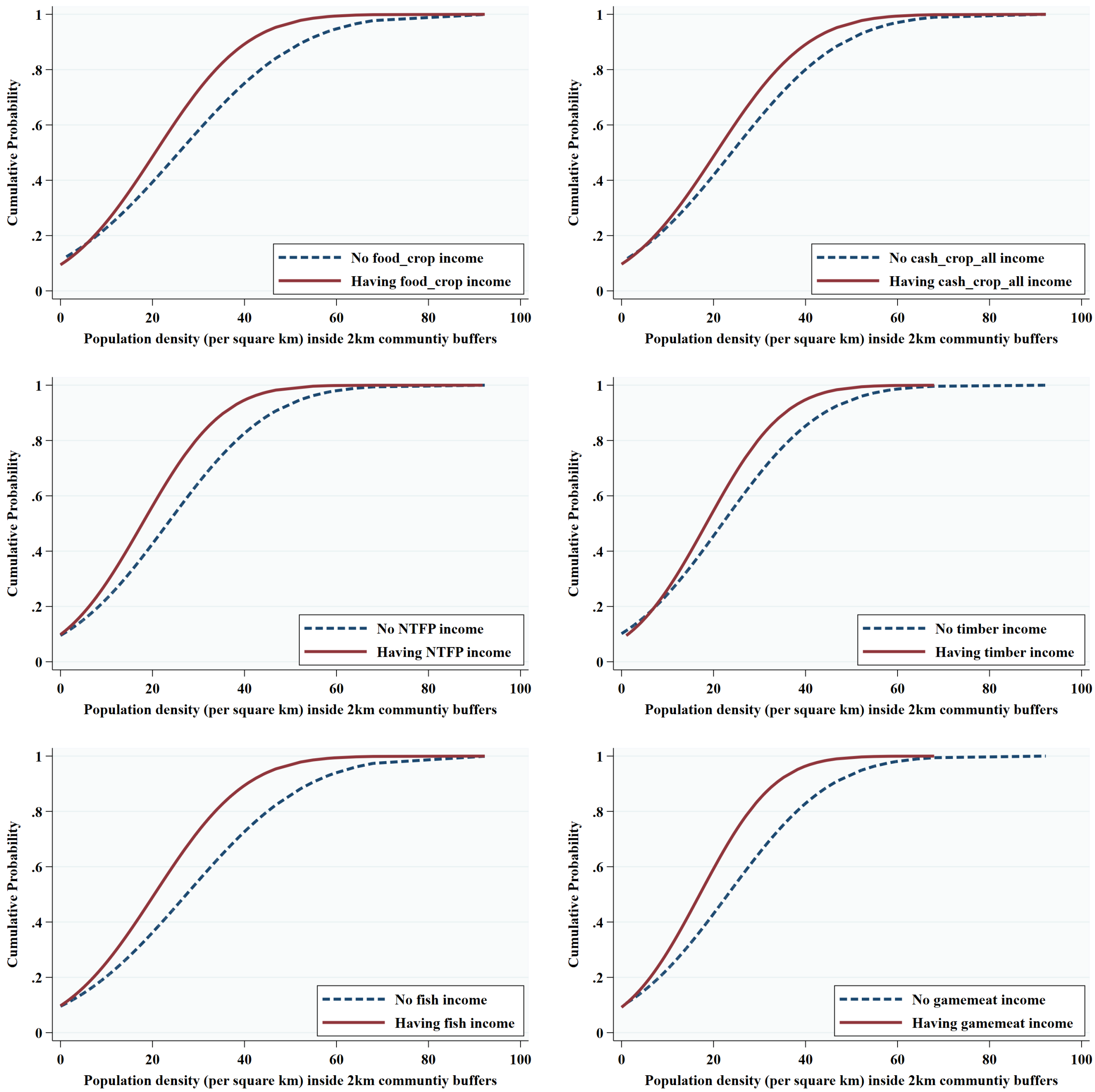


Figure A.8: Population Density (2km buffer) and Distribution of Activities (Household-level)—Agriculture vs. Forest Products vs. Wild Animal Extractions

Notes: The population information is from the PARLAP community census. We divide the total population in the $x$km buffer by the area of buffer ($x^2\pi$) to calculate the population density in the $x$km buffer. The total population in the $x$km buffer surrounding a community is measured by summing populations from communities whose centroid are inside the buffer.

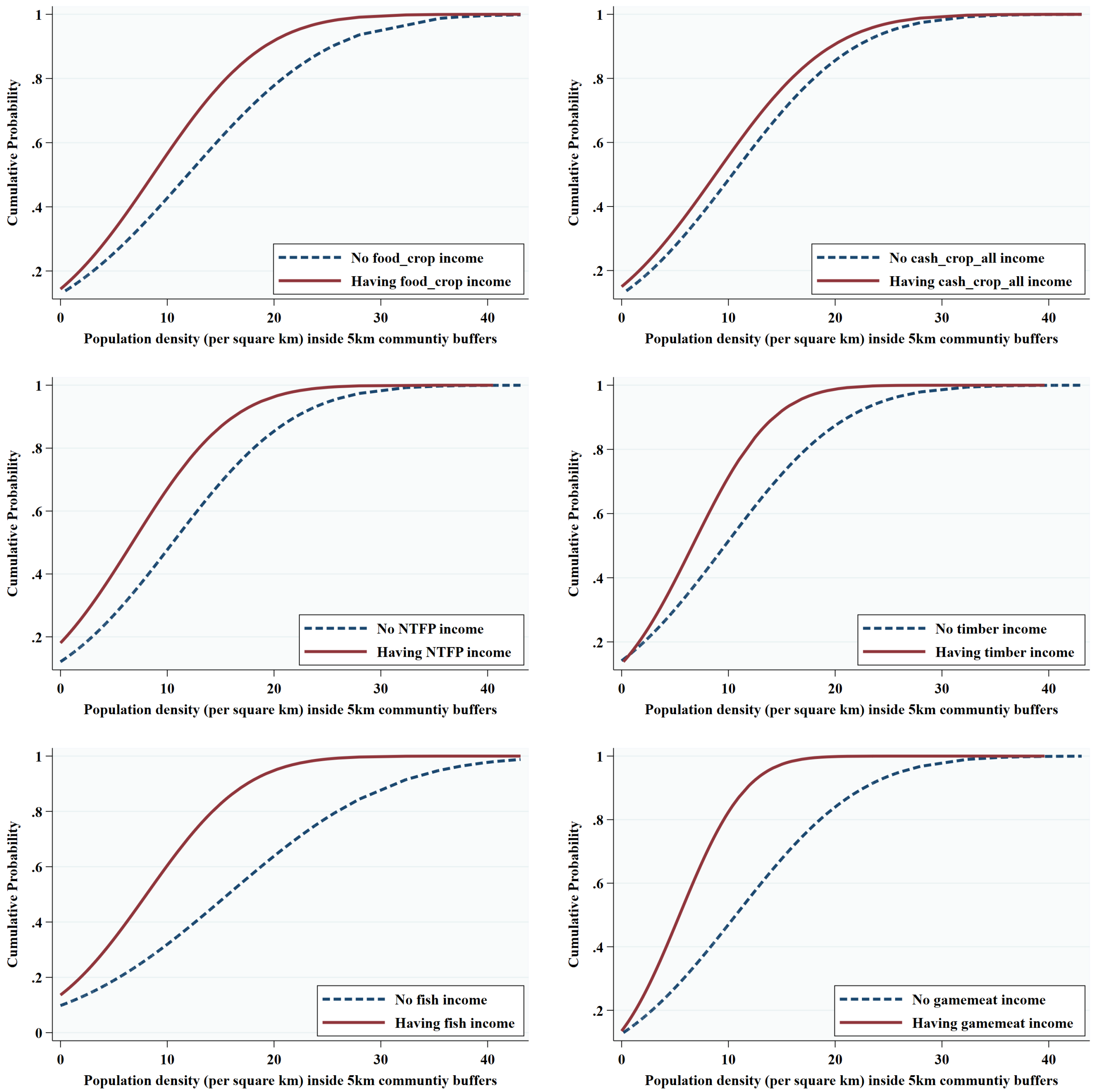


Figure A.9: Population Density (5km buffer) and Distribution of Activities (Household-level)—Agriculture vs. Forest Products vs. Wild Animal Extractions

Notes: The population information is from the PARLAP community census. We divide the total population in the $x$km buffer by the area of buffer ($x^2\pi$) to calculate the population density in the $x$km buffer. The total population in the $x$km buffer surrounding a community is measured by summing populations from communities whose centroid are inside the buffer.

Figure A.10: Transport Modes in the Peruvian Amazon

Notes: Canue (top-left) is the most traditional transport mode. Peque-peque (top-right) is the most widely-available transport mode with engine. Rapido (bottom-left) is an express motor boat. Lancha (bottom-right) is the largest boat type to carry people and cargo. The latter two types are faster, but they are not commonly available in all communities.

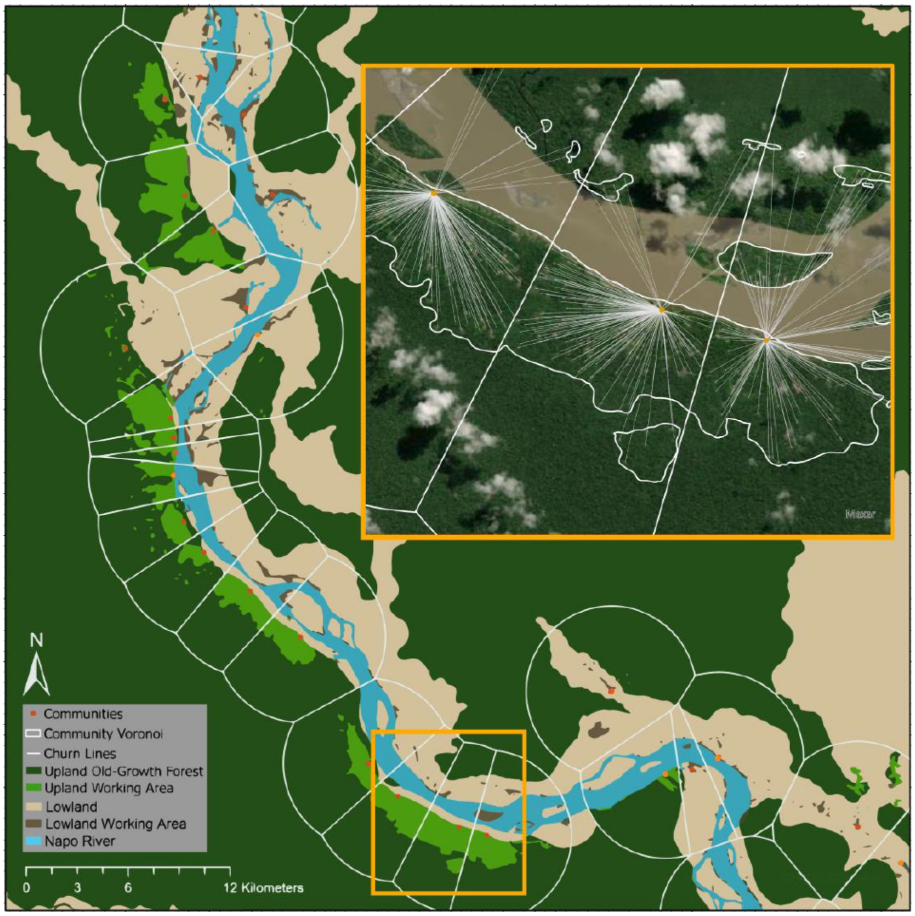


Figure A.11: Voronoi Polygons and Working Areas around the Census Communities

Notes: To proxy community boundaries for agricultural land use, we partition land in the study area into voronoi polygons. We define community boundaries as being up to 5 km from the centroid of the communities in the CC data. Within each community voronoi polygon, we detect all patches of agricultural fields and secondary forests through satellite images. We then sum them up to calculate the working area (land footprint) of each community. See Coomes et al. (2021) for more details. This community-level land footprint information is used for estimating the congestion externality in forest clearing.

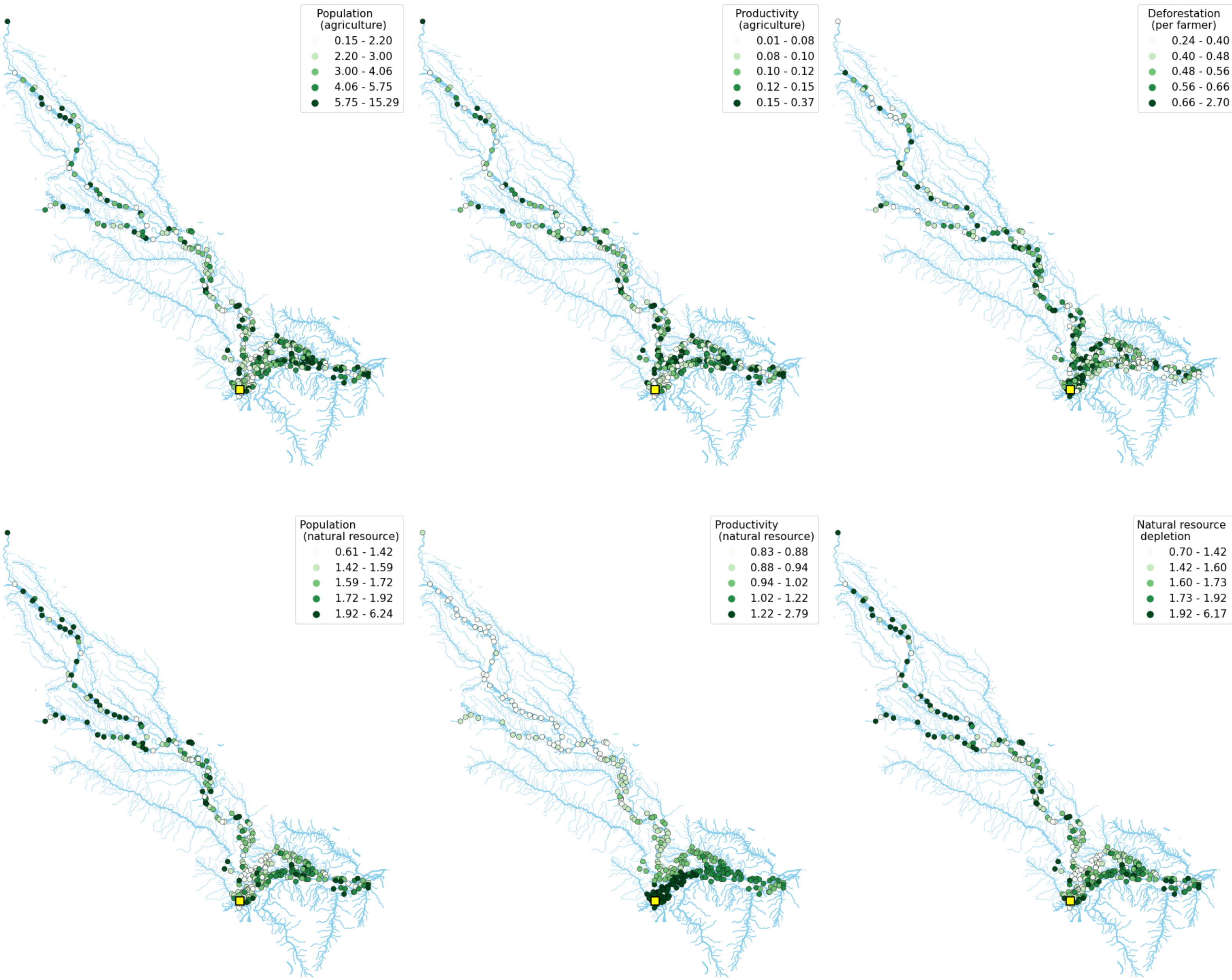


Figure A.12: Counterfactual Outcomes without the Agglomeration Externality (Napo-Amazon)

Notes: These maps present the counterfactual outcomes of shutting down the agglomeration externality in agriculture. Values shown in the legend of the circle dots are relative values in the counterfactual scenarios in terms of those in the benchmark spatial equilibrium in rural locations. The legend is based on quantiles. The yellow square represents the urban center.

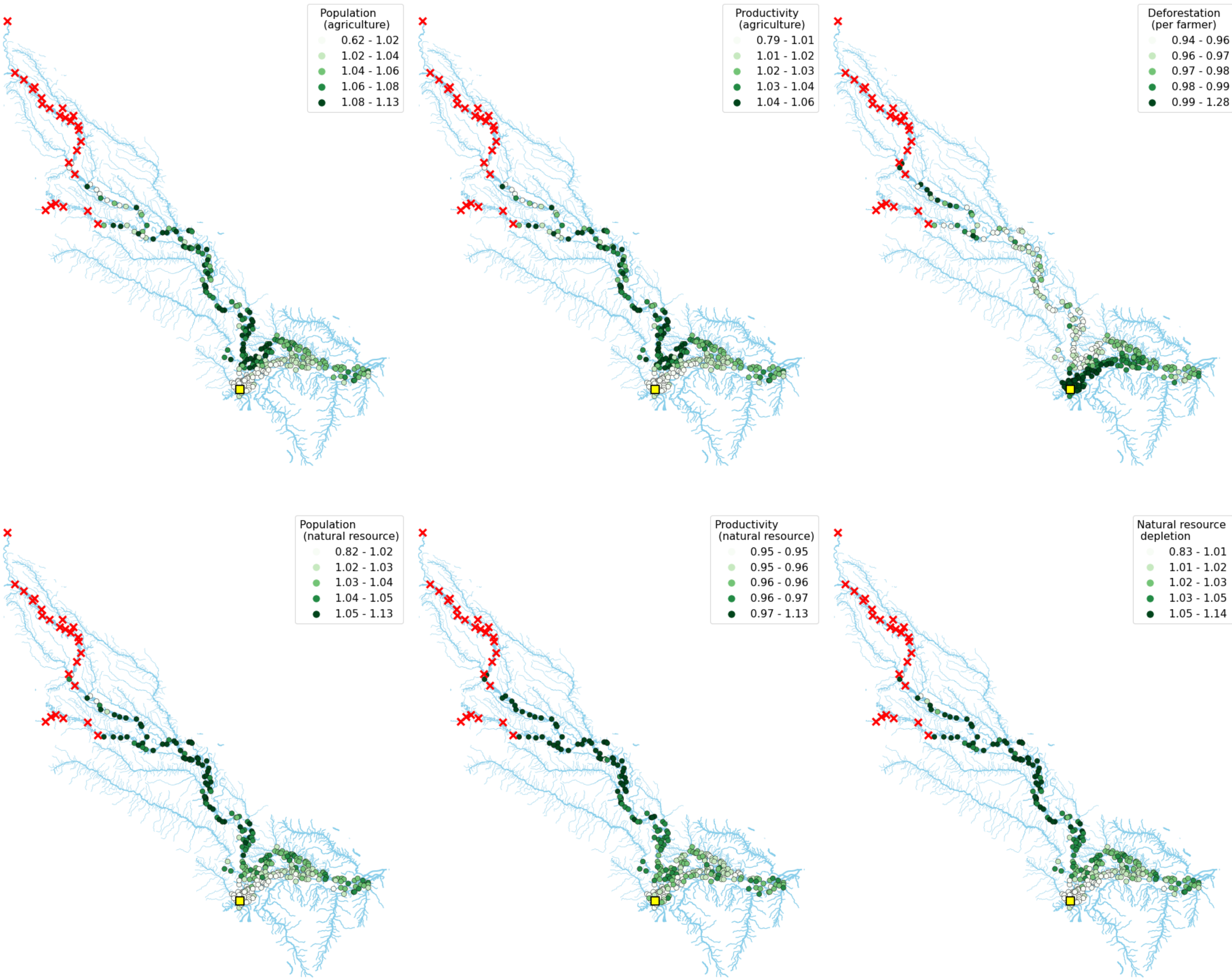


Figure A.13: Counterfactual Outcomes of Protecting the Rural Frontier (Napo-Amazon)

Notes: These maps present the counterfactual outcomes of a place-based protection policy of controlling the expansion of rural frontier. This experiment chooses rural locations to be treated in order, starting with those farthest from the urban center, until the treated population reaches 2.5% of the total rural population in each basin. The red **x** marks indicate the treated locations. Values shown in the legend of the circle dots are relative values in the counterfactual scenarios in terms of those in the benchmark spatial equilibrium in rural locations. The legend is based on quantiles. The yellow square represents the urban center.

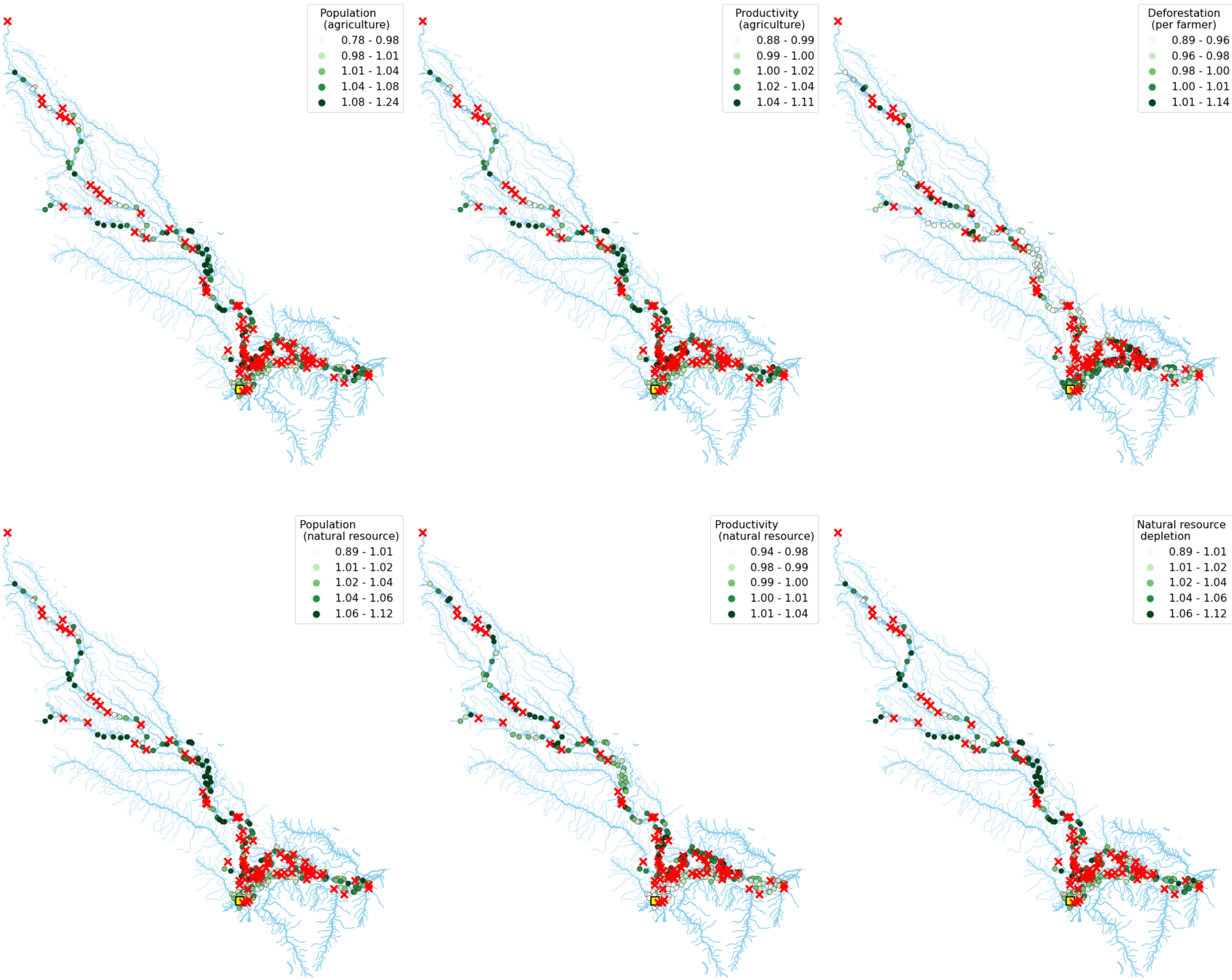


Figure A.14: Counterfactual Outcomes without Small Communities (Napo-Amazon)

Notes: These maps present the counterfactual outcomes of a place-based protection policy that targets the smallest communities. This experiment chooses rural locations to be treated in order, starting with the location with the smallest population size, until the treated population reaches 2.5% of the total population in each basin. The red **x** marks indicate the treated locations. Values shown in the legend of the circle dots are relative values in the counterfactual scenarios in terms of those in the benchmark spatial equilibrium in rural locations. The legend is based on quantiles. The yellow square represents the urban center.

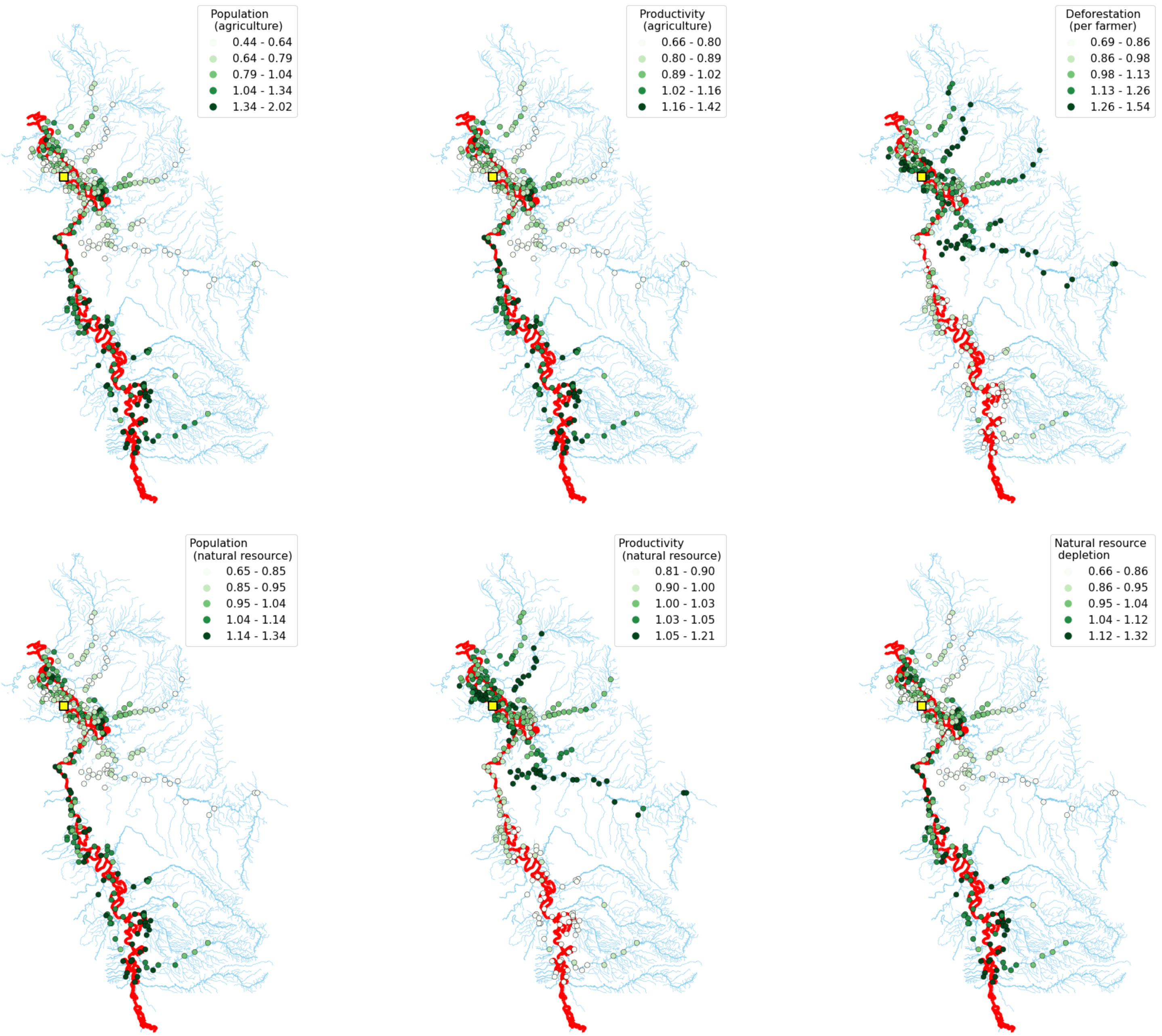


Figure A.15: Counterfactual Outcomes with Improved Transport Infrastructure along River Order 2 (Upper Ucayali)

Notes: These maps present the counterfactual outcomes of improving the transport infrastructure in a way that hinterlands are connected to the central area of the basin. The light blue lines indicate the river lines without the improvement of transport infrastructure. The red lines indicate the river lines where the transport infrastructure is upgraded. Values shown in the legend of the circle dots are relative values in the counterfactual scenarios in terms of those in the benchmark spatial equilibrium in rural locations. The legend is based on quantiles. The yellow square represents the urban center.

Table A.1: Natural Resource Endowments, Calibrated Productivity, and Community Land Footprint

(A)	Number of species found around a community				
	(1)	(2)	(3)	(4)	(5)
	Total	Fish	Timber	NTFP	Game
$\log(\tilde{A}_{o,Nr})$ (calibrated)	0.206***	0.0220	0.386***	0.0488**	0.380***
	(0.0306)	(0.0334)	(0.0407)	(0.0204)	(0.0437)
Mean (Dep. var.)	2.025	3.161	1.788	0.552	1.958
SD (Dep. var.)	1.145	1.163	1.676	0.893	1.636
R^2	0.059	0.150	0.126	0.349	0.213
Observations	909	909	909	909	909
(B)	Number of species found around a community				
	(1)	(2)	(3)	(4)	(5)
	Total	Fish	Timber	NTFP	Game
log (land footprint)	0.0171	-0.0653*	-0.0209	0.0126	0.0752
	(0.0381)	(0.0383)	(0.0533)	(0.0245)	(0.0478)
Mean (Dep. var.)	2.021	3.147	1.796	0.555	1.956
SD (Dep. var.)	1.147	1.183	1.677	0.894	1.637
R^2	0.014	0.163	0.059	0.336	0.146
Observations	906	906	906	906	906
(C)	Number of species found around a community				
	(1)	(2)	(3)	(4)	(5)
	Total	Fish	Timber	NTFP	Game
log (depth of land footprint)	0.0458	-0.0358	-0.135*	-0.0248	0.0911
	(0.0587)	(0.0541)	(0.0771)	(0.0342)	(0.0728)
Mean (Dep. var.)	2.070	3.168	1.873	0.550	1.964
SD (Dep. var.)	1.141	1.159	1.676	0.885	1.626
R^2	0.026	0.170	0.059	0.364	0.146
Observations	811	811	811	811	811
Basin FE	Yes	Yes	Yes	Yes	Yes
Geographic controls	No	No	No	No	No

Notes: Robust standard errors in parentheses. The unit of anaylysis is a community in the PARLAP Community Census (CC) in 2014. In panel (B), the land footprint represents the community-level land footprint within a voronoi polygon around the settlement, detected in satellite images. In panel (C), the land footprint depth represents the distance from the river to the furthest inland point in the community-level land footprint.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.2: River Networks, Initial Communities, and Current Populations

	$\log(N_{o,Ag})$			Community existence (1940)	
	(1)	(2)	(3)	(4)	(5)
$\log(RNA_o)$	0.758***		0.711***	-0.0145	0.0699
	(0.223)		(0.218)	(0.0254)	(0.0726)
Community existence (1940)		0.740***	0.730***		
		(0.0983)	(0.0980)		
Basin FE	Yes	Yes	Yes	Yes	Yes
Geographic controls	Yes	Yes	Yes	No	Yes
Mean (Dep. var.)	4.322	4.322	4.322	0.194	0.194
SD (Dep. var.)	1.192	1.192	1.192	0.395	0.395
R^2	0.154	0.195	0.206	0.094	0.117
Observations	893	893	893	904	899

Notes: Robust standard errors in parentheses. The sample includes 1 square km grid cells that have positive populations. Geographical controls include a dummy of high river orders (4 and 5), distance to the urban center, distance to the river, squared distance to the river, interaction terms of these two variables with a river cell dummy, elevation, river confluences, flood vulnerability, geology measures, and open water access measures.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.3: Correlations between Instrumental Variables and Geographic Fundamentals

(A) River Network Access						
	(1)	(2)	(3)	(4)	(5)	(6)
	Water share: non-main channel	River confluence: 1st×2nd or 2nd×3rd	River confluence: 3rd×4th	Flood vulnerability	Pleistocene soil share	Tertiary soil share
$\log(RNA_o)$	0.00751	0.0470	-0.0743	-0.217	-0.0444	-0.0498
	(0.0161)	(0.0470)	(0.0618)	(0.307)	(0.0362)	(0.0474)
Mean (Dep. var.)	0.030	0.077	0.083	1.606	0.021	0.211
SD (Dep. var.)	0.087	0.266	0.277	1.606	0.115	0.344
R^2	0.068	0.095	0.137	0.130	0.057	0.735
Observations	899	899	899	899	899	899
(B) Early human settlements						
	(1)	(2)	(3)	(4)	(5)	(6)
	Water share: main channel	Water share: non-main channel	Flood vulnerability	Floodplain soil share	Pleistocene soil share	Tertiary soil share
Community existence (1940)	0.0263	-0.00352	0.218	0.00191	0.00777	-0.0108
	(0.0193)	(0.00698)	(0.142)	(0.0268)	(0.0122)	(0.0257)
Mean (Dep. var.)	0.109	0.030	1.606	0.584	0.021	0.211
SD (Dep. var.)	0.203	0.087	1.606	0.359	0.115	0.344
R^2	0.162	0.038	0.116	0.250	0.030	0.243
Observations	899	899	899	899	899	899
Basin FE	Yes	Yes	Yes	Yes	Yes	Yes
Geographic controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Robust standard errors in parentheses. The sample includes 1 square km grid cells that have positive populations. In panel (A), geographical controls include a dummy of high river orders (4 and 5), distance to the urban center, distance to the river, squared distance to the river, interaction terms of these two variables with a river cell dummy, elevation, water share of main channel rivers, and floodplain soil share. In panel (B), geographical controls include a dummy of high river orders (4 and 5), distance to the urban center, distance to the river, squared distance to the river, interaction terms of these two variables with a river cell dummy, elevation, and river confluences.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.4: Agglomeration Externality in Agriculture

	The calibrated value of $\log(\tilde{A}_{o,Ag})$							
	All locations				$N_o < 1000$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\log(N_{o,Ag})$	0.676***	0.440**	0.514***	0.501***	0.735***	0.384**	0.509***	0.464***
	(0.0207)	(0.171)	(0.0809)	(0.0790)	(0.0196)	(0.169)	(0.124)	(0.109)
Basin FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
IV: RNA	No	Yes	No	Yes	No	Yes	No	Yes
IV: Historical	No	No	Yes	Yes	No	No	Yes	Yes
Mean (Dep. var.)	-0.096	-0.096	-0.096	-0.096	-0.172	-0.172	-0.172	-0.172
SD (Dep. var.)	4.578	4.578	4.578	4.578	4.614	4.614	4.614	4.614
First stage F-stat		11.502	56.653	31.005		15.298	35.632	22.822
Hansen's J test p-value				0.648				0.472
Observations	893	893	893	893	852	852	852	852

Notes: Robust standard errors in parentheses. The sample includes 1 square km grid cells that have positive populations. We use $\log(RNA_o)$ (IV: RNA) and the initial community existence in 1940 (IV: Historical) as instruments for $\log(N_{o,Ag})$. Geographical controls include a dummy of high river orders (4 and 5), distance to the urban center, distance to the river, squared distance to the river, interaction terms of these two variables with a river cell dummy, elevation, river confluences, flood vulnerability, geology measures, and open water access measures.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.5: Agglomeration Externality in Agriculture

	The calibrated value of $\log(\tilde{A}_{o,Ag})$						
	IV						OLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\log(N_{o,Ag})$	0.434***	0.519***	0.519***	0.521***	0.509***	0.501***	0.676***
	(0.0920)	(0.0789)	(0.0788)	(0.0789)	(0.0789)	(0.0790)	(0.0207)
log (Elevation)		2.341***	2.354***	2.324***	2.360***	2.397***	2.252***
		(0.171)	(0.175)	(0.179)	(0.176)	(0.177)	(0.176)
River confluence (1st×2nd or 2nd×3rd)			0.0155	0.0186	0.0206	0.0309	0.0180
			(0.0958)	(0.0964)	(0.0982)	(0.0995)	(0.0969)
River confluence (3rd×4th)			-0.0356	-0.0339	-0.0246	-0.0266	0.0173
			(0.0724)	(0.0723)	(0.0730)	(0.0733)	(0.0618)
Flood vulnerability (1-4)				-0.0115	-0.00947	-0.0123	-0.0154
				(0.0136)	(0.0137)	(0.0137)	(0.0130)
Water share: non-main channel					0.0806	0.123	-0.00146
					(0.238)	(0.238)	(0.203)
Water share: main channel					0.161	0.185	0.189
					(0.122)	(0.121)	(0.120)
Floodplain soil share						0.127**	0.126**
						(0.0625)	(0.0575)
Pleistocene soil share						0.175	0.333
						(0.222)	(0.227)
Basin FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean (Dep. var.)	-0.094	-0.096	-0.096	-0.096	-0.096	-0.096	-0.096
SD (Dep. var.)	4.576	4.578	4.578	4.578	4.578	4.578	4.578
First stage F-stat	28.030	29.419	29.974	29.634	30.770	31.005	
Observations	894	893	893	893	893	893	893

Notes: Robust standard errors in parentheses. The sample includes 1 square km grid cells that have positive populations. We use $\log(RNA_o)$ and the initial community existence in 1940 as instruments for $\log(N_{o,Ag})$. Other controls include distance to the urban center, distance to the river, squared distance to the river, and interaction terms of these two variables with a river cell dummy.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.6: Density Externality in Forest Clearing

	log (per capita land footprint)					
	All locations		$N_o < 1000$		$N_o < 500$	
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	IV	OLS	IV	OLS	IV
$\log(N_{o,Ag})$	-0.650***	-0.522***	-0.654***	-0.552***	-0.674***	-0.545***
	(0.0307)	(0.0940)	(0.0323)	(0.109)	(0.0346)	(0.123)
Basin FE	Yes	Yes	Yes	Yes	Yes	Yes
Geographic controls	Yes	Yes	Yes	Yes	Yes	Yes
Mean (Dep. var.)	0.929	0.929	0.956	0.956	0.981	0.981
SD (Dep. var.)	1.231	1.231	1.218	1.218	1.223	1.223
First stage F-stat		34.198		28.141		23.709
Hansen's J test p-value		0.987		0.896		0.969
Observations	895	895	878	878	847	847

Notes: Robust standard errors in parentheses. The unit of anaylysis is a community in the PARLAP Community Census (CC) in 2014. We use $\log(RNA_o)$ and the initial community existence in 1940 as instruments for $\log(N_{o,Ag})$. Geographical controls include a dummy of high river orders (4 and 5), distance to the urban center, distance to the river, squared distance to the river, interaction terms of these two variables with a river cell dummy, elevation, river confluence, flood vulnerability, geology measures, and open water access measures for a grid cell where each census community belongs.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.7: Congestion Externality in Natural Resource Extraction with Spatial Spillovers

	The calibrated value of $\log(\tilde{A}_{o,Nr})$									
	IV									OLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$\log(N_{o,Nr})$	-2.127**	-1.385	-0.879	-0.581	0.0960	0.278	0.573***	0.634***	0.606***	0.950***
	(1.075)	(0.939)	(0.688)	(0.558)	(0.322)	(0.280)	(0.208)	(0.179)	(0.184)	(0.0516)
$\log(\sum_{d\mid D_{o,d}\leq 2km} N_{d,Nr})$		-0.573*	-0.0343	0.0235	-0.0745	-0.0705	-0.0648	-0.0745	-0.0856	-0.0663
		(0.331)	(0.282)	(0.236)	(0.143)	(0.122)	(0.0883)	(0.0822)	(0.0838)	(0.0611)
$\log(\sum_{d\mid D_{o,d}\leq 5km} N_{d,Nr})$			-0.596***	-0.286	-0.138	-0.130	-0.120*	-0.111*	-0.112*	-0.132***
			(0.189)	(0.183)	(0.106)	(0.0888)	(0.0637)	(0.0597)	(0.0613)	(0.0425)
$\log(\sum_{d\mid D_{o,d}\leq 10km} N_{d,Nr})$				-0.337**	0.0345	0.0364	0.0579	0.0425	0.0322	0.0140
				(0.141)	(0.107)	(0.0885)	(0.0625)	(0.0596)	(0.0639)	(0.0378)
$\log(\sum_{d\mid D_{o,d}\leq 25km} N_{d,Nr})$					-0.470***	-0.357***	-0.327***	-0.294***	-0.285***	-0.165***
					(0.0918)	(0.0837)	(0.0584)	(0.0560)	(0.0571)	(0.0283)
$\log(\sum_{d\mid D_{o,d}\leq 50km} N_{d,Nr})$						-0.195***	-0.0318	-0.0548	-0.0480	-0.0619**
						(0.0610)	(0.0581)	(0.0526)	(0.0548)	(0.0242)
$\log(\sum_{d\mid D_{o,d}\leq 75km} N_{d,Nr})$							-0.280***	-0.0758	-0.0407	-0.0989***
							(0.0779)	(0.125)	(0.142)	(0.0352)
$\log(\sum_{d\mid D_{o,d}\leq 100km} N_{d,Nr})$								-0.258*	-0.439*	-0.263***
								(0.141)	(0.231)	(0.0498)
$\log(\sum_{d\mid D_{o,d}\leq 150km} N_{d,Nr})$									0.187	0.0970*
									(0.171)	(0.0567)
Basin FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean (Dep. Var.)	0.337	0.337	0.337	0.337	0.337	0.337	0.337	0.337	0.337	0.337
SD (Dep. Var.)	2.862	2.862	2.862	2.862	2.862	2.862	2.862	2.862	2.862	2.862
Observations	894	894	894	894	894	894	894	894	894	894

Notes: Robust standard errors in parentheses. The sample includes 1 square km grid cells that have positive populations. We use $\ln RNA_o$ and $\{\ln \sum_{d\mid D_{o,d}\leq x} RNA_d\}$ for $x \in \mathcal{X}$ as instruments when endogenous variables include $\log(N_{o,Nr})$ and $\{\ln \sum_{d\mid D_{o,d}\leq x} N_{d,Nr}\}$ for $x \in \mathcal{X}$. Geographical controls include a dummy of high river orders (4 and 5), distance to the urban center, distance to the river, squared distance to the river, interaction terms of these two variables with a river cell dummy, elevation, river confluence, flood vulnerability, geology measures, and open water access measures.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.8: Community Population and Trade Environment

	Availability of a river trader		Community population being contracted		Contractors living in the community	
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	IV	OLS	IV	OLS	IV
$\log(N_{o,Ag})$	0.0102	0.0430*	0.0412***	0.0988**	0.0489***	0.118***
	(0.00693)	(0.0244)	(0.0125)	(0.0395)	(0.0101)	(0.0350)
Basin FE	Yes	Yes	Yes	Yes	Yes	Yes
Geographic controls	Yes	Yes	Yes	Yes	Yes	Yes
Mean (Dep. var.)	0.031	0.031	0.182	0.182	0.056	0.056
SD (Dep. var.)	0.173	0.173	0.386	0.386	0.231	0.231
First stage F-stat		24.84462		26.12503		23.28224
Observations	906	906	891	891	853	853

Notes: Robust standard errors in parentheses. The unit of anaylysis is a community in the PARLAP Community Census (CC) in 2014. We use $\log(RNA_o)$ and the initial community existence in 1940 as instruments for $\log(N_{o,Ag})$. Geographical controls include a dummy of high river orders (4 and 5), distance to the urban center, distance to the river, squared distance to the river, interaction terms of these two variables with a river cell dummy, elevation, river confluence, flood vulnerability, geology measures, and open water access measures for a grid cell where each census community belongs.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.9: Community Population and Contracts for Trading Products

	Community population being contracted for							
	Maize		Rice		Fish		Timber	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
$\log(N_{o,Ag})$	0.0120*	0.0599**	0.0275***	0.0254				
	(0.00637)	(0.0300)	(0.00780)	(0.0164)				
$\log(N_{o,Nr})$					0.0282***	0.0283	0.0325***	-0.248*
					(0.00788)	(0.0440)	(0.00967)	(0.138)
Basin FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean (Dep. var.)	0.034	0.034	0.020	0.020	0.029	0.029	0.112	0.112
SD (Dep. var.)	0.180	0.180	0.141	0.141	0.168	0.168	0.316	0.316
First stage F-stat		14.59049		14.59049		8.799949		8.799949
Observations	891	891	891	891	892	892	892	892

Notes: Robust standard errors in parentheses. The unit of anaylysis is a community in the PARLAP Community Census (CC) in 2014. We use $\log(RNA_o)$ and the initial community existence in 1940 as instruments for $\log(N_{o,Ag})$. Geographical controls include a dummy of high river orders (4 and 5), distance to the urban center, distance to the river, squared distance to the river, interaction terms of these two variables with a river cell dummy, elevation, river confluence, flood vulnerability, geology measures, and open water access measures for a grid cell where each census community belongs.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.10: Community Population and Form of Crop Seed Acquisition

(A)	Obtain crop seeds from			
	Others in the community		City or other communities	
	(1)	(2)	(3)	(4)
	OLS	IV	OLS	IV
$\log(N_{o,Ag})$	0.0104	0.0497	-0.00290	-0.0908*
	(0.0106)	(0.0371)	(0.0150)	(0.0514)
Mean (Dep. var.)	0.859	0.859	0.445	0.445
SD (Dep. var.)	0.348	0.348	0.497	0.497
First stage F-stat		24.875		24.875
Observations	907	907	907	907
(B)	Obtain crop seeds from outside the community via			
	Market transactions		Non-market transactions	
	(1)	(2)	(3)	(4)
	OLS	IV	OLS	IV
$\log(N_{o,Ag})$	0.00242	0.0228	0.0385*	-0.0683
	(0.0148)	(0.0420)	(0.0216)	(0.0662)
Mean (Dep. var.)	0.900	0.900	0.286	0.286
SD (Dep. var.)	0.300	0.300	0.452	0.452
First stage F-stat		12.049		12.049
Observations	402	402	402	402
Basin FE	Yes	Yes	Yes	Yes
Geographic controls	Yes	Yes	Yes	Yes

Notes: Robust standard errors in parentheses. The unit of anaylysis is a community in the PARLAP Community Census (CC) in 2014. We use $\log(RNA_o)$ and the initial community existence in 1940 as instruments for $\log(N_{o,Ag})$. Geographical controls include a dummy of high river orders (4 and 5), distance to the urban center, distance to the river, squared distance to the river, interaction terms of these two variables with a river cell dummy, elevation, river confluence, flood vulnerability, geology measures, and open water access measures for a grid cell where each census community belongs.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.11: Community Population and Crop-Specific Form of Seed Acquisition

	Obtain seeds from a city or other communities for							
	Maize		Plantain		Rice		Yuca	
	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV	(7) OLS	(8) IV
$\log(N_{o,Ag})$	0.00682	-0.238*	0.00373	-0.200	0.0104	0.0636	0.00121	0.0904
	(0.0150)	(0.143)	(0.0138)	(0.122)	(0.00963)	(0.0756)	(0.0103)	(0.0819)
Basin FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean (Dep. var.)	0.393	0.393	0.305	0.305	0.126	0.126	0.165	0.165
SD (Dep. var.)	0.489	0.489	0.461	0.461	0.332	0.332	0.372	0.372
First stage *F*-stat		13.77193		13.77193		13.77193		13.77193
Observations	907	907	907	907	907	907	907	907

Notes: Robust standard errors in parentheses. The unit of anaylysis is a community in the PARLAP Community Census (CC) in 2014. We use $\log(RNA_o)$ and the initial community existence in 1940 as instruments for $\log(N_{o,Ag})$. Geographical controls include a dummy of high river orders (4 and 5), distance to the urban center, distance to the river, squared distance to the river, interaction terms of these two variables with a river cell dummy, elevation, river confluence, flood vulnerability, geology measures, and open water access measures for a grid cell where each census community belongs.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

B Data Appendix

B.1 PARLAP

Prices

The PARLAP CC data collects the following price information in the census communities.

Producer prices. The producer price is the gate price for products when people sell their products in the community. The producer price is collected for three agricultural products (plantain, yuca, and rice) and four natural resource products (aguaje fruit, fresh fish, salted fish, and salt bush meat). If a product cannot be sold in the community, its price is missing. It may not always be the case that these products are produced in the community where they are sold people from other communities can also sell their products.

Consumer prices. Consumer price is the price for commodities when people buy them in the community. The consumer price is collected for the following products: milled rice, sugar, cooking oil, soap, batteries, and kerosene. All these commodities including milled rice come from places outside the community (towns). If they cannot buy a product in the community, its price is missing.

Sectoral employment shares

The PARLAP CC data collects information about a census community's current economic activities and those when the community was established in the current location. The CC data collects the following three types of information about economic activities. First, it asks whether people in the community currently engage (or engaged in the past) in each of the following broadly-defined activities: agriculture, livestock, fishing, hunting, non-timber forest products (NTFPs), timber extraction, aquarium fish, petroleum industry, tourism, the state activity, and religious mission. Second, if people in the community engage (or engaged) in an economic activity categorized above, then the principal products produced in the community are listed. Third, among the broadly-defined economic activities that people in the community engage (or engaged) in, ranks of engagement in these activities are collected. We exploit the first and third information to construct the employment shares of agriculture and natural resource extractions in each community.[41] Natural resource extractions include the activities of fishing, hunting, non-timber forest products (NTFPs), timber extraction, and aquarium fish.

B.2 Survey of travel time and transportation costs

The PARLAP team conducted the survey of travel time and transportation costs. The survey was done during December 2017-January 2018 in Loreto and during February 2018-June 2019 in Ucayali. All monetary values are in Soles.

The survey provides travel time, passenger fees, gasoline, freights (one bundle of platano and one sack of 50kg) by season (low water/high water) and direction (upstream/downstream) for about 20 selected river routes with multiple modes of transport. Modes of transport (boat types) include lancha, peque-peque, canue, and rapido. Figure A.10 shows these transport modes. The survey also provides the cost for transporting 50kg package for 500 meter by

[41] The other activities constitute only a negligible shares and thus we do not consider them.

land transport at the selected communities and towns that are covered in the above travel time module.

B.3 Other Data and Variables

Other Data

We additionally use the following data.

Peru National Household Survey (ENAHO). ENAHO is an annual living standard survey, collected by the Institute Nacional de Estadistica e Informatica (INEI) in Peru. The data is publicly available from 2004 to present. To keep consistent periods and geography with other data, we use the data from Loreto and Ucayali departments in the Peruvian Amazon collected in 2013, 2014, 2015, 2016, and 2017. ENAHO contains detailed household-level information of consumption expenditure and we use this data to estimate demand parameters.

Other geographical variables

We provide definitions of some geographical control variables.

Distance to the urban center. This variable measures the river-equivalent distance from each grid cell to the urban center in each basin. We treat downstream and upstream rivers equivalently for this measure.

Distance to the river. This variable measures distance from the grid cell centroid to the nearest river point.

River orders. Higher river orders mean more splits from the central river. Our river network information contains river orders 1–5. We construct a dummy variable which takes one if the grid cell faces a river whose river order is 4 or 5. We choose these river orders because it tends to be difficult for large ships to navigate along these river lines.

River confluences. We construct two dummy variables regarding river confluences: (i) a dummy which takes one if the grid cell faces a confluence between the 1st and 2nd order rivers or a confluence between the 2nd and 3rd order rivers; (ii) a dummy which takes one if the grid cell faces a confluence between the 3rd and 4th order rivers;

Flood vulnerability. This variable is primarily based on the Global Inundation Extent from Multi-Satellites (GIEMS) dataset. It measures, at a fine spatial resolution, areas that do not flood and does not include water bodies; areas of Open Water; areas under Minimum Annual Flood Extent; areas under Maximum Annual Flood Extent; areas under Long Term Annual Flood Extent. See Fluet-Chouinard et al. (2015) for how the flood extents are determined. We construct a variable which takes the following five values: 0 = the grid cell does not flood and does not include water bodies; 1 = the grid cell with the highest areas share of Open Water; 2= the grid cell with the highest area share of Minimum Annual Flood Extent; 3 = the grid cell with the highest area share of Maximum Annual Flood Extent; 4 = the grid cell with the highest area share of Long Term Annual Flood Extent.

Water share: main channel. This variable captures the area share of main channel of rivers in the grid cell, determined from Landsat imagery (Kalacska et al. 2022).

Water share: non-main channel. This variable captures the area share of non-main channel open water in the grid cell, determined from Landsat imagery (Kalacska et al. 2022). This measure is regarded as an important predictor for species habitat. The non-main channel of open water tends to be found off the main channel of rivers, including lakes, abandoned side channels of rivers, streams, and other forms of standing water on the floodplain.

Soil characteristics. The original data source is La Carta Geológica Nacional del Peru (1:100,000) published by the Instituto Geológico Minero y Metalúrgico (INGEMMET). We proxy the floodplain soil share by the area share of young (Holocene) parent material in the grid cell. We also construct the area share of Pleistocene soil type in the grid cell. The Tertiary soil type (older than Pleistocene) is the omitted category. These characteristics are associated with soil fertility and agricultural productivity. See Coomes et al. (2022) for a detailed discussion.

C Estimation and Quantification Appendix

C.1 Empirical Tests of the Simplified Model with a Density Externality

We consider a simplified version of the main model in section 4 in the following three dimensions. First, we consider one sector with a continuum of goods (that implicitly pool both agricultural and natural resource goods), in contrast to the multi-sector model. This sector has land and labor as inputs while the land production process is same as the main model. Second, we consider symmetric (or quasi-symmetric) trade costs between locations, in contrast to asymmetric ones in the main model. The quasi-symmetric cost of transporting a good from o to d is defined as $\tau_{od} = \tilde{\tau}_{od}\tau_o^A\tau_d^B$ where $\tilde{\tau}_{od} = \tilde{\tau}_{do}$. That is, the term that depends on both o and d is symmetric. The simplest example is $\tau_o^A = 1/\text{elevation}_o$ and $\tau_d^B = \text{elevation}_d$. Third, we incorporate the density externality from population in the own location but without spatial spillovers across locations.

Given this setup, following a similar derivation as Donaldson and Hornbeck (2016), we can derive the measure of consumer's accessibility to low-price products ("Consumer Market Access"):

$$CMA_d \equiv P_d^{-\theta} = \kappa_2 \sum_o A_o N_o^{\mu\theta} \kappa_1^{\theta} (w_o \tau_{od})^{-\theta}$$

and the measure of firm's accessibility to consumers with low CMA ("Firm Market Access"):

$$FMA_o \equiv \sum_d \tau_{od}^{\theta} CMA_d^{-1} Y_d$$

Under the quasi-symmetric trade cost with $\tau_o^A = (\tau_o^B)^{-1}$, we can define the Market Access measure with the following relationship :

$$MA_o \equiv FMA_o = \rho(\tau_o^B)^{2\theta} CMA_o \quad \exists \rho > 0$$

Solving the balanced trade condition, we can use the Market Access measure to express community population:

$$N_o = \kappa_3 A_o^{\frac{1}{1-\theta\tilde{\mu}}} (\tau_o^B)^{-\frac{2(1+\theta)}{1-\theta\tilde{\mu}}} MA_o^{\frac{1+2\theta}{\theta(1-\theta\tilde{\mu})}}$$

deforestation and per capita deforestation:

$$\begin{aligned} DF_o &= \kappa_4 A_o^{\frac{1-\mu_L}{1-\theta\tilde{\mu}}} (\tau_o^B)^{-\frac{2(1+\theta)(1-\mu_L)}{1-\theta\tilde{\mu}}} MA_o^{\frac{(1+2\theta)(1-\mu_L)}{\theta(1-\theta\tilde{\mu})}} \\ \frac{DF_o}{N_o} &= \kappa_4 A_o^{\frac{-\mu_L}{1-\theta\tilde{\mu}}} (\tau_o^B)^{\frac{2(1+\theta)\mu_L}{1-\theta\tilde{\mu}}} MA_o^{-\frac{(1+2\theta)\mu_L}{\theta(1-\theta\tilde{\mu})}} \end{aligned}$$

nominal and real incomes:

$$
\begin{aligned}
Y_o &= \kappa_5 A_o^{\frac{1}{1-\theta\tilde{\mu}}} \left(\tau_o^B\right)^{-\frac{2\theta(1+\tilde{\mu})}{1-\theta\tilde{\mu}}} MA_o^{\frac{2+\tilde{\mu}}{\theta(1-\theta\tilde{\mu})}} \\
Y_o^R \equiv \frac{Y_o}{P_o} &= \kappa_6 A_o^{\frac{1}{1-\theta\tilde{\mu}}} \left(\tau_o^B\right)^{-\frac{2\theta(1+\tilde{\mu})}{1-\theta\tilde{\mu}}} MA_o^{\frac{1+2\theta}{\theta(1-\theta\tilde{\mu})}}
\end{aligned}
$$

where μ_L is the parameter governing congestion externality in forest clearing, μ is the parameter governing agglomeration externality given the obtained land, and $\tilde{\mu} \equiv \mu - \mu_L(1-\gamma)$ captures the net agglomeration.

The one-sector model implies the following empirical specification:

$$\ln y_{o,t} = \beta_0 + \beta_1 \ln MA_{o,t} + \theta X_o + \phi_{B,t} + \epsilon_{o,t} \tag{C.1}$$

where $y_{o,t}$ is an outcome measure of human settlements and forest cover at grid cell o, X_o is a vector of cell-level geographical controlsand $\phi_{B,t}$ represents basin $\times$ year fixed effects. The Market Access measure is approximated as:

$$MA_{o,t} \approx \sum_d N_{d,t}(\tau_{od})^{-\theta} = \sum_d N_{d,t}(D_{od}^{\delta})^{-\theta} \tag{C.2}$$

where $o, d \in$ 1km $\times$ 1km grid cells within 5km from a river (with order 1-6), $N_{d,t}$ is the population in d at period t, τ_{od} is the trade cost, and D_{od} represents river-equivalent distance (kilometer) along the least-cost route. To calculate this measure, we set a composite value of parameters $\delta \cdot \theta = 1.3182$ drawing from Donaldson (2018). We report results of both OLS and IV regressions. The IV regressions use the following River Network Access (RNA) measure, solely defined by river shapes, as an instrument variable for the Market Access:

$$RNA_o = \sum_{d \in RC} \tau_{od}^{\theta}$$

where RC is a set of all river cells in the basin (cells that contain a river) whether or not they have positive populations. We will exploit the variation in RNA to estimate density externality parameters in a later section. In interpreting the results of the Market Access regression, however, we confine our discussion to be correlational from a conservative perspective as it suffices to motivate the model.[42]

Table C.1 reports the relationship between the Market Access and human settlements. The Market Access is significantly associated with both community formation and population size. The size of the estimated coefficient based on the IV specification under a reasonable range of θ implies that congestion force is likely to be dominating on net. The multi-sector model in a later section will nevertheless emphasize sector-specific agglomeration and congestion externalities.

Table C.3 reports the relationship between the Market Access and forest cover change over decades.[43] The Market Access is significantly associated with forest disturbance, forest loss,

[42] We also check that the reduced form results are robust with different sets of controls, the RNA and MA measures adjusted by river order at origin, considering river networks only up to river order 4 or 5, and the MA measure with earlier period (1940) population.

[43] Using static measures of forest covers, Table C.2 reports that the Market Access is negatively associated with forest areas and positively associated with non-forest areas.

and forest recovery. The following three points are noteworthy. First, interpretation of the positive associations with forest disturbance and forest loss is straightforward. Locations with higher market access might attract more people to reside and clear forest for agricultural land use. Second, the Market Access is also associated with forest recovery, an inverse measure of forest loss. This result reflects the practice of sustainable land use in the shifting cultivation cycle. This result is also consistent with the recent finding by Coomes et al. (2021) that forest cover relative to population engaging in small-scale agriculture is stable over time. Finally, the rightmost column reports that the Market Access is negatively associated with per capita deforestation.[44] This observation is consistent with the statement of Fact 2A and the presence of congestion force in forest clearing. To summarize, the empirical results in this section imply the following fact.

Fact 2B: Human settlements and forest cover changes (both forest loss and recovery) increase with Market Access.

C.2 Parameters without Solving the Model

Downstream-River-Equivalent Distance

We use our original records of travel times and freight costs (Appendix B.2).

Upstream-river distance. We calibrate λ_{up} by taking the average ratio of upstream-river travel time to downstream-river travel time by peque-peque across all the travel routes available in the survey. We obtained $\hat{\lambda}_{up} = 1.282$.

Land distance. For comparing the freight costs on land and river transports, we focus on communities in the land transport data where the river transport data (either as origin or destination community) is also available. Since we also use the information on the distance on river network, we further focus on river routes that are found in our river network data in the four basins. There are two routes that satisfy these criteria: (1) Mazan-Santa Clotilde (Loreto) and (2) Puerto Alegre-Vinuncuro (Ucayali). For each of these two routes, we first use the observed river cost, $\hat{\delta}_M$ (obtained in the previous step in the main text), and the following relationship:

$$\text{River transport cost (observed)} \approx p_{od} - p_{oo} = (\tau_{od} - 1)p_{oo} = (D_{od,river}^{\hat{\delta}_M} - 1)p_{oo}$$

to back out hypothetical $\hat{p}_{oo}$. In these two routes, the reported river costs are same in both directions and thus we assume $D_{od,river} = D_{do,river}$. We next use the following relationship:

$$\text{Land transport cost (observed)} \approx [(D_{od',land} \times \lambda_{land})^{\hat{\delta}_M} - 1]\hat{p}_{oo}$$

to obtain $\hat{\lambda}_{land}$ where $D_{od',land} = 0.5$(km) and in each route we use the average value of the 500m land transport cost between those collected in origin and destination locations. We then take the average value of $\hat{\lambda}_{land}$ between those obtained from the two routes. We obtained $\hat{\lambda}_{land} = 36.767$.

[44] The forest loss is between 1985 and 2015 and the denominator population is from the INEI population census in 2007.

Table C.1: Market Access and Human Settlements

	Community existence		log (population)	
	(1)	(2)	(3)	(4)
	OLS	IV	OLS	IV
log(MA)	0.0151***	0.0134***	0.0705***	0.0646***
	(0.000563)	(0.000533)	(0.00277)	(0.00259)
log (Elevation)	0.000839**	-0.000397	0.00441**	0.000290
	(0.000397)	(0.000460)	(0.00201)	(0.00227)
River confluence (1st×2nd or 2nd×3rd)	0.0687***	0.0690***	0.348***	0.348***
	(0.00739)	(0.00739)	(0.0371)	(0.0372)
River confluence (3rd×4th)	0.0253***	0.0251***	0.114***	0.114***
	(0.00249)	(0.00248)	(0.0115)	(0.0115)
Flood vulnerability (1-4)	0.00244***	0.00243***	0.0108***	0.0108***
	(0.000171)	(0.000171)	(0.000816)	(0.000817)
Floodplain soil share	0.0000232***	0.0000245***	0.000107***	0.000111***
	(0.00000448)	(0.00000446)	(0.0000220)	(0.0000219)
Pleistocene soil share	-0.0000285***	-0.0000294***	-0.000160***	-0.000163***
	(0.00000446)	(0.00000447)	(0.0000205)	(0.0000205)
Water share: non-main channel	0.000465***	0.000472***	0.00234***	0.00237***
	(0.0000712)	(0.0000713)	(0.000351)	(0.000352)
Water share: main channel	0.0000192	0.0000230	0.0000950	0.000108
	(0.0000407)	(0.0000406)	(0.000200)	(0.000200)
Basin × Year FE	Yes	Yes	Yes	Yes
River Order FE	Yes	Yes	Yes	Yes
R^2	0.038	0.015	0.036	0.014
Mean (Dep. Var.)	0.008	0.008	0.037	0.037
SD (Dep. Var.)	0.089	0.089	0.431	0.431
Observations	403938	403938	403938	403938

Notes: Robust standard errors in parentheses. The sample includes 1 square km grid cells within 5km from rivers (up to 6th order). Other controls include distance to the river, squared distance to the river, and interaction terms of these two variables with a river cell dummy.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C.2: Market Access and Forest Cover

	Forest area		Non-forest area		Per-capita non-forest area	
	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV
log(MA)	-0.206***	-0.208***	1.970***	1.586***	0.126**	-0.0726
	(0.00607)	(0.00491)	(0.0243)	(0.0441)	(0.0621)	(0.216)
log (Elevation)	-0.174***	-0.176***	1.057***	0.753***	-0.996***	-1.626**
	(0.00702)	(0.00723)	(0.0459)	(0.0554)	(0.312)	(0.723)
River confluence (1st×2nd or 2nd×3rd)	-0.0544	-0.0541	0.383***	0.434***	-0.201*	-0.211*
	(0.0385)	(0.0385)	(0.125)	(0.121)	(0.121)	(0.122)
River confluence (3rd×4th)	-0.0660***	-0.0662***	0.218**	0.176*	-0.156	-0.153
	(0.00889)	(0.00886)	(0.0920)	(0.0919)	(0.138)	(0.140)
Flood vulnerability (1-4)	-0.0124***	-0.0124***	0.261***	0.257***	0.0155	0.00816
	(0.00130)	(0.00130)	(0.0102)	(0.0102)	(0.0224)	(0.0233)
Floodplain soil share	-0.137***	-0.136***	1.585***	1.591***	-0.155	-0.164
	(0.00470)	(0.00471)	(0.0373)	(0.0375)	(0.117)	(0.119)
Pleistocene soil share	0.0209***	0.0207***	-1.362***	-1.408***	-0.0415	-0.113
	(0.00387)	(0.00384)	(0.0604)	(0.0607)	(0.267)	(0.282)
Water share: non-main channel	-2.597***	-2.597***	7.176***	7.292***	-1.401***	-1.382***
	(0.0720)	(0.0723)	(0.270)	(0.263)	(0.477)	(0.487)
Water share: main channel	-2.962***	-2.962***	3.664***	3.694***	-0.567**	-0.677**
	(0.0877)	(0.0878)	(0.116)	(0.114)	(0.260)	(0.284)
Basin FE	Yes	Yes	Yes	Yes	Yes	Yes
River Order FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.252	0.181	0.205	0.090	0.141	0.032
Mean (Dep. var.)	13.642	13.642	6.553	6.553	7.737	7.737
SD (Dep. var.)	0.554	0.554	4.781	4.781	1.312	1.312
Observations	132369	132369	132369	132369	1189	1189

Notes: Robust standard errors in parentheses. The sample includes 1 square km grid cells within 5km from rivers (up to 6th order). Other controls include distance to the urban center, distance to the river, squared distance to the river, and interaction terms of these two variables with a river cell dummy.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C.3: Market Access and Forest Cover Change

	Forest disturbance		Forest loss		Forest recovery		Per-capita forest loss	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
log(MA)	1.411***	1.742***	0.375***	0.366***	0.402***	0.520***	-0.282***	-0.733***
	(0.0247)	(0.0338)	(0.00570)	(0.00832)	(0.00662)	(0.00963)	(0.0873)	(0.273)
log (Elevation)	0.345***	0.606***	-0.00510	-0.0117	0.313***	0.400***	-0.532	-1.897**
	(0.0521)	(0.0551)	(0.0110)	(0.0121)	(0.0115)	(0.0130)	(0.403)	(0.878)
River confluence (1st×2nd or 2nd×3rd)	0.205	0.160	-0.0151	-0.0138	0.0766*	0.0604	-0.342*	-0.382**
	(0.194)	(0.195)	(0.0375)	(0.0374)	(0.0437)	(0.0444)	(0.187)	(0.189)
River confluence (3rd×4th)	0.0247	0.0606	0.0487**	0.0478**	0.0128	0.0252	-0.287	-0.283
	(0.0817)	(0.0820)	(0.0193)	(0.0193)	(0.0217)	(0.0216)	(0.184)	(0.186)
Flood vulnerability (1-4)	0.0882***	0.0912***	0.0419***	0.0418***	0.0686***	0.0695***	-0.0246	-0.0404
	(0.00901)	(0.00902)	(0.00214)	(0.00214)	(0.00263)	(0.00263)	(0.0289)	(0.0303)
Floodplain soil share	1.111***	1.106***	0.386***	0.386***	0.536***	0.532***	0.0957	0.0965
	(0.0302)	(0.0302)	(0.00753)	(0.00753)	(0.00850)	(0.00850)	(0.156)	(0.156)
Pleistocene soil share	-0.109***	-0.0699*	-0.168***	-0.169***	-0.0195	-0.00684	0.342	0.185
	(0.0384)	(0.0385)	(0.00931)	(0.00934)	(0.0139)	(0.0140)	(0.290)	(0.311)
Water share: non-main channel	2.677***	2.577***	0.580***	0.583***	0.582***	0.539***	-0.927**	-0.846*
	(0.360)	(0.363)	(0.0662)	(0.0662)	(0.0783)	(0.0799)	(0.444)	(0.494)
Water share: main channel	1.237***	1.212***	-0.0594*	-0.0582	-1.099***	-1.115***	-0.981***	-1.189***
	(0.209)	(0.210)	(0.0361)	(0.0360)	(0.0333)	(0.0339)	(0.345)	(0.365)
Basin FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
River Order FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.118	0.065	0.175	0.104	0.218	0.105	0.128	0.002
Mean (Dep. var.)	1.606	1.606	0.668	0.668	1.043	1.043	-2.819	-2.819
SD (Dep. var.)	3.462	3.462	0.919	0.919	1.055	1.055	1.716	1.716
Observations	132369	132369	134646	134646	134646	134646	1203	1203

Notes: Robust standard errors in parentheses. The sample includes 1 square km grid cells within 5km from rivers (up to 6th order). Other controls include distance to the urban center, distance to the river, squared distance to the river, and interaction terms of these two variables with a river cell dummy. We are taking logarithms for the dependent variables.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Demand Parameters

We estimate elasticity of substitution between varieties within each sector (σ) and between sectoral composite products ($\bar{\sigma}$), using detailed household-level information on expenditures and unit values (interpreted as buying prices) over time from the Peru National Household Survey (Appendix B.3).

Elasticity of substitution between varieties. We first estimate the elasticity of substitution between varieties in each sector. We estimate the following empirical specification, derived in part from the expenditure share of each variety (1) in the model and also using the household-level expenditure information from outside the model:

$$\ln(\tilde{\alpha}_{o,K,t,h}(j)) = \beta_0 + (1-\sigma)\ln p_{o,K,t,h}(j) + \beta_1 X_{o,t,h} + \phi_K + \phi_t + \epsilon_{o,K,t,h}(j) \tag{C.3}$$

where $\tilde{\alpha}_{o,K,t,h}(j)$ is the expenditure share of household h on good j (classified by ENAHO) of sector K (classified by us) at period t in location o, $X_{o,t,h}$ is a vector of household-level demographic variables at t in o, and ϕ_K and ϕ_t are the sector fixed effect and period fixed effects.[45] $p_{o,K,t,h}(j)$ is a unit value of good j of sector K that household h at t in o pays. This measure is obtained by deriving the value of expenditure on good j by the quantity of its expenditure and can thus be interpreted as a buying price. We instrument $\ln p_{o,K,t,h}(j)$ by $\ln RNA_o$ (where RNA_o is defined by (17)) which is an exogenous price shifter due to a trade mechanism but plausibly uncorrelated with local preference shocks given controls. Table C.4 reports the results in left columns. The point estimate with the instrument implies the estimated value to be $\hat{\sigma} = 2.401$. Since $\hat{\sigma} > 1$, this estimate implies that consumption demands of varieties within the agricultural or natural resource sector are substitute.

Elasticity of substitution between sectoral composite goods. We next estimate the elasticity of substitution between sectoral goods, using the estimated $\hat{\sigma}$ above to approximate the price index measure $\hat{P}_{o,K,t,h} = [\sum_j P_{o,K,t,h}^{(1-\hat{\sigma})}]^{1/(1-\hat{\sigma})}$. We estimate the following empirical specification, derived in part from the expenditure share of each sectoral good (2) and also using the household-level expenditure information from outside the model:

$$\ln(\tilde{\alpha}_{o,K,t,h}) = \bar{\beta}_0 + (1-\bar{\sigma})\ln \hat{P}_{o,K,t,h} + \bar{\beta}_1 X_{o,t,h} + \bar{\phi}_K + \bar{\phi}_t + \epsilon_{o,K,t,h} \tag{C.4}$$

where $\tilde{\alpha}_{o,K,t,h}$ is the expenditure share of household h on sector K goods at period t in location o. We implement LASSO to select instruments from exogenous productivity shifters in addition to $\ln RNA_o$.[46] Table C.4 reports the results in right columns. The point estimate with the instruments implies the estimated value to be $\hat{\bar{\sigma}} = 0.752$. Since $\hat{\bar{\sigma}} < 1$, this estimate implies that consumption demands across composites of agricultural and natural resource sectors are com-

[45] The household-level demographic variables include household size, number of adult members, and number of male members in each household. The period fixed effects include both year fixed effects and the fixed effects of interview month. The interview month matters because the expenditure variable is based on the household's expenditure within 15 days prior to the interview date.

[46] We add exogenous productivity shifters to instruments because the price index, an aggregated measure across varieties, is likely to contain varieties produced in each location and thus local productivity shifters in theory have direct influence on it. This empirical design is internally consistent in that the candidate productivity shifters for LASSO are same as the controls used to estimate density externalities from inverted productivity composites in a later stage. We use adaptive LASSO to select a tuning parameter. We did not include productivity shifters for estimating the elasticity of substitution between varieties because each location does not produce many varieties observed in ENAHO and thus local productivity shifters have poor explanatory power for prices disaggregated at varieties.

plementary. Note also that the estimated value is larger than a general parameter value (0.5) of the elasticity of substitution across sectoral goods of agriculture and manufacturing in the literature of structural transformation (e.g., Ngai and Pissarides 2007). This difference is reasonable because both agricultural and natural resource goods are food items and the complementarity between them would be weaker than that between food and non-food consumption.

Table C.4: Elasticity of Substitution across Varieties and Sectors

	log(Expenditure share)			
	Across varieties		Across sectors	
	(1) OLS	(2) IV	(3) OLS	(4) IV
$\log(p_{o,K}(j)),\ K = Ag, Nr$	-0.326***	-1.401***		
	(0.00810)	(0.540)		
$\log(P_{o,K}),\ K = Ag, Nr$			0.00626	0.248**
			(0.0371)	(0.126)
Basin FE	Yes	Yes	Yes	Yes
Mean (Dep. var.)	-4.419	-4.419	-1.068	-1.068
SD (Dep. var.)	1.366	1.366	0.447	0.446
First stage F-stat		19.148		13.652
Observations	58115	58115	3276	3270

Notes: Robust standard errors in parentheses. The coefficients correspond to $1-\sigma$ and $1-\bar{\sigma}$ regarding elasticity of substitution across varieties and sectors, respectively. For estimating the elasticity of substitution between sectors, we use the implied σ from the IV estimation of the elasticity of substitution to construct the price index measure. The estimation sample includes variety-level or sector-level expenditures of households from ENAHO during 2013-2017. We control for year fixed effects, month fixed effects, a dummy of agricultural sector, and household-level demographic variables. Houesehold-level demographic variables include household size, number of adult members, and number of male members in each household. For estimating the elasticity of substitution between varieties, we use log of RNA as an IV. For estimating the elasticity of substitution between sectors, we implement LASSO to select IVs from exogenous productivity shifters in addition to log of RNA.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

C.3 Algorithms for the Model Inversion and Simulation

We provide the algorithm for the model inversion described in section 5.2. The inversion problem itself is standard. The purpose here is to present the algorithm that mitigates the computational burden. The model inversion takes place in the following steps:

Step 1. Set initial guesses of the productivity composites in all sectors in all locations up to a normalization. In particular, we normalize the productivity such that $\frac{\sum_{o\in\tilde{\mathcal{R}}}\tilde{A}_{o,Ag}+\sum_{o\in\tilde{\mathcal{R}}}\tilde{A}_{o,Nr}+A_{u,M}}{2\tilde{\mathcal{R}}+1} = 10$. Set initial guess of wages as well.

Step 2. Given the productivities guesses in step 1, obtain wages that balance the overall trade by solving the following fixed point problem:

$$w_o = \frac{\sum_{K \in \{Ag,Nr\}} \sum_{d \in \tilde{\mathcal{I}}} \pi_{od,Ag} \tilde{\alpha}_{d,Ag} w_d N_d + \sum_{d \in \tilde{\mathcal{I}}} \tilde{\alpha}_{d,M} w_d N_d}{N_o} \quad \forall\, o \in \tilde{\mathcal{I}}$$

This step is not conceptually necessary to solve the whole problem, but it facilitates computations in the next step. We implement a fixed point iteration to solve this problem, starting with the wage guesses in step 1.

Step 3. Solve a nested fixed point problem.

The outer problem is to solve the following fixed point problem of equilibrium wages:

$$w_o = \frac{w_u \left[\sum_{K=Ag,Nr,M} P_{o,K}^{1-\bar{\sigma}} \right]^{\frac{1}{1-\bar{\sigma}}}}{\left[\sum_{K=Ag,Nr,M} P_{u,K}^{1-\bar{\sigma}} \right]^{\frac{1}{1-\bar{\sigma}}}} \quad \forall\, o \in \tilde{\mathcal{I}} \tag{C.5}$$

that can be derived from (15).

The inner problem is to solve the following fixed point problem of the sectoral productivity composites:

$$\begin{aligned} \tilde{A}_{o,K} &= \frac{w_o N_{o,K}}{\sum_{d \in \tilde{\mathcal{I}}} \frac{(w_o \tau_{od,K})^{-\theta}}{\sum_{o' \in \mathcal{R}} \tilde{A}_{o',K} (w_{o'} \tau_{o'd,K})^{-\theta}} \tilde{\alpha}_{d,K} w_d N_d} \quad K \in \{Ag, Nr\},\, o \in \mathcal{R} \\ A_{u,M} &= \frac{P_{u,M}^{-\bar{\sigma}}}{N_{u,M}} \sum_{d \in \tilde{\mathcal{I}}} \frac{\tau_{ud,M}^{1-\bar{\sigma}} w_d N_d}{\sum_{K=Ag,Nr,M} P_{u,K}^{1-\bar{\sigma}}} \end{aligned} \tag{C.6}$$

that can be derived from (11)–(13).

Set guesses of the productivity composites from step 1. Set guesses of the wages from step 2. We solve for the sectoral productivities by applying a fixed point iteration algorithm for (C.6). Using the guessed wages and the productivities solved in the inner problem, we compute the left-hand side of (C.5) to update wages. If the updated wages are sufficiently close to the guessed wages, then stop. Otherwise, replace guessed wages with the updated wages and update wages again. We use a fixed point iteration algorithm for this outer problem as well. In particular, for both the inner and outer problems, we employ the Anderson acceleration method (Walker and Ni 2011).

To simulate the model under counterfactual scenarios, we first obtain the calibrated values of productivity fundamentals in all sectors in all locations, using the inverted productivity composites and the estimated density externality parameters. Given the productivity fundamentals, we then solve for endogenous sectoral populations and wages in all locations using those in the benchmark spatial equilibrium as initial guesses, but with an additional constraint of the fixed number of total populations in the basin. We employ the same fixed point iteration algorithm as in the inversion.

References for the Appendix

Coomes, Oliver T, Margaret Kalacska, Yoshito Takasaki, Christian Abizaid, and Tristan Grupp, "Smallholder agriculture results in stable forest cover in riverine Amazonia," *Environmental Research Letters*, 2021.

__ **, Yoshito Takasaki, and Christian Abizaid**, "Sparing of Amazonian old-growth forests with floodplain access," *Nature Sustainability*, 2022, pp. 1–8.

Donaldson, Dave, "Railroads of the Raj: Estimating the impact of transportation infrastructure," *American Economic Review*, 2018, *108* (4-5), 899–934.

__ **and Richard Hornbeck**, "Railroads and American economic growth: A "market access" approach," *The Quarterly Journal of Economics*, 2016, *131* (2), 799–858.

Fluet-Chouinard, Etienne, Bernhard Lehner, Lisa-Maria Rebelo, Fabrice Papa, and Stephen K Hamilton, "Development of a global inundation map at high spatial resolution from topographic downscaling of coarse-scale remote sensing data," *Remote Sensing of Environment*, 2015, *158*, 348–361.

Kalacska, Margaret, J Pablo Arroyo-Mora, Oliver T Coomes, Yoshito Takasaki, and Christian Abizaid, "Multi-Temporal Surface Water Classification for Four Major Rivers from the Peruvian Amazon," *Data*, 2022, *7* (1), 6.

Ngai, L Rachel and Christopher A Pissarides, "Structural change in a multisector model of growth," *American economic review*, 2007, *97* (1), 429–443.

Walker, Homer F and Peng Ni, "Anderson acceleration for fixed-point iterations," *SIAM Journal on Numerical Analysis*, 2011, *49* (4), 1715–1735.