Policy Research Working Paper 10415

# The Effects of Transportation Infrastructure on Deforestation in the Amazon

## A General Equilibrium Approach

*Rafael Araujo*
*Juliano Assunção*
*Arthur Amorim Bragança*

WORLD BANK GROUP
Environment, Natural Resources and Blue Economy Global Practice
April 2023

Policy Research Working Paper 10415


## Abstract

Investments in transportation infrastructure can impact the environment beyond their immediate surroundings. This paper builds an interregional trade model to estimate the general equilibrium effects of changes in infrastructure on deforestation. Using panel data on the evolution of the transportation network in Brazil and land use data in the Amazon, the paper estimates the model and finds sizable effects of infrastructure on deforestation. Model simulations show that ignoring general equilibrium underestimates the impacts of deforestation by one-quarter. The paper also shows that the model can be used for evaluation of the deforestation induced by individual projects, which is an essential input for public policies.


This paper is a product of the Environment, Natural Resources and Blue Economy Global Practice. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://www.worldbank.org/prwp. The authors may be contacted at aamorimbraganca@worldbank.org.

# The Effects of Transportation Infrastructure on Deforestation in the Amazon: A General Equilibrium Approach*

Rafael Araujo† Juliano Assunção‡ Arthur Amorim Bragança§

JEL: *F18, Q17, Q56, O13*

Keywords: *Deforestation, Amazon, Transportation, General Equilibrium*

*We would like to thank Rodrigo Adão, Francisco Costa, Teevrat Garg, Marcelo Sant'anna, and seminar participants at Climate Policy Initiative, FGV-EPGE, USP, RIDGE, and SBE for valuable comments and suggestions. We are grateful to Helena Arruda, Daniel Barbosa, Mateus Morais, and Brenda Prallon for excellent research assistance. This material is based on work supported by Norway's International Climate and Forest Initiative (NICFI) and the Gordon & Betty Moore Foundation. The views expressed here do not necessarily reflect those of these organizations or those of the World Bank or their member countries. All errors are our own.

†Climate Policy Initiative (e-mail: carlquist.rafael@gmail.com).

‡PUC-Rio and Climate Policy Initiative (e-mail: juliano.assuncao@cpiglobal.org).

§World Bank (e-mail: aamorimbraganca@worldbank.org).

# 1 Introduction

Forest conservation is critical to tackling global warming (Baccini et al., 2012; Lawrence and Vandecar, 2015; Gatti et al., 2021). Indeed, in developing countries like Brazil or Indonesia, deforestation corresponds to the bulk of $CO_2$ emissions, with its control being the backbone of their environmental policies and international environmental commitments (IPCC, 2017). However, one key question is whether there is a trade-off between these environmental commitments and policies focused on promoting economic growth.

Investments in transportation infrastructure have long been considered a pillar for economic growth in developing countries for their potential for reducing international and intra-national trade costs (Atkin and Donaldson, 2015; Costinot and Donaldson, 2016; Donaldson and Hornbeck, 2016; Jedwab and Storeygard, 2017; Fajgelbaum and Redding, 2022). Nonetheless, assessing the effects of these investments on deforestation has proven difficult, given their endogenous placement and potential general equilibrium effects. Even when exogenous investment rules enable the identification of causal effects (e.g., Asher et al. (2019)), the focus on local effects implies it is not possible to understand the general equilibrium effects of investments in transportation infrastructure on deforestation.[1]

This paper develops a tractable framework to estimate the general equilibrium of investments in transportation infrastructure on deforestation in the Amazon. We extend Donaldson and Hornbeck (2016)'s inter-regional trade model to enable farmers to produce on two types of land: consolidated land and recently deforested land. In this setting, we show that the effect of transportation infrastructure on deforestation is captured by a log-linear relationship between deforestation and a sufficient statistic known in the literature as market access that measures how well connected each region is to all regions, especially

[1]See Chomitz and Gray (1996), Pfaff (1999), Cropper et al. (1999), Pfaff et al. (2007), Reis (2008), Gibson (2011), and Bebbington et al. (2018) for more general evidence on the relationship between transportation infrastructure and deforestation using cross-sectional and panel data.

densely populated ones. We then estimate the model exploring newly constructed panel data on deforestation and the evolution of the transportation infrastructure in Brazil since 1990.

The main challenge of estimating the model is to measure market access adequately. Market access is a function of three elements: bilateral trade costs between regions, the distribution of the population, and the elasticity of trade with respect to transportation costs. To measure bilateral transportation costs, we use GIS information on roads, railroads, rail stations, waterways, and ports, as well as administrative and survey data on freights to estimate, for each decade, the costs of transporting goods between all pairs of municipalities in Brazil and from each municipality to the nearest port (a proxy of access to international markets). The richness and flexibility of our transportation network allow for multi-modal paths (e.g., using roads plus railroads to transport goods between two regions) and non-linear transportation costs (e.g., trans-shipment costs between modes of transportation). We combine these matrices of bilateral transportation costs with official data on population and a calibrated trade elasticity to build a measure of market access for each municipality-by-decade pair.

We then regress deforestation constructed using satellite-based information on deforestation from Mapbiomas (2019) on this measure of market access to obtain our model's key elasticity. The main threat to identifying this model-based regression is the endogenous placement of transportation infrastructure. We leverage our panel structure to flexibly control for time-invariant municipality characteristics and the time-varying effects of geographic factors, an approach not possible in applications using cross-sectional data (e.g., Souza-Rodrigues (2018)). We also explore the network structure of the data to isolate the variation in market access coming from distant regions of each unit. This procedure enables us to account for the time-varying local unobservables that might drive infrastructure building (e.g., Donaldson and Hornbeck (2016)).

We find that a 1% increase in market access increases deforestation by roughly 0.5%. Quantitatively, this elasticity implies that one standard deviation increase in market access increases deforestation by 0.5 standard deviations. This effect is almost identical across different estimation strategies (OLS and 2SLS) and is not sensitive to calibrating the trade elasticity with other values found in the literature. We further find that deforestation implied by the model is highly correlated with deforestation observed in our data, indicating that our model predicts deforestation quite well.

To assess the importance of accounting for general equilibrium effects, we run a simulation exercise inside the model. We simulate 1,000 random roads inside the Brazilian Amazon and compute, for each simulation, the deforestation levels implied by the model. First, we include the road in the network and re-compute market access. Second, we use our estimates to compute the counterfactual deforestation predicted by the model. We then compare the results of the simulation with the results that would be obtained by a reduced form approach, using the municipalities crossed by the road as the treatment group and their neighbors as the control group. This procedure is equivalent to implementing a difference-in-differences estimator with treatment defined by the exposure to the new road as in Asher et al. (2019).

We find that ignoring general equilibrium effects would underestimate deforestation, on average, by one-quarter. The bias can be observed even in small projects since what determines the deforestation footprint is not the road itself, but what it connects. These results point to the perils of assuming that the outcome of one region is not affected by the treatment of other regions – the Stable Unit Treatment Value Assumption (SUTVA) – in a scenario where infrastructure placement creates feedback effects and changes optimal transportation paths in the whole infrastructure network.

One feature of our framework is that it can be used to evaluate the deforestation induced by individual projects. We illustrate this by evaluating the effects of the *Ferrogrão* railroad

– a highly controversial project planned to be built in the Amazon. We follow an identical procedure to the one used to evaluate the effects of simulated roads. This project generates 400 $km^2$ of deforestation, unevenly distributed around the *Ferrogrão*'s outline and extending beyond the project's immediate vicinity. This result highlights the limitations of the criteria used for evaluating infrastructure projects in Brazil as, currently, consultations with local populations and social and environmental impact assessments only consider the municipalities crossed by the project.[2]

Our work relates primarily to the literature estimating the effects of transportation infrastructure on deforestation (Chomitz and Gray, 1996; Cropper et al., 1999; Laurance et al., 2001; Bebbington et al., 2018; Damania et al., 2018; Vilela et al., 2020; Asher et al., 2019). We depart from previous work by investigating the general equilibrium effects of transportation infrastructure on deforestation. Our paper also adds to the literature on the drivers of Amazon deforestation (Assunção et al., Forthcoming, 2019, 2020; Fetzer and Marden, 2017; Souza-Rodrigues, 2018; Burgess et al., 2019; Araujo et al., 2020; Bragança and Dahis, 2022) and tropical deforestation in general (Burgess et al., 2012; Prem et al., 2020; Hsiao, 2021). Our results are also connected to the extensive literature which explores the trade-offs between economic development and environmental protection (Foster and Rosenzweig, 2003; Alix-Garcia et al., 2013; Damania et al., 2018; Garg and Shenoy, 2021).

The rest of the paper is organized as follows. Section 2 presents the model. Section 3 describes our data. Section 4 discusses identification and presents the estimation results. Section 5 discusses the importance of general equilibrium effects. Section 6 discusses the counterfactual results for the *Ferrogrão* project. Section 7 concludes.

[2] Antonaccio and Chiavari (2021); Cozendey and Chiavari (2021) provide an overview of the regulatory process of infrastructure building in Brazil with a focus on the Amazon

# 2 Theoretical Framework

In this section, we build an inter-regional trade model that enables us to evaluate the general equilibrium effects of transportation infrastructure on deforestation. Our model follows closely the model proposed by Donaldson and Hornbeck (2016). We extend their model by allowing farmers to produce in "consolidated lands" or to clear forests to produce in "frontier lands". We allow for different productivity shocks for the two types of land. We interpret these productivity differences as arising from differences in soil productivity or possible expropriation in "frontier lands". In particular, we see consolidated lands as plots in which agricultural production has been taking place for a long time where the soil is mature for production and property rights are well defined. For frontier lands, we see them as plots covered with native vegetation where the soil is not mature and there are expropriation risks linked to the lack of well defined property rights.

Our model keeps the tractability of Donaldson and Hornbeck (2016)'s original model, delivering a closed form expression connecting deforestation with a properly defined measure of market access that captures the connectivity of each region of the economy to all other regions, especially to densely populated ones. This expression summarizes the general equilibrium effects of the entire transportation network on land use.

## 2.1 Environment

The economy consists of a set of regions indexed by $o \in O$. Agents living in region $o$ supply ineslastically one unit of labor, earn wage $w^o$, and allocate consumption through a CES utility function over a continuum of varieties of agricultural goods $a(j)$ with $j \in [0, A]$. Agricultural goods can be traded across regions. Trade between regions $o$ and $d$ is subject to an iceberg transportation cost $\tau_{od}$. Agents are indifferent between the location of the producers of the good, buying from the municipality offering the lowest price. We denote by $p_o(j)$ the price of agricultural good $j$ faced by an agent in region $o$. The indirect utility $V_o$ of an agent living in $o$ is thus:

$$V^o = \frac{w_o}{P_o}, \tag{1}$$

in which $(P_o)^{1-\sigma} = \int_0^A p_o(j)^{1-\sigma} dj$ is the perfect price index of the goods consumed in municipality $o$.

Each agricultural variety is produced by perfectly competitive producers using labor, land, and capital as inputs. We assume production can be represented by a Cobb-Douglas production function with constant returns to scale. We let producers choose whether to use two different types of land $T$: consolidated ($C$) or frontier ($F$). Let $q^T$ denote the price of land of type $T$ and $r$ denote the capital price. The marginal cost of a producer operating in region $o$ is given by:

$$MC_o(j|T) = \frac{{q_o^T}^{\alpha} {w_o}^{\gamma} {r_o}^{1-\alpha-\gamma}}{z_o^T(j)}, \tag{2}$$

in which $z_o^T(j)$ is a productivity shock specific to variety $j$, region $o$ produced using land type $T$. Equation (2) is key to our model. It assumes the same production function is used to produce in the two types of land. Thus, the difference in marginal costs and equilibrium factor intensities between different types of land is driven by differences in land prices ($q^T$) and productivity shocks ($z^T$). This reflects the central trade-off between prices and productivity that producers face when choosing whether to use consolidated or frontier lands. We assume that capital and labor are freely mobile, which implies that $r_o = r, \forall o \in O$ and $V_o = V, \forall o \in O$.

## 2.2 Land Choices

The first step to characterize the equilibrium of the model is to derive conditions in which producers will operate in different types of land. Because consumers are indifferent to varieties produced in different types of land, the producer will operate in the land with lower marginal cost. Thus, a producer will operate in consolidated land instead of frontier land if and only if

$$\frac{z_o^F(j)}{z_o^C(j)} < \left(\frac{q_o^F}{q_o^C}\right)^{\alpha}$$

The expression above states that producers operate in consolidated lands whenever the productivity differences more than offset the price differences transformed by the land share in production.

We assume that the productivity shocks are drawn from a bivariate Fréchet distribution with CDF given by $F_o(z^C, z^F) = \exp(-(A_o^l {z^C}^{-\theta} + A_o^F {z^F}^{-\theta}))$. There are three important parameters in this distribution: $\theta$, $A_o^F$, and $A_o^C$. The parameter $\theta$ is negatively related to the dispersion of productivity shocks. Thus, lower (higher) $\theta$ implies more (less) dispersion and more (less) incentives to trade goods between regions. This parameter is often referred to as the trade elasticity in the literature (Eaton and Kortum, 2002). The parameters $A_o^F$ and $A_o^C$ controls the position of the marginal distributions for each type of land. Notice this bivariate Fréchet distribution implies independence of productivity shocks across different types of land.

The Fréchet distribution is commonly used in trade models because it facilitates the computation of equilibrium trade between regions (e.g., Eaton and Kortum (2002)). Indeed, one important feature of the bivariate Fréchet distribution is that, given prices, we can compute the probability that a farmer will choose consolidated instead of frontier land. We denote this probability by $\bar{p}_o$. We state this result in the following lemma.

**Lemma 1.** *The probability that a farmer will choose consolidated land is given by:*

$$\bar{p}\left(\frac{q_o^F}{q_o^C}\right) = P\left(\frac{z_o^F(j)}{z_o^C(j)} < \left(\frac{q_o^F}{q_o^C}\right)^{\alpha}\right) = \frac{1}{1 + \frac{A_o^F}{A_o^C}\left(\frac{q_o^F}{q_o^C}\right)^{-\theta\alpha}}$$

*Proof.* See Appendix A. ■

## 2.3 Prices and Exports

The second step to characterize the equilibrium of the model is to compute the price distribution of each region and the exports between each pair of regions.

First, we derive the price distribution. Because producers are perfectly competitive, the price $p_{o,d}(j)$ of the good $j$ produced in region $o$ and offered in region $d$ is the marginal cost of the good $j$ in region $o$ multiplied by the iceberg trade cost between these regions. Moreover, because consumers are indifferent between goods produced in different regions and different types of land, they will purchase from the cheapest source.

We compute the price distribution in region $o$ in three steps. We begin by showing that the price distribution of varieties produced in $o$ offered in $d$ is a univariate Fréchet distribution, despite the productivity shocks being distributed according to a bivariate Fréchet distribution (see Lemma A.1 in Appendix A). Following Eaton and Kortum (2002), we then show that the distribution of the prices of the varieties produced in $o$ sold for $d$ in equilibrium is identical to the distribution of offered varieties (see Lemma A.2 in Appendix A). We further show that the prices distributions of the varieties produced in $o$ and sold in $d$ in different types of land $T$ is identical to the distribution of the varieties produced in $o$ and sold in $d$ as a whole (see Lemma A.3 in Appendix A). Indeed, the sole difference between these distributions being the length of the varieties produced.

Using the three results discussed above, it is possible to write the price distribution in municipality $d$ as:

$$(P_d)^{-\theta} = x \sum_{o \in O} (\tau_{o,d} w_o{}^{\gamma})^{-\theta} \left( A_o^l (q_o^l)^{-\theta\alpha} + A_o^i (q_o^i)^{-\theta\alpha} \right) \equiv CMA_d, \tag{3}$$

in which $x$ is a constant.[3]

Following Redding and Venables (2004), we refer the transformed price index in (3) as the

[3]Our specification implies $x = \left[\Gamma\left(\frac{\theta+1-\sigma}{\theta}\right)\right]^{\frac{-\theta}{1-\sigma}} r^{-\theta(1-\alpha-\gamma)}$.

consumer market access of region $d$. The $CMA$ is a weighted sum of productivity-adjusted costs of production in each origin $o$ that supplies the destination $d$. It is denoted "consumer market access' because it measures the access of consumers in a region to cheap products.

Second, we derive the exports between each pair of origins and destinations. We begin by noting that each region of origin exploits its comparative advantage through the length of varieties it sells to each destination. This happens because, as in Eaton and Kortum (2002), the distribution of goods prices that region $d$ actually buys from region $o$, is the same as of the overall distribution of prices in $d$. It is then possible to obtain total exports from $o$ to $d$ ($X_{od}$) by multiplying the price distribution and the number (mass) of goods sold between regions. Lemma A.4 in Appendix A derives the length of varieties a region $o$ exports to a region $d$.

Using this lemma, we obtain the following expression for the exports from municipality $o$ to municipality $d$:

$$X_{od} = x \left(w_o^{\gamma} \tau_{od}\right)^{-\theta} \left(A_o^C (q_o^C)^{-\theta\alpha} + A_o^F (q_o^F)^{-\theta\alpha}\right) (CMA_d)^{-1} X_d, \tag{4}$$

in which $X_d = \sum_d X_{od}$ and $x$ is a constant.

## 2.4 Equilibrium

Market clearing implies the total output of a region ($Y_o$) is equal to the total demand for its products ($\sum_d X_{od}$) and that agents must be indifferent to living in all municipalities ($\bar{U} = V_o = w_o / P_o$). Using these conditions and the expressions for prices (equation 3) and exports (equation 4), we obtain the following log-linear expression connecting output, land prices, and measures of market access:

$$\log Y_o = \log x + \gamma \log CMA_o + \log\left(A_o^C (q_o^C)^{-\theta\alpha} + A_o^F (q_o^F)^{-\theta\alpha}\right) + \log FMA_o, \tag{5}$$

in which $FMA_o \equiv \sum_d \left[\tau_{od}{}^{-\theta}(CMA_d)^{-1}Y_d\right]$. The term $FMA$ is a sum of the size of destinations inversely weighted by the costs of shipping goods to these destinations ($\tau_{od}^{\theta}$) and their competitiveness ($CMA$). It is denoted "firm market access" because it measures the access that firms face to sell their products.

The concepts of "firm market access" and "consumer market access" are closely intertwined. Donaldson and Hornbeck (2016) prove that it is possible to write $FMA_o = \rho CMA_o$, in which $\rho$ is a constant. We use the term "market access" ($MA$) to denote $MA_o = FMA_o = \rho CMA_o$.[4]

To close the model, we need to substitute for $Y_o$, $q_o^C$, and $q_o^F$ in equation (5). We begin by noting that the Cobb-Douglas production function implies $Y_o = \left(q_o^C L_o^C + q_o^F L_o^F\right)/\alpha$, that is, the $\alpha$ share of output goes to the land factor. We then note that the Fréchet distribution implies that the share of the land rents which goes to consolidated land is equal to the probability a producer uses consolidated land ($\bar{p}_o$). This result follows from the fact that the price distribution of varieties is identical (up to scaling constant) for consolidated and frontier land. Furthermore, the proportion of goods exported using consolidated land is $\bar{p}_o$. This result means that the income from consolidated and frontier land are connected through the following lemma:

**Lemma 2.** *Total income accrued to frontier land equals total income accrued to consolidated land adjusted by the relative probability producers operate in each type of land. Thus,*

$$\bar{p}_o q_o^F L_o^F = (1-\bar{p}_o) q_o^C L_o^C$$

*Proof.* See Appendix A. ■

The last element of the model is a specification of land supply for each type of land. We

[4] This implies market access can be written as $MA_o \equiv \rho \sum_d \left[\tau_{od}{}^{-\theta}(MA_d)^{-1}Y_d\right]$. Substituting for population ($\gamma Y_o = w_o N_o$ and $\bar{U} = V_o = w_o/P_o$), it is possible to re-write this expression as $MA_o \equiv \frac{\bar{U}\rho^{\frac{1}{\theta}+1}}{\gamma}\sum_d \left[\tau_{od}{}^{-\theta}(MA_d)^{-\frac{1+\theta}{\theta}}N_d\right]$.

assume that the supply of consolidated land is fixed, that is, $L_o^C = \bar{L}_o^C$. This assumption implies that our model collapses to Donaldson and Hornbeck (2016)'s model if there are no frontier lands.

We motivate the existence of a positively sloped supply curve for frontier lands with a simple setting of heterogeneous cost of deforestation. Due to heterogeneity in topography and forest density of different plots of land, the marginal cost of clearing land for agricultural production in region $o$ is increasing in the amount of land to be cleared. Thus, given a price of frontier land, $q_o^F$, a plot of land will cleared if the clearing cost does not exceed this price. Suppose that the probability of the marginal cost of clearing land is lower than $q_o^F$ is $B_o^{-1} q_o^{F\frac{1}{\eta}}$. Thus, the relationship between land price and the total amount of land which is cleared is $q_o^F = B_o \left(L_o^F\right)^\eta$.

Using the land supply curves and the expressions for $q_o^C$ and $q_o^F$, we obtain a closed form relationship between deforestation and market access:

$$(\eta + 1 + \eta\theta\alpha) \log L_o^F = \log \frac{x A_o^F}{B_o \rho^\gamma \bar{U}^{\gamma\theta}} + (1+\gamma) \log MA_o \tag{6}$$

As discussed before, market access is a function of trade costs ($\tau_{od}$), population ($N_d$) and three model parameters ($\theta$, $\bar{U}$ and $\rho$). It is hard to measure some of these parameters (e.g, $\bar{U}$) in the data. Thus, as Donaldson and Hornbeck (2016), we consider the following first-order approximation of market access in the empirical work:

$$MA_o \cong \sum_d \tau_{od}^{-\theta} N_d \tag{7}$$

The expression above significantly reduces the challenges for computing market access as it is a function only of observable data and the trade elasticity that can be calibrated using values from the literature (e.g., Eaton and Kortum (2002)).

Equations (6) and (7) are the main equations we use throughout the empirical analysis. They show that transportation costs between regions ($\tau_{od}$) influence deforestation solely through their effect on market access, that is, that market access is a sufficient statistic for the effects of transportation costs on deforestation. Therefore, it is possible to use them to evaluate quantitatively the effects of transportation infrastructure on deforestation in a general equilibrium setting.

# 3 Data Construction

## 3.1 Market Access

To compute market access, we combine newly constructed data on bilateral transportation costs over time ($\tau_{odt}$), population ($N_{ot}$), and a trade elasticity parameter ($\theta$). Below we detail how we measure each of these components.

**Transportation Costs**. We collect data from different sources to construct measures of bilateral trade costs between all pairs of municipalities and between municipalities and the nearest port with access to international markets over time ($\tau_{odt}$).

To construct a panel of the main roads network, we collect data on federal roads in Brazil from the Ministry of Transportation for the years 1990, 2000, and 2010. Figure 1, panels A, B, and C show the evolution of the roads network throughout the decades. We also have data on their traffic conditions, with each road being classified as paved or unpaved. It is possible to note that even though most roads are paved in Brazil, a significant proportion of roads are unpaved for the Amazon region.

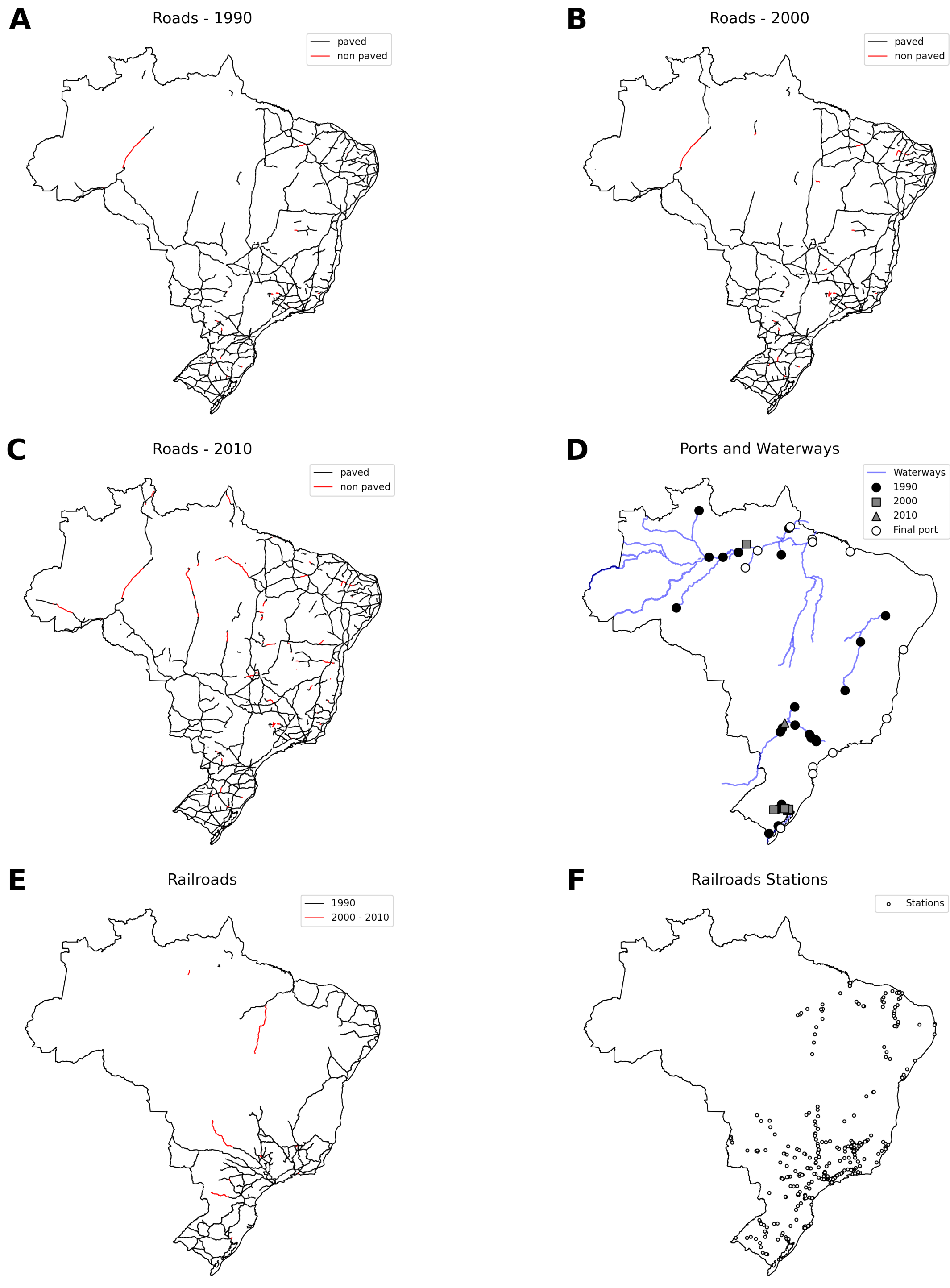
We collect data on railroads, navigable rivers, railroad stations, and ports. We allow agents to access waterways and railroads only through ports and railroads stations after paying a loading (trans-shipment) cost. This is a simple way to allow for non-linearity in transportation costs. Data on railroads is available from the Ministry of Transportation for the years 1990, 2000, and 2010. We overlap the railroad system of each year with the present data of railroad stations from the Ministry of Transportation to determine the location of stations throughout time. The waterways data does not vary across time, but we get time-series variation in transportation costs on water using the construction of new ports. We classify ports into two categories: final ports and intermediaries ports. Final ports have direct access to international markets and enough infrastructure for sea ships. Intermediary ports are the ones that are used as a way to access the waterway, to then access a final port, or to change transportation mode again to roads or railroads. Figure 1 shows the evolution of our transportation network.

We further collected data on the transportation cost of soy from the Group of Research and Extension in Agroindustrial Logistics of the College of Agriculture Luiz de Queiroz (SIFRECA) from 2008 to 2014 (ESALQ-LOG, 2008-2014). This data set provides surveyed transportation costs per ton of product between multiple destinations. We also collect data on yearly soy prices from the Center of Advanced Studies for Applied Economics of the College of Agriculture Luiz de Queiroz (CEPEA) (ESALQ-LOG, 2008-2014).

To compute transportation costs, we convert our transportation data into a graph (network) structure. In this graph, we use the Dijkstra's shortest path algorithm to find the least-cost path connecting two nodes in this graph (Dijkstra, 1959). Our graph structure is very flexible. It allows for multi-modal paths, ensuring agents can combine waterways, railroads, and roads to ship goods between nodes of our graph. It also incorporates non-linearity by restricting access to railroads and waterways to nodes with stations or ports and adding trans-shipment costs to move goods into and out of stations and ports. Figure 2 illustrates how the conversion from a map to a graph happens. The graph is built by breaking down the transportation network into small steps and assigning connections between those steps.

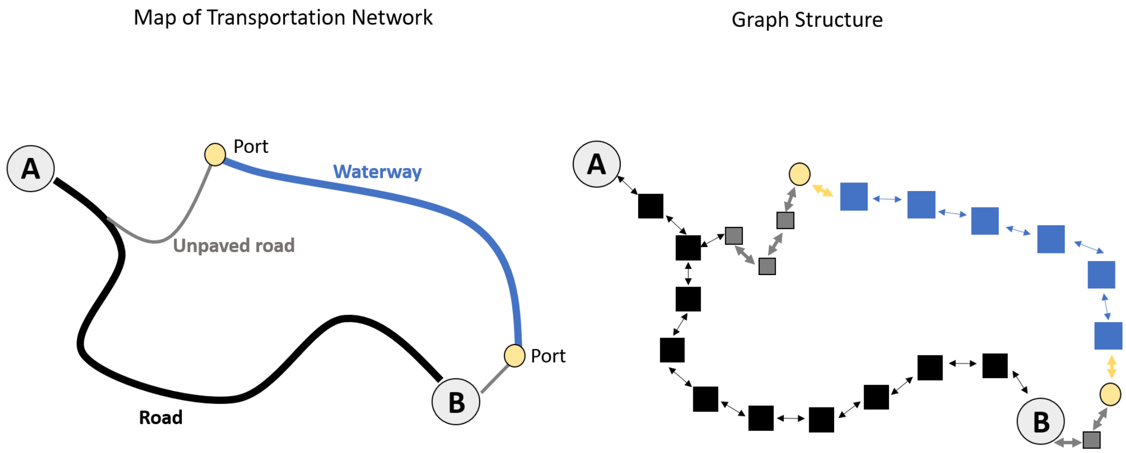
One key challenge in building the graph structure is assigning a cost for traversing each

Figure 1: Transportation Network




**Notes:** This figure describes the transportation network used in the paper. Panels A-C depict the federal roads by type of pavement for 1990, 2000, and 2010; Panel D depicts the location of waterways, ports, and the year of construction of the ports; Panel E depicts the railroads and their period of construction. Data is aggregated in "until 1990" and "after 1990" as the construction was minimal in the last decades. Panel F depicts the railroad stations. We overlap this map of stations with the map of railroads to determine existing stations for each year.

Figure 2: Converting map into graph (network)




**Notes:** This figure shows two markets – A and B – connected by a system of roads, a waterway, and ports. This transportation network is converted to a graph, composed of nodes (squares and circles) and vertices (arrows). Notice that the cost of moving from a port to a waterway is different than the other costs (the yellow line), representing the flexibility of the application in incorporating different transshipment costs on the transportation cost model.

type of node. We have a total of twelve types of nodes in our graph: paved roads (inside and outside of the Amazon), unpaved roads (inside and outside of the Amazon), no roads (inside and outside of the Amazon), protected areas (inside and outside of the Amazon), railroads, waterways, railroad stations, and ports. We choose these costs based on Araujo et al. (2020), incorporating heterogeneous costs of roads in the Amazon, as in Souza-Rodrigues (2018). For the trans-shipment costs in ports and railroad stations, we use the average maximum values allowed to be charged by the operator of a railroad compared with the average cost to transport agricultural goods by roads as in ESALQ-LOG (2008-2014).[5]

We use the following costs to traverse each type of node: paved road (inside the Brazilian Amazon), 10 (20); unpaved road (inside the Brazilian Amazon), 20 (40); no roads (inside the Brazilian Amazon), 50 (100); protected areas (inside the Brazilian Amazon), 100 (200); railroads, 5; waterways, 5; trans-shipment costs, 200 (see Table C.1). Notice that these

[5]We use the most recent concession contracts available at the Brazilian National Land Transport Agency (ANTT).

values are scale-invariant – their relative values determine the shortest paths chosen by the algorithm.

We then apply the Dijkstra's shortest path algorithm to compute the transportation cost between all possible pairs of municipalities and between all municipalities and final ports for each year. This procedure results in a unit-free measure of bilateral costs called *cost_graph*. To transform this measure into a measure of iceberg transportation costs, we fit the following linear model:

$$cost_{odt} = \alpha + \beta cost\_graph_{odt} + \epsilon_{odt},$$

in which $cost_{odt}$ is the proportional (iceberg) cost of transporting one ton of soy between municipalities $o$ and $d$ in year $t$ – freight cost divided by product price – from the ESALQ-LOG (2008-2014). Table C.2 reports the results. The $R^2$ of this regression is 0.63.

We use the coefficients of this regression to convert all our graph costs to iceberg costs, that is, we set $\tau_{odt} = 1 + \tilde{\alpha} + \tilde{\beta} cost\_graph_{odt}$.

**Population.** We use municipality-level data from the demographic census for 1991, 2000, and 2010 collected by the Brazilian Institute of Geography and Statistics (IBGE). This provides us direct measures of the size of all municipalities in Brazil for each decade. Measuring the size of international markets is more challenging as it requires constructing population-equivalent measures of the importance of these markets for producers located in the Amazon.

There are two approaches for incorporating international markets in the construction of market access. The first one, used by Donaldson and Hornbeck (2016), inflates the population of regions with direct access to ports to reflect the importance of consumers in other countries. The second one, used by Baum-Snow et al. (2020), includes another region in the model with the population chosen to reflect the importance of the consumers in other countries. We follow the second approach because it enables us to perform a useful decomposition of the effects of access to national and international markets. We therefore extend our expression of market access (7) to:

$$MA_{o,t} \cong \tau_{opt}^{-\theta} N_{p,t} + \sum_{d} \tau_{odt}^{-\theta} N_{d,t}, \tag{8}$$

which $\tau_{opt}^{-\theta}$ denotes the iceberg cost from region $o$ to a port with access to international markets and $N_{p,t}$ denotes the international markets equivalent population at time $t$. We set the equivalent population of international markets as the total exports divided by the Brazilian GDP per capita for each decade. Table C.3 in the Appendix gives the population totals by decade. Finally, we include an additional cost of 15% on top of the transportation cost to account for time and bureaucracy costs of shipping products internationally, as in Baum-Snow et al. (2016).

**Trade Elasticity.** Estimating the trade elasticity ($\theta$) in our setting is impossible because we do not observe data on trade. Thus, we calibrate this parameter using numbers from the literature. In our preferred specification, we set $\theta = 8.2$, a value close to both the preferred estimate in Eaton and Kortum (2002) and the main calibrated value in Donaldson and Hornbeck (2016). We explore the robustness of our results to different values of $\theta$ reported in the literature (Eaton and Kortum, 2002; Costinot et al., 2012; Simonovska and Waugh, 2014; Head and Mayer, 2014).

**Units.** The number of municipalities observed in our data changes over time due to the creation of new municipalities. We deal with this issue using the concept of minimum comparable areas (AMC), neighboring municipalities that can be consistently compared across time.[6] This leaves us with 4,297 minimum comparable areas for Brazil and 426 for the Amazon for 1990-2019. For simplicity, we denote these minimum comparable areas as municipalities throughout the text.

[6]See Ehrl (2017) for further information on how an AMC is defined. In our setting we use AMCs defined in the year 1991.

**Market Access.** After gathering the information on $\tau_{odt}$, $N_{ot}$, and $\theta$, we build the market access variable. Figure 3 shows the distribution of market access in 1990 and the difference in market access between 2010 and 1990 for Brazil (panels A and B) and the Amazon (panels C and D). We normalize market access by its maximum value, so it is bound between zero and one. The data highlights the isolation of the Amazon. The average market access in the Amazon varied between 35% and 40% of the average market access for the rest of the country from 1990-2010. Table 1, columns 1 to 3 show that the market access in the Amazon increased 15 percentage points between 1990 and 2010. This was accompanied by an increase of 8 percentage points in the dispersion of market access among municipalities in the Amazon.

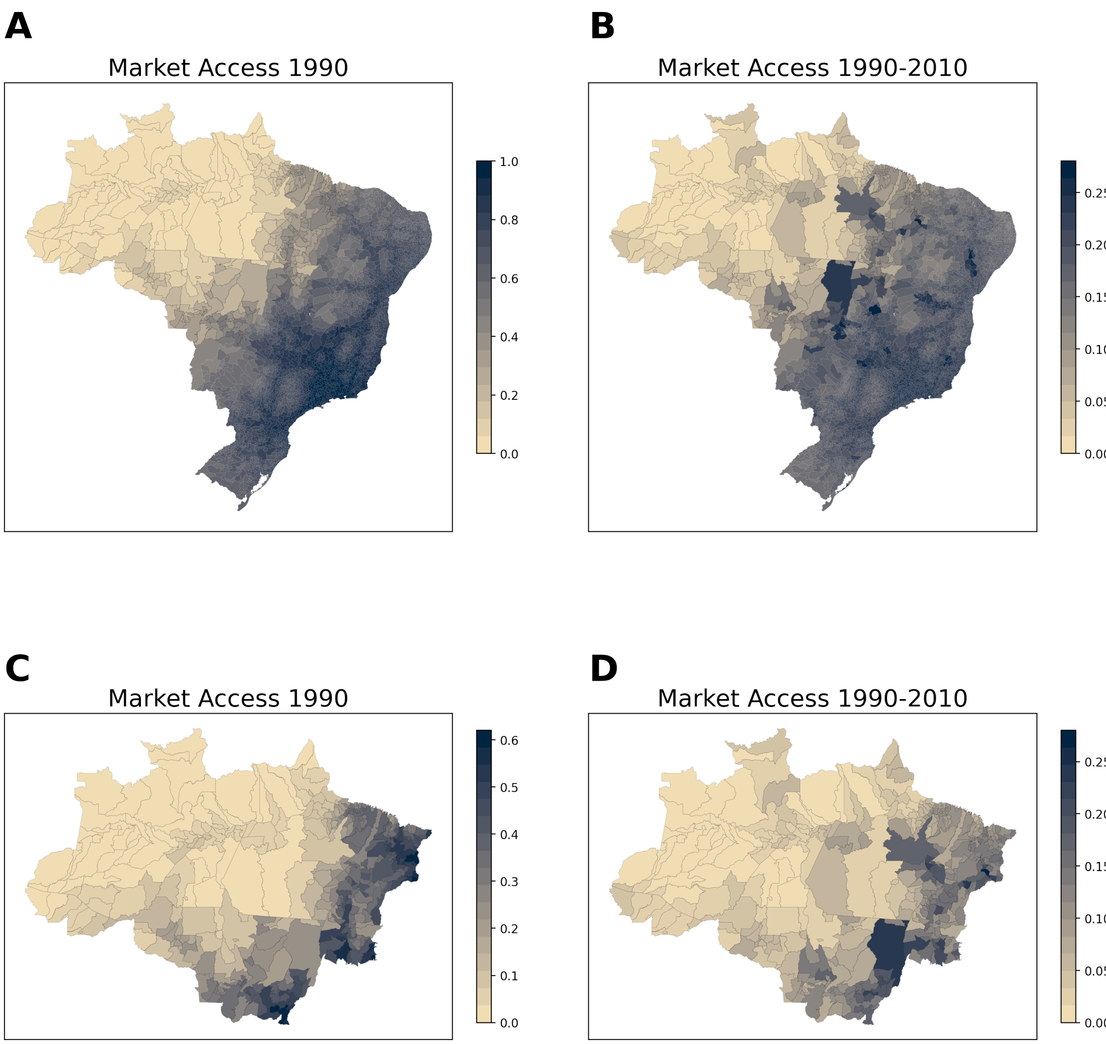
## 3.2 Deforestation

We use data from Mapbiomas (2019) to measure deforestation. This data enables us to measure deforestation for a more extended period than what would be available with other commonly used data sources such as Hansen et al. (2013). Using satellite images and ground truth observations, Mapbiomas classifies, for the years between 1985 and 2019, each pixel of 30 meters in a range of land uses. For each pixel, we identify the first year, if ever, that the pixel was deforested. We then sum the total area of the pixels deforested in each municipality-decade pair.[7]

Table 1, columns 4 to 6 reports summary statistics on deforestation. The dynamics of forest clearing in the Amazon changed drastically during these decades. Deforestation was high until the beginning of the 2000s. It then fell abruptly following the implementation of the Action Plan for Prevention and Control of the Legal Amazon Deforestation (Assunção et al., 2015, 2020; Assunção et al., 2022; Assunção et al., Forthcoming; Bragança and Dahis, 2022), changes in macroeconomic conditions (Assunção et al., 2015), and supply chain

[7] In the period that overlaps Hansen et al. (2013) and Mapbiomas (2019) data - from 2001 to 2019 - the $R^2$ of a regression of Hansen et al. (2013)'s deforestation on a constant and Mapbiomas (2019)'s deforestation is 0.97.

Figure 3: Market Access




**Notes:** This figure depicts the evolution of market access variable in the period 1990-2010. To facilitate visualization, we divide the market access of each municipality by the highest market access observed in the period. Panels A and B display data for all municipalities in Brazil, while Panels C and D display data only for municipalities in the Amazon. Panels A and C report market access in the beginning of the period studied in the paper (1990), Panels B and D report the change in market access between 1990 and 2010.

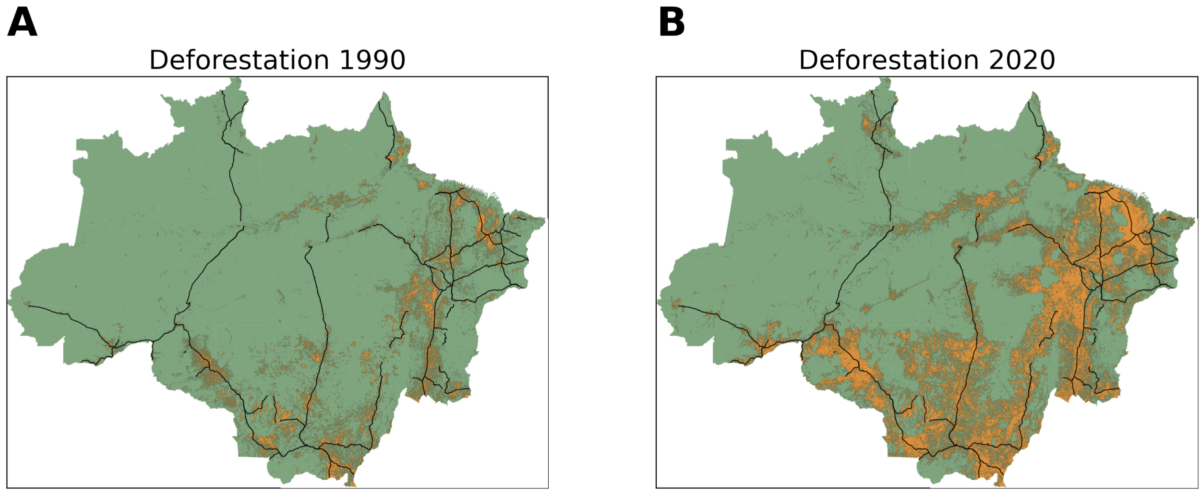
initiatives (Heilmayr et al., 2020; Villoria et al., 2022). Deforestation increased again at the end of the 2010s following a reversal of conservation policies (Burgess et al., 2019).

## 3.3 Geography and Productivity

Our empirical model uses latitude, longitude, distance to *Brasília*, distance to the coast, and suitability for cultivating soy as controls. Latitude and longitude are the coordinates of the municipality's main district. Distance to the coast and distance to *Brasília* are the distance of the main district to the national capital and to the coast. Suitability is the average suitability for cultivating soy for the pixels from the Global Agro-Ecological Zones from the Food and Agriculture Organization of the United Nations (FAO GAEZ version 3, Agricultural Suitability for rain fed crops utilizing high level of inputs) falling in the municipality. Table 1 reports descriptive statistics for these variables. There is considerable cross-sectional variation in them.

We close our data section by discussing the spatial correlation between roads and deforestation observed in the data. Figure 4, panels A and B reports the spatial distribution of deforestation and roads at the beginning and the end of our study period. Deforestation occurs close to roads in both periods.

Figure 4: Roads and Deforestation




**Notes:** This figure depicts the evolution of deforestation and its spatial correlation with roads. Panels A and B show the cumulative deforestation footprint (in orange) for 1990 and 2020. The black lines are the federal roads as of 2010.

# 4 Identification and Estimation Results

## 4.1 General Setting

Our empirical framework explores differences in the evolution of market access and deforestation across municipalities in the Amazon to estimate the key elasticity of our theoretical model.

The fact that our theoretical model is static and our data exhibits cross-sectional and temporal variation warrants some notes on dynamics. First, the immediate effects of changes in market access on deforestation might differ from equilibrium effects for two reasons: (1) it takes time for agents to adjust to changes in market access; (2) improvements in transportation infrastructure might have transitory effects on deforestation (see Asher et al. (2019) for a discussion on this). Second, we measure deforestation more frequently than we can measure market access. In this context, we build our estimation using long differences connecting the cumulative deforestation in a decade with market access measured

Table 1: Descriptive Statistics

| | **Market Access** | | | **Deforestation** ($km^2$) | | |
|---|---|---|---|---|---|---|
| | 1990 | 2000 | 2010 | 1990-1999 | 2000-2009 | 2010-2019 |
| mean | 0.32 | 0.39 | 0.47 | 612.97 | 585.79 | 281.82 |
| std | 0.20 | 0.24 | 0.28 | 1264.74 | 1494.71 | 631.06 |
| 25% | 0.12 | 0.14 | 0.2 | 66.69 | 45.12 | 31.98 |
| 50% | 0.35 | 0.42 | 0.51 | 161.29 | 125.53 | 91.84 |
| 75% | 0.49 | 0.6 | 0.71 | 612.08 | 433.56 | 291.48 |

| | **Geography** | | | | | |
|---|---|---|---|---|---|---|
| | Soybeans (kg/ha) | Distance to Brasília (km) | Distance to coast (km) | Area ($km^2$) | Latitude | Longitude |
| mean | 3424.58 | 1404.2 | 882.87 | 8046.16 | -6.51 | -52.3 |
| std | 492.91 | 559.48 | 684.09 | 12628.76 | 4.83 | 7.47 |
| 25% | 3236.0 | 996.27 | 229.53 | 1336.96 | -10.02 | -57.93 |
| 50% | 3578.0 | 1420.89 | 847.23 | 3702.49 | -5.39 | -49.5 |
| 75% | 3702.0 | 1661.34 | 1372.81 | 9368.16 | -2.63 | -46.81 |

**Notes:** This table reports descriptive statistics for the variables used in the estimation of our model. The upper panel shows descriptive statistics for market access and deforestation for each decade. The bottom panel shows descriptive statistics for the (time-invariant) geographic characteristics used in the estimation. All statistics are calculated for all the 426 minimum comparable areas (municipalities) in the Amazon during the period 1990-2019.

at the beginning of that decade. For this, we estimate the following empirical analog of equation (6):

$$\log y_{o,t} = \alpha + \beta \log MA_{o,t_I} + \phi_t X_o + \gamma_o + \gamma_{s,t} + \epsilon_{o,t}, \tag{9}$$

in which $y_{o,t}$ is the total deforestation observed in decade $t$, $MA_{o,t_I}$ is market access at the beginning of decade $t$, $X_o$ is a vector of time-invariant controls (cubic polynomials on latitude and longitude, distance to *Brasília*, distance to the coast, suitability for cultivating soy), $\gamma_o$ is a municipality fixed effect, $\gamma_{s,t}$ is a state $\times$ year fixed effect, and $\epsilon_{o,t}$ is an idiosyncratic error term. Notice that $MA_{o,t_I}$ is endogenous by construction because it depends on the region's population and therefore is co-determined by the region's land use. To deal with this problem, we do not consider the region's population in the computation of its market access.

It is important to notice that we use data for all regions in Brazil when building market access. However, as we are interested in modelling deforestation, we estimate equation (9) using data on deforestation and market access just for the municipalities located in Brazil's Amazon. Since our measure of market access considers consumers located outside the Amazon, we are considering not only the importance of trade *within* the Amazon, but also the importance of trade *between* the Amazon with the other regions of Brazil, and trade between the Amazon and other countries. This is important because the Amazon's population is small, around 15% of Brazil's population.

## 4.2 Results and a Discussion on Identification

Table 2, columns 1 to 3 report OLS estimates of equation (9). Column 1 includes municipality fixed effects, state-year fixed effects, and third-degree polynomials of latitude and longitude. Column 2 further includes distance to the coast and distance to *Brasília* (the national capital) interacted with time dummies as controls. Column 3 adds the suitability for cultivating soy interacted with times dummies as controls. We weight observations by municipality area (excluding protected areas) to recover the effects on the typical hectare and cluster standard errors at the municipality level to deal with serial correlation in the error term.

We find that a 1% increase in market access increases deforestation by 0.5%. Quantitatively, this elasticity implies that one standard deviation increase in market access increases deforestation by 0.5 standard deviations. The inclusion of different sets of controls does not influence the estimates.

One potential problem with OLS estimation of Equation (9) is the potential correlation between market access with non-observed local productivity shocks. Our OLS specifications deal with this problem by flexibly controlling for time-invariant municipality characteristics and the time-varying effects of geographic factors. However, there might still be components of these productivity shocks not absorbed by the fixed effects and controls. Following the literature (e.g., Donaldson and Hornbeck (2016) and Jedwab and Storeygard (2021)), we explore variation in market access coming from changes in transportation costs far from the region of interest to deal with this issue.[8] The identification hypothesis behind this instrument is that changes in transportation costs and population further than a distance $d$ from a region are not correlated with its own productivity shocks.

Table 2, columns 4 to 6 report the results from 2SLS obtained using this measure of distant market access as an instrument for market access. We set $d = 400$km. First-stage regressions show a strong correlation between distant market access and actual market access. Second-stage regressions indicate that the elasticity of deforestation to market access obtained using 2SLS is remarkably similar to one obtained using OLS. This finding diminishes concerns that local shocks drive the relationship between market access and deforestation reported in Table 2, columns 1 to 3.

[8] This means that we eliminate from the computation of market access of each municipality its neighbors located within a radius of $d$ kilometers.

Table 2: Market Access and Deforestation

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | **log (Deforestation)** | | | | | |
| log(Market Access) | 0.45*** | 0.51*** | 0.47*** | 0.47*** | 0.52*** | 0.49*** |
| | (0.13) | (0.13) | (0.13) | (0.13) | (0.13) | (0.13) |
| $R^2$ (within) | 0.16 | 0.16 | 0.17 | 0.16 | 0.16 | 0.17 |
| Observations | 1,278 | 1,278 | 1,278 | 1,278 | 1,278 | 1,278 |
| | **First stage: log(Market Access)** | | | | | |
| log(Market Access, $d = 400$km) | | | | 0.96*** | 0.96*** | 0.96*** |
| | | | | (0.002) | (0.002) | (0.002) |
| F Statistic | | | | 869,866 | 883,539 | 892,669 |
| Observations | | | | 1,278 | 1,278 | 1278 |
| Lat-Long | Yes | Yes | Yes | Yes | Yes | Yes |
| Distance | No | Yes | Yes | No | Yes | Yes |
| Soil | No | No | Yes | No | No | Yes |

**Notes:** This table reports the results of estimating Equation (9). All specifications include municipality and state-year fixed effects. Columns 1-3 report the results of OLS specifications. Columns 4-6 report the results of a 2SLS specifications obtained using market access excluding observations within a buffer of radius $d =$ 400km as an instrument for market access. Columns 1 and 4 include cubic polynomials of latitude and longitude as controls ('lat-long'). Columns 2 and 5 include distance to the coast and distance to *Brasília* as additional controls ('distance'). Columns 3 and 6 include suitability for cultivating soy as an additional control ('soil'). All controls are interacted with time dummies. The regressions are estimated for all the 426 minimum comparable areas (municipalities) in the Amazon during the period 1990-2019. Standard errors clustered at the municipality level are reported in parenthesis. *** p<0.01; ** p<0.05; * p<0.10.

Table 3 explores the robustness of these results to other instruments. Column 2 uses as the instrument a measure of market access constructed by eliminating all municipalities within the same state in its computation. The results do not change. Column 3 uses a measure of market access built holding population in 1990 fixed as the instrument. Again, results do not change, implying that changes in transportation costs are important to identify our elasticity, that is, our estimate is not driven purely by population changes. This finding is important as the relevant counterfactuals of interest are the ones in which the transportation network changes. Finally, column 4 uses as the instrument a measure of domestic market access obtained by removing international markets. The elasticity obtained is once more identical to the ones obtained in the other specifications, pointing out the relevance of changes in domestic market access to identification. Table C.5 in the Appendix shows the robustness of our results when using only the domestic market access to identification across a range of specifications of controls.

Table 4 provides evidence the robustness of our results to different values of the trade elasticity ($\theta$) reported in the literature (see Table C.4 for a list of references and their estimated/used trade elasticity). We find similar elasticities for different values of $\theta$, both with and without using the constrained version of market access as an instrument. It is important to note that a change in the trade elasticity ($\theta$) changes the dispersion of the market access variable yielding a different elasticity of market access and deforestation. Nonetheless, as Table 4 shows, the counterfactual effect of changing one standard deviation of the new market access variable is remarkably close regardless of the trade elasticity.

Different weighting procedures do not influence the results qualitatively (see Table C.6 in the Appendix). Point estimates of the elasticity obtained weighting by the square root of municipality area or by not using weights are larger than the ones obtained in our preferred specification (see Solon et al. (2015) for a discussion on different weighting procedures). However, it is not possible to rule out that the coefficients obtained using different weighting schemes are equal. Thus, if anything, the results from Table C.6 suggest that

Table 3: Market Access and Deforestation, Alternative Instruments

| | $d = 400$km | Out-of-state | Fixed pop. | Dom. market |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | **log (Deforestation)** | | | |
| log(Market Access) | 0.49*** | 0.48*** | 0.46*** | 0.52*** |
| | (0.13) | (0.13) | (0.13) | (0.14) |
| $R^2$ (within) | 0.17 | 0.17 | 0.17 | 0.17 |
| Observations | 1,278 | 1,278 | 1,278 | 1,278 |
| | **First stage: log(Market Access)** | | | |
| log(Alt. Market Access) | 0.96*** | 0.97*** | 1.02*** | 0.81*** |
| | (0.01) | (0.01) | (0.01) | (0.01) |
| F Statistic | 892,669 | 1,441,286 | 369,575 | 27,950 |
| Observations | 1,278 | 1,278 | 1,278 | 1,278 |

**Notes:** This table reports the results of estimating Equation (9) using different instruments. All specifications include municipality and state-year fixed effects as well as controls for geography (cubic polynomials on latitude and longitude, the distance to the coast and to *Brasília*, and suitability for cultivating soy) interacted with year dummies. In each column, market access is instrumented by a different variable: in column 1 by a constrained market access measure which excludes observations within a 400km buffer; in column 2 by a constrained market access measure which excludes observations within the same state; in column 3 by a market access measure constructed holding population at its 1990 level; in column 4 by domestic market access, that is, by a measure of market access obtained setting the equivalent population of international markets to zero. The regressions are estimated for all the 426 minimum comparable areas (municipalities) in the Amazon during the period 1990-2019. Standard errors clustered at the municipality level are reported in parenthesis. *** p<0.01; ** p<0.05; * p<0.10.

Table 4: Market Access on Deforestation for Different $\theta$'s

| | $\theta = 8.2$ | $\theta = 6.5$ | $\theta = 4$ | $\theta = 8.2$ | $\theta = 6.5$ | $\theta = 4$ |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | **log (Deforestation)** | | | | | |
| log(Market Access) | 0.47*** | 0.59*** | 0.64*** | 0.49*** | 0.58*** | 0.51*** |
| | (0.13) | (0.16) | (0.19) | (0.13) | (0.16) | (0.19) |
| $R^2(within)$ | 0.17 | 0.17 | 0.17 | 0.17 | 0.17 | 0.17 |
| Observations | 1,278 | 1,278 | 1,278 | 1,278 | 1,278 | 1,278 |
| | **First stage: log(Market Access)** | | | | | |
| log(Market Access, $d = 400$km) | | | | 0.96*** | 0.79*** | 1.12*** |
| | | | | (0.01) | (0.01) | (0.01) |
| F Statistic | | | | 892,669 | 18,968 | 30,131 |
| Observations | | | | 1,278 | 1,278 | 1,278 |
| std(log Market Access) | 1.07 | 0.87 | 0.74 | 1.07 | 0.87 | 0.74 |
| effect of +1 std | 0.50 | 0.51 | 0.48 | 0.52 | 0.50 | 0.38 |

**Notes:** This table reports the results of estimating Equation (9) for different trade elasticities ($\theta$). All specifications include municipality and state-year fixed effects as well as controls for geography (cubic polynomials on latitude and longitude, the distance to the coast and to *Brasília*, and suitability for cultivating soy) interacted with year dummies. Columns 1-3 report the results of OLS specifications. Columns 4-6 report the results of a 2SLS specifications obtained using market access excluding observations within a buffer of radius $d = 400$km as an instrument for market access. The regressions are estimated for all the 426 minimum comparable areas (municipalities) in the Amazon during the period 1990-2019. Standard errors clustered at the municipality level are reported in parenthesis. *** p<0.01; ** p<0.05; * p<0.10.

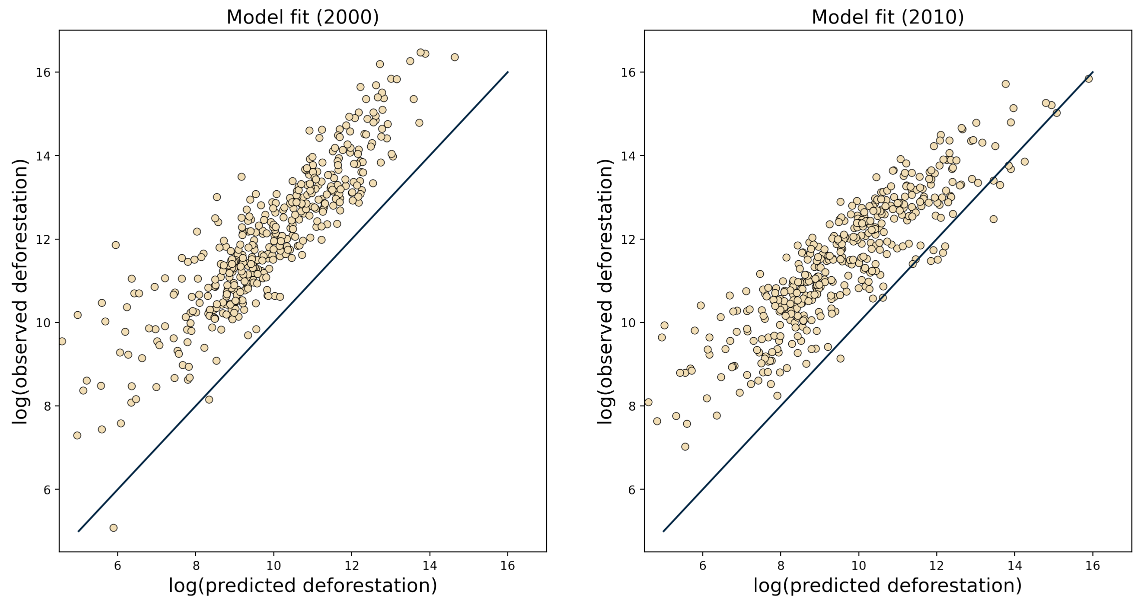
our preferred estimates are underestimating the impacts of deforestation.

**Price elasticity.** Our estimates can be used to compute the price elasticity of the land supply, an essential parameter for evaluating numerous public policies.

We begin by obtaining the price elasticity of frontier land. As shown in equation (7), the elasticity of deforestation to market access is a function of factor shares ($\alpha$ and $\gamma$), the trade elasticity ($\theta$), and the elasticity of the supply of frontier land ($1/\eta$). We calibrate the factor shares and the trade elasticity using common values from the literature to compute the elasticity of frontier land implied by our estimates. We assume that the share of land in production ($\alpha$) is 0.2 and the share of labor ($\gamma$) is 0.5 as in Valentinyi and Herrendorf (2008). Combining these numbers with the trade elasticity used to measure market access ($\theta = 8.2$), we find that the elasticity of frontier land implied by our empirical estimates is between 1.20-1.36. The elasticities of frontier land implied by the estimates obtained using other trade elasticities found in the literature are slightly larger (1.49 for $\theta = 6.5$ and 1.66 for $\theta = 4$). The values are close to the elasticities estimated and used in the literature (Costinot and Donaldson, 2016; Gouel and Laborde, 2018; Pellegrina and Sotelo, 2021).

The total land supply in the model is the sum of consolidated and frontier land. Moreover, the supply of consolidated land is fixed, implying that the elasticity of land supply is simply the elasticity of frontier land multiplied by its share in total land supply. Our data shows the share of frontier land in total land is about 0.33 throughout the decades used in our empirical exercise. Hence, the elasticity of land supply implied by our estimates is between 0.40-0.45 for our baseline $\theta$ and between 0.50-0.55 for our alternative $\theta$. These values are larger than the 0.17-0.26 elasticity for Brazil found by Roberts and Schlenker (2013). Empirically, this is consistent with the fact that the Amazon is the region of the country with more land to be incorporated and, therefore, a more elastic land supply. Methodologically, as discussed by Scott (2014), the static model estimated with annual data used by Roberts and Schlenker (2013) can underestimate the long-run land supply

Figure 5: Model fit




**Notes:** This figure depicts the relationship between the (log) predicted deforestation by the model and the (log) observed deforestation. Predicted deforestation is computed combining the change in market access in the decade and the estimated elasticity of market access on deforestation. Each dot represents a municipality. The dark line shows the 45° degree line. The left panel reports the relationship for the 2000 decade; the right panel reports the relationship for the 2010 decade. The $R^2$ of a regression of the (log) predicted deforestation on the (log) observed deforestation is 0.79 for both decades.

elasticity.

**Goodness of Fit.** Using the estimated elasticity of deforestation to market access, we evaluate the goodness of fit of our model. To do this, we compare the observed deforestation in the 2000s and 2010s with the predicted deforestation from the model, given the market access change between 1990-2000 and 2000-2010, respectively. [9] Figure 5 reports the results. While underestimating deforestation for most municipalities, our model matches the observed deforestation quite well. A regression of the (log of) observed deforestation on a constant and the (log of) predicted deforestation yields a coefficient estimate of 0.73 (0.01 standard error) for the first decade and 0.85 (0.02 standard error) for the second decade. Both regressions have an $R^2$ of 0.79.

[9] This approach implicitly assumes the changes in market access in the Amazon region do not influence the welfare of the workers in the country as a whole ($\bar{U}$ is fixed). The population of the Amazon is less than 15% of Brazil's population. Therefore, this does not seem a too stringent hypothesis.

### 4.3 Caveats and Extensions

**State roads.** Due to data availability, we do not include state roads in the computation of transportation costs. Nonetheless, these roads are usually built following the general structure designed by federal roads, implying that these roads do not change the structure of market access considerably. To provide some evidence on this, we explore data on state roads for 2010 (the only available year, see Figure C.1 for information on their spatial distribution) and compare measures of market access built excluding and including these roads. A regression of the log of market access with federal roads on a constant and the log of market access with federal and state roads yields an $R^2$ of 0.96. Given the strong correlation between these measures, it is unlikely that panel data on state roads would dramatically affect our empirical results.

**Two Sector Model.** One limitation of our theoretical model is that it ignores other sectors. As shown in Appendix B, it is possible to derive a model with two sectors (manufacturing and agriculture) that nests our one-sector model. The model retains tractability, delivering a log-linear expression connecting deforestation with measures of agricultural and non-agricultural market access (see equation (B.11)). However, as noted in Donaldson and Hornbeck (2016), agricultural and non-agricultural market access measures are typically strongly correlated and, thereby, hard to identify separately. Indeed, in our setting, a regression of the log of market access using rural population on the log of market access using urban population yields an $R^2$ between 0.86 and 0.90 depending on the decade. Given this correlation, one possible interpretation of our estimates is that they reflect the overall effect of increasing market access in all sectors.

**Correlated shocks.** Lind and Ramondo (2023) show the importance of correlated shocks in trade models like ours. As shown in Appendix B, it is possible to derive a model where the productivity shocks for each type of land – consolidated or frontier – are correlated. This model, that nests our main model, delivers a non-linear closed-form solution connecting deforestation to market access (see equation (B.12)). Unfortunately, it is impossible to identify the parameters from this expression using the variation in deforestation and market access of our empirical work. However, it is worth noting that equation (B.12) implies that the more correlated the productivity shocks, the higher elasticity of land supply and, therefore, the effects of investments in transportation infrastructure on deforestation. Thus, it is possible to interpret the effects obtained under the hypothesis of uncorrelated shocks across different types of land as a lower bound of the effects obtained under the hypothesis of correlated shocks across different types of land.

# 5 The Importance of General Equilibrium Effects

General equilibrium effects create a complex connection between the location of investments on transportation infrastructure and the location of its impacts. It means not only that regions distant from an investment might be affected by it but also that more distant regions might be more affected than closer regions. General equilibrium forces generate a violation of the Stable Unit Treatment Value Assumption (SUTVA) in empirical settings that use distance to an investment in transportation infrastructure to define treatment and control units. This implies that these designs will bias even the "local" effects of transportation infrastructure on deforestation.

To assess the importance of these general equilibrium effects when estimating local effects of transportation infrastructure, we leverage the structure of the model to simulate deforestation effects of randomly placed roads added to the 2010 transportation network. We then compare these model-implied effects with the local effects that would be estimated using a difference-in-differences strategy analogous to the one used by Asher et al. (2019).

We proceed as follows. First, we simulate a total of 1,000 roads (see Figure 6, panel A). Second, we compute the market access change generated by adding each of the simulated roads. Third, we use the elasticity of deforestation with respect to market access to calculate the counterfactual deforestation associated with each of the simulated roads.[10] Fourth, we compare the simulated effects with the effects that would be recovered using a difference-in-differences design that defines treatment and control municipalities based on their distance to the randomly placed road. We use the municipalities crossed by the road as the treatment group and their neighboring municipalities as the control group (see Figure 6, panel B).

We find this difference-in-differences underestimates the local effects of roads on deforestation. Figure 6, panel C reports the percentage of the true local effect of deforestation captured by the reduced form approach. On average, not accounting for general equilibrium effects would result in underestimating by one quarter the local effects of roads on deforestation. However, there are a significant share of simulations with much higher bias.

Figure 6, panel D depicts the correlation between this bias and the length of the randomly drawn roads. The correlation between these variables is quite small with the bias varying considerably across the distribution of road length. This finding shows that the length of roads is not a good proxy for the importance of general equilibrium effects.

Absent in Figure 6 is a small percentage of simulations $(< 1\%)$ where the bias is strong enough to flip the sign of the effect. This happens when the simulated effect on deforestation is lower in the municipalities crossed by the road than in their neighbors. This can occur because the effect of a road on deforestation is conditional on the rest of the entire transportation network. Indeed, depending on the access conditions of a proposed road, the neighbors can deforest more than the municipalities directly affected by the road's

[10] In our model, the population of different regions is also influenced by their market access. Thus, investments in transportation infrastructure directly affect a region's market access by changing its transportation costs and indirectly by changing its population. We ignore the effects on population when computing our counterfactuals. Conceptually, incorporating these effects would increase the effects of individual projects. However, access to distant population centers is the primary driver of market access in Brazil's Amazon. Thus, empirically, incorporating the effects on population is unlikely to influence our counterfactuals significantly.

outline. In this case, the reduced-form approach could mislead the researcher to conclude that the road has an effect of decreasing deforestation.

# 6 The Deforestation Effects of Individual Projects

Our framework can be used to evaluate the deforestation effects of projects currently under planning. This type of *ex-ante* evaluation is relevant for public policies for different reasons. First, it helps to determine the potential cost-benefit of different projects under analysis, improving project selection. Second, it helps to map the localities potentially affected by a project, guiding consultations with the local populations and the implementation of mitigation measures.[11].

As an example, we build a counterfactual scenario for the construction of the *Ferrogrão* railroad (Figure 7, panel A). *Ferrogrão*'s construction is meant to facilitate the logistics of producers from the state of *Mato Grosso*. In 2020, Mato Grosso was responsible for 15% of Brazil's agricultural output. Its producers export about 70% of their production using ports in the South and Southeast region of Brazil located more than 2,000 kilometers from the state. *Ferrogrão*'s construction will reduce transportation costs considerably by enabling these producers to export through ports in the North of Brazil.

To compute the effects of the *Ferrogrão* project, we modify our transportation network to include the proposed railroad and use our estimates to compute its effects on deforestation using a procedure identical to the one used to compute the effects of randomly drawn roads in the previous section. We find that the *Ferrogrão* construction is expected to increase total deforestation by 400 $km^2$ in the following decade.

We monetize this deforestation with parameters currently used to fund conservation projects in the Amazon (Fund, 2018). Specifically, we assume a forest carbon stock of 48,510 tCO2

[11] For an overview of the regulatory process of infrastructure building in Brazil, especially in the Amazon, see Antonaccio and Chiavari (2021); Cozendey and Chiavari (2021)

per $km^2$ and a carbon price of USD 5 per $tCO_2$. We find that this deforestation will generate an environmental cost value of US$ 97 million. This is a lower bound of the true environmental cost as it does not consider other environmental costs (e.g., eco-system services) and uses a carbon value that is far from recent estimates of the social cost of carbon, usually starting at USD 50 (EPA, 2016).

Figure 7, panel B shows the environmental cost is not concentrated in municipalities immediately along the railroad, being dispersed in municipalities throughout the mid north of the state of *Mato Grosso*. It also highlights the importance of location of the proposed stations in determining the geography of the project's impacts, emphasizing the perils of using the distance to the project's outline to determine potential impacts as is currently done in Brazil's regulation.

# 7 Conclusion

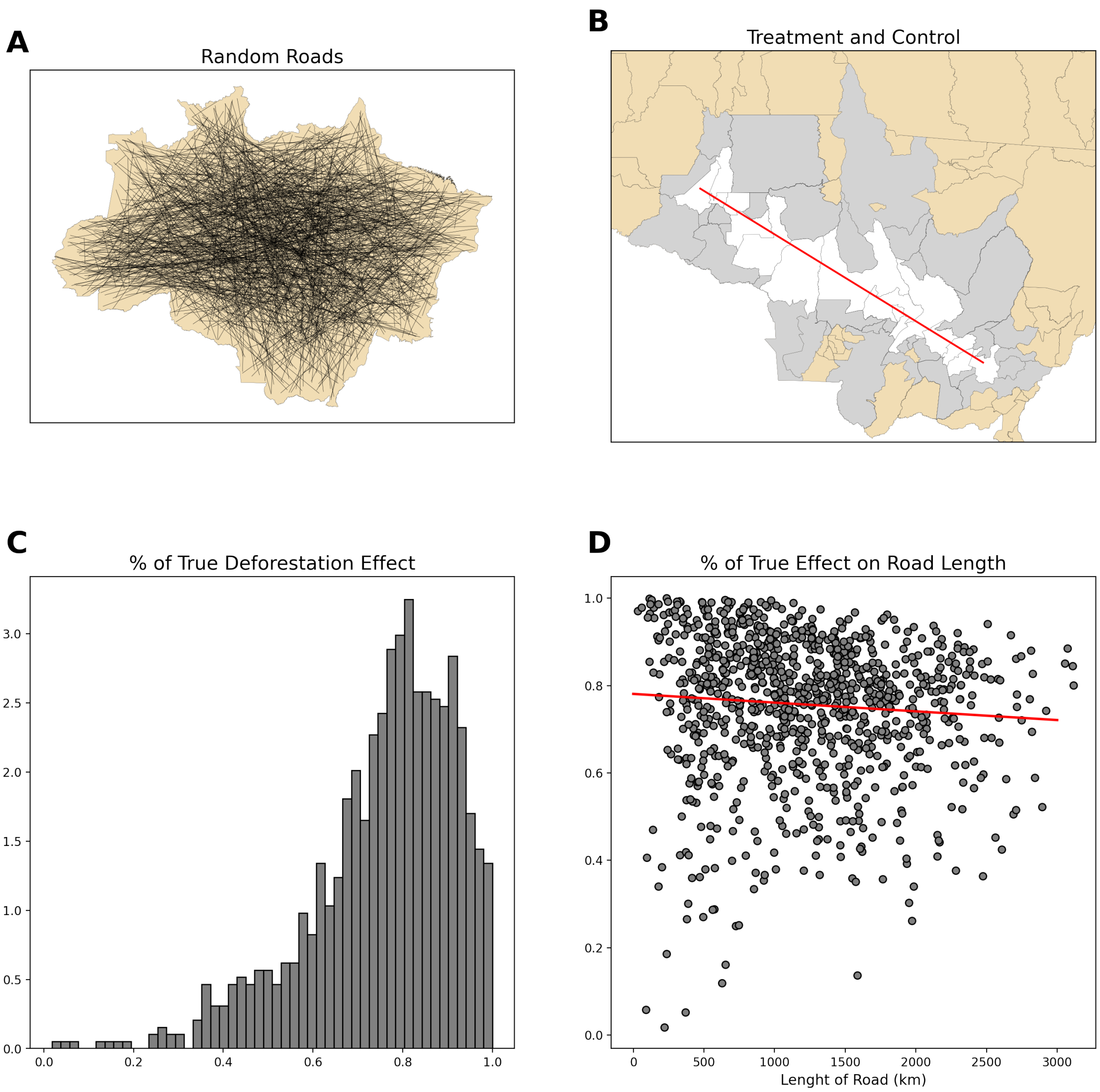
The development of transportation infrastructure is a pillar for economic development (Atkin and Donaldson, 2015; Costinot and Donaldson, 2016; Donaldson and Hornbeck, 2016; Jedwab and Storeygard, 2017; Fajgelbaum and Redding, 2022). Nonetheless, the efficient placement of infrastructure depends on an accurate assessment of its potential environmental costs (Damania et al., 2018; Bebbington et al., 2018; Asher et al., 2019). In this paper, we develop a framework to assess the deforestation cost of infrastructure projects in a general equilibrium setting. Specifically, we build and estimate an inter-regional trade model that connects deforestation and transportation costs through a properly defined metric of market access.

We obtain four main results. First, we estimate that a 1% increase in market access increases deforestation by roughly 0.5%. Second, we use this elasticity to predict deforestation within sample and find that our model explains deforestation remarkably well. Third, we simulate the construction of 1,000 random roads in the Amazon and find that ignoring general equilibrium effects would underestimate the effect of these roads by one-quarter. Fourth, we use our model and estimates to predict the impact of the *Ferrogrão* railroad – a highly controversial project planned to be built in the Amazon – and find that it will generate substantial environmental impacts, mostly in municipalities not crossed by the project.

Methodologically, we not only provide evidence of the importance of incorporating general equilibrium effects in evaluations of infrastructure investments, but also demonstrate the possibility of incorporating these effects without losing the tractability of regression-based approaches. It is worth noticing that the comprehensiveness and flexibility of our transportation network allow for studying a wide range of counterfactuals. Our framework can be used to study the effects of investments in transportation infrastructure as done in the paper, the effects of regulations (e.g., price controls or taxes on specific types of transportation modes), and the effects of inefficiencies (e.g., heterogeneity in times to process trans-shipment in different ports). Thus, our work provides a useful tool to improve transportation policies.

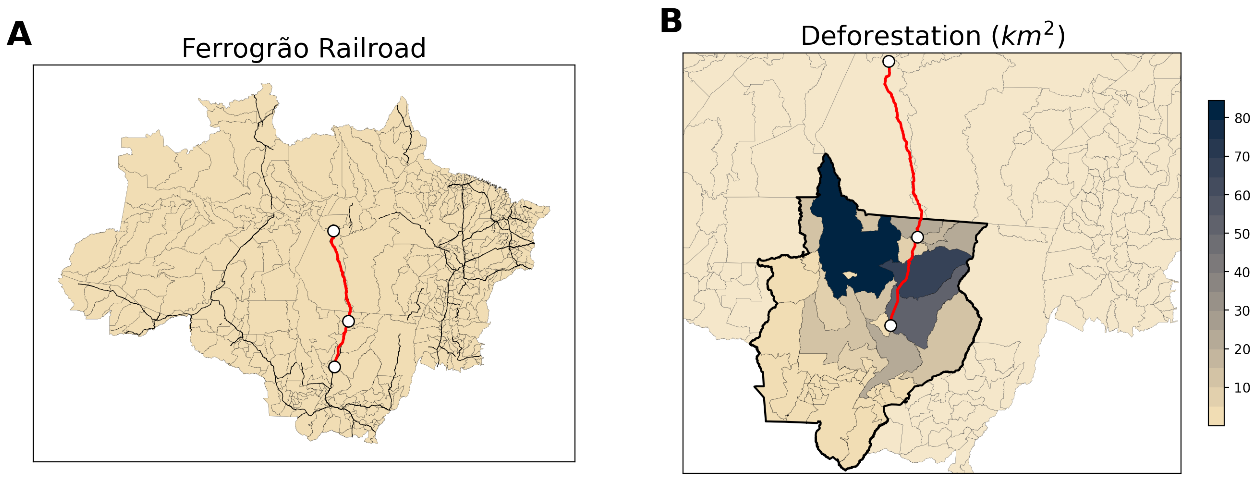
Empirically, our results document the importance of improvements in transportation infrastructure in explaining the dynamics of deforestation in the Amazon throughout the last three decades. This contributes to the growing literature documenting the drivers of deforestation in the tropics (Assunção et al., 2015, 2020; Assunção et al., 2022; Araujo et al., 2020; Burgess et al., 2019; Bragança and Dahis, 2022; Assunção et al., Forthcoming). Future work evaluating how to mitigate the negative impacts of transportation infrastructure on deforestation is fundamental to enable the Amazon to reduce its isolation without generating irreversible environmental losses.

Figure 6: General Equilibrium and Local Effects




**Notes:** Panel A shows the 1,000 random roads generated; Panel B provides an example of the reduced form framework used to estimate the local effects of each road. The road is shown i red, the treatment group (municipalities crossed by the road) in white, and the control group (neighbors of the municipalities crossed by the road) in gray; Panel C depicts the distribution of the share of the true deforestation effect captured by the reduced form framework; Panel D reports the correlation between this share and road length.

Figure 7: Ferrogrão and Deforestation




**Notes:** Panel A depicts the location of the Ferrogrão railroad project (in red), its three stations (in white), and roads as of the year 2010 (in black). Panel B depicts the deforestation impact of the project. The region delimited by the black polygon is the region that will have its market access affected by the construction of the railroad.

# References


**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.

**Antonaccio, Luiza and Joana Chiavari**, "Strengthening Environmental Studies for Federal Land Infrastructure Concessions," *Climate Policy Initiative*, 2021.

**Araujo, Rafael, Francisco Costa, and Marcelo Sant'Anna**, "Efficient forestation in the Brazilian Amazon: Evidence from a dynamic model," *Working paper*, 2020.

**Asher, Sam, Teevrat Garg, and Paul Novosad**, "The ecological footprint of transportation infrastructure," *Economic Journal*, 2019.

**Assunção, Juliano, Clarissa Gandour, and Romero Rocha**, "DETERring deforestation in the Brazilian Amazon: environmental monitoring and law enforcement," *Americal Economic Journal: Applied Economics*, Forthcoming.

**_ , _ , and Rudi Rocha**, "Deforestation slowdown in the Brazilian Amazon: prices or policies?," *Environment and Development Economics*, 2015, *20* (6), 697–722.

**_ , _ , Romero Rocha, and Rudi Rocha**, "The effect of rural credit on deforestation: evidence from the Brazilian Amazon," *The Economic Journal*, 2020, *130* (626), 290–330.

**_ , Robert McMillan, Joshua Murphy, and Eduardo Souza-Rodrigues**, "Optimal environmental targeting in the amazon rainforest," Technical Report, National Bureau of Economic Research 2019.

**Assunção, Juliano, Robert McMillan, Joshua Murphy, and Eduardo Souza-Rodrigues**, "Optimal Environmental Targeting in the Amazon Rainforest," *The Review of Economic Studies*, 10 2022. rdac064.

**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.

**Baccini, AGSJ, SJ Goetz, WS Walker, NT Laporte, Mindy Sun, Damien Sulla-Menashe, Joe Hackler, PSA Beck, Ralph Dubayah, MA Friedl et al.**, "Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps," *Nature climate change*, 2012, *2* (3), 182–185.

**Baum-Snow, Nathaniel, J Vernon Henderson, Matthew A Turner, Qinghua Zhang, and Loren Brandt**, *Highways, market access and urban growth in China*, SERC, Spatial Economics Research Centre, 2016.

**_ , _ , _ , _ , and _ ,** "Does investment in national highways help or hurt hinterland city growth?," *Journal of Urban Economics*, 2020, *115*, 103124.

**Bebbington, Anthony J, Denise Humphreys Bebbington, Laura Aileen Sauls, John Rogan, Sumali Agrawal, César Gamboa, Aviva Imhof, Kimberly Johnson, Herman Rosa, Antoinette Royo et al.**, "Resource extraction and infrastructure threaten forest cover and community rights," *Proceedings of the National Academy of Sciences*, 2018, *115* (52), 13164–13173.

**Bragança, Arthur and Ricardo Dahis**, "Cutting special interests by the roots: Evidence from the Brazilian Amazon," *Journal of Public Economics*, 2022, *215*, 104753.

**Burgess, Robin, Francisco Costa, and Benjamin A Olken**, "The Brazilian Amazon's double reversal of fortune," *Working paper*, 2019.

**_ , Matthew Hansen, Benjamin A Olken, Peter Potapov, and Stefanie Sieber**, "The political economy of deforestation in the tropics," *The Quarterly journal of economics*, 2012, *127* (4), 1707–1754.

**Caliendo, Lorenzo and Fernando Parro**, "Estimates of the Trade and Welfare Effects of NAFTA," *The Review of Economic Studies*, 2015, *82* (1), 1–44.

**Chomitz, K. M. and D. A. Gray**, "Roads, Land Use, and Deforestation: A Spatial Model Applied to Belize," *The World Bank Economic Review*, 09 1996, *10.*

**Costinot, Arnaud and Dave Donaldson**, "How large are the gains from economic integration? theory and evidence from us agriculture, 1880-1997," Technical Report, National Bureau of Economic Research 2016.

**_ , _ , and Ivana Komunjer**, "What goods do countries trade? A quantitative exploration of Ricardo's ideas," *The Review of economic studies*, 2012, *79* (2), 581–608.

**Cozendey, Gabriel and Joana Chiavari**, "Environmental Viability of Land Transport Infrastructure in the Amazon," *Climate Policy Initiative*, 2021.

**Cropper, Maureen, Charles Griffiths, and Muthukumara Mani**, "Roads, population pressures, and deforestation in Thailand, 1976-1989," *Land Economics*, 1999, pp. 58–73.

**Damania, Richard, Jason Russ, David Wheeler, and Alvaro Federico Barra**, "The road to growth: Measuring the tradeoffs between economic growth and ecological destruction," *World Development*, 2018, *101*, 351–376.

**Dijkstra, Edsger W**, "A note on two problems in connexion with graphs," *Numerische mathematik*, 1959, *1* (1), 269–271.

**Donaldson, Dave and Richard Hornbeck**, " Railroads and American Economic Growth: A "Market Access" Approach," *The Quarterly Journal of Economics*, 02 2016, *131* (2), 799–858.

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

**Ehrl, Philipp**, "Minimum comparable areas for the period 1872-2010: an aggregation of Brazilian municipalities," *Estudos Econômicos (São Paulo)*, 2017, *47* (1), 215–229.

**EPA**, "Social Cost of Carbon," *Environmental Protection Agency (EPA): Washington, DC, USA*, 2016.

**ESALQ-LOG**, "SIFRECA yearbooks," *Piracicaba, Brazil*, 2008-2014.

**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.

**Fetzer, Thiemo and Samuel Marden**, "Take what you can: property rights, contestability and conflict," *The Economic Journal*, 2017, *127* (601), 757–783.

**Foster, Andrew D and Mark R Rosenzweig**, "Economic growth and the rise of forests," *The Quarterly Journal of Economics*, 2003, *118* (2), 601–637.

**Fund, Amazon**, "Amazon Fund: Technical Note 293/2018," *http://www.fundoamazonia.gov.br/en*, 2018.

**Garg, Teevrat and Ajay Shenoy**, "The Ecological Impact of Place-Based Economic Policies," *American Journal of Agricultural Economics*, 2021, *103* (4), 1239–1250.

**Gatti, Luciana V, Luana S Basso, John B Miller, Manuel Gloor, Lucas Gatti Domingues, Henrique LG Cassol, Graciela Tejada, Luiz EOC Aragão, Carlos Nobre, Wouter Peters et al.**, "Amazonia as a carbon source linked to deforestation and climate change," *Nature*, 2021, *595* (7867), 388–393.

**Gibson, Xiangzheng Deng; Jikun Huang; Emi Uchida; Scott Rozelle; John**, "Pressure cookers or pressure valves: Do roads lead to deforestation in China?," *Journal of Environmental Economics and Management*, 2011, *61.*

**Gouel, Christophe and David Laborde**, "The crucial role of international trade in adaptation to climate change," Technical Report, National Bureau of Economic Research 2018.

**Hansen, Matthew C, Peter V Potapov, Rebecca Moore, Matt Hancher, Svetlana A Turubanova, Alexandra Tyukavina, David Thau, Stephen V Stehman, Scott J Goetz, Thomas R Loveland et al.**, "High-resolution global maps of 21st-century forest cover change," *science*, 2013, *342* (6160), 850–853.

**Head, Keith and Thierry Mayer**, "Gravity Equations: Workhorse, Toolkit, Cookbook," *Handbook of International Economics, Vol. 4*, 2014.

**Heilmayr, Robert, Lisa L Rausch, Jacob Munger, and Holly K Gibbs**, "Brazil's Amazon soy moratorium reduced deforestation," *Nature Food*, 2020, *1* (12), 801–810.

**Hsiao, Allan**, "Coordination and commitment in international climate action: evidence from palm oil," *Job market paper*, 2021.

**IPCC**, "The Fourth Assessment Report of the Intergovernmental Panel on Climate-Change," *Geneva, Switzerland*, 2017.

**Jedwab, Remi and Adam Storeygard**, "The Average and Heterogeneous Effects of Transportation Investments: Evidence from sub-Saharan Africa 1960-2010," Technical Report, Tufts University 2017.

**Jedwab, Rémi and Adam Storeygard**, "The Average and Heterogeneous Effects of Transportation Investments: Evidence from Sub-Saharan Africa 1960–2010," *Journal of the European Economic Association*, 06 2021.

**Laurance, William F, Mark A Cochrane, Scott Bergen, Philip M Fearnside, Patricia Delamônica, Christopher Barber, Sammya D'angelo, and Tito Fernandes**, "The future of the Brazilian Amazon," *Science*, 2001, *291* (5503), 438–439.

**Lawrence, Deborah and Karen Vandecar**, "Effects of tropical deforestation on climate and agriculture," *Nature climate change*, 2015, *5* (1), 27–36.

**Lind, Nelson and Natalia Ramondo**, "Trade with correlation," *American Economic Review*, 2023, *113* (2), 317–353.

**Mapbiomas**, "Mapbiomas project. Collection 4.0," *http://www.mapbiomas.org accessed: 01.10.2018*, 2019.

**Pellegrina, Heitor S and Sebastian Sotelo**, "Migration, Specialization, and Trade: Evidence from Brazil's March to the West," Technical Report, National Bureau of Economic Research 2021.

**Pfaff, Alexander, Juan Robalino, Robert Walker, Steven Aldrich, Marcellus Caldas, Eustaquio Reis, Stephen Perz, Claudio Bohrer, Eugenio Arima, William Laurance et al.**, "Road investments, spatial spillovers, and deforestation in the Brazilian Amazon," *Journal of Regional Science*, 2007, *47* (1), 109–123.

**Pfaff, Alexander S.P.**, "What Drives Deforestation in the Brazilian Amazon?: Evidence from Satellite and Socioeconomic Data," *Journal of Environmental Economics and Management*, 1999, *37.*

**Prem, Mounu, Santiago Saavedra, and Juan F Vargas**, "End-of-conflict deforestation: Evidence from Colombia's peace agreement," *World Development*, 2020, *129*, 104852.

**Redding, Stephen and Anthony J Venables**, "Economic geography and international inequality," *Journal of International Economics*, 2004, *62* (1), 53–82.

**Reis, Diana Weinhold; Eustaquio**, "Transportation costs and the spatial distribution of land use in the Brazilian Amazon," *Global Environmental Change*, 2008, *18.*

**Roberts, Michael J and Wolfram Schlenker**, "Identifying supply and demand elasticities of agricultural commodities: Implications for the US ethanol mandate," *American*

*Economic Review*, 2013, *103* (6), 2265–2295.

**Scott, Paul**, "Dynamic discrete choice estimation of agricultural land use," 2014.

**Simonovska, Ina and Michael E Waugh**, "The elasticity of trade: Estimates and evidence," *Journal of international Economics*, 2014, *92* (1), 34–50.

**Solon, Gary, Steven J Haider, and Jeffrey M Wooldridge**, "What are we weighting for?," *Journal of Human resources*, 2015, *50* (2), 301–316.

**Souza-Rodrigues, Eduardo**, "Deforestation in the Amazon: A Unified Framework for Estimation and Policy Analysis," *The Review of Economic Studies*, 12 2018.

**Valentinyi, Akos and Berthold Herrendorf**, "Measuring factor income shares at the sectoral level," *Review of Economic Dynamics*, 2008, *11* (4), 820–835.

**Vilela, Thais, Alfonso Malky Harb, Aaron Bruner, Vera Laísa da Silva Arruda, Vivian Ribeiro, Ane Auxiliadora Costa Alencar, Annie Julissa Escobedo Grandez, Adriana Rojas, Alejandra Laina, and Rodrigo Botero**, "A better Amazon road network for people and the environment," *Proceedings of the National Academy of Sciences*, 2020, *117* (13), 7095–7102.

**Villoria, Nelson, Rachael Garrett, Florian Gollnow, and Kimberly Carlson**, "Leakage does not fully offset soy supply-chain efforts to reduce deforestation in Brazil," *Nature Communications*, 2022, *13* (1), 5476.

# Appendix to “The Effects of Transportation Infrastructure on Deforestation in the Amazon”

## Table of Contents

# A Proofs of the Theoretical Model

**Lemma 1**. *The probability that a farmer will choose consolidated land is given by:*

$$\bar{p}\left(\frac{q_o^F}{q_o^C}\right) = P\left(\frac{z_o^F(j)}{z_o^C(j)} < \left(\frac{q_o^F}{q_o^C}\right)^{\alpha}\right) = \frac{1}{1+\frac{A_o^F}{A_o^C}\left(\frac{q_o^F}{q_o^C}\right)^{-\theta\alpha}}$$

*Proof.*

$$\begin{aligned}\bar{p}\left(\frac{q_o^F}{q_o^C}\right) &= P\left(\frac{z_o^F(j)}{z_o^C(j)} < \left(\frac{q_o^F}{q_o^C}\right)^{\alpha}\right) \\ &= \int_0^{\infty}\int_0^{\left(\frac{q_o^F}{q_o^C}\right)^{\alpha} z_o^C(j)} \frac{\partial^2 F_o}{\partial z_o^F(j)\partial z_o^C(j)} dz_o^F(j)dz_o^C(j) \\ &= \frac{1}{1+\frac{A_o^F}{A_o^C}\left(\frac{q_o^F}{q_o^C}\right)^{-\theta\alpha}}\end{aligned} \tag{A.1}$$

■

**Lemma 2**. *Total income accrued to frontier land equals total income accrued to consolidated land adjusted by the relative probability producers operate in each type of land. Thus,*

$$\bar{p}_o q_o^F L_o^F = (1-\bar{p}_o) q_o^C L_o^C$$

*Proof.* Offered prices from consolidated land and frontier land follows the same distribution. The only difference on income from both types of land comes from differences of the length of varieties that is sold. As shown in A.4, the ratio of the length of varieties is given by $\frac{\pi_{o,d}^C}{\pi_{o,d}^F} = \frac{\bar{p}}{1-\bar{p}}$ Therefore,

$$\bar{p}_o \alpha q_o^F L_o^F = (1-\bar{p}_o)\alpha q_o^C L_o^C \tag{A.2}$$

■

**Lemma A.1.** *Offered price distribution from region $o \in R$ to region $d \in O$ is a univariate Frechet* *distribution.*

*Proof.*

$$
\begin{aligned}
G_{o,d}(p) &= P(p_{o,d}(j) < p) \\
&= P\left(\min\left\{\tau_{od}\frac{q_o^{C^{\alpha}} w_o{}^{\gamma} r^{1-\alpha-\gamma}}{z_o^C(j)}, \tau_{od}\frac{q_o^{F^{\alpha}} w_o{}^{\gamma} r^{1-\alpha-\gamma}}{z_o^F(j)}\right\} < p\right) \\
&= 1 - P\left(\tau_{od}\frac{q_o^{C^{\alpha}} w_o{}^{\gamma} r^{1-\alpha-\gamma}}{z_o^C(j)} > p, \tau_{od}\frac{q_o^{F^{\alpha}} w_o{}^{\gamma} r^{1-\alpha-\gamma}}{z_o^F(j)} > p\right) \\
&= 1 - P\left(z_o^C(j) < \tau_{od}\frac{q_o^{C^{\alpha}} w_o{}^{\gamma} r^{1-\alpha-\gamma}}{p}, z_o^F(j) < \tau_{od}\frac{q_o^{F^{\alpha}} w_o{}^{\gamma} r^{1-\alpha-\gamma}}{p}\right) \\
&= 1 - \exp\left(-(\tau_{od} w_o^{\gamma} r^{1-\alpha-\gamma})^{-\theta}\left(A_o^C (q_o^C)^{-\theta\alpha} + A_o^F (q_o^F)^{-\theta\alpha}\right) p^{\theta}\right)
\end{aligned}
\tag{A.3}
$$

■

**Lemma A.2.** *The price distribution for what region $d \in O$ actually buys inherits the form of the distribution of offered prices.*

*Proof.*

$$
\begin{aligned}
G_d(p) &= P(p_d(j) < p) \\
&= P\left(\min_{o\in R}\left\{\min\left\{\tau_{od}\frac{q_o^{C^{\alpha}} w_o{}^{\gamma} r^{1-\alpha-\gamma}}{z_o^C(j)}, \tau_{od}\frac{q_o^{F^{\alpha}} w_o{}^{\gamma} r^{1-\alpha-\gamma}}{z_o^F(j)}\right\}\right\} < p\right) \\
&= 1 - \prod_{o\in O} P\left(\min\left\{\tau_{od}\frac{q_o^{C^{\alpha}} w_o{}^{\gamma} r^{1-\alpha-\gamma}}{z_o^C(j)}, \tau_{od}\frac{q_o^{F^{\alpha}} w_o{}^{\gamma} r^{1-\alpha-\gamma}}{z_o^F(j)}\right\} > p\right) \\
&= 1 - \prod_{o\in O} P\left(\tau_{od}\frac{q_o^{C^{\alpha}} w_o{}^{\gamma} r^{1-\alpha-\gamma}}{z_o^C(j)} > p, \tau_{od}\frac{q_o^{F^{\alpha}} w_o{}^{\gamma} r^{1-\alpha-\gamma}}{z_o^F(j)} > p\right) \\
&= 1 - \prod_{o\in O} \exp\left(-\left(A_o^C\left(\tau_{od}\frac{q_o^{C^{\alpha}} w_o{}^{\gamma} r^{1-\alpha-\gamma}}{p}\right)^{-\theta} + A_o^F\left(\tau_{od}\frac{q_o^{F^{\alpha}} w_o{}^{\gamma} r^{1-\alpha-\gamma}}{p}\right)^{-\theta}\right)\right) \\
&= 1 - \prod_{o\in O} \exp\left(-(\tau_{od} w_o^{\gamma} r^{1-\alpha-\gamma})^{-\theta}\left(A_o^C (q_o^C)^{-\theta\alpha} + A_o^F (q_o^F)^{-\theta\alpha}\right) p^{\theta}\right) \\
&= 1 - \exp\left(\sum_{o\in O} -(\tau_{od} w_o^{\gamma} r^{1-\alpha-\gamma})^{-\theta}\left(A_o^C (q_o^C)^{-\theta\alpha} + A_o^F (q_o^F)^{-\theta\alpha}\right) p^{\theta}\right)
\end{aligned}
\tag{A.4}
$$

■

**Lemma A.3.** *The price distribution that region $o \in O$ offers region $d \in O$ conditional on being produced in consolidated land is the same distribution as unconditional offered prices.*

*Proof.* To facilitate visualization define for now $c = \left(\frac{q_o^F}{q_o^C}\right)^\alpha$ and $s = \frac{\tau_{od} {q_o^C}^\alpha w_o^\gamma r^{1-\alpha-\gamma}}{p}$

$$
\begin{aligned}
\bar{G}_{o,d}^C(p) &= P\left(p_{o,d}(j) < p | \frac{z_o^F(j)}{z_o^C(j)} < \left(\frac{q_o^F}{q_o^C}\right)^\alpha\right) \\
&= P\left(z_o^C > \frac{\tau_{od} {q_o^C}^\alpha w_o^\gamma r^{1-\alpha-\gamma}}{p} | \frac{z_o^F(j)}{z_o^C(j)} < \left(\frac{q_o^F}{q_o^C}\right)^\alpha\right) \\
&= 1 - \frac{1}{\bar{p}_o} \int_0^s \int_0^{\left(\frac{q_o^F}{q_o^C}\right)^\alpha z_o^C} \frac{\partial^2 F_o}{\partial z_o^F \partial z_o^C} dz_o^F dz_o^C \\
&= 1 - \frac{1}{\bar{p}_o} \int_0^s \left[\frac{\partial F_o}{\partial z_o^C}\left(\left(\frac{q_o^F}{q_o^C}\right)^\alpha z_o^C, z_o^C\right) - \lim_{t \to 0} \frac{\partial F_o}{\partial z_o^C}\left(t, z_o^C\right)\right] dz_o^C \\
&= 1 - \frac{1}{\bar{p}_o} \int_0^s \left[A_o^C \theta \left(A_o^C + A_o^F c^{-\theta}\right)\right] {z_o^C}^{-\theta-1} \exp\left[-(A_o^C + A_o^F c^{-\theta}) {z_o^C}^{-\theta}\right] dz_o^C \\
&= 1 - \frac{1}{\bar{p}_o} \left[\frac{A_o^C}{A_o^C + A_o^F c^{-\theta}} \exp\left[-(A_o^C + A_o^F c^{-\theta}) s^{-\theta}\right]\right] = \\
&= 1 - \exp\left[-(A_o^C + A_o^F c^{-\theta}) s^{-\theta}\right] \\
&= 1 - \exp\left[-(\tau_{od} w_o^\alpha r^{1-\alpha-\gamma})^{-\theta} (A_o^C (q_o^C)^{-\theta\alpha} + A_o^F (q_o^F)^{-\theta\alpha}) p^\theta\right]
\end{aligned}
\tag{A.5}
$$

■

**Lemma A.4.** *Exports and prices.*

*Proof.* Derivation of the exports from region $o$ to region $d$.

The length of varieties (or proportion) that region $o \in O$ exports to $d \in O$ is given by

$$
\begin{aligned}
\pi_{o,d} &= P(p_{od}(j) < \min\{p_{s,d}(j) : s \neq o\}) \\
&= \int_0^\infty \prod_{s \neq o} [1 - G_{s,d}(p)] \, dG_{o,d}(p) = \\
&= \frac{\phi_{od}}{\Phi_d},
\end{aligned}
\tag{A.6}
$$

in which

$$\phi_{od} = (\tau_{od} w_o^{\gamma} r^{1-\alpha-\gamma})^{-\theta} \left( A_o^C (q_o^C)^{-\theta\alpha} + A_o^F (q_o^F)^{-\theta\alpha} \right)$$

$$\Phi_d = \sum_{s \neq o} \left( (\tau_{sd} w_s^{\gamma} r^{1-\alpha-\gamma})^{-\theta} \left( A_s^C (q_s^C)^{-\theta\alpha} + A_s^F (q_s^F)^{-\theta\alpha} \right) \right)$$

As $G_{o,d}(p)$ differs from $G_{o,d}^c$ and $G_{o,d}^f$ only by a constant factor, conditioning on consolidated or frontier land will result in the same above integral above, up to a multiplicative constant. Therefore

$$\pi_{o,d}^C = \bar{p} \frac{\phi_{od}}{\Phi_d}$$

$$\pi_{o,d}^F = (1 - \bar{p}) \frac{\phi_{od}}{\Phi_d}$$

■

# B Correlated Shocks and a Manufacturing Sector

In this section we build an inter-regional trade model with two sectors - agricultural and manufacturing – and correlated productivity shocks. The proofs are built base on Donaldson and Hornbeck (2016),Lind and Ramondo (2023), and Eaton and Kortum (2002).

## B.1 Environment

Our economy is composed of a set $O = \{U\} \cup \{R\}$ of regions which we understand as being either rural ($o \in R$) or urban ($o \in U$).

The agents in region $o \in O$ supply ineslastically one unit of labor, earn wage $w^o$, and allocate consumption through a CES utility function over a continuum of varieties from goods produced by the agricultural sector located in rural regions - denoted by $a(j)$ with $j \in [0, A]$ - and a continuum of varieties produced by the manufacturing sector located in urban regions - denoted by $m(j)$ with $j \in [0, M]$.

Each pair of origin-destination regions can trade with each other the goods produced by the two sectors. We will denote a origin region by the letter $o \in O$ and a destination region by $d \in O$. An agent living in municipality $o$ solves the following maximization problem

$$\max_{\{a_j\},\{m_j\}} \left[\int a(j)^{\frac{\sigma-1}{\sigma}} dj\right]^{\mu\frac{\sigma}{\sigma-1}} \left[\int m(j)^{\frac{\sigma_m-1}{\sigma_m}} dj\right]^{(1-\mu)\frac{\sigma_m}{\sigma_m-1}} \tag{B.1}$$

subject to

$$\int p_o(j)a(j)dj + \int p_o^m(j)m(j)dj = w^o \tag{B.2}$$

Where $p_o(j)$ denotes the price of agricultural good $j$ on municipality $o$, as does $p_o^m(j)$ for the manufacturing good $j$. Thus, the indirect utility of an agent living in $o \in O$ is given by

$$V^o = \frac{w^o}{(P_o)^{\mu}(P_o^m)^{1-\mu}} \tag{B.3}$$

Where $(P_o)^{1-\sigma} = \int_0^A p_o(j)^{1-\sigma}dj$ and $(P_o^m)^{1-\sigma_m} = \int_0^M p_o^m(j)^{1-\sigma_m}dj$ are the perfect price indexes.

We assume that the productivity shocks of the two types of land in the agricultural sector $(z_o^T(j))$ are drawn from a bivariate Fréchet distribution with CDF given by $F_o(z^C, z^F) = \exp(-(A_o^C z^{C^{-g\theta}} + A_o^F z^{F^{-g\theta}})^{\frac{1}{g}})$. Here, $g$ measures the degree of dependence between the two shocks.

In urban regions, the marginal cost of producing one unit of good $m(j)$ is

$$MC_o(j) = \frac{{q_o}^{\alpha_m}{w_o}^{\gamma_m}r^{1-\alpha_m-\gamma_m}}{z_o^m(j)} \tag{B.4}$$

Where $z_o(j)$ denotes the productivity shock specific for the manufacturing variety produced in region $o \in U$. In the manufacturing sector the productivity shock is drawn from an univariate Frechet $F_o(z) = \exp(-M_o z^{-\theta_m})$.

## B.2 Prices and trade flows

### Manufacturing

The manufacturing sector follows the same derivations for the one sector model in Donaldson and Hornbeck (2016).

The price index of manufactured goods at region $d \in O$ is given by[1]

$$(P_d^m)^{-\theta_m} = x_m \sum_{o \in U} M_o(\tau_{o,d}^m {q_o}^{\alpha_m} {w_o}^{\gamma_m})^{-\theta_m} \equiv CMA_d^m \tag{B.5}$$

[1]Here $x_m = \left[\Gamma\left(\frac{\theta_m+1-\sigma_m}{\theta_m}\right)\right]^{\frac{-\theta_m}{1-\sigma_m}} r^{-\theta_m(1-\alpha_m-\gamma_m)}$

Trade flow from $o \in U$ to $d \in O$

$$X^m_{od} = x_m M_o (\tau^m_{o,d} {q_o}^{\alpha_m} {w_o}^{\gamma_m})^{-\theta_m} (CMA^m_d)^{-1} X^m_d \tag{B.6}$$

And the condition of equilibrium in the manufacturing sector is

$$Y^m_o = \sum_d X^m_{od} = x_m M_o ({q_o}^{\alpha_m} {w_o}^{\gamma_m})^{-\theta_m} \sum_d {\tau^m_{o,d}}^{-\theta_m} (CMA^m_d)^{-1} X^m_d \tag{B.7}$$

**Agriculture**

For the agricultural sector he have the price index of agricultural goods at region $d \in O$ is given by

$$(P_d)^{-\theta} = x \sum_{o \in R} (\tau_{od} {w_o}^{\gamma})^{-\theta} \left( A^C_o (q^C_o)^{-g\theta\alpha} + A^F_o (q^F_o)^{-g\theta\alpha} \right)^{\frac{1}{g}} \equiv CMA_d \tag{B.8}$$

Trade flow from $o \in R$ to $d \in O$

$$X_{od} = x \left( \tau_{od} w^{\gamma}_o \right)^{-\theta} \left( A^C_o (q^C_o)^{-g\theta\alpha} + A^F_o (q^F_o)^{-g\theta\alpha} \right)^{\frac{1}{g}} (CMA_d)^{-1} X_d \tag{B.9}$$

The equilibrium in agricultural markets is given by

$$Y_o = \sum_d X_{od} = x \left( w^{\gamma}_o \right)^{-\theta} \left( A^C_o (q^C_o)^{-g\theta\alpha} + A^F_o (q^F_o)^{-g\theta\alpha} \right)^{\frac{1}{g}} \sum_d {\tau_{od}}^{-\theta} (CMA_d)^{-1} X_d \tag{B.10}$$

## B.3 Equilibrium

Making the same substitutions as in the main model specification, we arrive at the final equation connecting market access (rural and urban) with land use.

$$
(\eta + 1 + \eta\theta\alpha)\log L_o^F + \left(\frac{g-1}{g}\right)\log\left[\left[\frac{A_o^C L_o^F}{A_o^F \bar{L}_o^C}\right]^{\frac{1}{1+g\theta\alpha}} \frac{\bar{L}_o^C}{L_o^F} + 1\right] =
$$

$$
\log\frac{x\alpha A_o^{F\frac{1}{g}}}{\rho^{\mu\gamma}\rho_m^{\frac{(1-\mu)\theta\gamma}{\theta_m}}\bar{U}^{\theta\gamma}B_o} + (1+\mu\gamma)\log MA_o + \frac{(1-\mu)\theta\gamma}{\theta_m}\log MA_o^m
$$

Notice that a model without a manufacturing sector is nested within the one presented above, by setting $\mu = 1$ we eliminate this sector. Notice also that a model with independent productivity shocks of the agricultural sector is nested, we just need to set $g = 1$.

Therefore a model with independent shocks and a manufacturing sector would yield:

$$
\begin{aligned}
(\eta + 1 + \eta\theta\alpha)\log L_o^F = \log\frac{x\alpha A_o^F}{\rho^{\mu\gamma}\rho_m^{\frac{(1-\mu)\theta\gamma}{\theta_m}}\bar{U}^{\theta\gamma}B_o} \\
+ (1+\mu\gamma)\log MA_o + \frac{(1-\mu)\theta\gamma}{\theta_m}\log MA_o^m
\end{aligned}
\tag{B.11}
$$

And a model without a manufacturing sector but with correlated shocks:

$$
\begin{aligned}
(\eta + 1 + \eta\theta\alpha)\log L_o^F + \left(\frac{g-1}{g}\right)\log\left[\left[\frac{A_o^C L_o^F}{A_o^F \bar{L}_o^C}\right]^{\frac{1}{1+g\theta\alpha}} \frac{\bar{L}_o^C}{L_o^F} + 1\right] = \\
\log\frac{x\alpha A_o^{F\frac{1}{g}}}{\rho^{\gamma}\bar{U}^{\theta\gamma}B_o} + (1+\gamma)\log MA_o
\end{aligned}
\tag{B.12}
$$

# C Additional Results

Table C.1: Cost parameters in the graph structure

| | |
|---|---|
| Paved road | 10 |
| Paved road in the Amazon | 20 |
| | |
| Unpaved road | 20 |
| Unpaved road in the Amazon | 40 |
| | |
| Railroad | 5 |
| Waterway | 5 |
| Transshipment cost | 200 |
| | |
| Land without road | 50 |
| Land without road in the Amazon | 100 |
| | |
| Protected area without road | 100 |
| Protected area without road in the Amazon | 200 |

**Notes:** This table shows the value used in the transportation network graph structure. The value correspond to the cost of traversing a node of a specific type of transportation infrastructure. The transshipment cost is paid for agents to access railroads and waterways. The important aspect for the optimal path algorithm is the proportion among the values and not their magnitude. For example, multiplying all values by 10 would yield the same optimal paths, with a total cost 10 times higher.

Table C.2: Convert graph to iceberg cost

| | Dep Var. is $cost_{odt}$ |
|---|---|
| $1000 \times cost\_graph_{odt}$ | 0.002*** |
| | (0.0001) |
| const. | 0.0127*** |
| | (0.0032) |
| Obs | 1,200 |
| R2 | 0.63 |

**Notes:** This table reports the results of regressing iceberg costs ($cost_{odt}$) on raster costs ($cost\_graph_{odt}$) as explained in the main text. The graph costs are quite high in levels because we store the graphs with integers in order to save storage space when computing the optimal paths. Therefore, we multiply coefficients by 1000 to facilitate visualization. p<0.01; ** p<0.05; * p<0.10.

Table C.3: Population by decade

| Decade | International | Domestic |
|---|---|---|
| 1990 | 13.78 | 146.82 |
| 2000 | 17.78 | 169.79 |
| 2010 | 25.33 | 190.75 |

**Notes:** This table reports the population (in millions) of the domestic market and the equivalent population representing the international market used in each decade.

Table C.4: Trade elasticity in the literature

| Paper | Preferred $\theta$ |
|---|---|
| Eaton and Kortum (2002) | 8.28 |
| Donaldson and Hornbeck (2016) | 8.22 |
| Caliendo and Parro (2015) | 8.64 |
| Costinot et al. (2012) | 6.53 |
| Simonovska and Waugh (2014) | 4.10 |
| Head and Mayer (2014) | 6.74 |

**Notes:** This table summarizes the estimated values for the trade elasticity ($\theta$) found in the economics literature.

Table C.5: Estimation Results of Domestic Market Access on Deforestation

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | **log (Deforestation)** | | | |
| log(Market Access) | 0.43*** | 0.49*** | 0.45*** | 0.46*** |
| | (0.11) | (0.12) | (0.12) | (0.12) |
| $R^2(within)$ | 0.16 | 0.16 | 0.17 | 0.17 |
| Observations | 1,278 | 1,278 | 1,278 | 1,278 |
| | **First stage: log(Market Access)** | | | |
| log(Market Access, $d = 400$km) | | | | 0.90*** |
| | | | | (0.004) |
| F Statistic | | | | 190,224 |
| Observations | | | | 1,278 |
| Lat-Long | Yes | Yes | Yes | Yes |
| Distances | No | Yes | Yes | Yes |
| Soil | No | No | Yes | Yes |

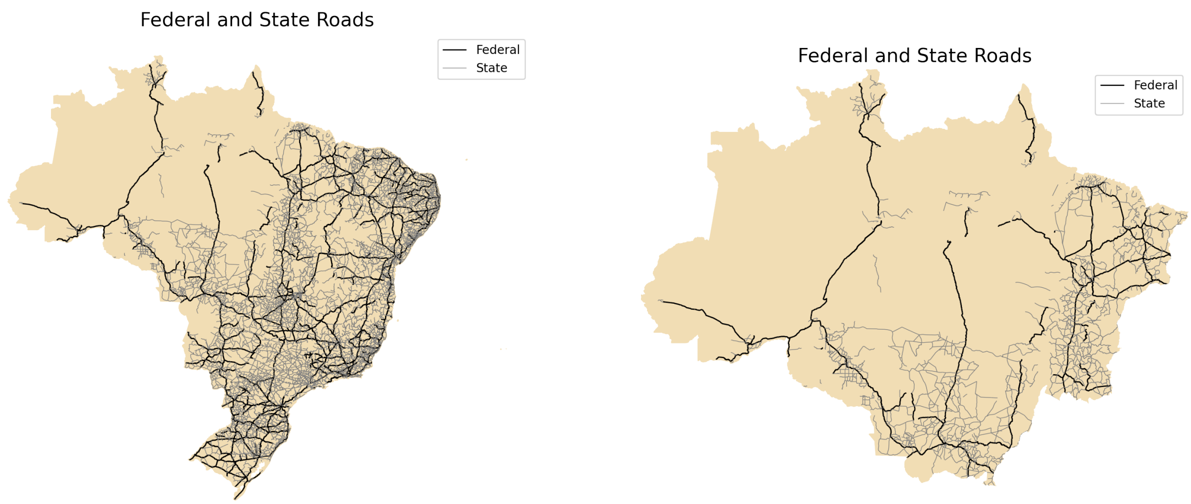
**Notes:** This table reports the results for estimating Equation (9) using only domestic markets to build the measure of market access. All specifications include municipality and state-year fixed effects. Column 1 includes cubic polynomials of latitude and longitude interacted with time dummies as controls. Column 2 add distance to the coast and distance to *Brasília*) interacted with time dummies as controls. Columns 3 and 4 add suitability to cultivate soy interacted with time dummies as controls. Columns 1-3 report the results of OLS specifications. Column 4 reports the results of a 2SLS specification obtained using market access excluding observations within a buffer of radius $d = 400$km as an instrument for market access. Standard errors clustered at the municipality level are reported in parenthesis. *** p<0.01; ** p<0.05; * p<0.10.

Table C.6: Market Access on Deforestation for Different Weights

| | area | $\sqrt{area}$ | None | None |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| log(Market Access) | 0.47*** | 0.60*** | 0.86*** | 0.69*** |
| | (0.13) | (0.16) | (0.20) | (0.22) |
| Area $\times$ log(Market Access) | | | | 0.01** |
| | | | | (0.006) |
| $R^2(within)$ | 0.17 | 0.17 | 0.17 | 0.17 |
| Observations | 1,278 | 1,278 | 1,278 | 1,278 |

**Notes:** This table reports the results of estimating Equation (9) using different weighting procedures. All specifications include municipality fixed effects, state-year fixed effects, geographic variables (cubic polynomials on latitude and longitude, the distance to the coast and to *Brasília*, and suitability for cultivating soy) interacted with year dummies as controls. Column 1 weights observations by the municipality area as in our preferred specification; column 2 weights observations by the squared root of the area; columns 3 does not weight the observations; Column 4 does not weight the observations, but includes area interacted with the market access as an additional control. Standard errors clustered at the municipality level are reported in parenthesis. *** p<0.01; ** p<0.05; * p<0.10.

Figure C.1: State and Federal Roads




**Notes:** These maps show the location of state roads in Brazil (left-panel) and the Amazon (right-panel) in the year 2010. The $R^2$ of a regression of market access constructed including state roads and market access constructed excluding state roads is 0.96.