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

Shunsuke Tsuda
Brown University
Job Market Paper

Yoshito Takasaki
The University of Tokyo

Mari Tanaka
Hitotsubashi University

This paper is updated often. Please check HERE for the latest version.

October 7, 2022

Abstract

This paper investigates the ways to resolve trade-offs between human and ecological well-being in tropical forests. We build a multi-sector spatial model of rainforest communities that features concentration and dispersion forces of the population. The model is estimated by exploiting exogenous river networks in the Peruvian Amazon without road access and a novel dataset from rural communities. We find the agglomeration externality in agricultural production and that this gain outweighs the dispersion force in access to land, leading to higher productivity with smaller deforestation per farmer by concentrating. We also find the congestion externality with spatial spillovers in natural resource extraction. We provide evidence that the agglomeration is primarily driven by the economies of scale in transport technology and agricultural intensification. A quantification exercise reveals that the agglomeration externality contributes substantially to improving human welfare and reducing deforestation, while increasing natural resource depletion through a general equilibrium effect. Counterfactual experiments demonstrate that well-targeted place-based protection policies and river infrastructure investments are complementary to improve human-ecological well-being. Protecting the rural frontier works primarily to conserve biodiversity by increasing the congestion force in natural resource extraction within more compact areas for human settlements. Improving transport infrastructure that integrates a hinterland reduces total deforestation by making the agglomeration benefits spread more evenly across the basin.

Keywords: Agriculture, Conservation, Externalities, Natural resources, Rainforests, Spatial models

*Tsuda is grateful to Andrew Foster and Matt Turner for continuous guidance and support and to Dan Björkegren and Jesse Shapiro for detailed discussions and feedback. We thank Dave Donaldson, Mateusz Filipski, John Friedman, Peter Hull, Hiro Ishise, Heitor Pellegrina, Ellen McCullough, Yuhei Miyauchi, Kaivan Munshi, Takeshi Murooka, Mike Neubauer, Matt Pecenco, and audiences at Brown University, University of Georgia, and Osaka University for helpful comments. We acknowledge the PARLAP project collaborators Oliver Coomes, Christian Abizaid, and Margaret Kalacska for their invaluable inputs. Nicholas Cadieux and Tristan Grupp provided excellent research assistance. Tsuda thanks Margarita Gafaro for providing community-level identifiers of the Peruvian Agricultural Census. Tsuda acknowledges financial support from a dissertation fellowship at Brown University, funding from the Murata Science Foundation, and research support from CREPE. 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

Amazon rainforest is home to much of the world's biodiversity and natural resources. Rapid deforestation and biodiversity loss, caused by human activities, are pressing issues. Many populations there have low living standards with limited modern technologies in remote areas. Balancing the conservation of rainforests with improving the living standards of local populations is a global challenge.

Policymakers face two types of fundamental trade-offs. The first trade-off is between rainforest conservation and improving welfare of local populations. For example, raising the cost of forest clearing may reduce deforestation, but it may also reduce agricultural income and human welfare. The second trade-off is between reducing different types of environmental costs. For example, if the cost of forest clearing is increased, then local populations may simply shift their activities from agriculture to other extractive activities such as fishing and hunting, which could lead to bio-diversity loss. The broad research question of this paper is whether it is possible to design a policy that eliminates these trade-offs.

To answer this question, this paper builds a multi-sector spatial general equilibrium model that highlights the rainforest population's trade-off between concentration and dispersion and apply it to a dataset from major river basins in the Peruvian Amazon. We estimate externality parameters that explain how population density affects productivity in each sector by exploiting exogenous river shapes as a source of identification. We implement counterfactual experiments to evaluate policy impacts on human welfare, deforestation, and other natural resource depletions. The experiments reveal that protection policies and the improvement of transport infrastructure are complementary to resolve the fundamental trade-offs.

We compile a novel dataset covering major river basins of the Peruvian Amazon, a suitable setting to study fundamental human-nature interactions from the following two aspects. First, most of the population engage in traditional lifeways with limited modern technologies and market access in remote areas, attributing resource extractions to small-scale farmers. Second, exogenous river networks constitute almost the sole transportation routes in this area, facilitating the identification of key structural parameters. The dataset consists of community-level censuses, original community- and household-level surveys from communities in major river basins, and grid cell-level remote sensing data, including forest cover over time.

Using the dataset, we start with three stylized facts that motivate the theoretical model. First, both concentration and dispersion of populations and community locations are observed. Second, human settlements (in terms of both community formation and population size) and forest cover changes are increasing in Market Access, but not proportionally so. Locations with higher market access might attract more people to reside and clear forest for agricultural land use. At the same time, *per capita* deforestation (or land footprint) is decreasing in population size. This fact is consistent with the presence of congestion force in forest clearing. Third, population density is associated with household-level activity choice between agriculture and natural resource extraction sectors. In particular, while agriculture is widely observed in both concentrated and dispersed areas, households tend to engage in natural resource extractions more in locations with lower population densities in surrounding areas.

Motivated by the stylized facts, we build a quantitative spatial model that highlights the rainforest population's trade-off: natural resource endowments are richer in sparse areas while dense areas have higher market access and agglomeration benefits. We incorporate two rural sectors—agricultural sector (crop production), which leads to deforestation, and natural resource extraction sector (fishing, hunting, and forest products), which leads to bio-diversity loss—and an urban sector. The model analyzes trade of these goods based on comparative advantage across multiple locations (grid cells) in a general equilibrium framework. We explicitly incorporate an agglomeration externality for productivity in the agricultural sector. That is, agricultural productivity depends on location fundamental and population density in that location. For productivity in the natural resource extraction sector, we incorporate a congestion externality allowing for spatial spillovers. That is, natural resource extraction productivity depends on location fundamental and the population density in surrounding locations.

The model aims at explaining how agglomeration and congestion externalities function in determining the spatial distribution of economic activities in the rainforest river networks. Calibrating the productivities that rationalize the observed sectoral employments and implementing GMM estimation that exploits exogenous river shapes, we estimate sector-specific density externalities. We find the presence of agglomeration externality in the agricultural production *on net*, despite the presence of congestion externality in clearing forest for gaining land for cropping. We also find the presence of congestion externality with spatial spillovers in the natural resource extraction.

While interpreting the congestion externality in natural resource extraction is straightforward given its rivalrous nature, interpreting the agglomeration externality in agriculture is not trivial. Mechanism analysis reveals that the economies of scale in transport technologies play a key role for the agglomeration benefit. We also find supportive evidence that the economies of scale in accessing inputs and technologies in the cropping process are also behind the agglomeration externality.

We find that the agglomeration externality has quantitatively large effects on improving human welfare and reducing deforestation, at the expense of other natural resource endowments. Without the agglomeration externality, deforestation increases in about 20-60%. The human welfare is about 8-13% lower without the agglomeration. On the other hand, the natural resource depletion also decreases in about 1-3% without the agglomeration externality.

We implement two types of counterfactual experiments, resettlement policies and improvement of transport infrastructure, aiming for finding a 'win-win policy' that improves human welfare, reduces deforestation, and reduces natural resource extraction. Resettlement policies aim for answering the question: in the presence of density externalities, is it beneficial to concentrate the ecological footprint in fewer spots than having many small communities? If so, how do different ways of selecting communities to encourage resettlement and setting protected areas affect different outcomes? Improvement of transport infrastructure aims for reducing high trade costs in the environment of Amazon river networks. The infrastructure improvement involves building roads along the river lines or deepening rivers as the Amazon Waterway Project (a government plan with Chinese investment but not yet implemented) proposed. High trade costs primarily stem from asymmetry of transport costs due to river orientations, seasonality of transport costs due to water rise/down, and slow speed of river boats.

The comparison between different resettlement policies illustrates that policymakers face the ecological trade-off in mitigating different types of environmental costs: deforestation and other natural resource depletion. We compare different policies that directly treat the same number of populations in each basin. Among the policies, resettling population from the smallest communities has reduced deforestation most. On the other hand, setting protected areas by limiting the expansion of rural frontier has mitigated bio-diversity loss while the deforestation impact is smaller than the previous policy.

While improvement of transport infrastructure increases human welfare, we find that the direction of deforestation impact depends on where in the spatial structure of river networks the improvement takes place. Improving the transport infrastructure in a way that it connects hinterlands to the central area of a basin has reduced total deforestation in the basin. As hinterlands become more integrated in the trade network, we observe spatial reallocation of farmers toward remote areas from the benchmark equilibrium. The agglomeration externality in agricultural production amplifies this effect and remote areas have agricultural productivity gains. 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. On the other hand, improving the transport infrastructure only in densely populated areas has increased total deforestation in the basin, because of the opposite forces to the previous case. This comparison implies that policy interventions that make the areas enjoying agglomeration externalities more evenly across the basin are preferable in terms of reducing deforestation. In other words, policies causing more middle-sized communities evenly 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.

Finally, the win-win outcome is achieved by combining the protected areas that control rural frontier expansion with the improvement of transport infrastructure that connects hinterlands to the central area of basin. This intervention increases welfare for about 1.1-2.3%, decreases deforestation for about 5-7%, and decreases natural resource depletion for about 0.5-3%. The intuition behind this result follows from the previous arguments. This combined intervention makes the basin more compact for human settlements, which preserves bio-diversity in the presence of congestion externality with spatial spillovers, and makes the agglomeration benefits spread more evenly across the basin, which is a primary force in reducing deforestation.

Related literature. This paper contributes to three strands of literature.

First, this paper contributes to a large body of literature on the trade-off between economic development and environmental goals (see Jayachandran 2021; Jayachandran 2022 for review) with particular focuses on natural resource depletion in rural areas. 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; Szerman et al. 2022) where empirical evidence on whether higher agricultural productivity increases deforestation is mixed. We introduce a new channel, the agglomeration externality primarily stemming from the economies of scale in transport technologies, through which agricultural productivity increases and deforestation decreases in spatial general equilibrium. Notably, improving agricultural productivity through our channel can reduce deforestation even without the strong conditions under which recent papers (Abman et al. 2020; Szerman et al. 2022) have drawn the same conclusions: agriculture is the most land-intensive sector in our model; we are not imposing any assumption about factor market constraints; modern technologies for agricultural intensification are limited in the Peruvian Amazon. Our counterfactual experiments also relate to the policy discussion of combating deforestation such as protected areas and taxes (e.g., Alix-Garcia et al. 2013; Alix-Garcia et al. 2015; Araujo et al. 2020; Assunçao et al. 2022; Naughton-Treves et al. 2011; Robalino and Pfaff 2012; Sims and Alix-Garcia 2017; Souza-Rodrigues 2019). In addition, this paper joins the decades-long discussion on commons management (Dasgupta and Mäler 1995; Hardin 1968; Ostrom 1990) by incorporating spatial competition over common pool resources in the general equilibrium model. Finally, we emphasize that we incorporate distinct types of environmental costs, deforestation and other natural resource extractions such as fishing and hunting, which have often been studied in isolation, in a unified framework.

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 and Arkolakis 2014; Allen et al. 2020; Donaldson and Hornbeck 2016; Fajgelbaum and Redding 2022; Fujita et al. 1999; Miyauchi 2021; 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 exogenous (future) climate change on economic activities and welfare using spatial models at the 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) levels. On the other hand, investigation of endogenous environmental changes caused by economic activities in the spatial equilibrium framework is scarce in the literature with some recent exceptions focusing on polluting activities (Hollingsworth et al. 2022; Jégard 2021; Yamada 2020). 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., Pellegrina 2022; Porteous 2019; Sotelo 2020; Rivera-Padilla 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 and that it incorporates natural resource extraction, which has contrasting characteristics to agriculture, featuring different types of density externalities in these sectors.

Third and relatedly, 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 an urban setting in the literature, to our knowledge, 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 the 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 the 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.

2 Empirical Setting and Data

Amazonian rainforests are an important repository of bio-diversity and natural resources not only for local people but also for the rest of the world. Rapid deforestation and bio-diversity loss, caused by human activities, are pressing issues. In particular, the Peruvian Amazon is an almost ideal setting to study fundamental human-nature interactions in rainforests from the following two aspects. First, in this environment, most of the population engage in traditional lifeways with limited modern technologies and market access in remote areas. We take advantage of this feature for attributing resource extractions to small-scale farmers and focusing on density externalities caused by them. Second, river networks constitute almost the sole transportation routes in this region. We take advantage of this feature for identifying key structural parameters by exploiting the exogenously given river shape and structure.

Our study area consists of two administrative departments of Loreto and Ucayalli that cover an area of 4711,99 square kilometer or about 85% of the Peruvian Amazon. Iquitos and Pucallpa are urban centers of these regions, respectively. Figure 1 shows our study area. The area that our data covers consists of four major river basins of the Peruvian Amazon (Amazon, Napo, Pastaza and Ucayali), which together encompass a vast area of 117,680 square kilometer which is more than twice of the total area of Costa Rica. We will consider a general equilibrium in the model in each of these major river basins.

The main data and their sources are as follows. Appendix B describes more details and introduces other data.

2.1 Grid cell-level geographic information

Our primary unit of empirical analysis is at the 1 km × 1 km grid cell level. We construct grid cells covering the 4 basins, regardless of the presence of community and human settlements. We use a variety of grid cell-level geographic information.

Distance matrix. We select all the grid cells within 5km from a river (with order 1-6). We then implement the Dijkstra shortest path calculations between all pairs of the selected grid cells. We calculate total upstream and downstream distances on river and land distance for each shortest path. Cadieux et al. (2020) provides the detailed explanation on the Python algorithm to implement this.

Forest cover. We classify Landsat satellite imageries from 1985, 2001, and 2015 into forest, non-forest, and masked (cloud and water) classes by CLASlite (Asner et al. 2009)[1] and aggregate them into the grid cell-level information. The forest or non-forest area within the grid cell is measured by the number of 30m × 30m pixels that were classified as forest or non-forest. From these 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 presence of secondary forests. Secondary forests are forests re-growing at the fallow phase in the shifting cultivation system after initial clearing of primary forests and cropping.

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

2.2 Peruvian Amazon Rural Livelihoods and Poverty (PARLAP) Project

The PARLAP project[2] collects our original community- and household-level data from the four major river basins in the Peruvian Amazon for key information about the population. 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 consists of the following types of data.

Community Census (CC). CC is a basin-wide census that collects community-level information from almost all communities (919 communities in total) in the four river basins. The field

[1] We are currently working on updating the forest cover data from the Google Earth Engine.

[2] The detailed information about this project is found here: `https://parlap.geog.mcgill.ca/`.

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 includes 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 this data for various purposes throughout the paper. Appendix B describes more details. Henceforth we call this data the "CC data".

Community/Household Survey (CS/HS). The PARLAP also surveyed 235 communities (out of the 919 census communities), selected by a stratified random sampling based on ethnicity, location, initial resource endowments, history, and public services. From these 235 surveyed communities, about 4000 households were surveyed.[3] 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 call these data the "CS data" and the "HS data".

Unless otherwise noted, we aggregate variables from these data into the grid cells.

2.3 National censuses

We complement the dataset using the following censuses.

Peru Population and Housing Census. Institute Nacional de Estadistica e Informatica (INEI) in Peru conducted population censuses in 1981, 1993, 2007, and 2017. The census generally contains information of total population size, total number of households, and basic demographic characteristics in the entire country although the detailed contents vary across census years. We primarily use the data from 2007 and 2017. The data from 1993 is also under construction. 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 this data the "INEI population census".

Peruvian Agricultural Census (CENAGRO). The INEI collected this data in 1994 and 2012, conducting direct interviews with agricultural producers throughout Peru. We use the data from 2012. This data contains comprehensive farm-level and plot-level information about land use, land tenure, crop choices, irrigation, marketing, livestock practices, credit access, infrastructure and machineries, labor, input use, and demographic characteristics of the producer's household members. Since the community-level unique identifier of the location of each agricultural pro-

[3]If the number of households in a surveyed community is higher than 20, then we surveyed 20 households selected by the stratified random sampling based on wealth status. Otherwise, we surveyed all households in the community.

ducer is available, we can match this data from Loreto and Ucayali departments in the Peruvian Amazon with the others. We primarily use this data to investigate mechanisms underlying density externalities in agriculture implied by the structural model.

3 Stylized Facts: Human-Nature Interactions in Rainforests

This section presents motivating descriptive and reduced-form evidence.

3.1 Spatial Distribution of Communities and Populations

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

Figure 2 indicates the concentration and dispersion of populations. The circle dots represent centroid of 1km × 1km populated grid cells in rural locations ($n = 900$). The square symbol in each basin represents the urban center. The legend of the circle dots presents quantiles of population sizes where each of the five ranges contains nearly equal number of locations. It is apparent from the legend that, in all of the four basins, more than 80% of rural locations have populations smaller than 450. Moreover, about 40% of rural locations ave very small, with populations of up to about 150. On the other hand, less 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 of populated communities. These maps show establishment of communities in the Napo and Upper Ucayali basins over decades. Every two decades, new communities are observed in both sparse and concentrated areas. Relatedly, these maps also indicate the following fact.

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

We define the rural frontier as the distance from the urban center (represented by a square in the maps of Figure 2) in each river basin. Beyond the rural frontier lands are filled by rainforests and we do not assume human settlements. Figure A.1 and Figure A.2 illustrate that, during most of every two decades, the rural frontier expanded. This process is consistent with the explanation provided by the classical model of dynamic city formation (Fujita and Krugman 1995; Fujita et al. 1999) that argues that agricultural frontier expands as population grows. In contrast to the assumption in such a classical model that areas inside the agricultural frontier are filled with farmers and farmlands, we also observe new community formations by filling in the interior of the rural frontier 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 a question of what happens to the spatial distribution of economic activity and environmental costs when the expansion of the rural frontier is limited by some protection policy. A counterfactual experiment in a later section comes back to this question.

3.2 Market Access, Human Settlements, and Forest Cover

Figure 3 illustrates the significant relationship between community location, population size, and deforestation. We present the following two stylized facts.

Fact 2A: Human settlements and forest cover changes are increasing in Market Access

We consider a simplified version of the quantitative spatial model introduced in a later section to derive tractable reduced-form empirical specifications. The full detail of this simplified model is presented in Appendix C.1. The simplification in this section comes with the following three aspects. 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 presented later. Second, we consider symmetric trade costs between locations, in contrast to asymmetric ones in the main model. Third, we incorporate the density externality from population in the own location but without spatial spillover across locations.

The one-sector model implies the following empirical specification:

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

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 controls[4] and $\phi_{B,t}$ represents basin × year fixed effects. The Market Access measure is approximated as:

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

where $o, d \in$ 1km × 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 Access measure, which is solely defined by river shapes, as an instrument variable for the Market Access:

$$RA_o = \sum_{d \in RC} \tau_{od}^{\delta} \tag{3}$$

where RC is a set of all river cells in the basin (cells that contain a river) whether or not they have positive populations. Figure 4 and Figure A.3 visually illustrates strong first-stage fits. We will exploit the variation in this River Access measure to estimate density externality parameters in a later section. In interpreting the results of the Market Access regression, on the other hand,

[4] The geographical controls include a river cell dummy, distance to the river point and its square, interaction between the previous two variables, river confluences, elevation, flood experience, geology measures, and main and non-main water channels.

we confine our discussion to be correlational from a conservative perspective as it suffices to motivate the model.[5]

Table 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 2 reports the relationship between the Market Access and forest cover change over decades.[6] The Market Access is significantly associated with forest disturbance, forest loss, 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, which is a sort of the 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.[7] This observation is consistent with the presence of congestion force in forest clearing.

Fact 2B: Per capita land footprint is decreasing in population size

The result reported in the rightmost column of Table 2 is consistent with the statement of Fact 2B. However, the grid-cell level data has a limitation to understand the relationship between human settlements and forest cover because areas people clear forest and their residential locations (reflected in the census community location) might not be exactly same. Therefore, we also look at community-buffer-level data to robustly show this fact.

Figure 5 shows the population and its increasing relationship with deforestation and decreasing relationship with per capital deforestation. In order to address the limitation with the grid cell-level observations, we measure the deforestation within 5km buffers of points of the 2007 INEI population census communities. The total population in the 5km buffer surrounding a community is measured by summing populations from communities whose centroid are inside the buffer. The upper two figures use the deforestation between 1985 and 2015. The lower two figures use the non-forest area in 2015. If we assume that Amazon river basins were fully covered by rainforests at the very beginning (i.e., before any human settlements), then this could

[5] We also check that the reduced form results are robust with different sets of controls, the RA 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.

[6] Using static measures of forest covers, Table A.1 reports that the Market Access is negatively associated with forest areas and positively associated with non-forest areas.

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

also be a reliable measure of total net deforestation associated with human settlements until today. In either way of measuring the deforestation, the figure shows the pattern consistent with Fact 2B. This pattern is robust with different buffer sizes (Figure A.4; Figure A.5).

3.3 Spatial Distribution of Agriculture and Natural Resource Extraction

We 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 forest products and wild animal extractions. Forest products consist of non-timber forest products (NTFP) and timber. Wild animal extractions include fishing and game meat (hunting).

Exploiting the HS data for detailed housedhold-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 6 and Figure A.6–Figure A.8 illustrate these facts. Population density in a smaller buffer of a community becomes close to the population density of that community itself. A smaller population density in a large buffer of a community implies more sparseness of communities and less competition over natural resources. From these figures we observe that spatial distributions of natural resource extractions become more contrastable from agricultural activities as the buffer size enlarges (x =5km, 10km compared to x =1km, 2km).

4 Quantitative Spatial Model of Rainforest Communities

We construct a multi-sector spatial general equilibrium model based on comparative advantage and trade across communities in a river basin in Amazonian rainforests. The model is built on Michaels et al. (2011) and incorporates novel sector-specific density externalities and 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.

The model has three purposes. First, the model rationalizes the stylized facts presented in the previous section. Second, we model rainforest peasants's fundamental trade off 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 policy planner's two types of trade-offs. The first trade-off is between rainforest conservation and peasants' welfare. The second type of trade-off is between different elements of conservation, such as between deforestation and bio-diversity loss.

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 $o,d \in \mathcal{I}_b = \mathcal{R}_b \bigcup u_b$.[8] 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 population within the basin. 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 consumer in o obtains wage w_o by inelastically supplying one unit of labor and 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 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:

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

expenditure shares across varieties within each sector in each location:

$$\tilde{\alpha}_{o,K}(j) = \frac{P_{o,K}^{\sigma} p_{o,K}(j)^{(1-\sigma)}}{\sum_{K'=Ag,Nr,M} P_{o,K'}} \quad K = Ag, Nr \tag{4}$$

and expenditure shares across sectors in each location:

$$\tilde{\alpha}_{o,K} = \frac{\alpha_K^{\bar{\sigma}} P_{o,K}^{(1-\bar{\sigma})}}{\sum_{K'=Ag,Nr,M} \alpha_{K'}^{\bar{\sigma}} P_{o,K'}^{1-\bar{\sigma}}} \quad K = Ag, Nr, M \tag{5}$$

[8]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.

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

4.2 Production with Density Externalities

Agricultural production with congestion and agglomeration externalities

We denote total agricultural employment in o by $N_{o,Ag}$. 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 cultivation of variety j as follows:

$$L_o(j) = qN_{o,Ag}^{-\mu_L} \cdot N_{o,L}(j), \quad o \in \mathcal{R} \tag{6}$$

where $N_{o,L}$ represents employment for forest clearing. The first factor ($qN_{o,Ag}^{-\mu_L}$) represents the productivity which depends on the common factor q and on total agricultural employment in the location. We assume the common productivity q of forest clearing across locations given limited modern technologies (e.g., chainsaw). Parameter μ_L governs the congestion force in forest clearing.[9]

The second step is to produce agricultural goods given the accessed land. We define the production function of variety j of agricultural goods as follows:

$$Q_{o,Ag}(j) \quad = \quad z_{o,Ag}(j)N_{o,Ag}^{\mu_{Ag}} \cdot N_{o,C}(j)^{\gamma} L_o(j)^{(1-\gamma)}, \quad o \in \mathcal{R} \tag{7}$$

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.

Natural resource extraction with congestion externality across space

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

$$Q_{o,Nr}(j) \quad = \quad z_{o,Nr}(j)\Big[\sum_{d \in \mathcal{R}} D_{od}^{-\nu} N_{d,Nr}\Big]^{-\mu_{Nr}} \cdot N_{o,Nr}(j), \quad o \in \mathcal{R} \tag{8}$$

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

[9] More explanation TBD

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 employments 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 surrounding population in common pool natural resource extraction. Variable D_{od} represents the river-equivalent distance between cells o and d in the shortest path. Parameter ν governs the spatial decay in accessing to natural resources. The implicit assumption underlying this specification is that people travel longer distances for natural resource extractions (than for agriculture).[10]

Urban good production at the urban center

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

$$Q_{u,M} = A_{u,M} \cdot N_{u,M} \tag{9}$$

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

4.3 Prices and Trade

We denote by $p_{od,K}(j)$ the price of product j in sector K produced in o (origin) to purchase 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 $\tau_{uu,M} = 1$.

While the above specification of iceberg trade cost is standard in the literature, note that for $K \in \{Ag, Nr\}$ $\tau_{od,K} = \tau_{do,K}$ does not necessarily hold due to river orientations. That is, trade cost between a pair of locations may be asymmetric. The trade literature often assumes symmetric trade costs, but the asymmetry would especially matter in particular environments such as transports along a river or a high-slope road and in 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 Greek where

[10] More explanation TBD

the major transport mode was sailing. Trade costs could also be endogenous. See, for example, Brancaccio et al. (2020) and a later section for more detailed discussion.

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

$$D_{od} = D_{od,down} + \lambda_{up} D_{od,up} + \lambda_{land} D_{od,land} \tag{10}$$

where $D_{od,down}$, $D_{od,up}$, and $D_{od,land}$ are the lengths (kilometer) of going downstream and upstream on the river and the length of land travel. λ_{up} and λ_{land} are the parameters capturing the downstream-river-equivalent distance per one 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:

$$\tau_{od,K} = D_{od}^{\delta_K} \tag{11}$$

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, $p_{oo,Ag}(j) = \frac{w_o}{z_{o,Ag}(j) N_{o,Ag}^{\tilde{\mu}_{Ag}} \kappa_1}$ 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.[11] 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[12]:

$$\pi_{od,K} \;=\; \frac{\tilde{A}_{o,K}(w_o \tau_{od,K})^{-\theta}}{\sum_{o'\in\mathcal{R}} \tilde{A}_{o',K}(w_{o'}\tau_{o'd,K})^{-\theta}}, \quad o \in \mathcal{R},\, d \in \mathcal{I} \tag{12}$$

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

$$\begin{aligned} \tilde{A}_{o,Ag} &\equiv A_{o,Ag} N_{o,Ag}^{\tilde{\mu}_{Ag}\theta} \kappa_1^{\theta} \\ \tilde{A}_{o,Nr} &\equiv A_{o,Nr}\Big[\sum_{d\in\mathcal{R}} D_{od}^{-\nu} N_{d,Nr}\Big]^{-\mu_{Nr}\theta} \end{aligned} \tag{13}$$

For the urban good, all locations purchase it only from the urban center u.

[11] Specifically, $\kappa_1 = q^{(1-\gamma)}\gamma^{\gamma}(1-\gamma)^{(1-\gamma)}$ and in general κ_k for $k = 1, 2, ...$ will also consist of constant terms.

[12] = 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

4.4 Spatial Equilibrium

The spatial equilibrium is defined as the following conditions.

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

$$w_o N_{o,Ag} = \sum_{d \in \mathcal{I}} \pi_{od,Ag} \tilde{\alpha}_{d,Ag} w_d N_d \quad \forall\, o \in \mathcal{R} \tag{14}$$

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

$$w_o N_{o,Nr} = \sum_{d \in \mathcal{I}} \pi_{od,Nr} \tilde{\alpha}_{d,Nr} w_d N_d \quad \forall\, o \in \mathcal{R} \tag{15}$$

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

$$w_u N_{u,M} = w_u N_u = \sum_{d \in \mathcal{I}} \tilde{\alpha}_{d,M} w_d N_d \tag{16}$$

4. The overall labor market clears:

$$\overline{N} = \sum_{o \in \mathcal{R}} N_o + N_u = \sum_{o \in \mathcal{R}} \sum_{K \in \{Ag,Nr\}} N_{o,K} + N_{u,M} \tag{17}$$

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

$$V_o = \frac{w_o}{\left[\sum_{K=Ag,Nr,M} P_{o,K}^{1-\bar{\sigma}}\right]^{\frac{\bar{\sigma}}{\bar{\sigma}-1}}} = \overline{U} \quad \forall\, o \in \tilde{\mathcal{I}} \tag{18}$$

   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 geography, productivity fundamentals, exogenous parameters, and total population in the economy.

5 Estimating the Model

Table 3 summarizes the parameters of the model. We estimate the parameters and validate the model in the following sequential steps.

5.1 Estimating 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 collects prices of representative products of agricultural and natural resource sectors in the census communities (see Appendix B.2 for detail). 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 $(p^{Max}_{od,K} - D^{\delta_K}_{od,down})$ where $p^{Max}_{od,K} \equiv \max_j\{p_{d,K}(j)/p_{o,K}(j)\}$ is the maximum price ratio across all the products in each sector among location pairs that satisfy $D_{od,up} = D_{od,land} = 0$. We take the maximum price ratio for the following two reasons. First, this way mitigates the concern that observed price gaps of each product might underestimate trade costs, given our limited price information[13] and that price ratios between two locations are bounded above by trade costs between them (Eaton and Kortum 2002). Second, we assume $\tau_{od,K}$ to be the cost of transport by the most-widely available transport mode ('peque-peque', shown in the top-right picture of Figure A.9). However, the observed price ratio may be reflecting transports by other superior boat types (e.g., the bottom two pictures of Figure A.9) even if the location pair consists of actual origin and destination, 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. Higher values of trade costs 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.[14] We obtain $\hat{\lambda}_{land} = 36.767$, using records of transportation costs by land travels. Appendix D.1 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 $\tau_{od,K} = \hat{D}^{\hat{\delta}_K}_{od}$ where $\hat{D}_{od}$ is calculated by (23) using $\hat{\lambda}_{up}$ and $\hat{\lambda}_{land}$.

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

[14] The downstream-river-equivalent distance measure specified by (23) implicitly assumes the constant slope everywhere. Future work will construct a more sophisticated measure of the effective distance taking into account river orientations with slopes in each travel route.

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 D.1 describes the data and provides the detail procedure to estimate these parameters. We obtained $\hat{\sigma} = 2.401$ (substitutes since $\hat{\sigma} > 1$) and $\hat{\bar{\sigma}} = 0.752$ (complements since $\hat{\bar{\sigma}} < 1$).[15]

Other parameters

We assume the labor cost share in the agricultural production to be $\hat{\gamma} = 0.6$, 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 the productivity dispersion (the shape parameter) in the agricultural production and natural resource extraction to be $\hat{\theta} = 7.8$, drawn from Donaldson (2018) that studies agricultural trade in the context of developing economies.

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 that rationalize the observable data (sectoral employment share and total population in each location) as a spatial equilibrium. Given that the inverted productivity composites contain both exogenous and endogenous terms, the model inversion is technically equivalent to and easier than simulating the model.

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. See Appendix B.2 for the construction of the sectoral employment shares. 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 (14), (15), (16), and (18). The $2|\tilde{\mathcal{R}}|+1+|\tilde{\mathcal{I}}|$ unknowns are $\{\tilde{A}_{o,Ag}, o \in \tilde{\mathcal{R}}\}$, $\{\tilde{A}_{o,Nr}, o \in \tilde{\mathcal{R}}\}$, $A_{u,M}$, and $\{w_o, o \in \tilde{\mathcal{I}}\}$.

The model inversion reduces to the nested fixed point problem. The algorithm is described in Appendix D.2. By construction of the inversion problem, the model perfectly fits the data of sectoral employments. Proving the uniqueness of the solution of this inversion problem follows Michaels et al. (2011).

[15]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 $(\sigma, \bar{\sigma})$ are qualitatively robust.

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

5.3.1 Density externalities in agriculture

Net density externality in productivity composite in agriculture

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

$$\ln \tilde{A}_{o,Ag} = \beta_0 + \tilde{\mu}_{Ag}\theta \ln N_{o,Ag} + \mathbf{b}X_o + \phi_B + \epsilon_{o,Ag} \tag{19}$$

where $\ln N_{o,Ag}$ is logarithm of agricultural employment at location o and X_o is a vector of exogenous productivity shifters. $\ln N_{o,Ag}$ is likely to be endogenous, since demand of agricultural labor would be higher at locations that have higher productivity fundamentals but there might remain unobservable elements captured in $\epsilon_{o,Ag}$. In order to estimate the coefficient of our interest, $\tilde{\mu}_{Ag}\theta$, we instrument $\ln N_{o,Ag}$ by $\ln RA_o$ where RA_o is the River Access measure defined by (3).

In addition to the fact that river shape is exogenous and not affected by human settlements, we need the following identifying assumption. Controlling for own-location characteristics, productivity fundamentals are uncorrelated with accessibility to other locations. That is, market opportunity (due to the accessibility to other locations) affects productivity only through its effect on employment (and thus through externalities arisen), rather than through productivity fundamentals. The corresponding moment condition is $\mathbb{E}[\epsilon_{o,Ag} \ln RA_o] = 0$.

Figure 7 provides the intuition behind this identification assumption. 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. The ideal experiment to estimate the causal effect of population on productivity is to randomize population size across these three cells, which is not feasible to implement. We assume that these three cells have the same agricultural productivity fundamentals (e.g, soil quality, distance to the nearest river point, lake size nearby, elevation, climate conditions) that affect labor demand. The only difference between these cells in the figure is the River Access measure, i.e., 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, the River Access is higher at the red and brown cells than that at the pink cell. Comparing the red cell in the middle map with the brown cell in the right map, the River Access measure is higher at the brown cell because the additional cell next to the top-right is located closer to the brown cell. This variation in the River Access measure, stemming from the exogenous river shape in nature, is an exogenous shifter of market potential. All else equal, this variation is the only source that affects the difference in population size due to the trade mechanism. Even if there are unobservable productivity shifters remaining in the own location, there would be no theoretically-plausible reason to believe that they are associated with the variation in the River Access measure that can be generated by river shapes in locations very far away from the own location.

Table A.2 reports the IV estimation result. In our preferred specification, the point estimate $\tilde{\mu}_{Ag}\theta$ is 0.570 at the 1% level of statistical significance. Dividing this value by $\hat{\theta}$, we obtain the reported point estimate $\hat{\tilde{\mu}}_{Ag} = 0.071$ (standard error = 0.009) in Table 4. This result confirms the presence of agglomeration externality in the agricultural sector *on net*.[16] Note that this result is not being driven by a small set of locations with very large populations. We obtain similar results by restricting the sample locations to only those with small populations (e.g., < 1000). Table A.2 also shows that the OLS estimate is larger than the IV estimate. This observation is consistent with the view that agricultural employment is correlated with agricultural productivity unobservable by econometrician and thus the OLS estimate has an upward bias.

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

Recall 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.

In order 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 the grid cell-level data. In particular, we use the data of community-level working area after clearing forest, constructed by Coomes et al. (2021) and shown in Figure A.11.

From the land access function (6) and $N_{o,L} = (1-\gamma)N_{o,Ag}$ at the optimal input choice, we

[16] *1. Contrast the empirical result with one-sector model, in which the congestion externality dominates on net. 2. Comparison with the estimates of agglomeration externality in the urban economics literature. 3. Investigate the heterogeneity.*

obtain the following empirical specification:

$$\ln \frac{L_o}{N_{o,Ag}} = \beta_0 - \mu_L \ln N_{o,Ag} + \mathbf{b} X_o + \phi_B + \epsilon_{o,L} \tag{20}$$

and to estimate μ_L we follow the similar identification strategy as above. We use the same instrumental variable and the same set of geographical controls. We obtain the point estimate $\hat{\mu}_L = 0.498$ (standard error = 0.090).[17]

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) $\hat{\mu}_{Ag} = 0.271$ reported in Table 4.

5.3.2 Congestion externality in natural resource extraction

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:

$$\epsilon_{o,Nr} = \ln \tilde{A}_{o,Nr} + \mu_{Nr}\theta \ln\Big(\sum_d D_{od}^{-\nu} N_{d,Nr}\Big) - \mathbf{b} X_o - \phi_B \tag{21}$$

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:

$$\mathbb{E}[\epsilon_{o,Nr} \ln RA_o] = 0 \quad \text{and} \quad \mathbb{E}[\epsilon_{o,Nr} \ln(\sum_{d|D_{o,d} \leq x} RA_d)] = 0,\ x \in \mathcal{X} \tag{22}$$

Panel (B) of Table 4 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). We are also working on the iterative

[17]To do: *Investigate the heterogeneity in density externalities by population size. A simple example is that it would be more productive to cut trees if two people cooperate than doing it by one person, while there would not be congestion with only two people in the community. Therefore, the net congestion externality (congestion - agglomeration) would increase in population. Having this heterogeneity will not change the story and qualitative findings of this paper, but it will change the magnitude of quantitative results.*

GMM estimation (Hansen and Lee 2021) and the result will be available soon.

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:

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

and employ a linear specification to estimate μ_{Nr} and $\{\mu_{Nr,x}\}$ with instruments $\ln RA_o$ and $\ln \sum_{d|D_{o,d}\leq x} RA_d$ for $x \in \mathcal{X}$. Table A.6 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 population engaging in natural resource extraction in the own location becomes weaker both economically and significantly than 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 turns to even 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 is consistent with the view that employment of natural resource extraction is correlated with its productivity unobservable by econometrician 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 $\hat{\mu}_{Nr,x_{max}}$ becomes gradually decreasing. All of their signs are kept same up to $x_{max} = 100$, but this regular pattern diminishes with $x = 150$. This empirical pattern implies that the strength of congestion 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 $x = 100$.

5.4 Non-Targeted Data and Moments (Step 4)

We first note that, although section 3 (the stylized facts) and this section use different data, the findings are internally consistent between these two sections. Table 6 reports significant positive correlation between reported number of species found around a community (from the CC data) and the calibrated productivity of the natural resource sector. More detail will be added soon.

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 and natural resource extractions is straightforward given its rivalrous nature. On the other hand, 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 an urban setting.

6.1 Economies of Scale in Transport and Transaction Costs

One hypothesis is that the cost of transporting products from a community is decreasing in the community's population producing and exporting 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.9) 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 (e.g., 'rapido', shown in the bottom-left picture of Figure A.9). Another simple possibility could be that the average transport cost charged is decreasing in the amount of transported products given the same transport mode. In any ways, the actual transport costs could be endogenous depending on the exporting community's population. On the other hand, the model assumes exogenous river transport costs that 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.9).

However, 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 now that there is no agglomeration externality in agricultural productivity such that we replace (7) 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 (6). Suppose also that the trade cost of agricultural products is characterized as:

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

where $\tau_{od,Ag}$ is the exogenous term of trade cost defined by (11) and $\mu_{Ag} > 0$ so that the trade cost is decreasing in the agricultural employment. These assumptions lead to the following expenditure share across locations within the agricultural sector:

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

which is same as (12) where $\tilde{A}_{o,Ag}$ is same as (13). 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 allows us to test this hypothesis. Table 7 reports the results of estimating the scale effect on the availability of different transport modes in a community. First of all, the mean value of peque-peque availability is close to one (0.972) and much higher than that 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) and (6), a 1% increase in the agricultural employment increases the probability of lancha and rapido being available in the community by 0.17% and 0.06% at the 1% and 5% levels of statistical significance, respectively.[18]

Table A.9 shows the scale effect on the frequency of these transport modes available. According to column (2), a higher agricultural employment significantly increases the number of days when lancha is available to the community. The mean number of days available per week is 3.429. Lancha is a commercial large boat and 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. On the other hand, rapido is available almost everyday (6.164/7) among communities that have access to it (and the scale effect is insignificant). This observation would be more consistent with the case where the community owns it as a result of collective investment.

A related mechanism is that other transaction costs decrease in population size. According to Table A.10, probabilities that a river trader is available to the community, community population is being contracted for selling some product, or there are contractors living in the community are significantly higher with higher community population.[19] 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.[20]

6.2 Economies of Scale in Agricultural Intensification

Another hypothesis is that there are economies of scale in accessing agricultural inputs. The 2012 Peruvian Agricultural Census (CENAGRO) allows us to test this hypothesis. Table 8 reports the scale effect on the household-level use of various agricultural infrastructure, technologies, and inputs into land and crops in the communities in our study area.

The first important fact from this table is that modern technologies are 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/canue/speedboat having about 0.6. Moreover, the mean value is less than 0.01 for 13 items (out of 24).

[18] Foster and Gafaro (2017) also found a consistent finding.

[19] Table A.11 reports product-specific contracts.

[20] 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.

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 (insecticides, herbicides, and fungicides) as well as on their complementary equipment (sprayers) have higher point estimates than others at 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.

The CC data from PARLAP collects detailed information on the form of seed acquisition. It asks whether people in the community obtain their seeds or other planting materials from others in the community, from people in other communities, or from a city. It also asks the principle form of acquisition if they obtain seeds from other places. Table A.12 reports the results. The point estimates reported in panel (A) imply that as 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 other communities[21], though 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 significant vary in community population.[22] 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 resettlement and 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 primary outcome variables at each basin level are welfare (of the population), total deforestation, and natural resource depletion. We proxy the measure of natural resource depletion in each basin by the total expected production amount (not value) of the natural resource extraction sector in the spatial equilibrium, denoted by $Q(Nr)$. 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

[21] Table A.13 reports crop-specific results and that this effect largely stems from maize and plantain.

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

experiments we use the same simulation algorithm, described in Appendix D.3, that aims for reaching the spatial equilibrium closest to the benchmark equilibrium given the set of populated locations. We 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 Quantifying the Agglomeration Externality

We first quantify the effects of agglomeration externality on human welfare and environmental costs. To do so, 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 keep the congestion externalities in forest clearing and natural resource extractions.

Table 9 indicates that the agglomeration externality has large effects. Without the agglomeration externality, deforestation increases in about 20-60% but total agricultural production decreases in about 30-35% from the benchmark equilibrium. The human welfare is about 8-13% lower without the agglomeration. That is, policy interventions that enhance agglomeration benefits achieve the win-win outcome of decreasing deforestation and increasing human welfare. On the other hand, the natural resource depletion also decreases in about 1-3% without the agglomeration externality. That is, there remains another trade-off between two types of environmental costs: deforestation vs. bio-diversity loss.

Figure 8 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.[23] As the upper middle maps indicate, agricultural productivity has decreased 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. There is also a sectoral reallocation of workers: 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 the general equilibrium feedback effect. Since the agricultural productivity is declined, the economy needs more agricultural employment to satisfy the demand of agricultural goods of the population.

7.2 Resettlement Policies

In this section, given the fixed number of total population and estimated density externalities, we ask how alternative spatial allocations of human settlements affect welfare and environ-

[23] 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.

mental outcomes. To understand this question, we implement the following resettlement policy experiments. We simulate the model to derive endogenous outcomes given the additional condition imposed by each experiment. Note that the model is not incorporating the cost of (or subsidy to compensate for it) associated resettlements 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 population in the benchmark equilibrium in each basin.

(A) Not allowing new community formation

We start with a simple experiment of not allowing new community formation. Figure A.1 and Figure A.2 show establishment of communities in the Napo and Upper Ucayali basins over decades. Modeling the dynamic process of community formation is left for future research. 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 order, starting with newly established communities, until the treated population reaches 2.5% of total population in each basin. We then solve the model with the rest of communities established earlier given the total population.

Panel (A) of Table 10 reports the results. In the Napo basin, total deforestation decreases in 8.3% with small a decrease in welfare (0.2%). Similar results follow in the other three basins. This finding implies that, unlike enhancing the agglomeration externality, just reducing the number of populated locations entails the trade-off between human welfare and forest cover gain. Figure 9 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.

(B) Minimum population threshold to form a community

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

Panel (B) of Table 10 reports the results. The welfare outcome is almost same as the previous experiment. On the other hand, targeting residents of small communities for resettlement leads to a higher gain in preserving forest cover. In all of the four basins, this policy has reduced deforestation more that the previous experiment. In the Napo basin, this policy has reduced deforestation by 12.5%. The right five columns in (A) and (B) indicate that the outcomes of sectoral reallocation of the population and sectoral outputs are almost same as 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 presence of agglomeration externality. The 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 10 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.

(C) Protected areas that control rural frontier expansion

We consider another place-based policy of setting protected areas where there is a limit on the expansion of rural frontier (recall Fact 1B in section 3). 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 total rural population in each basin.

Panel (C) of Table 10 reports the results. The welfare outcome is almost same as the previous experiment. This experiment has also reduced deforestation with a similar mechanism as the previous experiment. Compared to the previous experiment that targets small-scale communities, however, the forest cover gain is smaller. In the Napo basin, this policy has reduced deforestation by 5%. On the other hand, this policy has reduced natural resource depletion while the previous one has increased it. The intuition is as follows. With the protected areas targeting the rural frontier, the overall scope of natural resource extraction activities is narrowed. The experiment increased the overall population density in the basin as it relocated the population closer to the urban center. This reallocation means that surrounding populations increase in most of populated areas. Therefore, overall the productivity of the natural resource sector would be negatively impacted due to 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 11 and Figure A.15 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. Comparing the bottom-middle map of Figure 11 with that of Figure 10 indicates the primary source of the contrasted outcomes from the two experiments that we described above. 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 this experiment, while it becomes larger than 0.97 for all of the remaining locations by the previous experiment. To summarize, the comparison between policies (B) and (C) illustrates that policymakers face the ecological trade-off in mitigating different types of environmental costs:

deforestation and other natural resource depletion.

7.3 Improvement of Transport Infrastructure

We consider improving the transport infrastructure. Recall (23) where we defined the downstream-river-equivalent kilometer along the lowest-cost route. With the improved transport infrastructure, the downstream-river-equivalent distance is expressed as follows:

$$D_{od} = D_{od,down} + \lambda_{up} D_{od,up} + \lambda_{land} D_{od,land} + \lambda_{road} D_{od,road} \tag{23}$$

where $D_{od,road}$ is the upgraded part of the transport network that was previously a river line. The infrastructure improvement has two elements. The first element is that there is no asymmetric cost in the road part. The second element is that the travel cost of the road part decreases relative to the downstream river. To capture this effect, we set $\lambda_{road} = 0.8$.

Table 11 reports the results of a variety of experiments. Reducing the trade costs increases human welfare in all the cases. However, effects on environmental costs are heterogeneous across different ways of improving the transport infrastructure.

Infrastructure investments that make the areas enjoying agglomeration externalities more evenly 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 11 reports the results of improving the transport infrastructure in a way that it connects hinterlands to the central area of basin (by targeting river lines with river order 2). In all of the four basins, total deforestation is decreased by this experiment. Figure 13 and Figure A.16 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 from the benchmark equilibrium (the top-left maps). The agglomeration externality in agricultural production amplifies this effect. 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 11 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). Dues 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 understand the importance of place-based targeting of infrastructure investments to mitigate environmental costs. Panel (C) of Table 11 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 11 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 evenly across the basin are preferable in terms of reducing deforestation. In other words, policies causing more middle-sized communities evenly 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 resettlement experiment 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 11; Figure 13; Figure A.16). 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 two maps in the left. This difference is due to the different nature of density externalities between agriculture and natural resource extraction. Dues 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 in 0.3% in Upper Ucayali but it decreases in the other three basins.

To combat the natural resource depletion, one might consider a sector-specific protection policy such as banning hunting in specific locations. However, its feasibility is unrealistic and even more difficult than resettlement policies. It would be more difficult to monitor people's activities than whether they 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 11 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.

7.4 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 11 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 contrary, the natural resource depletion is increased in the Upper Ucayali basin even with the same model, the same parameter values, and the similar study area (compared to other rainforest areas outside the Peruvian Amazon). This example illustrates the caveat in deriving conclusions from a single economy.

Therefore, this study states conclusions if and only if primary outcomes from all of the four basins are qualitatively same. Although it is not still perfect to argue the external validity, this way would improve the credibility of policy implications compared to those derived from a single economy. To achieve the win-win outcome, we consider a composite intervention that combines the resettlement policy with the improvement of transport infrastructure. Table 12 reports results of the composite counterfactual experiment that contains the following two components. The first component is it sets protected areas where there is a limit on the expansion of rural frontier by directly targeting 2.5% of total rural population in each basin. The second component is it improves the transport infrastructure in a way that it connects hinterlands to the central area of basin (by targeting river lines with river order 2). Each component is exactly same as a previous experiment. According the table, this policy is a win-win one because it increases welfare for about 1.1-2.3%, decreases deforestation for about 5-7%, and decreases natural resource depletion for about 0.5-3%. 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.[24]

[24]*Note that we have not incorporated the potential negative externality of developing transport infrastructure on natural resource endowments. It is not feasible to estimate this externality. Also, I have not found rigorous empirical estimates*

8 Conclusion

This paper models human settlements in Amazonian rainforests and studies trade-offs between human and ecological well-being quantitatively, using a dataset from rainforest communities in major river basins of the Peruvian Amazon. Counterfactual experiments based on the novel density externality parameters imply the complementarity between a protection policy and investments in transport infrastructure to improve human and ecological well-being in rural areas.

from the previous literature though Madhok (2022) is very related. Therefore, I will set different values of this additional parameter and show simulation results to determine a threshold parameter value at which the environmental costs exceed the economic gains from infrastructure investments. I would appreciate any ideas to approach this issue.

References

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 and Costas Arkolakis, "Trade and the Topography of the Spatial Economy," *The Quarterly Journal of Economics*, 2014, *129* (3), 1085–1140.

__ , __ , 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.

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

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 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.

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

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.

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

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.

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.

Hollingsworth, Alex, Taylor Jaworski, Carl Kitchens, and Ivan J Rudik, "Economic geography and the efficiency of environmental regulation," Technical Report, National Bureau of Economic Research 2022.

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.

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.

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.

Jégard, Martin, "An Optimal Distribution of Polluting Activities across Space," 2021.

Madhok, Raahil, "The Development-Biodiversity Tradeoff in India's Forests," 2022.

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

Miyauchi, Yuhei, "Matching and Agglomeration: Theory and Evidence from Japanese Firm-to-Firm Trade," *Working Paper*, 2021.

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.

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

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

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

Porteous, Obie, "High trade costs and their consequences: an estimated dynamic model of african agricultural storage and trade," *American Economic Journal: Applied Economics*, 2019, *11* (4), 327–66.

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.

Robalino, Juan A and Alexander Pfaff, “Contagious development: Neighbor interactions in deforestation,” *Journal of Development Economics*, 2012, *97* (2), 427–436.

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

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

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.

Yamada, Eiji, “A Spatial Equilibrium Analysis of Air Pollution in China,” Technical Report, JICA Research Institute 2020.

Figures and Tables

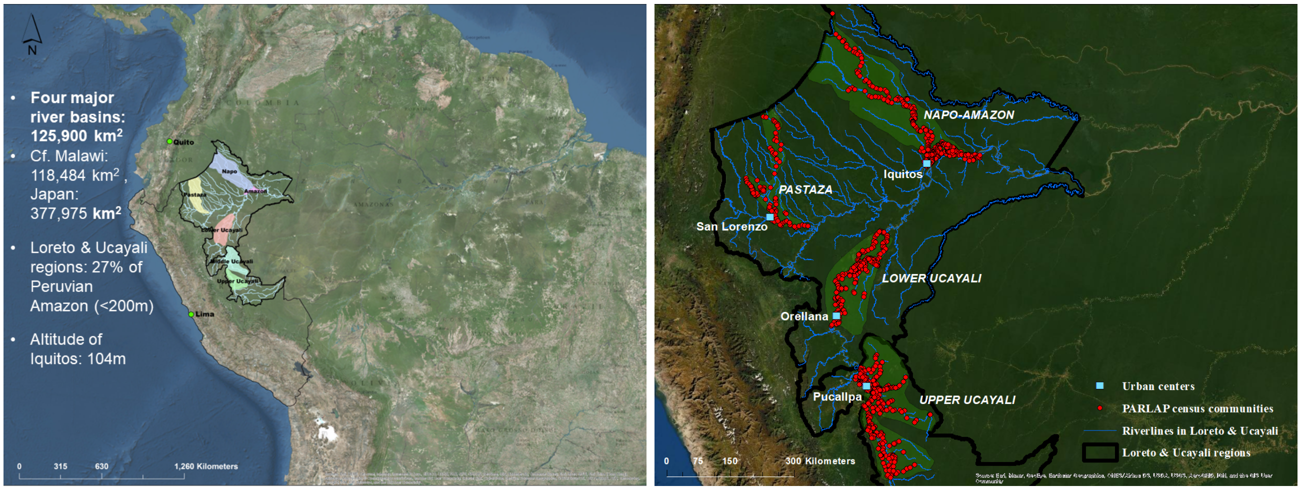

Figure 1: Study Area

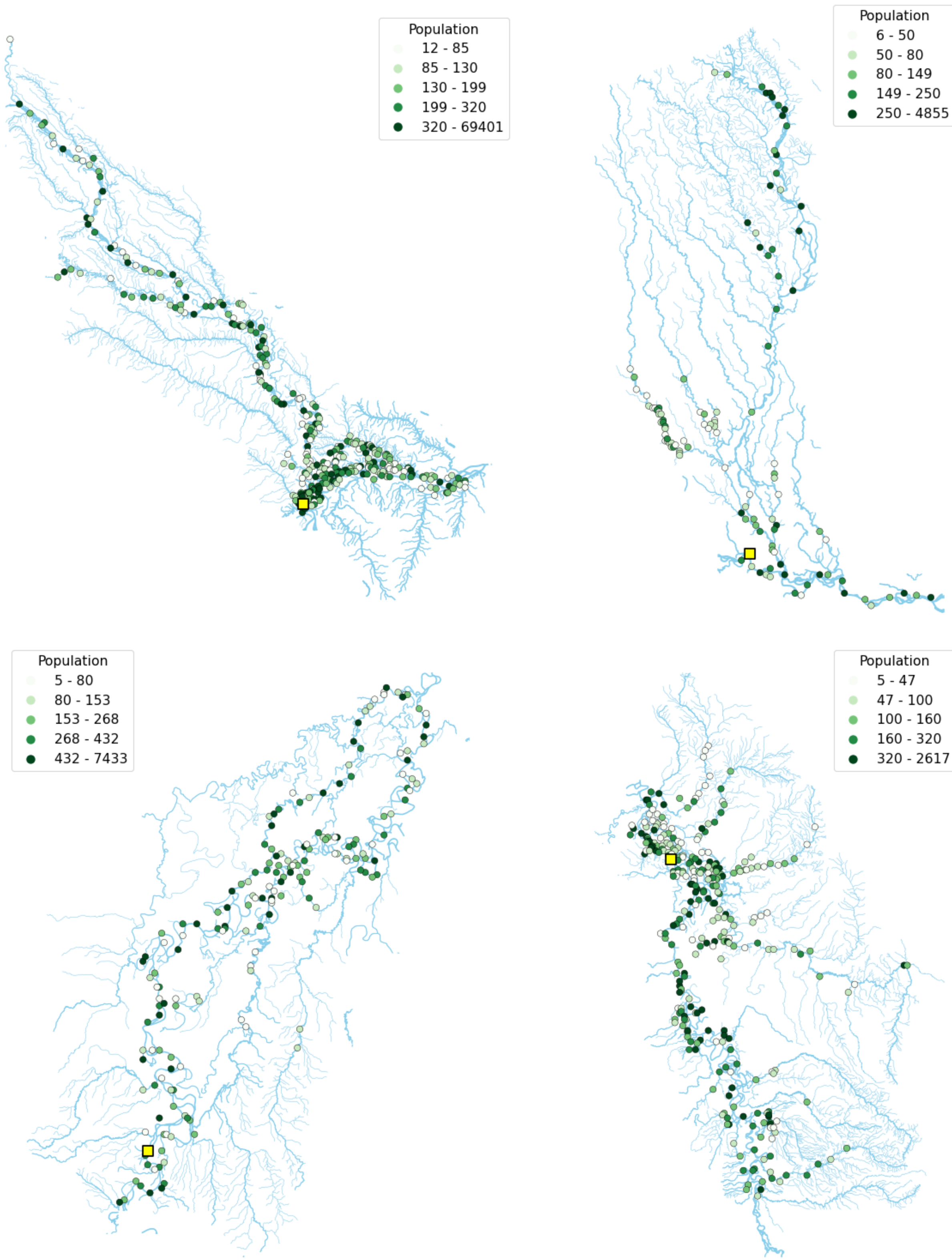

Figure 2: Population in Rural Locations

Notes: Each map shows population in rural locations in each basin (upper-left = Napo-Amazon; upper-right = Pastaza; bottom-left = Lower Ucayali; bottom-right = Upper Ucayali) of the study area. Light-blue squares represent the urban centers.

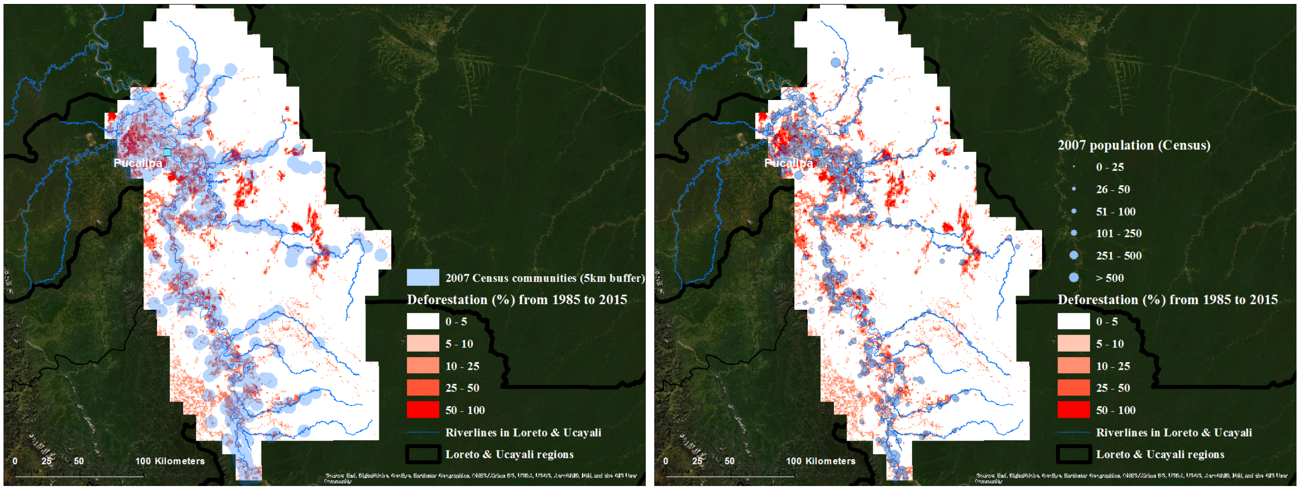

Figure 3: Communities, Populations, and Deforestation

Notes: The maps show the Upper-Ucayali basin.

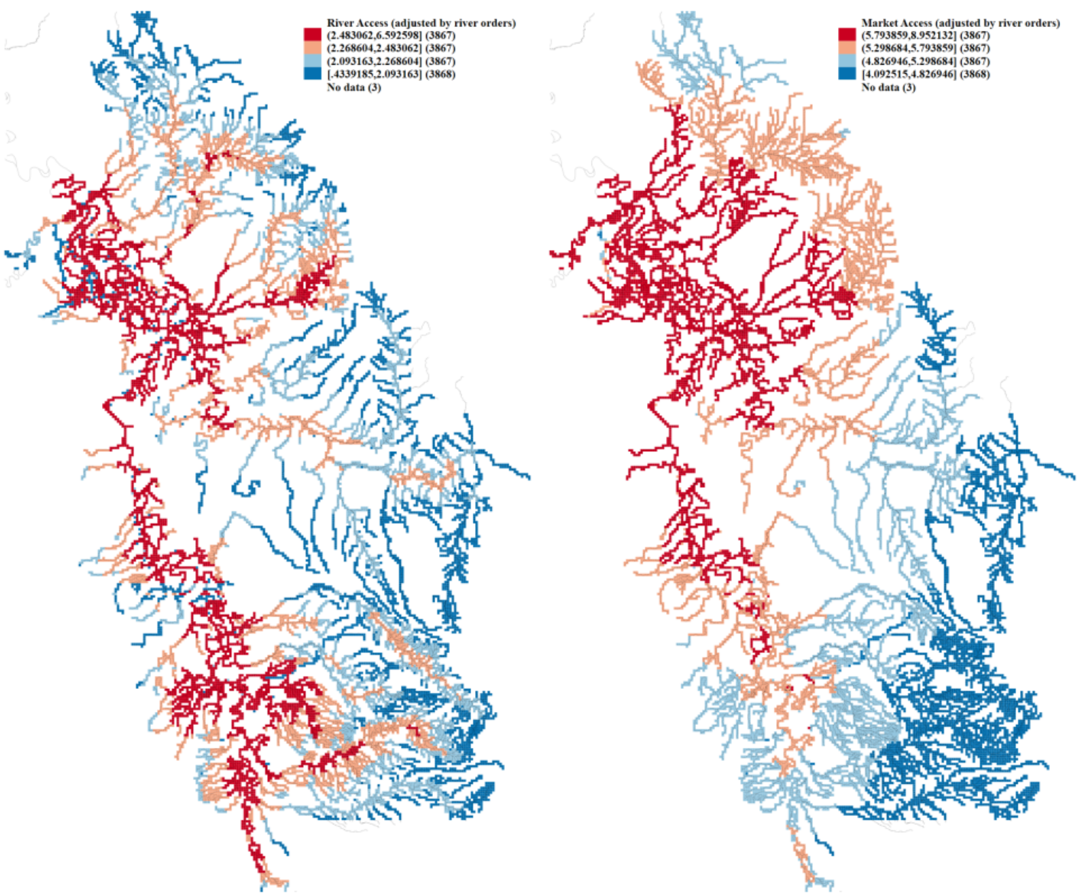

Figure 4: River Access and Market Access at River Cells (Upper Ucayali)

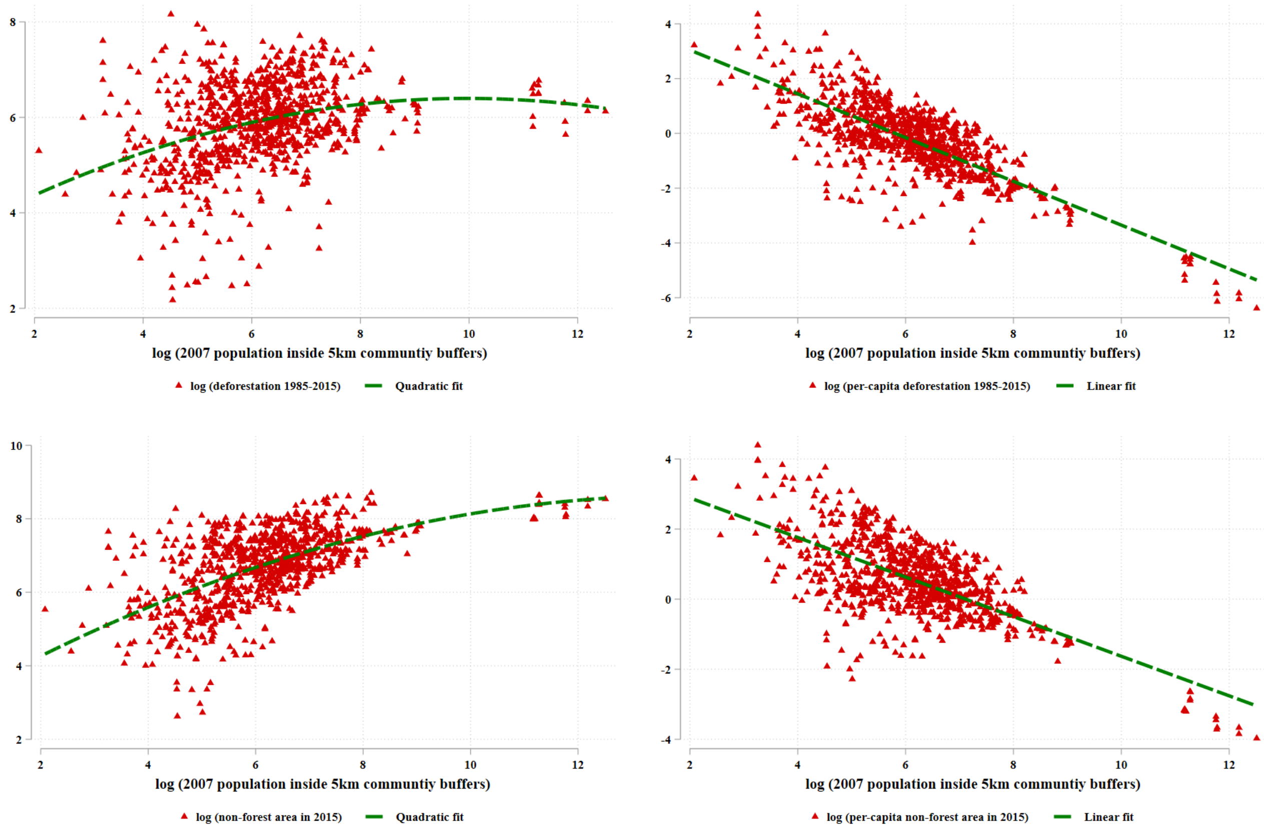

Figure 5: Populations and (Per Capita) Deforestation

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

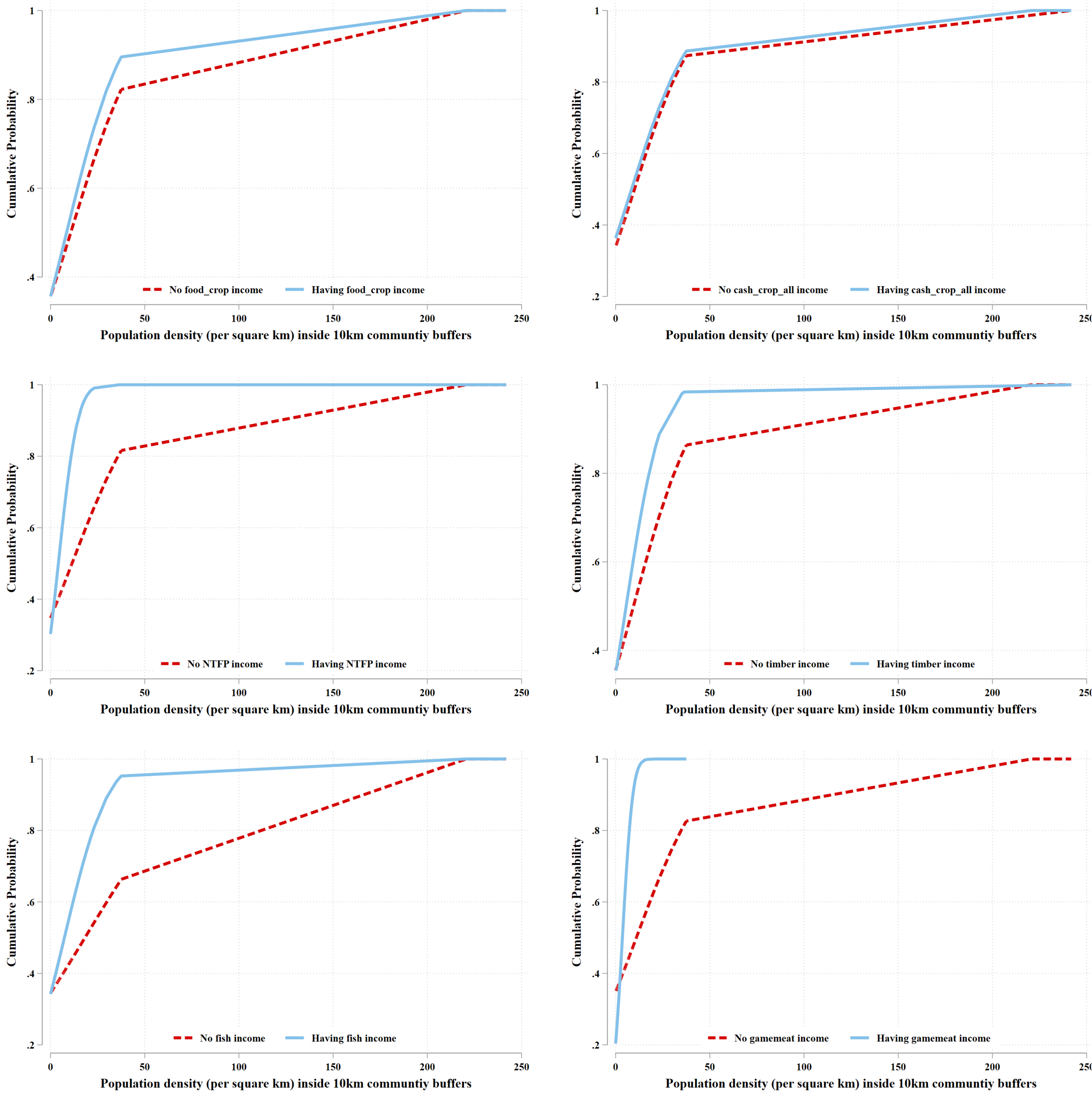

Figure 6: Population Density (10km buffer) and Distribution of Activities (Household-level)—Agriculture vs. Forest Products vs. Wild Animal Extractions

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.

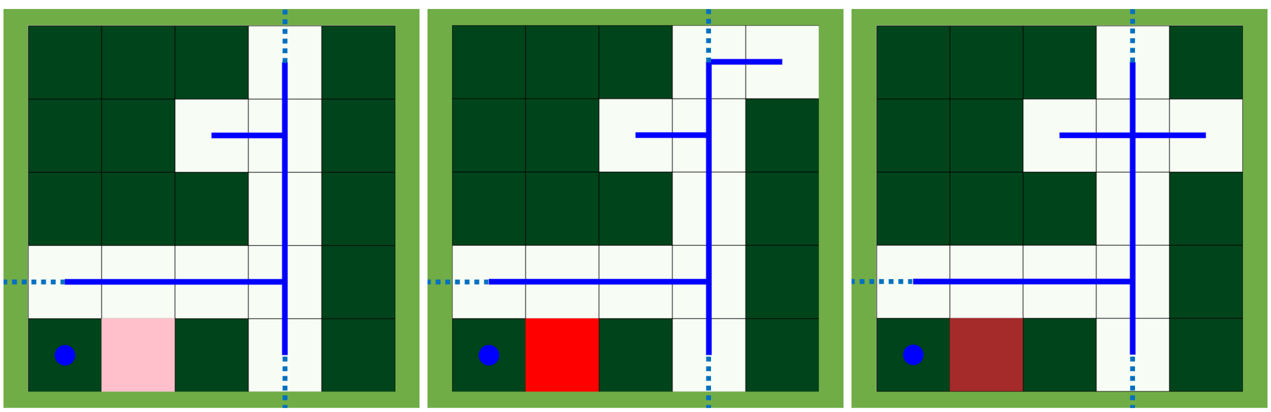

Figure 7: Intuition of Identifying the Density Externalities

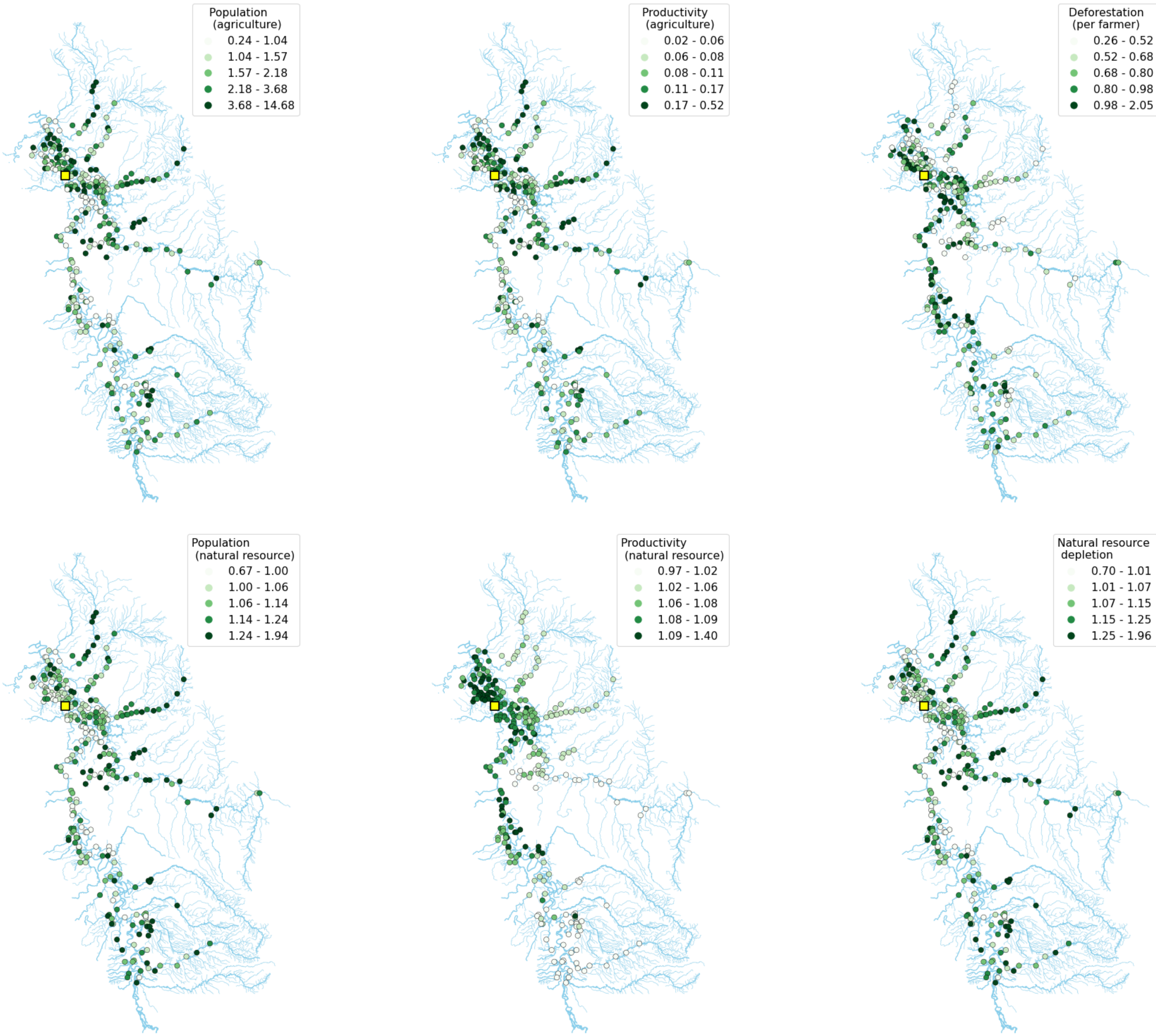

Figure 8: Counterfactual Outcomes without the Agglomeration Externality (Upper Ucayali)

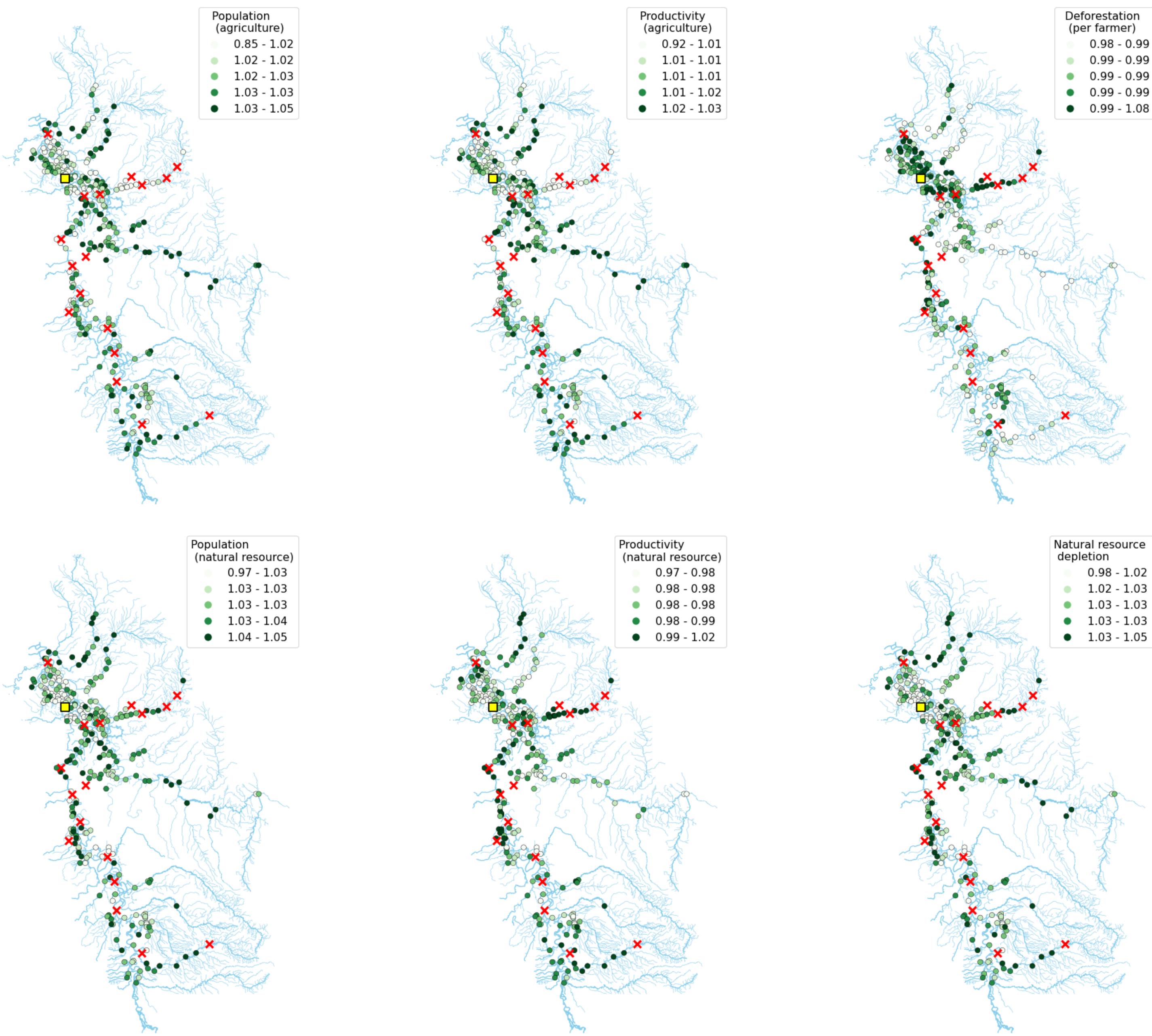

Figure 9: Counterfactual Outcomes without New Community Formation (Upper Ucayali)

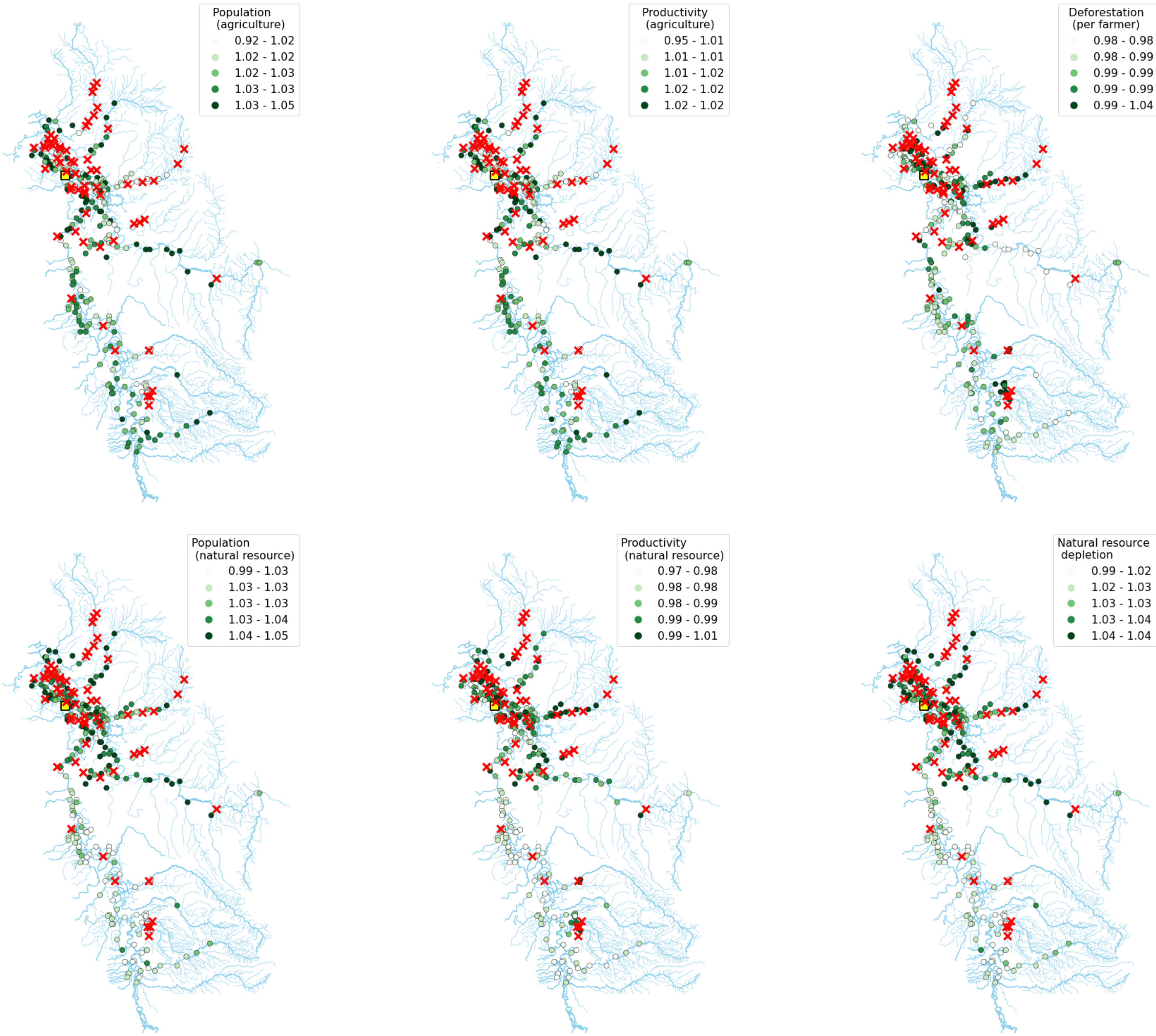

Figure 10: Counterfactual Outcomes without Small Communities (Upper Ucayali)

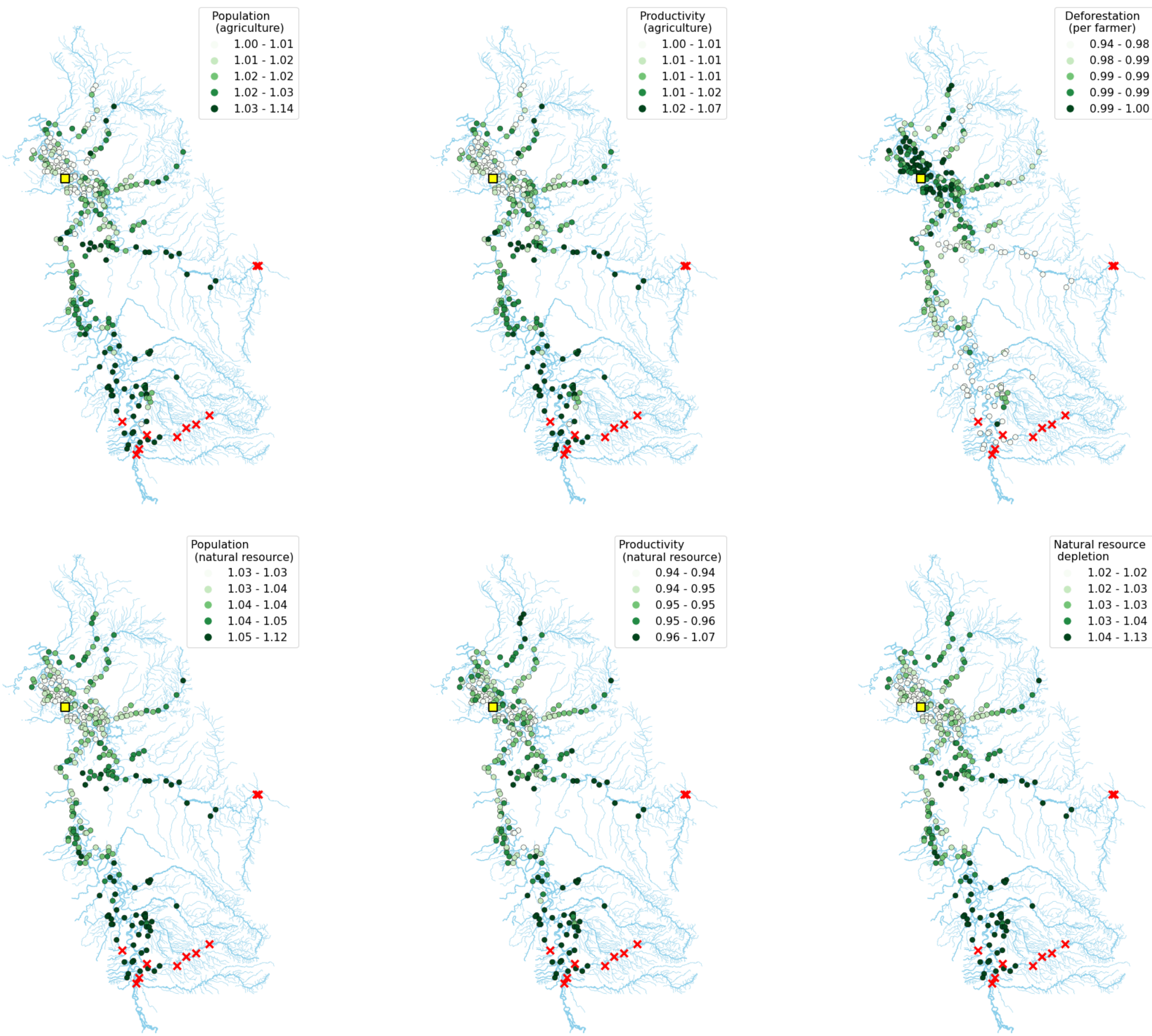

Figure 11: Counterfactual Outcomes with Limited Rural Frontier (Upper Ucayali)

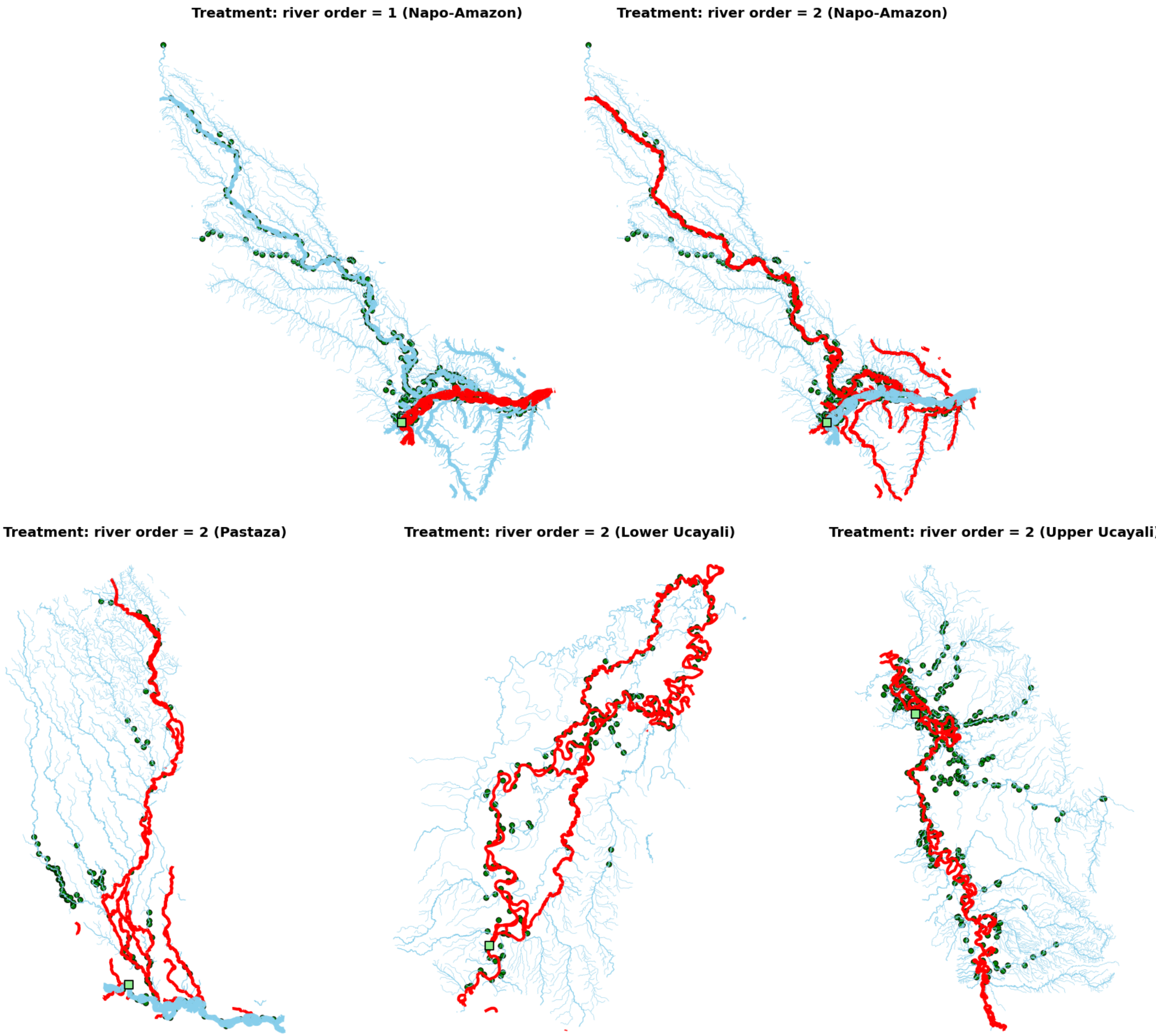

Figure 12: River Orders and Counterfactual Experiments of Improving Transport Infrastructure

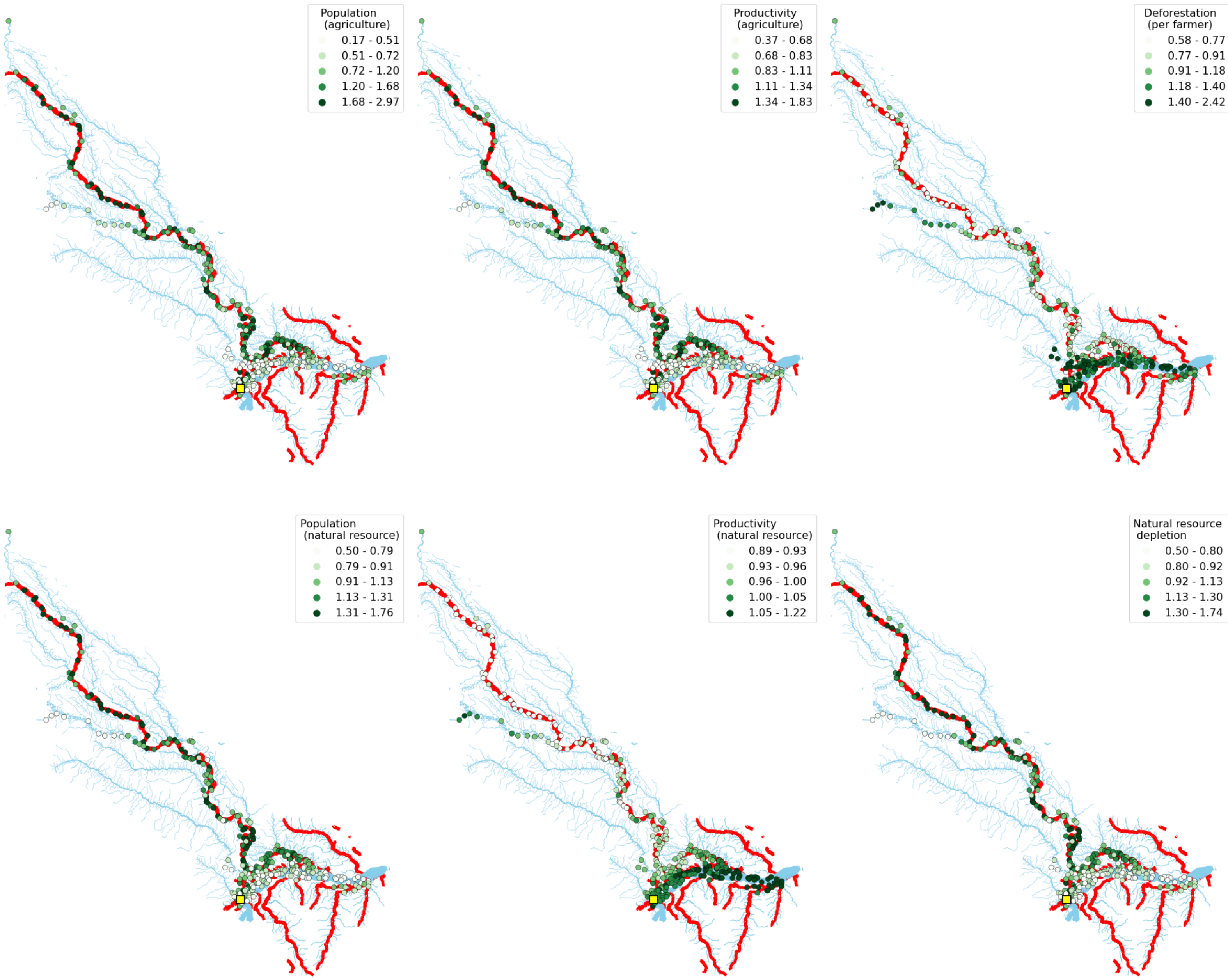

Figure 13: Counterfactual Outcomes with Improved Transport Infrastructure along River Order 2 (Napo-Amazon)

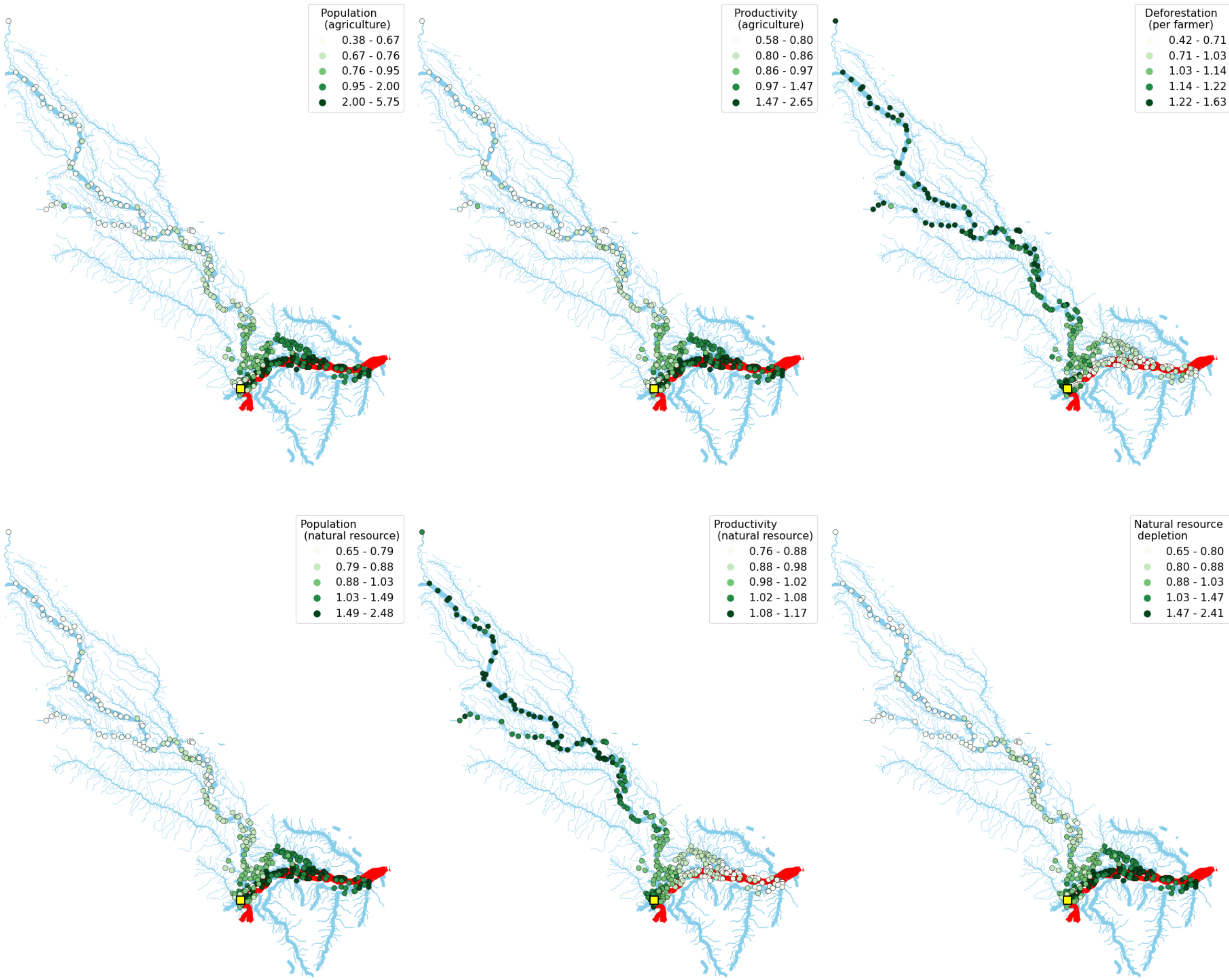

Figure 14: Counterfactual Outcomes with Improved Transport Infrastructure along River Order 1 (Napo-Amazon)

Table 1: Market Access and Human Settlements

	Community existence		log (population)	
	(1) OLS	(2) IV	(3) OLS	(4) 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)
Confluence (1 or 2)	0.0687***	0.0690***	0.348***	0.348***
	(0.00739)	(0.00739)	(0.0371)	(0.0372)
Confluence (3)	0.0253***	0.0251***	0.114***	0.114***
	(0.00249)	(0.00248)	(0.0115)	(0.0115)
Flood	0.00244***	0.00243***	0.0108***	0.0108***
	(0.000171)	(0.000171)	(0.000816)	(0.000817)
Holocene	0.0000232***	0.0000245***	0.000107***	0.000111***
	(0.00000448)	(0.00000446)	(0.0000220)	(0.0000219)
Pleistocene	-0.0000285***	-0.0000294***	-0.000160***	-0.000163***
	(0.00000446)	(0.00000447)	(0.0000205)	(0.0000205)
Non-Main Channel	0.000465***	0.000472***	0.00234***	0.00237***
	(0.0000712)	(0.0000713)	(0.000351)	(0.000352)
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 2: Market Access and Forest Cover Change

	Forest disturbance		Forest loss		Forest recovery		Per-capita forest loss	
	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV	(7) OLS	(8) 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)
Confluence (1 or 2)	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)
Confluence (3)	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	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)
Holocene	0.0111***	0.0111***	0.00386***	0.00386***	0.00536***	0.00532***	0.000957	0.000965
	(0.000302)	(0.000302)	(0.0000753)	(0.0000753)	(0.0000850)	(0.0000850)	(0.00156)	(0.00156)
Pleistocene	-0.00109***	-0.000699*	-0.00168***	-0.00169***	-0.000195	-0.0000684	0.00342	0.00185
	(0.000384)	(0.000385)	(0.0000931)	(0.0000934)	(0.000139)	(0.000140)	(0.00290)	(0.00311)
Non-Main Channel	0.0268***	0.0258***	0.00580***	0.00583***	0.00582***	0.00539***	-0.00927**	-0.00846*
	(0.00360)	(0.00363)	(0.000662)	(0.000662)	(0.000783)	(0.000799)	(0.00444)	(0.00494)
Main Channel	0.0124***	0.0121***	-0.000594*	-0.000582	-0.0110***	-0.0111***	-0.00981***	-0.0119***
	(0.00209)	(0.00210)	(0.000361)	(0.000360)	(0.000333)	(0.000339)	(0.00345)	(0.00365)
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 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.

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

Table 3: 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
θ	Trade elasticity	From the literature
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
$\{A_{o,Ag}\}$, $\{A_{o,Nr}\}$, A_M	Absolute advantages	Calibration

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

Table 4: Density Externalities in Agriculture and Natural Resource Extraction

Parameter	Point estimate	Standard errror	Description
(A) Agriculture			
$\tilde{\mu}_{Ag}$	0.071	0.009	$= \mu_{Ag} - (1-\gamma)\mu_L$
μ_L	0.498	0.090	Congestion in forest clearing
μ_{Ag}	0.271		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
	Hansen J stat = 2.905		
	J test p-value = 0.821		

Notes: Estimates of density externalities in agriculture (panel A) are based on the linear specification using $\ln RA_o$ as an instrument. Estimates of parameters governing congestion externality in natural resource extraction (panel B) are based on the non-linear GMM using $\ln RA_o$ and $\{\ln \sum_{d|D_{o,d} \leq x} RA_d\}$ for $x \in \mathcal{X} = \{2, 5, 10, 25, 50, 75, 100, 150\}$ as instruments.

Table 5: Density Externalities in Agriculture with Heterogeneity and Natural Resource Extraction

Parameter	Point estimate	Standard errror	Description
(A) Agriculture (among small communities)			
$\tilde{\mu}_{Ag}$	0.073	0.040	$= \mu_{Ag} - (1-\gamma)\mu_L$
μ_L	0.227	0.565	Congestion in forest clearing
μ_{Ag}	0.164		Agglomeration in agricultural production
(B) Agriculture (among large communities)			
$\tilde{\mu}_{Ag}$	0.043	0.025	$= \mu_{Ag} - (1-\gamma)\mu_L$
μ_L	0.524	0.239	Congestion in forest clearing
μ_{Ag}	0.272		Agglomeration in agricultural production
(C) Natural resource extraction			
μ_{Nr}	0.335	0.042	Congestion in natural resource extraction
ν	0.593	0.075	Spatial decay of congestion externality
	Hansen J stat = 2.905		
	J test p-value = 0.821		

Notes: Estimates of density externalities in agriculture (panel A) are based on the linear specification using $\ln RA_o$ as an instrument. Estimates of parameters governing congestion externality in natural resource extraction (panel B) are based on the non-linear GMM using $\ln RA_o$ and $\{\ln \sum_{d|D_{o,d} \leq x} RA_d\}$ for $x \in \mathcal{X} = \{2, 5, 10, 25, 50, 75, 100, 150\}$ as instruments.

Table 6: Natural Resource Endowments and Calibrated Productivity

	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)
Basin FE	Yes	Yes	Yes	Yes	Yes
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

Notes: Robust standard errors in parentheses. The unit of anaylysis is a community in the PARLAP Community Census (CC) in 2014.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Community Population and Availability of Transport Modes

	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
$\log(N_{o,Ag})$	0.0463***	0.166***	0.0477***	0.0290	0.0511***	0.0627**	-0.00517	0.00506
	(0.0111)	(0.0442)	(0.0115)	(0.0390)	(0.0107)	(0.0288)	(0.00569)	(0.0153)
Basin FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
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		24.313		24.313		24.313		24.313
Observations	906	906	906	906	906	906	906	906

Notes: Robust standard errors in parentheses. The unit of anaylysis is a community in the PARLAP Community Census (CC) in 2014. We use $\log(RA_o)$ and initial community existence (in 1910 and 1940) as instruments for $\log(N_{o,Ag})$. Geographical controls include a dummy of high river orders (4 and 5), 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, and geology measures for a grid cell where each census community belongs.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 8: 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
$\log(N_{o,Ag})$	-0.00309*	-0.00406	0.0000856	-0.000788	0.00410**	0.000584
	(0.00182)	(0.00423)	(0.000592)	(0.000878)	(0.00195)	(0.000822)
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	1090.954	1090.954	1090.954	1090.954	1090.954	1090.954
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
$\log(N_{o,Ag})$	0.000726	0.00215*	0.0207***	0.0300***	0.0106***	-0.00416*
	(0.00110)	(0.00111)	(0.00349)	(0.00361)	(0.00214)	(0.00237)
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	1090.954	1090.954	1090.954	1090.954	1090.954	1090.954
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
$\log(N_{o,Ag})$	-0.000784	-0.000261	-0.000179	-0.00102*	0.00194**	0.0195***
	(0.000501)	(0.000287)	(0.000297)	(0.000547)	(0.000805)	(0.00388)
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	1090.954	1090.954	1090.954	1090.954	1090.954	1090.954
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
$\log(N_{o,Ag})$	0.00590***	0.000397	-0.000964	-0.0101***	0.000997*	-0.0286***
	(0.00188)	(0.000419)	(0.000644)	(0.00320)	(0.000599)	(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	1090.954	1090.954	1090.954	1090.954	1090.954	1090.954
Observations	25827	25827	25827	25827	25827	25827
Basin FE	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 $\log(RA_o)$ and initial community existence (in 1910 and 1940) as instruments for $\log(N_{o,Ag})$. Geographical controls include a dummy of high river orders (4 and 5), 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, and geology measures for a grid cell where each census community belongs.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 9: Counterfactual Outcomes without the Agglomeration Externality

Basin	Welfare	Deforestation	Q (Ag)	Q (Nr)	N (Ag)	N (Nr)	N (Urban)
Napo	0.878	1.599	0.654	0.991	1.029	0.956	1.011
Pastaza	0.902	1.262	0.688	0.992	1.044	0.974	0.988
LowerUcayali	0.860	1.226	0.718	0.968	1.051	0.959	0.987
UpperUcayali	0.922	1.192	0.709	0.987	1.066	0.982	0.985

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

Table 10: Counterfactual Experiments of Resettlement Policies

Basin	Welfare	Deforestation	Q (Ag)	Q (Nr)	N (Ag)	N (Nr)	N (Urban)
(A) Not allowing new community formation							
Napo	0.998	0.917	1.008	1.002	1.002	1.002	0.997
Pastaza	0.997	0.954	1.004	1	0.999	1.003	0.997
LowerUcayali	0.997	0.971	0.996	0.995	0.998	1.002	0.998
UpperUcayali	0.999	0.971	1.005	1.001	1.001	1.002	0.999
(B) Not allowing for small communities							
Napo	0.997	0.875	1.01	1.002	1.002	1.003	0.996
Pastaza	0.996	0.94	1.003	1.002	0.999	1.003	0.996
LowerUcayali	0.998	0.895	1.011	1.003	1.001	1	0.994
UpperUcayali	0.999	0.931	1.009	1.005	1.001	1.003	0.999
(C) Protected areas: setting a limit on rural frontier							
Napo	0.998	0.95	0.999	0.997	1.002	1.003	0.997
Pastaza	0.999	0.976	0.999	1.001	1.001	1	0.998
LowerUcayali	0.998	0.959	1.003	0.997	1	1.001	0.996
UpperUcayali	0.998	0.98	1.001	0.992	1.001	1.004	0.999

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

Table 11: Counterfactual Experiments of Improving Transport Infrastructure

Basin	Welfare	Deforestation	Q (Ag)	Q (Nr)	N (Ag)	N (Nr)	N (Urban)
(A) Transport infrastructure improved ($\lambda_{road} = 0.8$) along river order = 2							
Napo	1.018	0.984	0.992	0.997	0.997	0.998	1.004
Pastaza	1.013	0.972	1.031	0.994	1.003	0.999	0.998
LowerUcayali	1.025	0.972	1.058	0.971	1.004	0.993	1.02
UpperUcayali	1.012	0.985	1.029	1.003	1.002	1.005	0.999
(B) Transport infrastructure improved ($\lambda_{road} = 0.8$) along river order = 1							
Napo	1.01	1.07	0.969	0.996	0.992	0.997	1.008
(C) Transport infrastructure improved ($\lambda_{road} = 0.8$) along river order = 1, 2							
Napo	1.029	1.06	0.963	0.994	0.99	0.995	1.011
(D) Symmetric trade cost ($\lambda_{up} = 1$)							
Napo	1.011	1.027	0.982	0.997	0.996	0.997	1.005
Pastaza	1.008	1.007	1.002	0.999	1.000	1.000	1.001
LowerUcayali	1.010	1.010	1.023	0.987	1.001	0.996	1.016
UpperUcayali	1.005	1.013	1.005	1.004	1.003	1.006	0.998
(E) Transport infrastructure improved ($\lambda_{road} = 0.8$) along river order = 2 with checkposts that prohibit transporting natural resource goods							
Napo	1.016	0.961	0.999	1.002	1	1.002	0.999
Pastaza	1.011	0.975	1.024	0.999	1.001	1.002	0.996
LowerUcayali	1.023	0.978	1.046	0.985	0.999	1	1.004
UpperUcayali	1.011	0.984	1.031	1.007	1.001	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 12: Counterfactual Outcomes of a Composite Intervention

Basin	Welfare	Deforestation	Q (Ag)	Q (Nr)	N (Ag)	N (Nr)	N (Urban)
Protected areas: setting a limit on rural frontier &							
Transport infrastructure improved (λ_{road} = 0.8) along river order = 2							
Napo	1.016	0.93	0.989	0.994	0.999	1.001	1.011
Pastaza	1.011	0.95	1.027	0.996	1.004	0.999	0.996
LowerUcayali	1.023	0.951	1.055	0.969	1.003	0.994	1.017
UpperUcayali	1.011	0.965	1.029	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.

— Appendix —

A Additional Figures and Tables

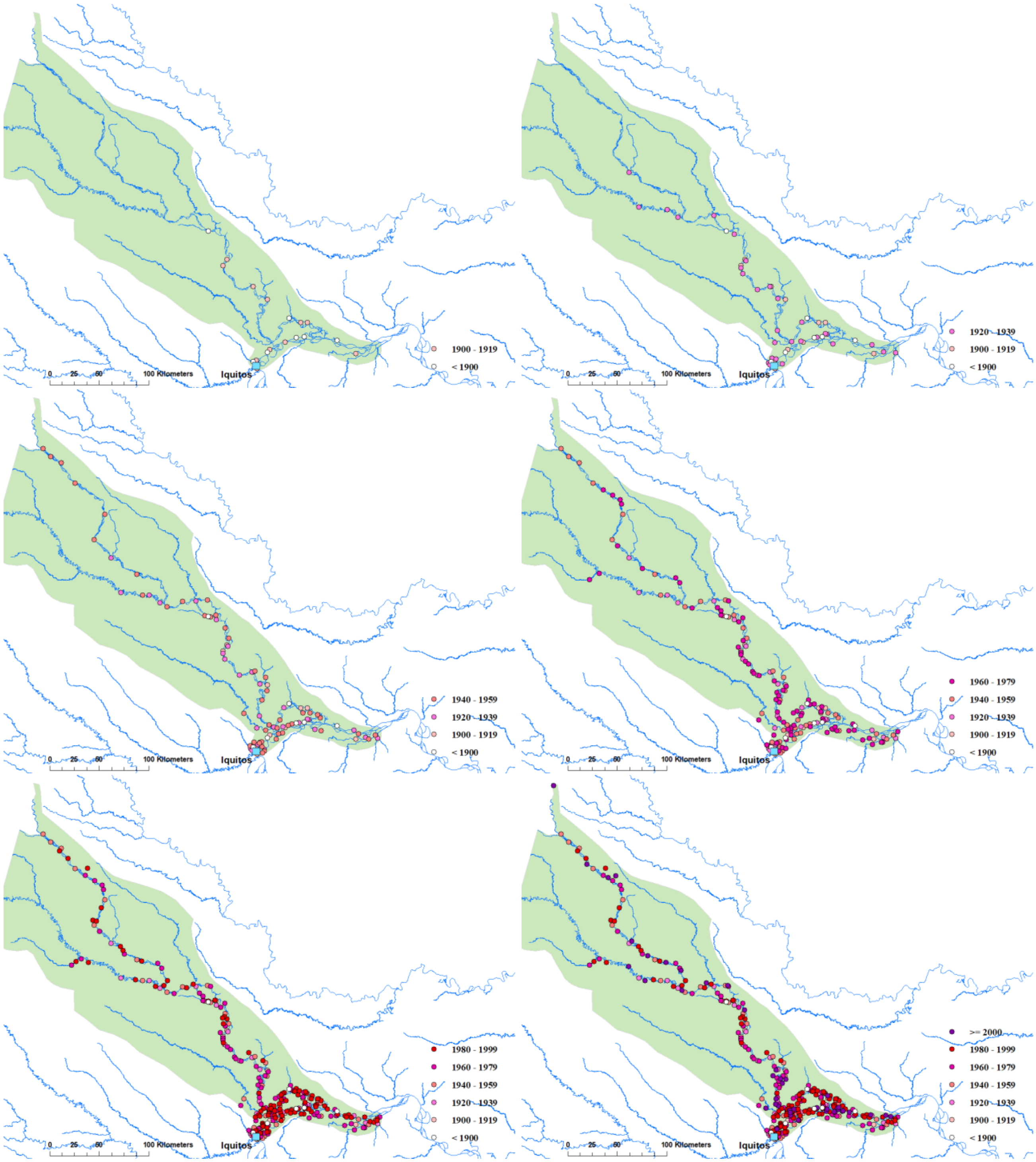

Figure A.1: Establishment of Rainforest Communities over Decades in the Napo-Amazon Basin

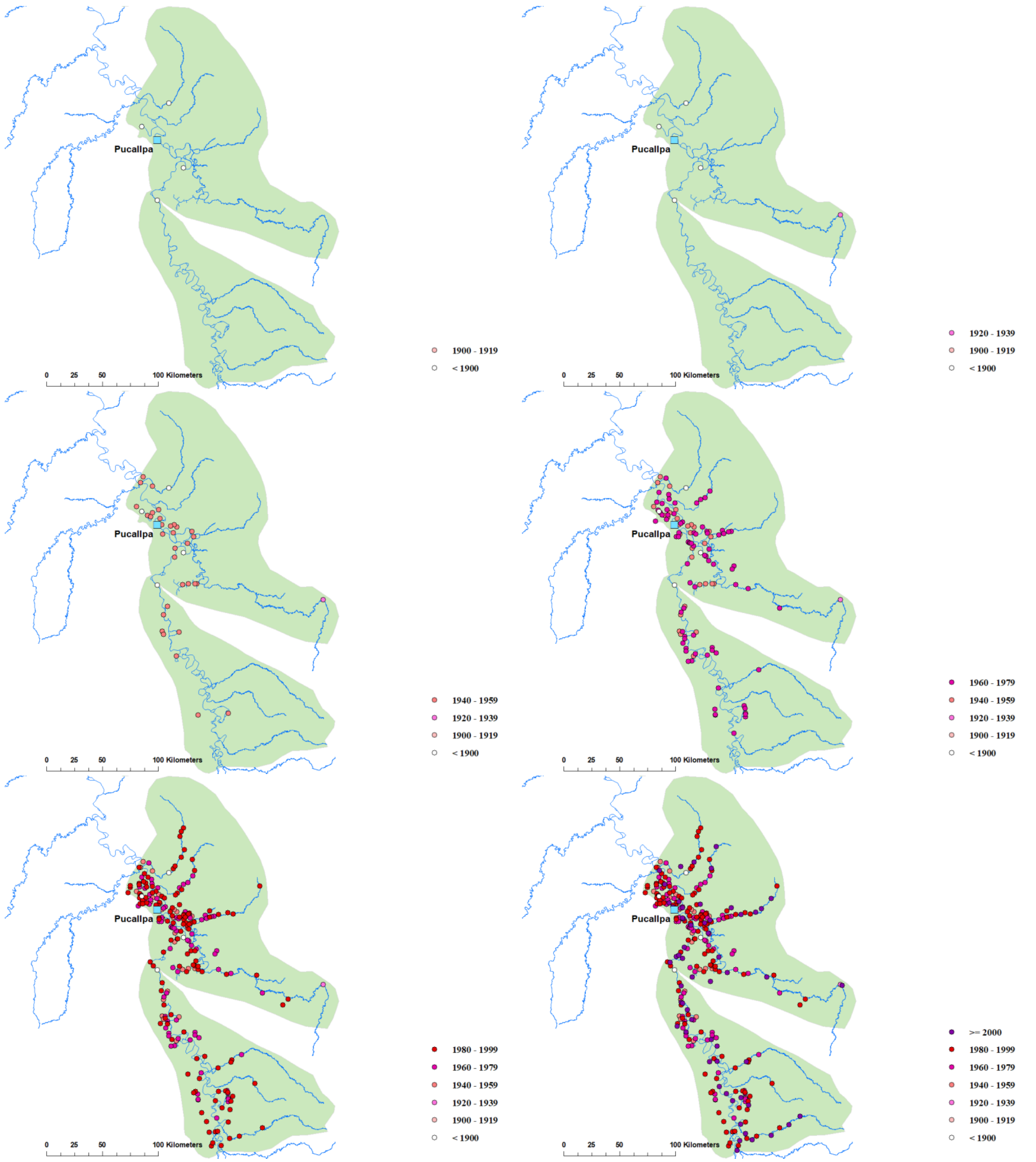

Figure A.2: Establishment of Rainforest Communities over Decades in the Upper Ucayali Basin

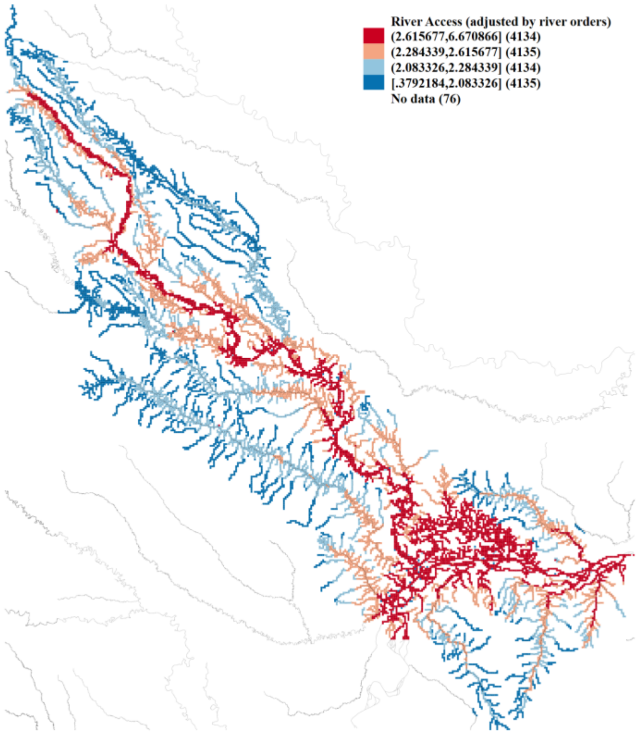

Figure A.3: River Access and Market Access at River Cells (Napo)

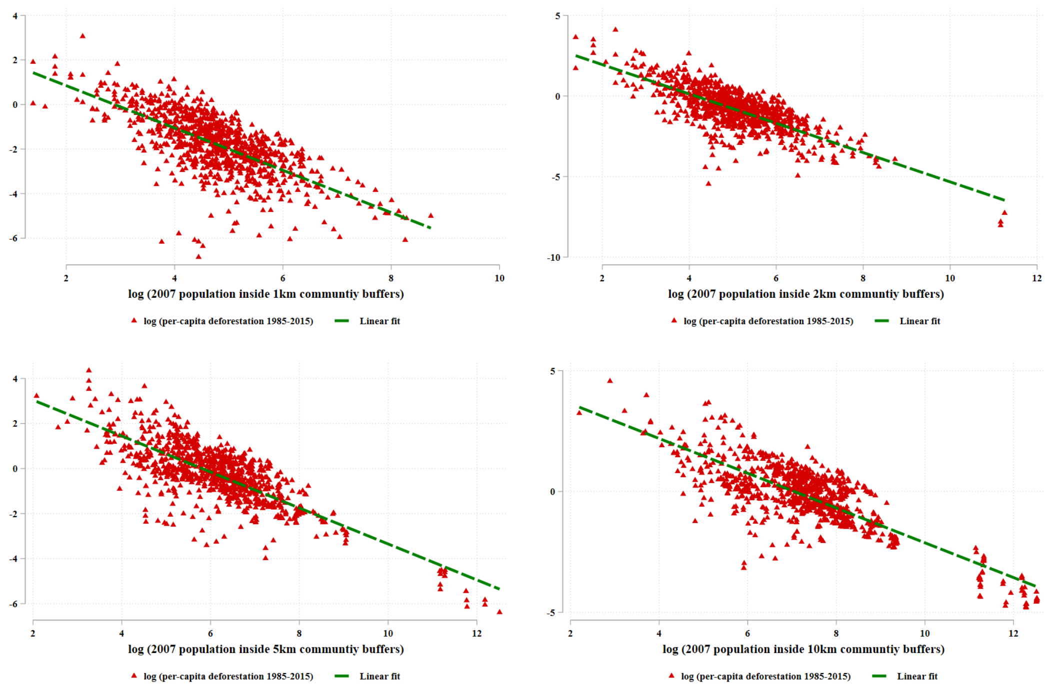

Figure A.4: Populations and Per Capita Deforestation with Different Buffers

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.

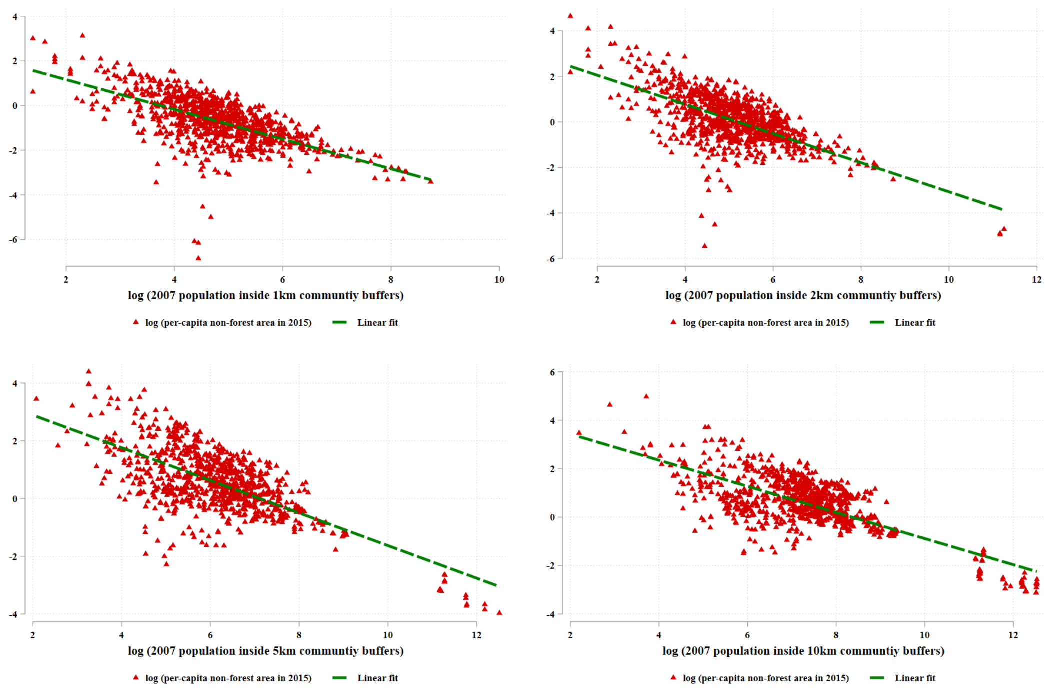

Figure A.5: Populations and Per Capita Non-Forest Area with Different Buffers

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.

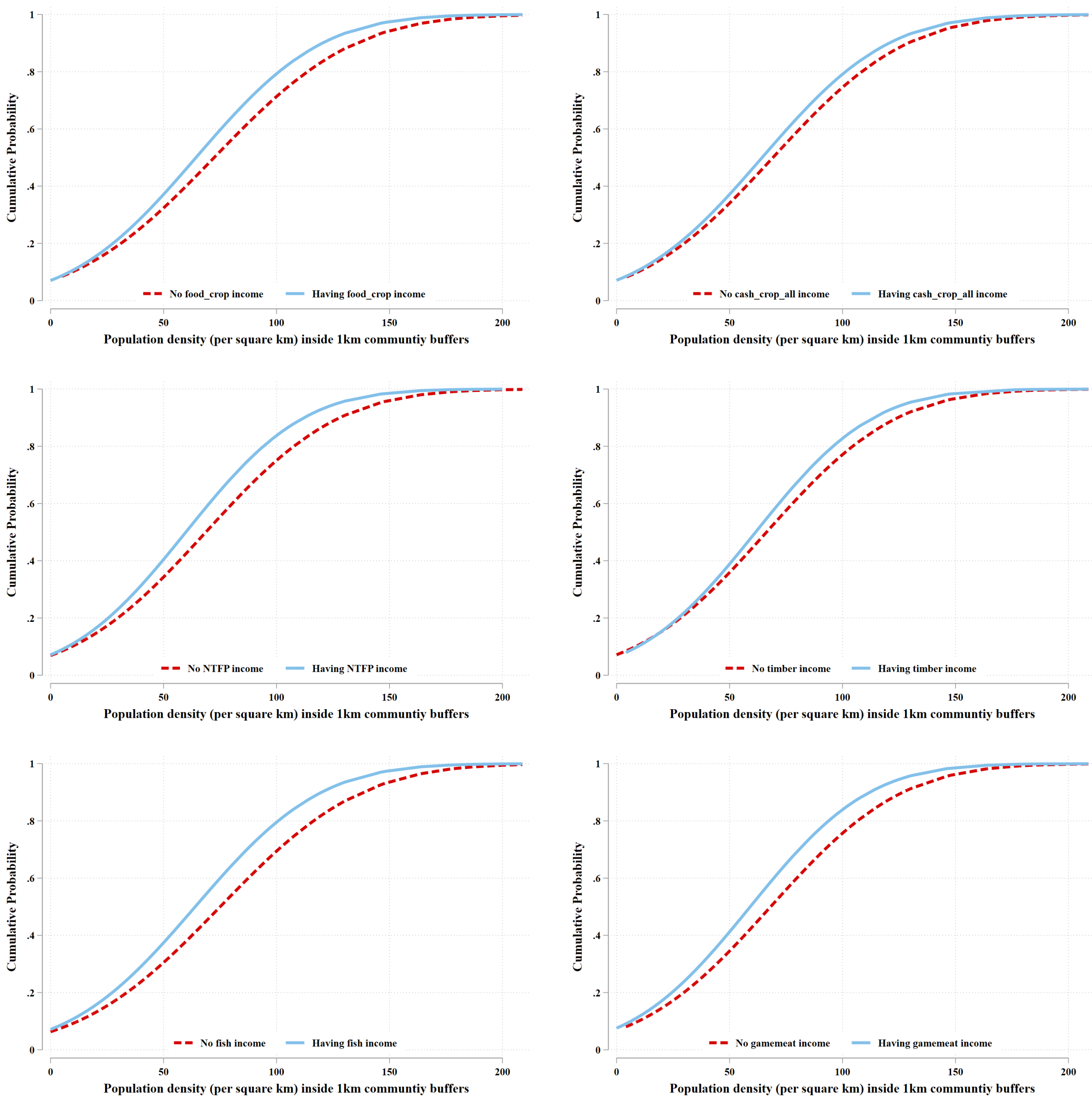

Figure A.6: Population Density (1km buffer) and Distribution of Activities (Household-level)—Agriculture vs. Forest Products vs. Wild Animal Extractions

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.

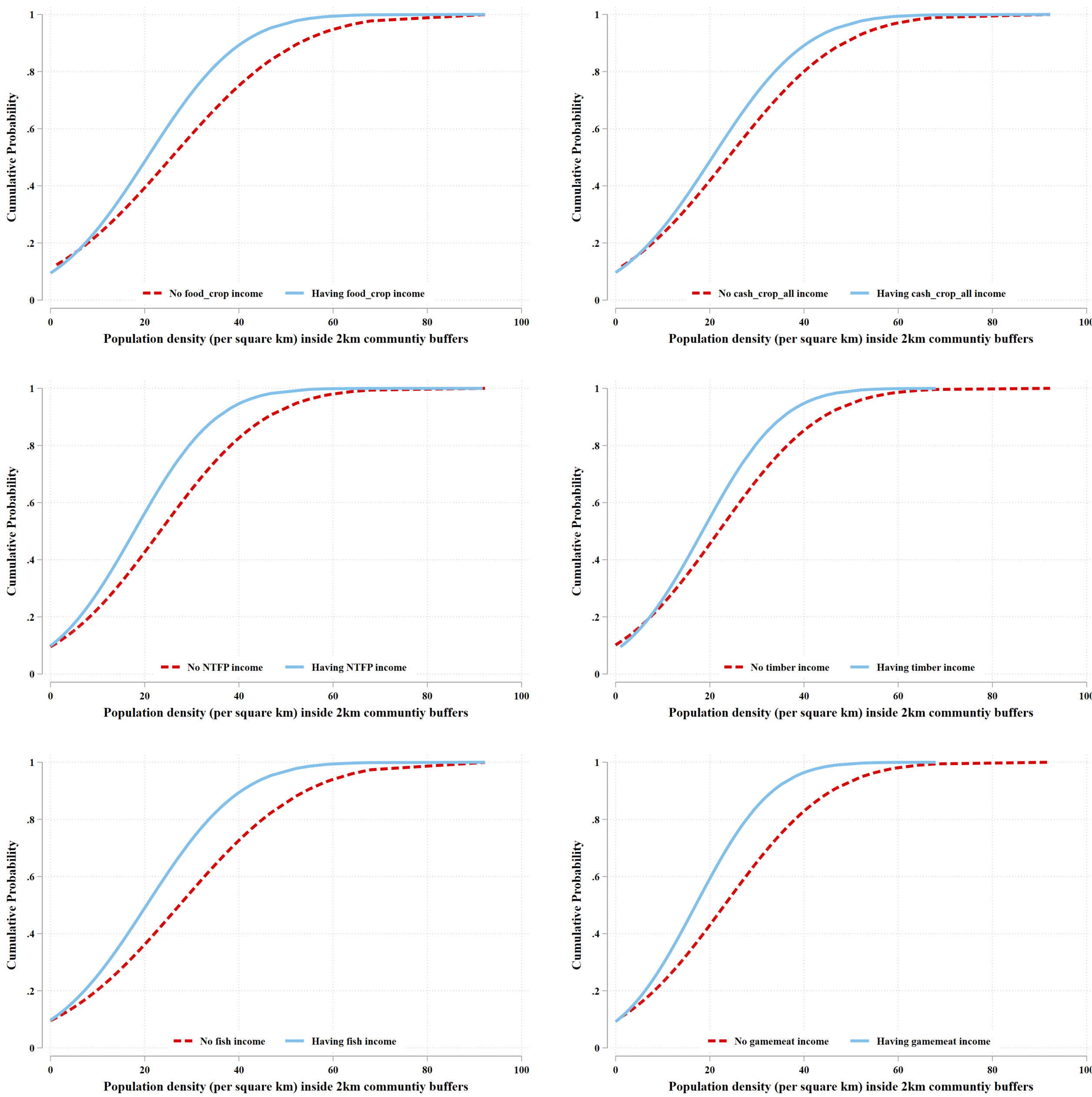

Figure A.7: Population Density (2km buffer) and Distribution of Activities (Household-level)—Agriculture vs. Forest Products vs. Wild Animal Extractions

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.

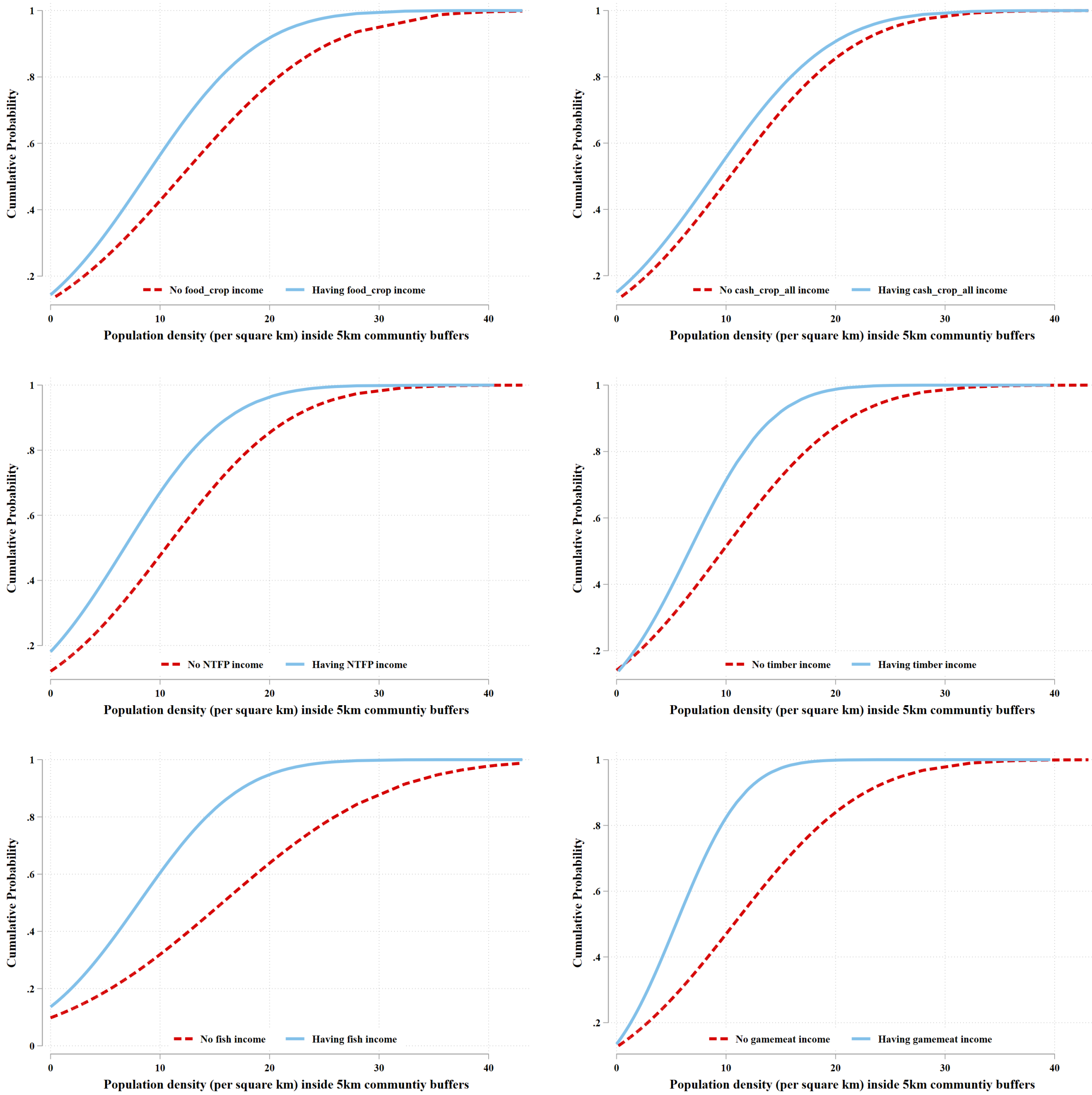

Figure A.8: Population Density (5km buffer) and Distribution of Activities (Household-level)—Agriculture vs. Forest Products vs. Wild Animal Extractions

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.9: 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.10: Cooperative nature of forest clearing

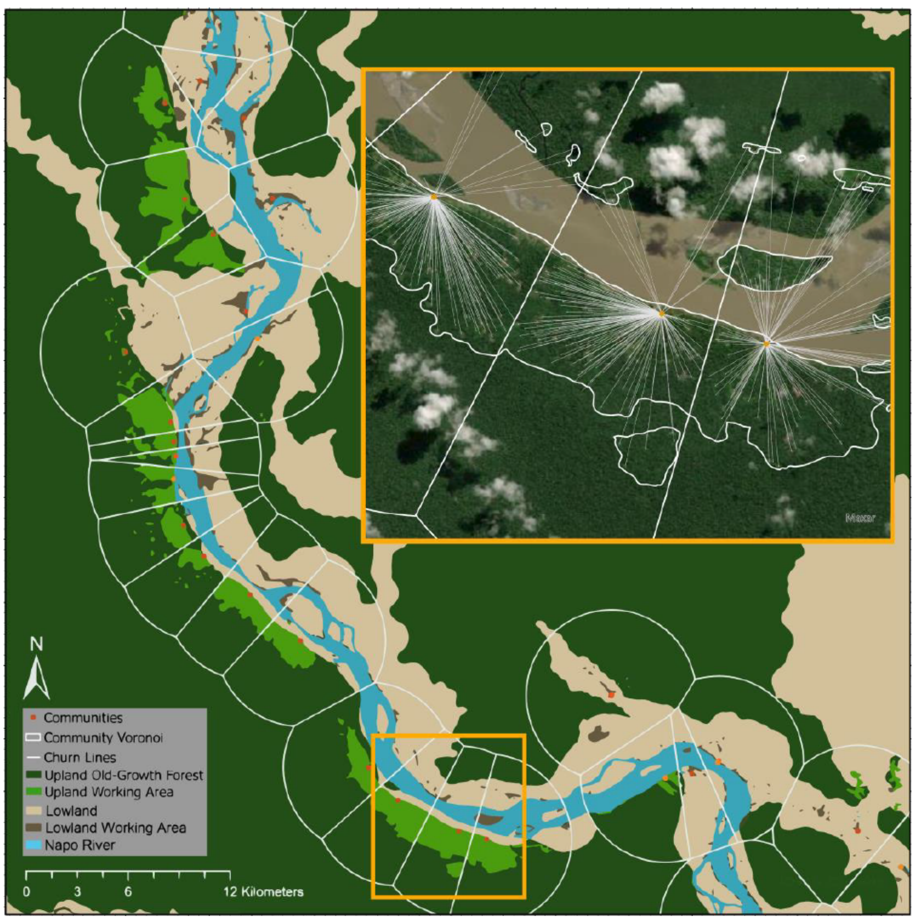

Figure A.11: Voronoi Polygons and Working Areas around the Census Communities
Source: Coomes et al. (2021)

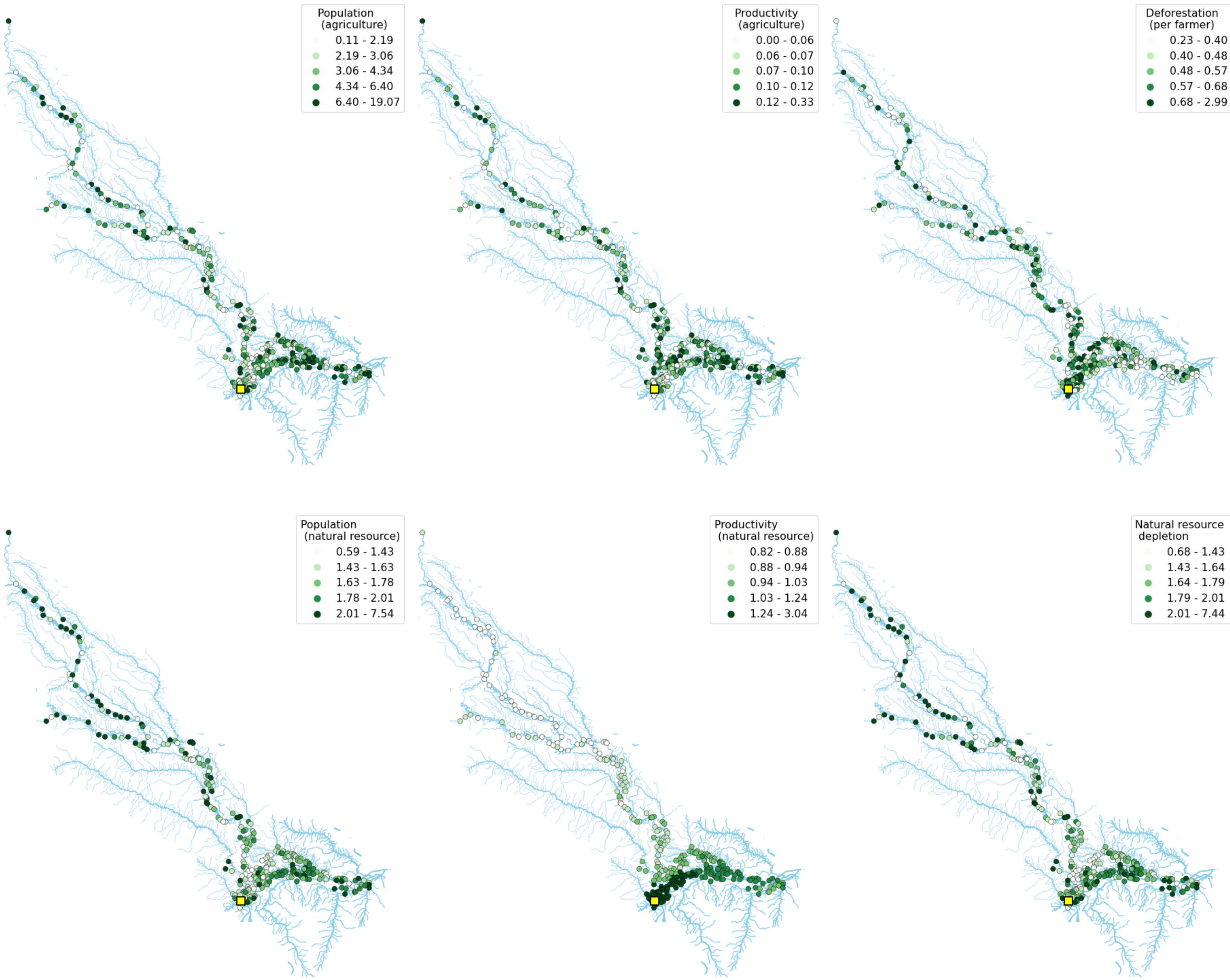

Figure A.12: Counterfactual Outcomes without the Agglomeration Externality (Napo-Amazon)

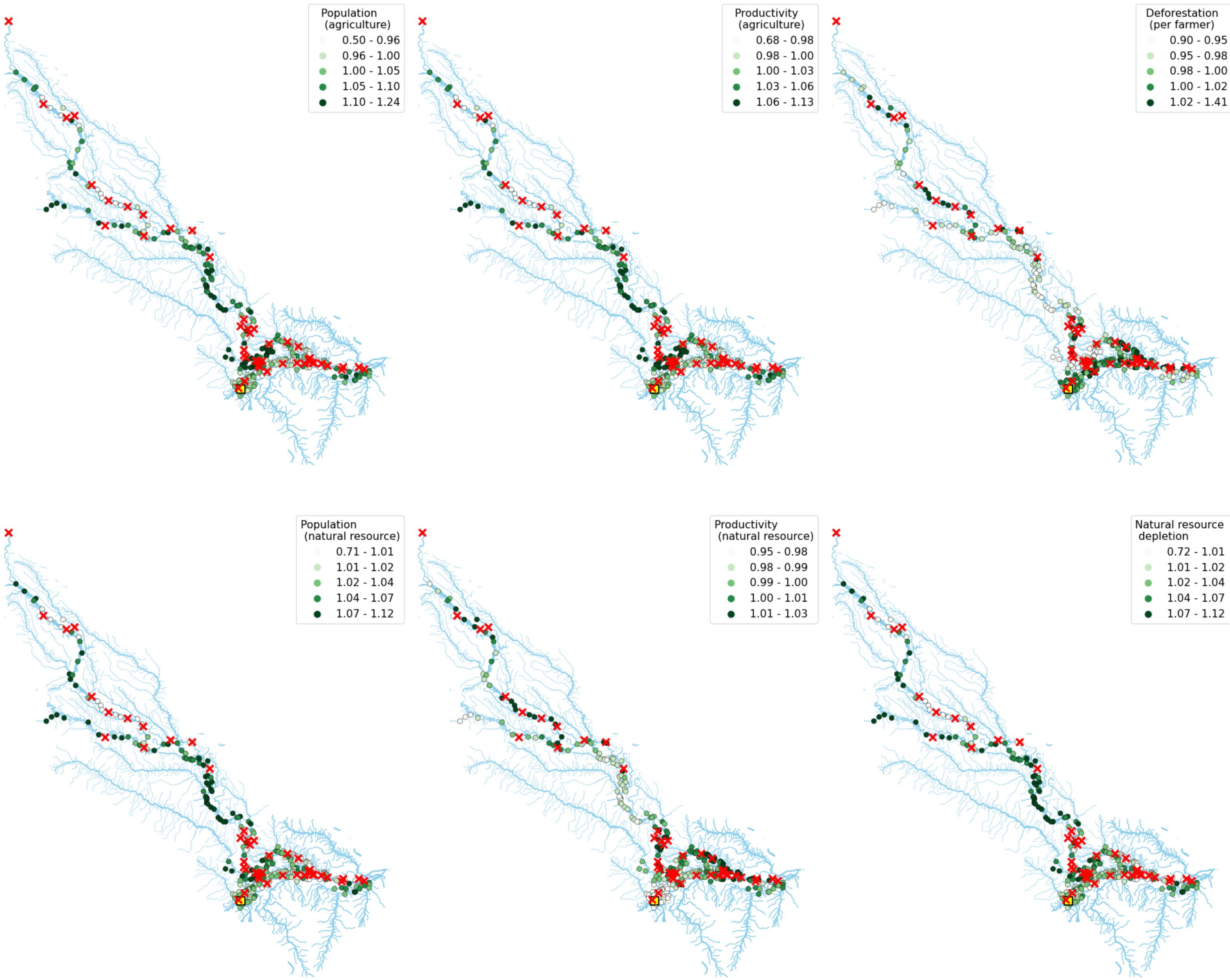

Figure A.13: Counterfactual Outcomes without New Community Formation (Napo-Amazon)

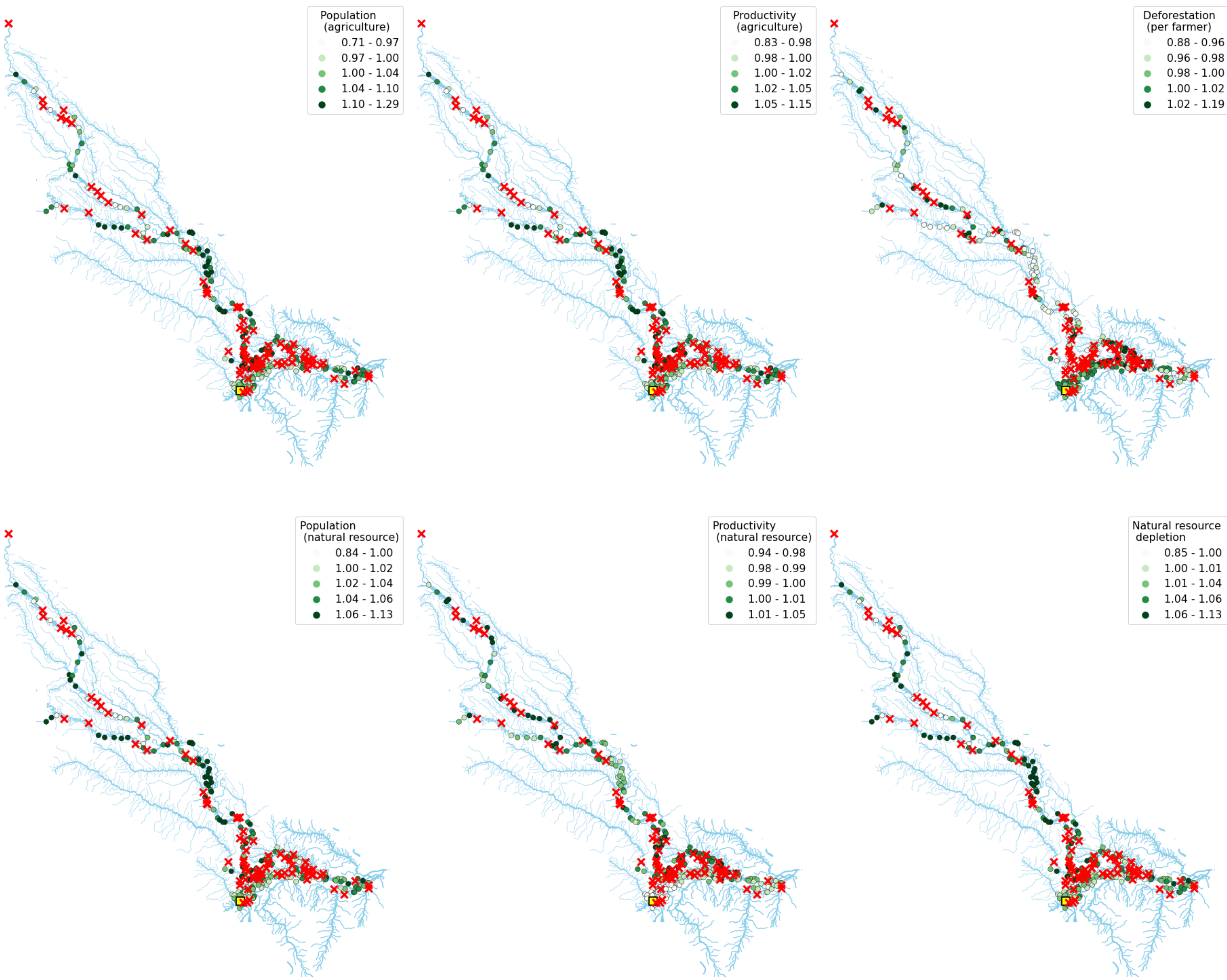

Figure A.14: Counterfactual Outcomes without Small Communities (Napo-Amazon)

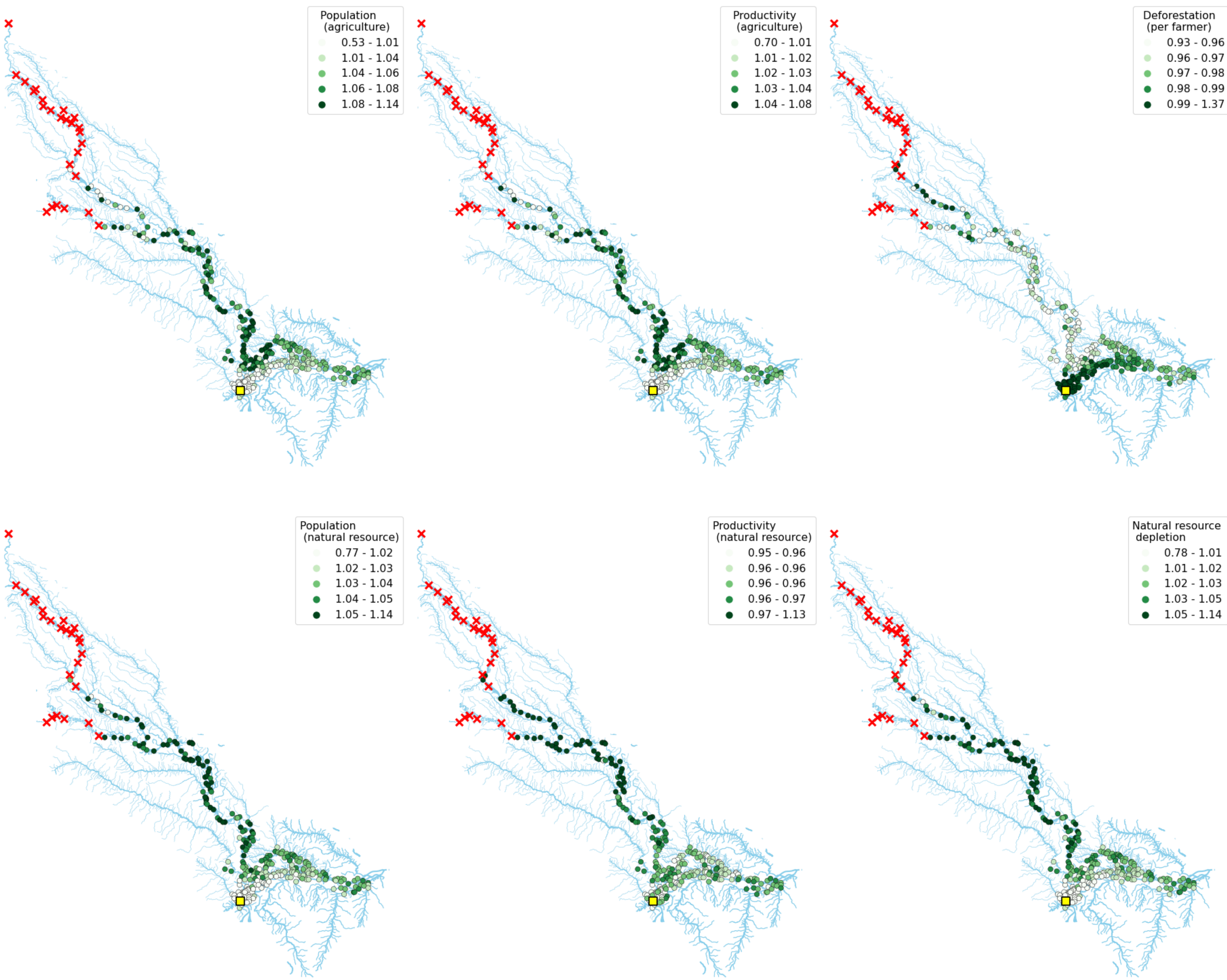

Figure A.15: Counterfactual Outcomes with Limited Rural Frontier (Napo-Amazon)

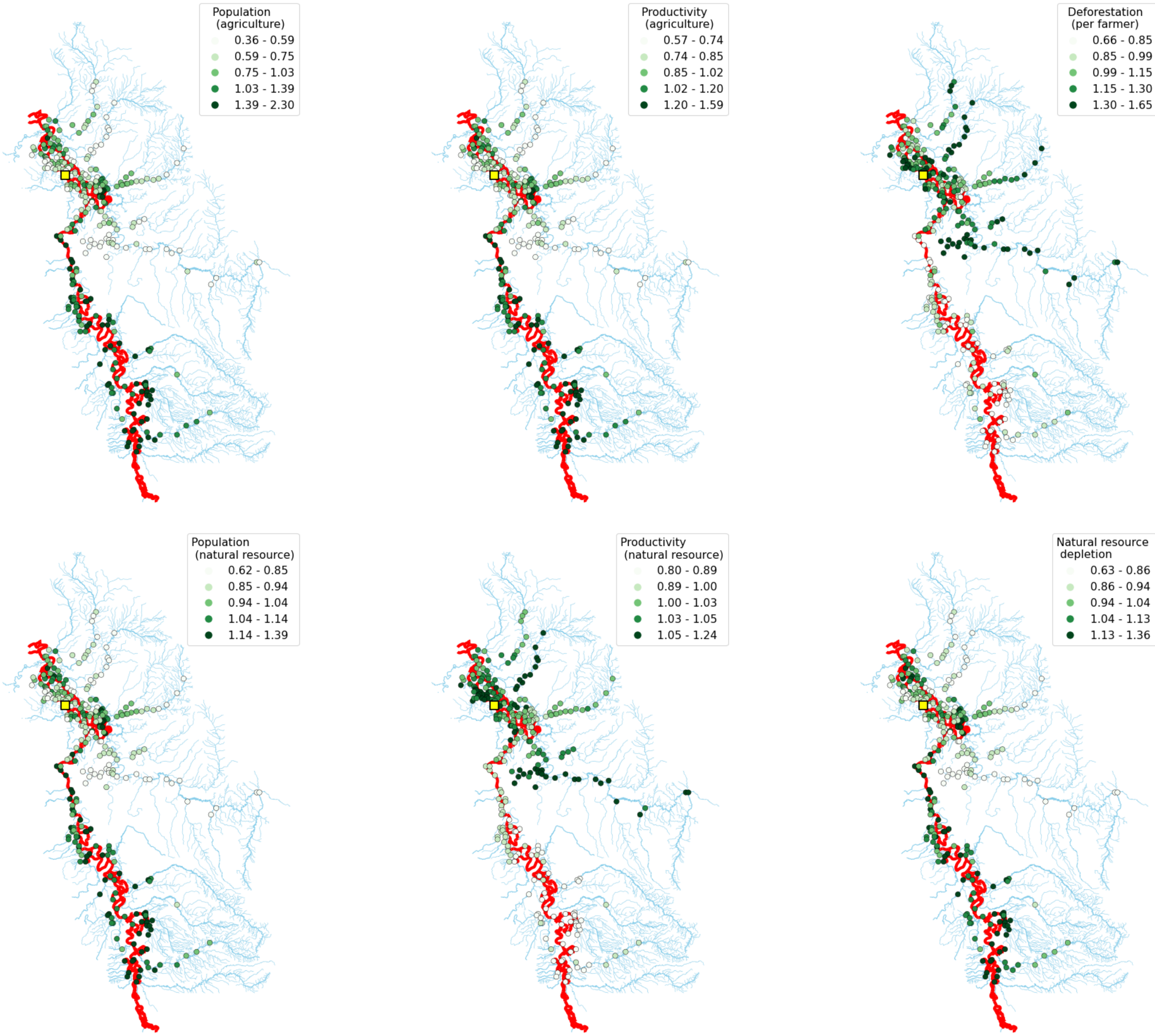

Figure A.16: Counterfactual Outcomes with Improved Transport Infrastructure along River Order 2 (Upper Ucayali)

Table A.1: 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)
Confluence (1 or 2)	-0.0544	-0.0541	0.383***	0.434***	-0.201*	-0.211*
	(0.0385)	(0.0385)	(0.125)	(0.121)	(0.121)	(0.122)
Confluence (3)	-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	-0.0124***	-0.0124***	0.261***	0.257***	0.0155	0.00816
	(0.00130)	(0.00130)	(0.0102)	(0.0102)	(0.0224)	(0.0233)
Holocene	-0.00137***	-0.00136***	0.0158***	0.0159***	-0.00155	-0.00164
	(0.0000470)	(0.0000471)	(0.000373)	(0.000375)	(0.00117)	(0.00119)
Pleistocene	0.000209***	0.000207***	-0.0136***	-0.0141***	-0.000415	-0.00113
	(0.0000387)	(0.0000384)	(0.000604)	(0.000607)	(0.00267)	(0.00282)
Non-Main Channel	-0.0260***	-0.0260***	0.0718***	0.0729***	-0.0140***	-0.0138***
	(0.000720)	(0.000723)	(0.00270)	(0.00263)	(0.00477)	(0.00487)
Main Channel	-0.0296***	-0.0296***	0.0366***	0.0369***	-0.00567**	-0.00677**
	(0.000877)	(0.000878)	(0.00116)	(0.00114)	(0.00260)	(0.00284)
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 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.2: Agglomeration Externality in Agriculture

	The calibrated value of $\log(\tilde{A}_{o,Ag})$					
	All locations		$N_{o,Ag} < 75$		$N_{o,Ag} \geq 75$	
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	IV	OLS	IV	OLS	IV
$\log(N_{o,Ag})$	0.676***	0.557***	0.794***	0.466	0.541***	0.323*
	(0.0205)	(0.0706)	(0.0459)	(0.318)	(0.0411)	(0.194)
Basin FE	Yes	Yes	Yes	Yes	Yes	Yes
Mean (Dep. var.)	-0.096	-0.096	-0.754	-0.754	0.546	0.546
SD (Dep. var.)	4.578	4.578	4.823	4.823	4.234	4.234
First stage F-stat		22.959		2.855		5.249
Observations	893	893	441	441	452	452

Notes: Robust standard errors in parentheses. The sample includes 1 square km grid cells that have positive populations. We use $\log(RA_o)$ and initial community existence (in 1910 and 1940) as instruments for $\log(N_{o,Ag})$. Geographical controls include a dummy of high river orders (4 and 5), 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, and geology measures.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.3: Agglomeration Externality in Agriculture by Basin

	The calibrated value of $\log(\tilde{A}_{o,Ag})$							
	Napo-Amazon		Pastaza		Lower Ucayali		Upper Ucayali	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
$\log(N_{o,Ag})$	0.700***	0.756***	0.703***	0.847***	0.414***	0.127	0.758***	0.682***
	(0.0276)	(0.125)	(0.0633)	(0.209)	(0.0632)	(0.136)	(0.00822)	(0.0240)
Mean (Dep. var.)	1.778	1.778	6.514	6.514	2.531	2.531	-5.887	-5.887
SD (Dep. var.)	1.348	1.348	1.493	1.493	0.848	0.848	1.336	1.336
First stage F-stat		4.001		5.424		8.847		24.227
Observations	310	310	110	110	170	170	303	303

Notes: Robust standard errors in parentheses. The sample includes 1 square km grid cells that have positive populations. We use $\log(RA_o)$ and initial community existence (in 1910 and 1940) as instruments for $\log(N_{o,Ag})$. Geographical controls include a dummy of high river orders (4 and 5), 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, and geology measures.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.4: Density Externality in Forest Clearing

	log (per capita deforestation)					
	All locations		$N_{o,Ag} < 75$		$N_{o,Ag} \geq 75$	
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	IV	OLS	IV	OLS	IV
$\log(N_{o,Ag})$	-0.653***	-0.498***	-0.658***	-0.251	-0.653***	-0.525**
	(0.0306)	(0.0899)	(0.0605)	(0.553)	(0.0672)	(0.239)
Basin FE	Yes	Yes	Yes	Yes	Yes	Yes
Mean (Dep. var.)	0.929	0.929	1.478	1.478	0.372	0.372
SD (Dep. var.)	1.231	1.231	1.149	1.149	1.049	1.049
First stage F-stat		24.002		2.252		7.290
Observations	895	895	451	451	444	444

Notes: Robust standard errors in parentheses. The unit of anaylysis is a community in the PARLAP Community Census (CC) in 2014. We use $\log(RA_o)$ and initial community existence (in 1910 and 1940) as instruments for $\log(N_{o,Ag})$. Geographical controls include a dummy of high river orders (4 and 5), 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, and geology measures for a grid cell where each census community belongs.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.5: Density Externality in Forest Clearing by Basin

	log (per capita deforestation)							
	Napo-Amazon		Pastaza		Lower Ucayali		Upper Ucayali	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
$\log(N_{o,Ag})$	-0.650***	-0.510**	-0.418***	-0.558**	-0.586***	-0.498***	-0.859***	-0.748***
	(0.0558)	(0.227)	(0.0660)	(0.264)	(0.0452)	(0.139)	(0.0566)	(0.115)
Mean (Dep. var.)	0.606	0.606	0.536	0.536	0.763	0.763	1.515	1.515
SD (Dep. var.)	1.093	1.093	1.171	1.171	1.005	1.005	1.296	1.296
First stage F-stat		4.093		5.949		8.791		25.187
Observations	313	313	115	115	169	169	298	298

Notes: Robust standard errors in parentheses. The unit of anaylysis is a community in the PARLAP Community Census (CC) in 2014. We use $\log(RA_o)$ and initial community existence (in 1910 and 1940) as instruments for $\log(N_{o,Ag})$. Geographical controls include a dummy of high river orders (4 and 5), 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, and geology measures for a grid cell where each census community belongs.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.6: 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
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 $\ln RA_o$ and $\{\ln\sum_{d\mid D_{o,d}\leq x} RA_d\}$ 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 river, squared distance to the river, interaction terms of these two variables with a river cell dummy, elevation, river confluence, flood vulnerability, and open water access measures.

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

Table A.7: Congestion Externality in Natural Resource Extraction with Spatial Spillovers ($N_{o,Nr} < 75$)

	The calibrated value of $\log(\tilde{A}_{o,Nr})$									
	IV									OLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$\log(N_{o,Nr})$	-7.908	-6.994	-8.308	-8.814	-2.829	-3.023	-0.725	-0.582	-0.652	1.040***
	(10.22)	(8.963)	(13.93)	(16.98)	(5.198)	(7.943)	(3.261)	(3.049)	(2.973)	(0.0478)
$\log(\sum_{d\|D_{o,d}\leq 2km} N_{d,Nr})$		-0.427	-0.732	-0.781	-0.400	-0.420	-0.190	-0.180	-0.191	-0.0108
		(0.675)	(1.435)	(1.709)	(0.594)	(0.854)	(0.361)	(0.337)	(0.323)	(0.0525)
$\log(\sum_{d\|D_{o,d}\leq 5km} N_{d,Nr})$			0.339	0.310	0.0837	0.0947	-0.0406	-0.0457	-0.0401	-0.136***
			(1.086)	(1.074)	(0.388)	(0.517)	(0.225)	(0.210)	(0.205)	(0.0374)
$\log(\sum_{d\|D_{o,d}\leq 10km} N_{d,Nr})$				0.0822	0.257	0.266	0.181	0.170	0.170	0.0692*
				(0.722)	(0.340)	(0.425)	(0.180)	(0.168)	(0.173)	(0.0412)
$\log(\sum_{d\|D_{o,d}\leq 25km} N_{d,Nr})$					-0.464*	-0.487	-0.339*	-0.320*	-0.320*	-0.156***
					(0.249)	(0.463)	(0.193)	(0.185)	(0.190)	(0.0368)
$\log(\sum_{d\|D_{o,d}\leq 50km} N_{d,Nr})$						0.0339	0.0224	0.00616	0.0114	-0.0764***
						(0.480)	(0.223)	(0.214)	(0.210)	(0.0271)
$\log(\sum_{d\|D_{o,d}\leq 75km} N_{d,Nr})$							-0.228	-0.131	-0.125	-0.0621*
							(0.151)	(0.224)	(0.238)	(0.0357)
$\log(\sum_{d\|D_{o,d}\leq 100km} N_{d,Nr})$								-0.129	-0.180	-0.256***
								(0.227)	(0.379)	(0.0675)
$\log(\sum_{d\|D_{o,d}\leq 150km} N_{d,Nr})$									0.0595	0.0839
									(0.317)	(0.0766)
Basin FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean (Dep. Var.)	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006
SD (Dep. Var.)	2.772	2.772	2.772	2.772	2.772	2.772	2.772	2.772	2.772	2.772
Observations	479	479	479	479	479	479	479	479	479	479

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

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

Table A.8: Congestion Externality in Natural Resource Extraction with Spatial Spillovers ($N_{o,Nr} \geq 75$)

	The calibrated value of $\log(\tilde{A}_{o,Nr})$									
	IV									OLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$\log(N_{o,Nr})$	-16.93	-27.22	-12.14	-10.80	-16.89	411.1	-5.857	-0.155	-1.222	1.099***
	(37.88)	(95.06)	(23.66)	(23.16)	(78.74)	(48521.9)	(29.13)	(2.006)	(4.029)	(0.0917)
$\log(\sum_{d\mid D_{o,d}\leq 2km} N_{d,Nr})$		3.496	3.559	3.235	5.057	-122.0	1.851	0.122	0.435	-0.271**
		(15.90)	(7.623)	(7.169)	(23.57)	(14406.9)	(8.843)	(0.667)	(1.302)	(0.107)
$\log(\sum_{d\mid D_{o,d}\leq 5km} N_{d,Nr})$			-2.593	-2.195	-3.538	79.60	-1.351	-0.287	-0.549	-0.0941
			(3.883)	(4.520)	(16.58)	(9403.4)	(5.904)	(0.485)	(0.990)	(0.0925)
$\log(\sum_{d\mid D_{o,d}\leq 10km} N_{d,Nr})$				-0.213	-0.790	20.52	-0.313	-0.159	-0.289	-0.131**
				(1.359)	(4.037)	(2428.9)	(1.658)	(0.263)	(0.526)	(0.0616)
$\log(\sum_{d\mid D_{o,d}\leq 25km} N_{d,Nr})$					0.887	-17.11	-0.134	-0.189	-0.0350	-0.169***
					(6.872)	(1965.8)	(1.488)	(0.294)	(0.606)	(0.0416)
$\log(\sum_{d\mid D_{o,d}\leq 50km} N_{d,Nr})$						-32.36	0.877	0.0618	0.124	-0.0322
						(3801.9)	(3.412)	(0.172)	(0.329)	(0.0379)
$\log(\sum_{d\mid D_{o,d}\leq 75km} N_{d,Nr})$							-0.806	0.295	0.780	-0.130**
							(1.624)	(0.661)	(1.482)	(0.0603)
$\log(\sum_{d\mid D_{o,d}\leq 100km} N_{d,Nr})$								-0.874	-1.909	-0.306***
								(0.810)	(2.244)	(0.0581)
$\log(\sum_{d\mid D_{o,d}\leq 150km} N_{d,Nr})$									0.593	0.111*
									(0.678)	(0.0587)
Basin FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean (Dep. Var.)	0.718	0.718	0.718	0.718	0.718	0.718	0.718	0.718	0.718	0.718
SD (Dep. Var.)	2.920	2.920	2.920	2.920	2.920	2.920	2.920	2.920	2.920	2.920
Observations	415	415	415	415	415	415	415	415	415	415

Notes: Robust standard errors in parentheses. The sample includes 1 square km grid cells that have positive populations. We use $\ln RA_o$ and $\{\ln \sum_{d\mid D_{o,d}\leq x} RA_d\}$ 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 river, squared distance to the river, interaction terms of these two variables with a river cell dummy, elevation, river confluence, flood vulnerability, and open water access measures.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.9: Community Population and Frequency of Transport Modes Available

	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
$\log(N_{o,Ag})$	0.161	1.035**	0.207**	0.342	0.0489	-0.326	0.230	0.657
	(0.116)	(0.454)	(0.0963)	(0.312)	(0.271)	(0.557)	(0.173)	(0.648)
Basin FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
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		7.664		9.799		4.297		3.286
Observations	330	330	276	276	61	61	144	144

Notes: Robust standard errors in parentheses. The unit of anaylysis is a community in the PARLAP Community Census (CC) in 2014. We use $\log(RA_o)$ and initial community existence (in 1910 and 1940) as instruments for $\log(N_{o,Ag})$. Geographical controls include a dummy of high river orders (4 and 5), 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, and geology 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 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
$\log(N_{o,Ag})$	0.0101	0.0478*	0.0406***	0.0905**	0.0492***	0.117***
	(0.00684)	(0.0246)	(0.0124)	(0.0402)	(0.00999)	(0.0351)
Basin FE	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		24.313		25.598		22.847
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 $\log(RA_o)$ and initial community existence (in 1910 and 1940) as instruments for $\log(N_{o,Ag})$. Geographical controls include a dummy of high river orders (4 and 5), 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, and geology 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 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
$\log(N_{o,Ag})$	0.0115*	0.0705**	0.0273***	0.0262				
	(0.00632)	(0.0319)	(0.00773)	(0.0171)				
$\log(N_{o,Nr})$					0.0282***	0.0284	0.0325***	-0.253*
					(0.00788)	(0.0442)	(0.00967)	(0.140)
Basin FE	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		13.857		13.857		8.656		8.656
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 $\log(RA_o)$ and initial community existence (in 1910 and 1940) as instruments for $\log(N_{o,Ag})$. Geographical controls include a dummy of high river orders (4 and 5), 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, and geology measures for a grid cell where each census community belongs.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.12: 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
$\log(N_{o,Ag})$	0.0102	0.0470	-0.00261	-0.0929*
	(0.0105)	(0.0370)	(0.0150)	(0.0516)
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		24.346		24.346
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
$\log(N_{o,Ag})$	0.00284	0.0236	0.0357	-0.0652
	(0.0145)	(0.0433)	(0.0217)	(0.0663)
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		11.332		11.332
Observations	402	402	402	402
Basin FE	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 $\log(RA_o)$ and initial community existence (in 1910 and 1940) as instruments for $\log(N_{o,Ag})$. Geographical controls include a dummy of high river orders (4 and 5), 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, and geology measures for a grid cell where each census community belongs.

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

Table A.13: Community Population and Crop-Specific Form of Seed Acquisition

	Obtain seeds from a city or other communities for							
	Maize		Plantain		Rice		Yuca	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
$\log(N_{o,Ag})$	0.00744	-0.246*	0.00315	-0.209*	0.00948	0.0650	0.00138	0.0953
	(0.0149)	(0.146)	(0.0138)	(0.125)	(0.00954)	(0.0775)	(0.0102)	(0.0834)
Basin FE	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		13.016		13.016		13.016		13.016
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 $\log(RA_o)$ and initial community existence (in 1910 and 1940) as instruments for $\log(N_{o,Ag})$. Geographical controls include a dummy of high river orders (4 and 5), 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, and geology measures for a grid cell where each census community belongs.

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

B Data Appendix

B.1 Forest Cover Measures

The detail is under construction.

B.2 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, timber extraction, non-timber forest products (NTFPs), hunting, 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 in the order of importance are collected. We exploit the first and third information to construct the sectoral employment shares in a community.

B.3 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.9 shows these transport modes. The survey also provides the cost for transporting 50kg package for 500 meter by land transport at the selected communities and towns that are covered in the above travel time module.

B.4 Other Data and Variables

Other Data

We additionally use the following data.

Peru National Household Survey (ENAHO). ENAHO is an annual living standard survey, also collected by the 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 variables

C Model Appendix

C.1 One-Sector Model with a Density Externality

We consider the 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 spillover 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:

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

deforestation and per capita deforestation:

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

nominal and real incomes:

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

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.

D Quantification Appendix

D.1 Parameters without Solving the Model

Downstream-River-Equivalent Distance

We use our original records of travel times and freight costs (Appendix B.3).

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 $\hat{\lambda}_{land} = 36.767$.

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 B.4).

Elasticity of substitution between varieties. We first estimate the elasticity of substitution between varieties in each sector. We estimate the following empirical specification implied by the expenditure share of each variety (4):

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

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.[25] $p_{o,K,t,h}(j)$

[25]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

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 $\ln RA_o$ (where RA_o is defined by (3)) which is an exogenous price shifter due to a trade mechanism but plausibly uncorrelated with local preference shocks given controls.

Table D.1 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 implied by the expenditure share of each variety (5):

$$\ln(\tilde{\alpha}_{o,K,t,h}) = \bar{\beta}_0 + (1-\bar{\sigma})\ln \hat{P}_{o,K,t,h} + \bar{\beta}_1 X_{o,t,h} + \bar{\phi}_K + \bar{\phi}_t + \epsilon_{o,K,t,h} \tag{D.2}$$

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 $\ln RA_o$.[26]

Table D.1 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 complementary. 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.

D.2 Algorithm for the Model Inversion

The detail is under construction.

D.3 Algorithm for the Model Simulation

The detail is under construction.

expenditure within 15 days prior to the interview date.

[26]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 powers for prices disaggregated at varieties.

Table D.1: Elasticity of Substitution across Varieties and Sectors

	log(Expenditure share)			
	Across varieties		Across sectors	
	(1)	(2)	(3)	(4)
	OLS	IV	OLS	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 RA 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 RA.

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