Transportation Infrastructure and Deforestation in the Amazon*

Rafael Araujo† Juliano Assunção‡ Arthur Bragança§

May 2025

Abstract

We examine the effects of transportation infrastructure on deforestation in the Amazon. We build an inter-regional trade model in which agricultural commodities can be produced either in cleared lands or in forest lands. The model delivers a closed-form expression connecting deforestation and market access. Using panel data on the evolution of the transportation network and land use we estimate sizable effects of infrastructure on deforestation. Model simulations indicate that ignoring transportation infrastructure's effects beyond the projects surroundings underestimates deforestation impacts by one quarter. We also show how our framework can be used to evaluate the deforestation induced by individual projects, an essential input for public policies.

JEL: *F18, Q56, O13, R13*

Keywords: *Deforestation, Amazon, Transportation, Market Access*

*This paper previously circulated as "The Effects of Transportation Infrastructure on Deforestation in the Amazon : A General Equilibrium Approach" (World Bank Policy Research Working Paper WPS10415). We would like to thank Rodrigo Adão, Bruno Conte, Francisco Costa, Teevrat Garg, João Paulo Pessoa, Juan Jose Miranda, Emanuel Ornelas, Marcelo Sant'anna, and seminar participants at Climate Policy Initiative, FGV-EPGE, USP, AERE Summer Conference, RIDGE, SBE, and World Bank Land Conference for valuable comments and suggestions. We are grateful to Helena Arruda, Daniel Barbosa, Mateus Morais, and Brenda Prallon for excellent research assistance. We also thank Andrew Foster for excellent editorial advice and three anonymous referees for helpful comments and suggestions. 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.

†São Paulo School of Economics, Getulio Vargas Foundation, FGV EESP (e-mail: rafael.araujo@fgv.br).

‡PUC-Rio and Climate Policy Initiative (e-mail: juliano@econ.puc-rio.br).

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

1 Introduction

Investments in transportation infrastructure are considered fundamental to promote economic development (Redding and Venables, 2004; Donaldson and Hornbeck, 2016; Donaldson, 2018; Jedwab and Storeygard, 2021; Fajgelbaum and Redding, 2022). Nonetheless, there is a long-standing literature pointing to their impacts on deforestation, especially in tropical countries (Chomitz and Gray, 1996; Pfaff, 1999; Damania et al., 2018; Asher et al., 2020). Properly understanding the *aggregate* impacts of these investments is key for improving project selection and designing mitigation measures. However, the current literature focuses on the *local* impacts of investments in transportation infrastructure on deforestation, neglecting effects beyond the project surroundings

In this paper, we develop a market access framework to estimate the *aggregate* effects of investments in transportation infrastructure on deforestation. We then apply it to evaluate the effects of infrastructure on deforestation in the Brazilian Amazon – one of the world's most important biomes in terms of carbon storage and biodiversity (Baccini et al., 2012; Lawrence and Vandecar, 2015; Mitchard, 2018).

We begin by building an inter-regional trade model connecting transportation costs and land use choices. Our model departs from the literature of transportation infrastructure in agricultural settings (e.g., Donaldson and Hornbeck (2016), Donaldson (2018), Sotelo (2020)) by letting farmers choose to produce in two different types of land: cleared lands and recently deforested forest lands. This choice is driven by differences between productivity and prices with their interplay determining the evolution of the stock (cleared land) and flow (deforestation) of agricultural lands. We show that, in this setting, the effect of transportation infrastructure on deforestation is captured by a log-linear relationship between deforestation and market access, a sufficient statistic measuring how well connected each region is to all other regions.

Market access is a function of three elements: bilateral trade costs, the distribution of the population, and the elasticity of trade with respect to 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 for 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., transshipment costs between modes of transportation). We combine these matrices of bilateral transportation costs with official data on population and a calibrated trade elasticity from the literature to build market access for each municipality-by-decade pair.

We then regress deforestation, constructed using satellite-based information on deforestation from Mapbiomas (2019), on 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 control for time-invariant municipality characteristics and the time-varying effects of geographic factors, an approach not possible in previous applications based on cross-sectional data (e.g., Souza-Rodrigues (2018)). We also build a instrumental variables estimator in which a measure of market access constructed using only variation in transportation costs and population coming from distant regions is used as an instrument for market access. This procedure enables us to deal with time-varying local unobservables that might drive infrastructure building (e.g., Donaldson and Hornbeck (2016)).

Our results indicate that a 1% increase in market access increases deforestation by roughly 0.5%. This effect is almost identical across different estimation strategies (OLS and 2SLS), is not sensitive to calibrating the trade elasticity with other values found in the literature, and is robust to different ways of computing transportation costs.[1]

[1] We also show that relaxing some of our model's core hypotheses (e.g., introducing dynamics, multiple sectors or correlated productivity shocks) would reduce the tractability of our framework while increasing the effects transportation infrastructure on deforestation.

One key feature of our market access framework is that it explicitly considers the effects of investments in transportation infrastructure that go beyond the investment's surroundings – the indirect effects. To assess the importance of accounting for these indirect effects, we simulate 1,000 random roads and compute, for each simulation, the deforestation levels implied by the model. We then compare the results of the simulation with the results that would be obtained by a difference-in-differences that ignored indirect effects such as the one used by Asher et al. (2020). We find that ignoring indirect effects would underestimate the deforestation footprint of infrastructure projects by one-quarter, on average. The results point to the perils of assuming that the treatment of one region does not affect the outcome of other regions (the Stable Unit Treatment Value Assumption – SUTVA) in a scenario where infrastructure placement creates feedback effects across the whole infrastructure network.

Another key feature of our market access 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. Our model predicts that this project will generate 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, impact assessments only consider the municipalities crossed by the project.

We compare the results we obtained with our model with two types of land with the results obtained with a model with one type of land.[2] While we also estimate a positive relationship between market access and deforestation in the model with one type of land, we find that the responses to improvements in transportation infrastructure it generates are more homogeneous than the responses generated by the model with two types of

[2]This model with one type of land is essentially Donaldson and Hornbeck (2016)'s model with a positively sloped land supply curve.

land. This matters quantitatively – for instance, the model with one type of land predicts a deforestation footprint of the *Ferrogrão* railroad nearly five times larger than the model with two types of land. This highlights the importance of modelling the heterogeneity across cleared and forest lands in our setting.

Our work is primarily related to the literature on tropical deforestation (see Balboni et al. (2023) for a review). Our contribution is fourfold. First, we add to the literature on transportation costs and deforestation. Previous work in this literature explored heterogeneity between regions located closer or further from recently built roads to estimate the local effects of improvements in transportation infrastructure on deforestation (e.g., Damania et al. (2018) in Congo and Asher et al. (2020) in India). However, these papers were silent on the aggregate impacts of transportation infrastructure on deforestation – a key input for project selection and design. We use a novel framework to provide evidence that transportation infrastructure is a major driver of aggregate deforestation on a key region for global conservation efforts – the Brazilian Amazon.

Second, we add to a growing body of research using trade models to examine land use decisions (e.g., Costinot et al. (2016), Costinot and Donaldson (2016), Sotelo (2020)). Farrokhi et al. (2023) and Dominguez-Iino (2021) study the impact of international trade on deforestation. We focus on within-country variation to study impacts of transportation infrastructure, highlighting the importance of both international and domestic markets. Restrepo and Mariante (2023) and Gollin and Wolfersberger (2023) build spatial models of deforestation in Brazil to study mechanisms of leakage of conservation policies and infrastructure building, with Restrepo and Mariante (2023) in particular focusing on intertemporal effects of conservation with a dynamic model. Methodologically these papers work with calibrated models while we estimate our model, allowing for a clear discussion of identification and inference. Our paper also shifts focus to quantifying the bias of reduced-form approaches that study impacts of transportation on deforestation and drawing policy evaluations regarding the deforestation footprint of individual projects, an essential input for public policies. This is similar to work in urban settings (e.g., Tsivanidis (2019)), that also use this framework to evaluate the effects of individual projects.

Third, our findings provide evidence on the mismatch of the deforestation footprint of investments in transportation infrastructure and current regulations on project selection and design, thereby contributing to the literature on the design of conservation policies in the Amazon (e.g., Fetzer and Marden (2017), Souza-Rodrigues (2018), Assunção et al. (2020), Baragwanath and Bayi (2020), Heilmayr et al. (2020), Assunção et al. (2022), Araujo et al. (2020), Assunção et al. (2023),Tsuda et al. (2023)).

Fourth, our findings provide additional evidence on potential trade-offs between economic development and environmental conservation in a key ecosystem, thereby contributing to the long-standing debate on the relationship between economic development and the environment (see Foster and Rosenzweig (2003) for seminal work on the relationship between economic development and deforestation and Jayachandran (2022) for a review of the literature).

In the rest of the paper Section 2 presents the model; Section 3 describes the data; Section 4 presents our estimates; Section 5 discusses indirect effects; Section 6 discusses the *Ferrogrão* project; Section 7 concludes.

2 Theoretical Framework

In this section, we build an inter-regional trade model that enables us to evaluate the *aggregate* effects of transportation infrastructure on deforestation. Our model extends the approach proposed by Donaldson and Hornbeck (2016) by allowing farmers to produce in "cleared lands" or to clear forests to produce in "forest 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, conversion costs, or possible expropriation in forest lands (e.g., through land grabbing).

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. This expression summarizes the local and indirect 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 inelastically 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.[3] Trade between regions o and d is subject to an iceberg transportation cost τ_{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: cleared (C) or forest (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

[3] Section 4.3 discusses an extension with a two-sector model.

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 cleared or forest lands. We assume that capital and labor are freely mobile, which implies that $r_o = r$ 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 cleared land instead of forest 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 cleared 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 $F_o(z^C, z^F) = \exp(-(A_o^C {z^C}^{-\theta} + A_o^F {z^F}^{-\theta}))$. The parameter θ is negatively related to the dispersion of productivity shocks. Thus, lower (higher) θ 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 control 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.[4] Notice that the parameters A_o^F and A_o^C can also be more broadly interpreted as incorporating differences in fixed costs of exploring each type of land or a certainty equivalent to receiving a fine for deforestation.

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 cleared instead of forest 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 cleared 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

[4] In the Caveats and Extension section (Section 4.3) we discuss the implications of a model derived with correlated shocks.

and different types of land, they will purchase from the cheapest source.

It is possible to write the price distribution in municipality d as:[5]

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

in which x is a constant.[6]

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

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

[5]To derive equation (3) we show: (1) 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 (Lemma A.1; (2) the distribution of the prices of the varieties produced in o sold for d is identical to the distribution of offered varieties (Lemma A.2); (3) 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 (Lemma A.3).

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

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) equals the total demand for its products ($\sum_d X_{od}$). Using this condition and the expressions for prices (equation 3) and exports (equation 4), we obtain the following log-linear expression connecting output, wages, land prices, productivity, and measures of market access:

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

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 (τ_{od}^{θ}) and their competitiveness (CMA). It is denoted "firm market access" because it measures the access that firms face to sell their products. Donaldson and Hornbeck (2016) prove that it is possible to write $FMA_o = \rho CMA_o$, in which ρ is a constant. We use the term "market access" (MA) to denote $MA_o = FMA_o = \rho CMA_o$.[7]

To close the model, we need to substitute for Y_o, w_o, q_o^C, and q_o^F in equation (5). We begin by noting that agents must be indifferent to living in all municipalities ($\bar{U} = V_o = w_o / P_o$) in equilibrium. Using the expression for prices (equation 3), this means that it is possible to replace wages by $w_o = \bar{U} \times (CMA_o)^{-1}$. Moreover, 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 α share of output goes to the land factor. Finally, we note that the Fréchet distribution implies that the share of the land rents which goes to cleared land is equal to the probability a producer uses cleared land ($\bar{p}_o$). We formalize it through the following lemma:

[7] 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 $MA_o \equiv \frac{\bar{U} \rho^{\frac{1}{\theta}+1}}{\gamma} \sum_d \left[\tau_{od}{}^{-\theta} (MA_d)^{-\frac{1+\theta}{\theta}} N_d \right]$.

Lemma 2. *Total income accrued to forest land equals total income accrued to cleared 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 assume that the supply of cleared 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 forest lands.

We motivate the existence of a positively sloped supply curve for forest 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 forest land, q_o^F, a plot of land will be 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 $B_o^{-1} q_o^{F\frac{1}{\eta}}$, where B_o is a region-specific parameter that captures heterogeneity in the forest land supply curve. 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:[8]

$$(\eta + 1 + \eta\theta\alpha) \log L_o^F = \log x^{''} + (1 + \gamma) \log MA_o \tag{6}$$

As discussed before, market access is a function of trade costs (τ_{od}), population (N_d) and

[8]Here $x^{''} = \frac{x\alpha A_o^F}{B_o \rho^\gamma \bar{U}^{\gamma\theta}}$. In Appendix A.1 we show the algebraic steps to arrive at this final expression.

three model parameters (θ, $\bar{U}$ and ρ). 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 (τ_{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 (τ_{odt}), population (N_{ot}), and a trade elasticity parameter (θ). 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 (τ_{odt}).

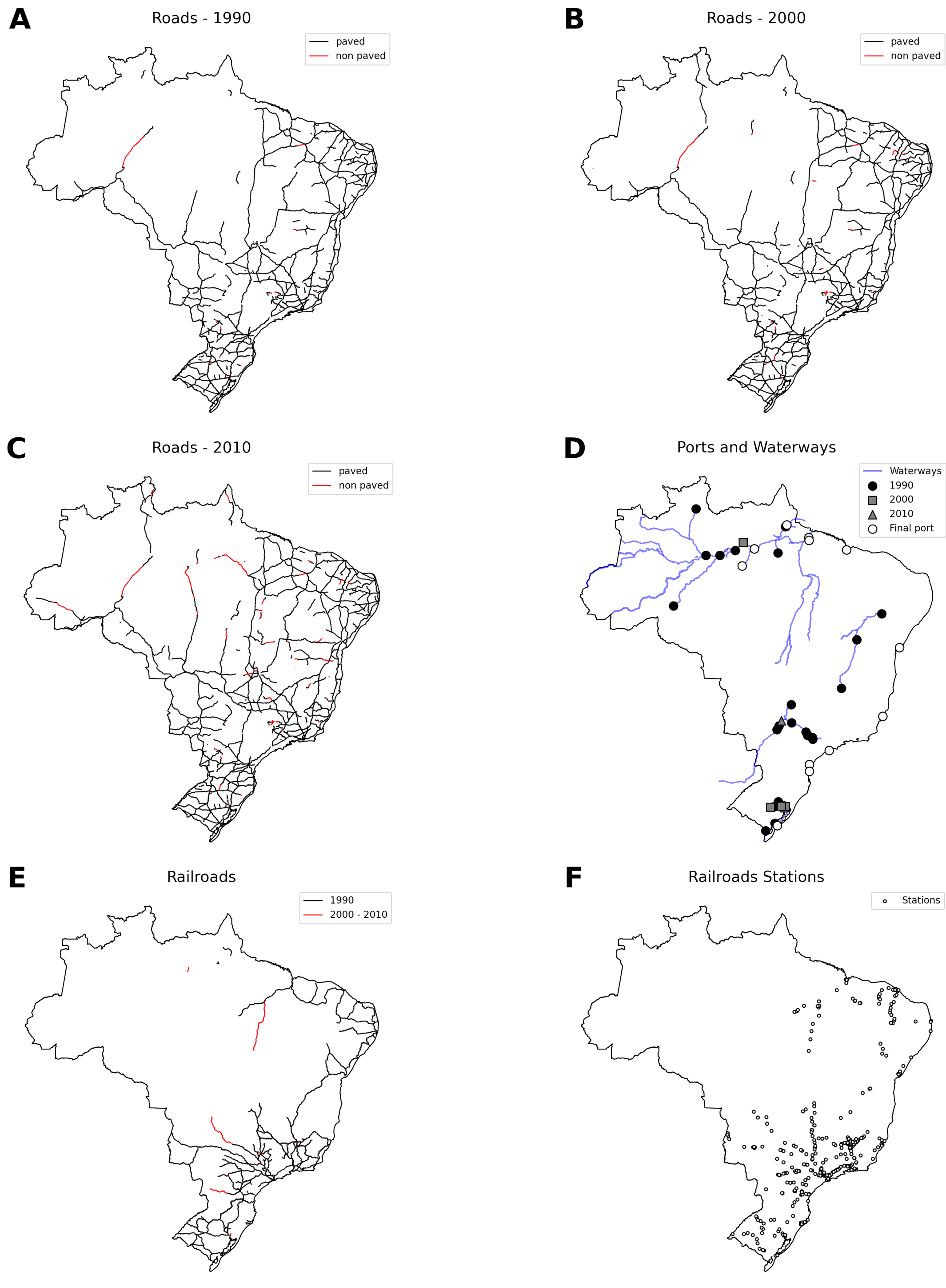
To construct a panel of the main roads network, we collect data on federal roads in Brazil from the Ministry of Transportation for the years 1990, 2000, and 2010. Figure 1, panels A, B, and C show the evolution of the roads network throughout the decades. We also have data on their traffic conditions, with each road being classified as paved or unpaved. 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).

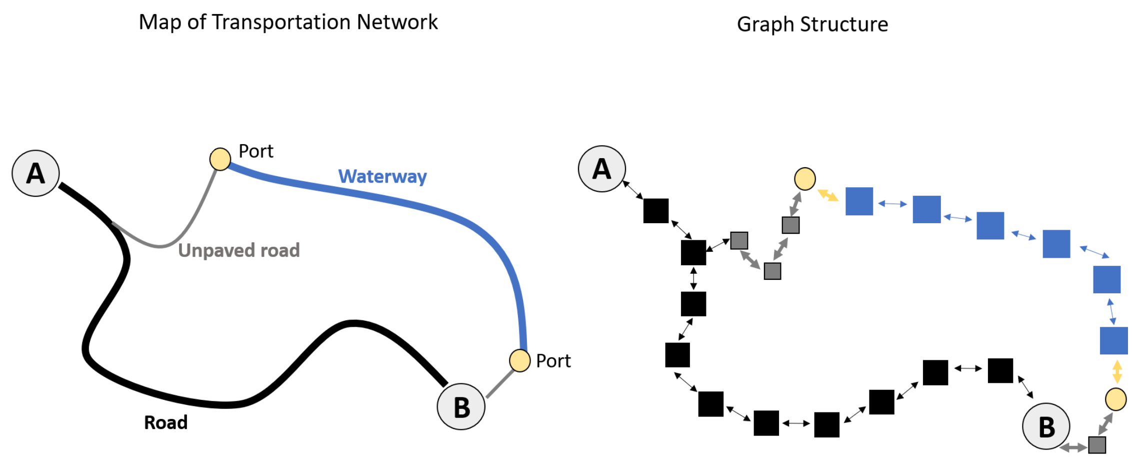
To compute transportation costs, we convert our transportation data into a graph (network) structure. In this graph, we use the Dijkstra's shortest path algorithm to find the

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.

least-cost path connecting two nodes (Dijkstra, 1959). Our graph structure 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 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).[9]

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 D.1). Notice that these values are scale-invariant – their relative values determine the shortest paths chosen by the algorithm. Compared with Souza-Rodrigues (2018), we are conservative with respect to impacts of roads on transportation cost, since our proportion of no road cost to road cost is between 5 and 20, while Souza-Rodrigues (2018)'s is between 20 and 40.[10]

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 D.2 reports the results. We use the coefficients of this regression to convert all our graph costs to iceberg costs, that is, we set $\tau_{odt} = 1 + \tilde{\alpha} + \tilde{\beta} cost\_graph_{odt}$. Both freight cost and the product price data - which is inclusive of shipping costs to the port - are measured in Brazilian Reais (BRL). Thus freight cost divided by product price yields a unit-free number measuring the proportion of the price lost by producers to pay

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

[10]The no road cost is the analog of wagon transportation in Donaldson and Hornbeck (2016). In our setting, it can be interpreted as the cost of transporting goods via small last mile roads, with poor quality and low traffic velocity.

for transportation costs. This measure maps to the definition of the iceberg costs (τ).

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 D.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. We calibrate the trade elasticity parameter (θ) 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 θ 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, neighboring municipalities that can be consistently compared across time.[11] 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.

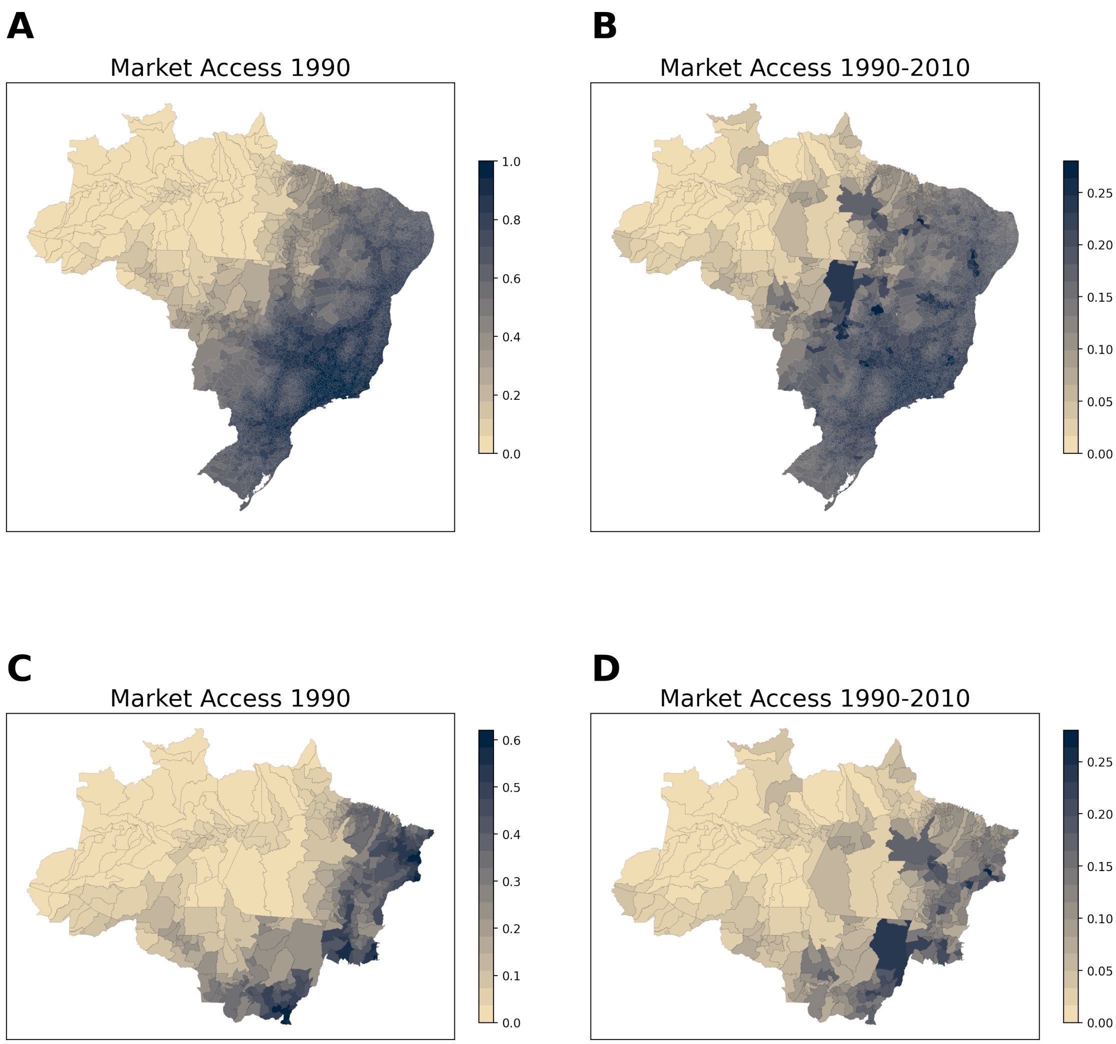
Market Access. After gathering the information on τ_{odt}, N_{ot}, and θ, we build the market access variable. Figure 3 shows the distribution of market access in 1990 and the difference in market access between 2010 and 1990 for Brazil (panels A and B) and the Amazon (panels C and D). We normalize market 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

[11] See Ehrl (2017) for further information on how an minimum comparable area is defined. In our setting we use comparable areas defined in the year 1991.

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.

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.[12]

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., 2023; Bragança and Dahis, 2022), changes in macroeconomic conditions (Assunção et al., 2015), and supply chain 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

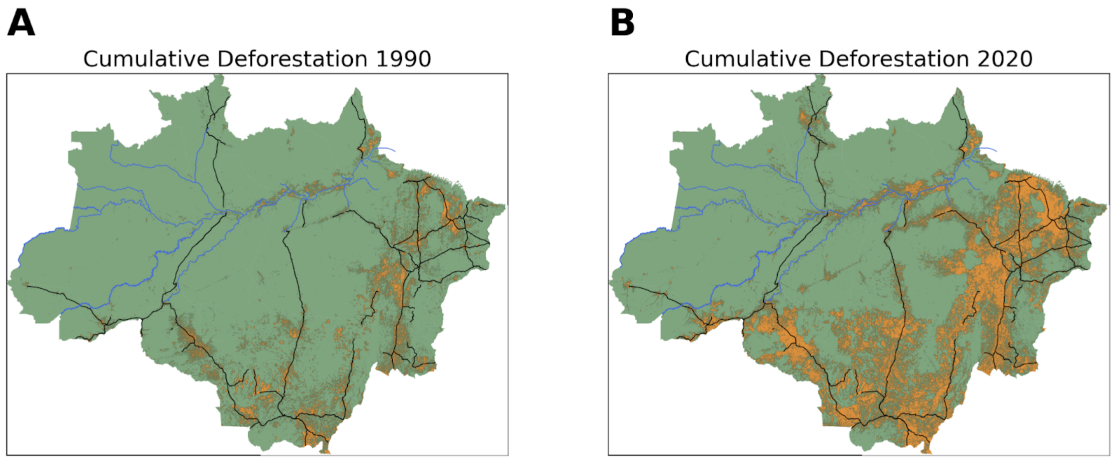
Our empirical model uses latitude, longitude, distance to *Brasília* (the national capital), distance to the coast, and suitability for cultivating soy as controls. Suitability is the average suitability for cultivating soy for the pixels from the Global Agro-Ecological Zones from the Food and Agriculture 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

[12] 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.

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. The blue lines show the main rivers of the Amazon basin.

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. (2020) for a discussion on this); (3) agents can anticipate changes in market access, which

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.

would result in observing an increase in deforestation before an increase in market access. This would, nonetheless, attenuate our estimates of the impacts of market access on deforestation. Second, we measure deforestation more frequently than we can measure market access. In this context, we build our estimation using long differences connecting the deforestation in a decade with market access measured 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 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*,[13] distance to the coast, suitability for cultivating soy), γ_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 own population in the computation of its market access.

Notice that we use the deforestation occurred in decade t ($\log(L_o^F)$) and not the cumulative deforestation ($\log(L_o^F + L_o^C)$) on the left side of equation (9). Using cumulative deforestation would be consistent with our model only if forest and cleared lands had the same productivity distribution – an unrealistic feature in the Amazon. We discuss results obtained using this alternative empirical model at the end of this section.

It is also 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

[13] We include this control because Brasilia has had an important impact in rearranging the Brazilian transportation infrastructure (see Morten and Oliveira (2024))

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 relatively 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 interacted with times dummies as controls. 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 an 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. Changing the market access of a median municipality to the 75th percentile increases deforestation by 16%. 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. A second potential problem, given the spatial correlation in both deforestation and market access, is that a omitted variable is driving the advance of the deforestation and infrastructure investment. Our OLS specifications deal with this problem by flexibly controlling for time-invariant municipality characteristics and the time-varying effects of geographic factors, including a cubic polynomial of latitude and longitude and distance to the national capital. 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.[14]

The identification hypothesis behind this instrument is that changes in market access further than a distance d from a region is not correlated with its own productivity shocks. In particular, we are concerned with a local productivity shock that could increase deforestation and create pressure for the development of transportation infrastructure, for example, political pressure to build a road after a region becomes more productive in fattening cattle. If this were the case our estimate would be biased upwards.

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. With this specification we aim to address the endogeneity of local productivity shocks and market access. Another concern regarding the potential impact of local shocks on market access is that it could subsequently attract more local infrastructure. Appendix Table D.5 replicates the instrumental strategy presented in Table 2, but includes as an additional control the total transportation infrastructure within the designated buffer area, using either a cubic polynomial or interacted with year dummies. Our findings show elasticities that are similar to, or slightly larger than, those estimated previously.

[14]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)				0.95***	0.95***	0.95***
				(0.002)	(0.002)	(0.002)
F Statistic				87,994	94,216	94,346
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 D.10 in the appendix presents our main results under alternative clustering strategies. All results remain significant at the 5% level.

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. Appendix Table D.6 shows the robustness of our results when using only the domestic market access to identification across a range of specifications of controls. These results contradicts a large prior body of research considering the potential effects of international trade on deforestation (see Copeland et al., 2022), but are in line with recent evidence for Brazil (Carreira et al., 2024).

Table 4 provides evidence of the robustness of our results to different values of the trade elasticity (θ) reported in the literature (see Appendix Table D.4 for a list of references and their estimated/used trade elasticity). We find similar elasticities for different values of θ, 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 (θ) 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 Appendix

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)	0.95***	0.95***	1.00***	0.88***
	(0.01)	(0.01)	(0.01)	(0.01)
F Statistic	94,346	132,155	54,860	2,815
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 θ'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)				0.95***	0.74***	1.15***
				(0.01)	(0.01)	(0.01)
F Statistic				94,346	2,730	5,099
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 (θ). 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.

Table D.7). 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 Appendix Table D.7 suggest that our preferred estimates are underestimating the impacts of deforestation.

Our study area is composed of the states of the Legal Amazon, a political denomination that includes the Brazilian Amazon biome and a portion of the Cerrado biome. Table D.9 in the Appendix shows that there is no heterogeneous effect of market access on deforestation across these two biomes and that the inclusion of this heterogeneity does not affect our main estimates.

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 forest land. As shown in equation (6), the elasticity of deforestation to market access is a function of factor shares (α and γ), the trade elasticity (θ), and the elasticity of the supply of forest land ($1/\eta$). We calibrate the factor shares and the trade elasticity using common values from the literature to compute the elasticity of forest land implied by our estimates. We assume that the share of land in production (α) is 0.2 and the share of labor (γ) 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 forest land implied by our empirical estimates is between 1.20-1.36. The elasticities of forest 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 cleared and forest land. Moreover, the supply of cleared land is fixed, implying that the elasticity of land supply is simply the elasticity of forest land multiplied by its share in total land supply. Our data shows the share of forest 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 θ and between 0.50-0.55 for our alternative θ. 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 elasticity.

Additional deforestation induced by changes in market access. Our estimates can be used to evaluate the additional deforestation induced by changes in market access. The expression $\beta \times \Delta \log MA_{o,t+1} \times L^F_{o,t}$ gives the model-implied additional deforestation induced by changes in market access for all changes in market access that do not affect utility levels ($\bar{U}$ – see Donaldson and Hornbeck (2016) for a discussion). We use this expression to compute the model-implied deforestation in 2000 and 2010 that was induced by changes in market access. We then compare it to observed deforestation levels. We find that, at the median municipality, additional deforestation induced by changes in market access corresponds to 13.0% (14.4%) of total deforestation observed in 2000 (2010). We also found that additional deforestation induced by changes in market access is typically below observed observed deforestation for more than 98% of the municipality-decade pairs. These numbers indicate that the model generates shocks in deforestation that are within the support of observed deforestation.[15]

[15] It is important to highlight that this exercise computes additional deforestation induced by changes in market access implied by our model and not total deforestation implied by our model. To compute total deforestation we would have to use the formula $(1 + \beta \times \Delta \log MA_{o,t+1}) \times L^F_{o,t}$ instead of $\beta \times \Delta \log MA_{o,t+1} \times L^F_{o,t}$. Appendix Figure D.2 plots the log of additional deforestation implied by our model (Panel A) and the log of total deforestation implied by our model (Panels C) against the log of observed deforestation for each decade. It shows that additional deforestation is typically below observed deforestation while total deforestation is much closely aligned with observed deforestation. It is important

4.3 Discussion and Extensions

Transport within municipality. When computing bilateral trade costs we build optimal trajectories between points located in each municipality. In our results presented so far, we use a representative point which is guaranteed to be within the municipality geometry. To assess the importance of the choice of points we run the following exercise: (1) draw a random point inside each municipality (2) add the cost of going from this point to the representative point (3) re-estimate the model. We bound the result assuming that the transportation within municipality is all done either by road or by land. We run this exercise 1,000 times for each specification. We find that the average elasticity of the specification in column 3 of Table 2 for this exercise is of 0.47 (std. 0.001) when assuming agents move by road inside municipality and of 0.51 (std. 0.02) when assuming agents move by land. This result show that within municipality transportation is not important enough to generate significant changes in our results.

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 D.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.

to highlight that the high correlation of the measures is largely attributable to the fact that deforestation is quite persistent. Indeed, because transportation costs vary smoothly in space, the heterogeneity in shocks in market access does not help to explain much of the heterogeneity shocks in deforestation. Future work could use more detailed transport network data to explore local variation in market access and its effects on deforestation.

Model with one type of land. The distinction between forest and cleared lands is a key feature with our model. Without this distinction, our theoretical model would collapse to Donaldson and Hornbeck (2016)'s original model with the addition of a positively sloped land supply curve. This model generates a log-linear relationship between cumulative deforestation up to decade t (instead of deforestation in decade t) and market access in decade t.

The elasticity of cumulative deforestation with respect to market access is between 0.17-0.19 (see Appendix Table D.8). Whether these coefficients entail different effects of a shock in market access on deforestation depends on the share of forest land. The model with one type of land will underestimate (overestimate) deforestation when the share of forest land is higher (lower) than 0.36.[16] The aggregate share of forest land in our data is 0.33. Therefore, the mean effect of a shock in market access on deforestation obtained in the model with one type of land is similar to the effect obtained in the original model. Nevertheless, the model with one type of land will miss important heterogeneity as it predicts that the effects of improvements in transportation infrastructure are comparable in regions (or periods) in which the share of forest lands is quite different. Section 6 provides an example of these differences.

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

[16] Let D be accumulated deforestation and s be the share of forest land, $(0.18\Delta\%MA)D > (0.5\Delta\%MA)sD \iff s < 0.36$

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 – cleared or forest – 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.

Labor mobility. Gollin and Wolfersberger (2023) and Restrepo and Mariante (2023) incorporate imperfect labor mobility in their model of deforestation. In Appendix C, we derive one extension of the model with imperfect labor mobility through taste/amenities shocks that results in a closed form solution implicitly connecting deforestation to market access. However, similar to the extension with correlated productivity shocks across types of land, it is not possible to estimate the resulting expression using a regression framework and to conduct counterfactuals using the sufficient statistic approach. Instead, we would need to calibrate all the parameters of the model and incorporate new data on the location parameters (A_o^F, A_o^C) to simulate this extension. Nonetheless, we use this extension to construct bounds of the difference between counterfactuals produced with perfect and imperfect mobility. These bounds indicate that imperfect mobility is unlikely to influence our results qualitatively or quantitatively. This is consistent with the existence of some features of our setting (long differences in our estimation and a setting with low population density and high mobility) that render perfect mobility a justifiable approximation in our setting. It is also consistent with the results reported in Table 3 that indicate that the identification of our sufficient statistic is coming from variation in transportation costs and not in population dynamics.

Dynamics. We choose a static model to keep the general equilibrium framework tractable. We then leverage our decennial long-differences to recover long-run elasticities of deforestation with respect to transportation infrastructure. The deforestation literature has used static models both with cross-section (Souza-Rodrigues, 2018) and panel data (Dominguez-Iino, 2021). While dynamic models of deforestation have been employed in the study of deforestation, they either abstract from general equilibrium effects (Scott, 2014; Araujo et al., 2020; Sant'Anna, 2021) or use calibration exercises to construct counterfactuals and evaluate the effects of policies Farrokhi et al. (2023); Restrepo and Mariante (2023). Although we are not able to explicitly evaluate the implications of a dynamic version of our model, it is worth highlighting the papers that explicitly compare static and dynamic models of land conversion (e.g., (Scott, 2014), Araujo et al. (2020), Sant'Anna (2021)) find that the land conversion elasticities are lower in static models than in dynamic models. Because this is the key elasticity guiding the results of our counterfactuals, this suggests that it is likely that our approach underestimates the effect of transportation infrastructure on deforestation. Consistent with this, we find that our model underestimates deforestation in most municipalities.

5 The Importance of Indirect Effects

Indirect 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. These indirect effects 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.

To assess the importance of these indirect effects when estimating local effects of transportation infrastructure, we leverage the structure of the model to simulate deforestation impact of randomly placed roads added to the 2010 transportation network. We then compare these model-implied deforestation with the local effects that would be estimated using a difference-in-differences strategy analogous to the one used by Asher et al. (2020).

We proceed as follows. First, we simulate a total of 1,000 roads (see Figure 5, panel A). Second, we compute the market access change generated by 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.[17] 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 5, panel B). The average change in market access for the 1,000 random roads is of 4% (95th percentile of 11%), which is below the average change in market access observed in the data of 17% (95th percentile of 21%). Thus, in this exercise we are not extrapolating changes in market access far from the magnitudes observed in the data.

We find this difference-in-differences underestimates the local effects of roads on deforestation. Figure 5, panel C reports the percentage of the true local effect of deforestation

[17] 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.

captured by the reduced form approach. On average, not accounting for indirect effects would result in underestimating by one quarter the local effects of roads on deforestation. However, there is a significant share of simulations with much higher bias.

Figure 5, 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 indirect effects.

Absent in Figure 5 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 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.[18]

As an example, we build a counterfactual scenario for the construction of the *Ferrogrão*

[18] 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)

railroad (Figure 6, 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. The average change in market access is of 1% (95th percentile of 5%), which is well below the average change in market access observed in the data of 17% (95th percentile of 21%). Thus, we are not extrapolating changes in market access far from the magnitudes observed in the data. We find that the *Ferrogrão* construction is expected to increase total deforestation by 400 km^2 in the following decade.

Interestingly, we estimate that the *Ferrogrão* construction is expected to generate almost five times this deforestation (1,967 km^2) using a model with only one type of land. This is largely due to the fact that the region affected by the *Ferrogrão* has more cleared lands than the typical region in the Amazon – a feature ignored by the model with one type of land. This result shows the importance of distinguishing between forest and cleared lands as done in our theoretical model.

We monetize this deforestation with parameters currently used to fund conservation projects in the Amazon (Amazon Fund, 2018). Specifically, we assume a forest carbon stock of 48,510 tCO2 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., ecosystem 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). With a carbon price of USD 50, even with an underestimated deforestation effect, the environmental cost would correspond to more than 20% of the fiscal cost of *Ferrogrão*. Finally, with a carbon price of USD 1,056 (Bilal and Känzig, 2024) the social cost of the project surpasses the fiscal cost four fold, even without accounting for other positive externalities of the forest.

Figure 6, 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 the locations 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., 2020). In this paper, we develop a framework to assess the aggregate deforestation cost of infrastructure projects. 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 indirect effects would underestimate the deforestation of these roads by one-quarter.

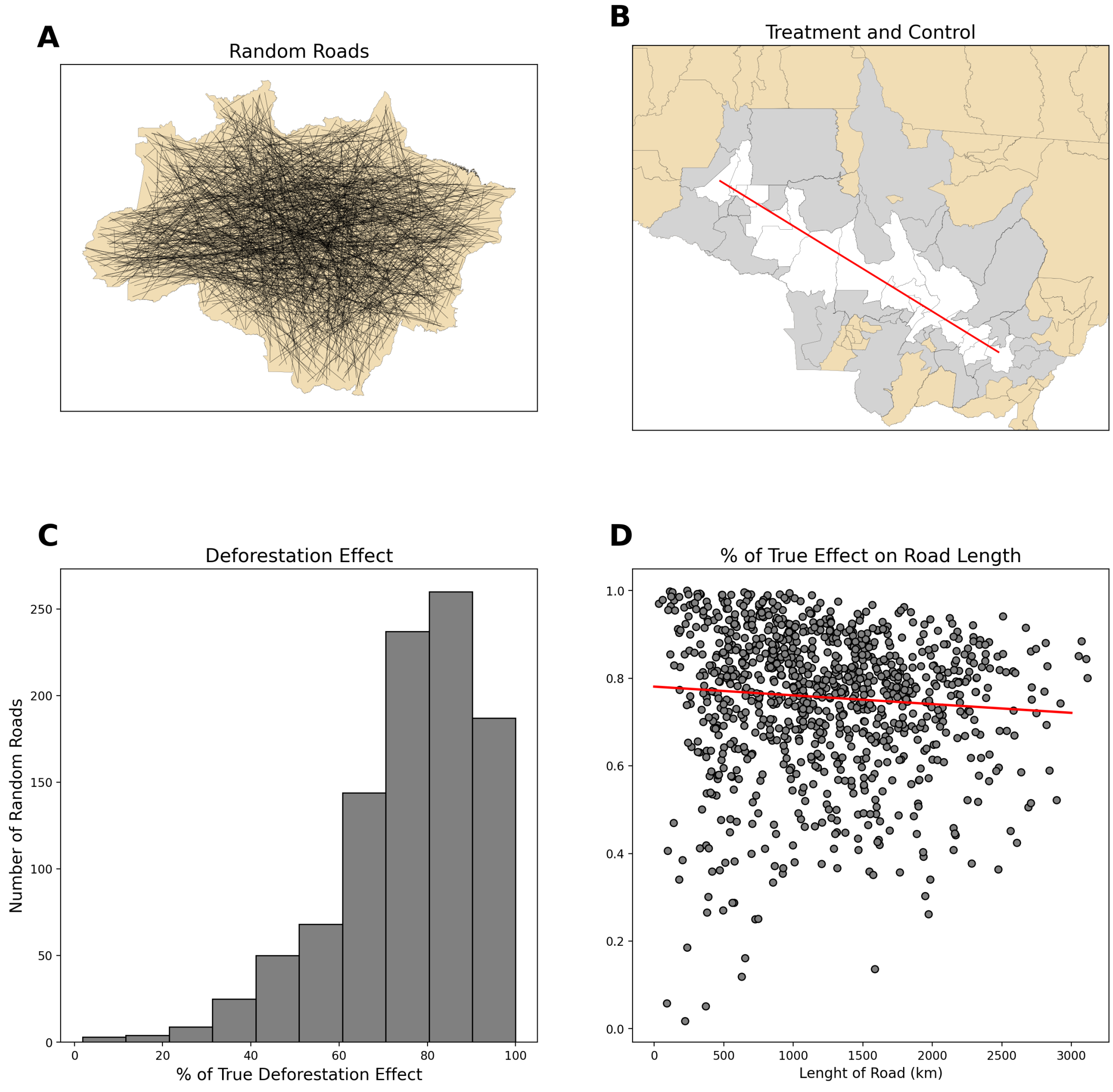
Fourth, we use our model and estimates to predict the impact of the *Ferrogrão* railroad – a highly controversial project planned to be built in the Amazon – and find that it will generate substantial environmental impacts, mostly in municipalities not crossed by the project.

Methodologically, we not only provide evidence of the importance of incorporating direct and indirect 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 and costs to process trans-shipment in different ports (Bonadio, 2021)). 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 Amazon (e.g., Fetzer and Marden (2017), Souza-Rodrigues (2018), Assunção et al. (2020), Baragwanath and Bayi (2020), Heilmayr et al. (2020), Assunção et al. (2022), Araujo et al. (2020), Assunção et al. (2023)). 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.

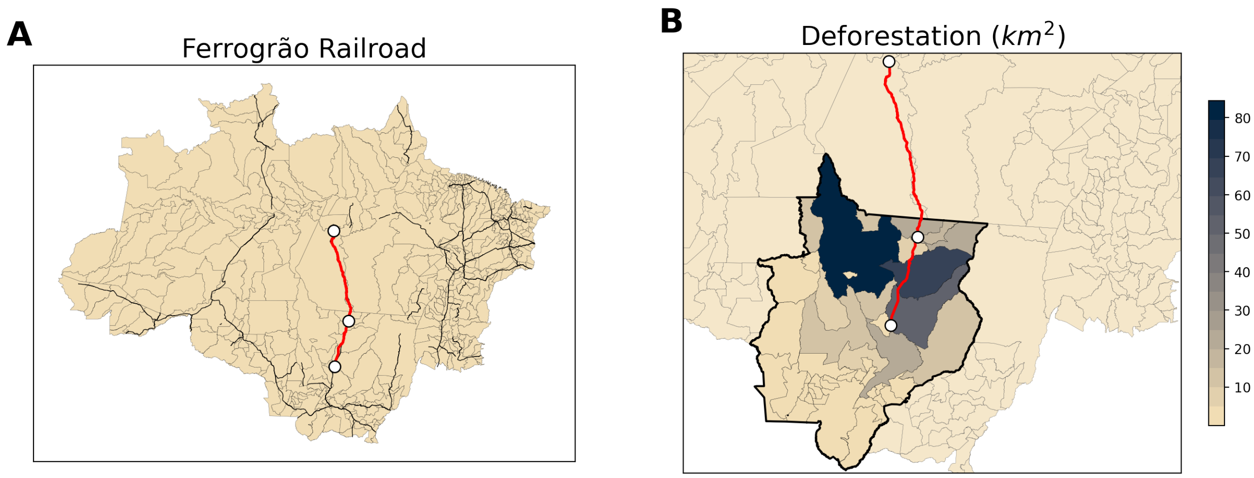
As with any paper, ours have limitations. In particular we do not consider other externalities of deforestation, such as, forest degradation and the possibility of a tipping point (Flores et al., 2024), energy generation (Stickler et al., 2013; Araujo, 2024), and mortality effects of increasing temperatures (Masuda et al., 2021). Neither do we consider other effects of market access, such as, inequality (Hochard and Barbier, 2017) and property value (Donaldson and Hornbeck, 2016). The composition of indirect effects with these other negative impacts of deforestation is an avenue for future research.

Figure 5: The Importance of Indirect 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 in 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 6: 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

Adao, Rodrigo, Costas Arkolakis, and Federico Esposito, "General equilibrium effects in space: Theory and measurement," Technical Report, National Bureau of Economic Research 2019.

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

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

Araujo, Rafael, "The value of tropical forests to hydropower," *Energy Economics*, 2024, *129*, 107205.

_ , 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 impact of transportation infrastructure," *The Economic Journal*, 2020, *130* (629), 1173–1199.

Assunção, Juliano, Clarissa Gandour, and Romero Rocha, "DETER-ing Deforestation in the Amazon: Environmental Monitoring and Law Enforcement," *American Economic Journal: Applied Economics*, 2023, *15* (2), 125–156.

_ , _ , 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.

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.

Balboni, Clare, Aaron Berman, Robin Burgess, and Benjamin A Olken, "The economics of tropical deforestation," *Annual Review of Economics*, 2023, *15*, 723–754.

Baragwanath, Kathryn and Ella Bayi, "Collective property rights reduce deforestation in the Brazilian Amazon," *Proceedings of the National Academy of Sciences*, 2020, *117* (34), 20495–20502.

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.

Bilal, Adrien and Diego R Känzig, "The Macroeconomic Impact of Climate Change: Global vs. Local Temperature," Technical Report, National Bureau of Economic Research 2024.

Bonadio, Barthélémy, "Ports vs. roads: infrastructure, market access and regional outcomes," 2021.

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.

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

Carreira, Igor, Francisco Costa, and Joao Paulo Pessoa, "The deforestation effects of trade and agricultural productivity in Brazil," *Journal of Development Economics*, 2024, *167*, 103217.

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.*

Copeland, Brian R., Joseph S. Shapiro, and M. Scott Taylor, "Chapter 2 - Globalization and the environment," in Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff, eds., *Handbook of International Economics: International Trade, Volume 5*, Vol. 5 of *Handbook of International Economics*, Elsevier, 2022, pp. 61–146.

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

— , — , **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.

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.

Diamond, Rebecca, “The determinants and welfare implications of US workers’ diverging location choices by skill: 1980–2000,” *American Economic Review*, 2016, *106* (3), 479–524.

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

Dominguez-Iino, Tomas, “Efficiency and redistribution in environmental policy: An equilibrium analysis of agricultural supply chains,” Technical Report, Technical report, working paper 2021.

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

— **and Richard Hornbeck**, “ Railroads and American Economic Growth: A “Market Access” Approach,” *The Quarterly Journal of Economics*, 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.

Farrokhi, Farid, Elliot Kang, Heitor S Pellegrina, and Sebastian Sotelo, "Deforestation: A global and dynamic perspective," *Cited on*, 2023, p. 6.

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

Flores, Bernardo M, Encarni Montoya, Boris Sakschewski, Nathália Nascimento, Arie Staal, Richard A Betts, Carolina Levis, David M Lapola, Adriane Esquível-Muelbert, Catarina Jakovac et al., "Critical transitions in the Amazon forest system," *Nature*, 2024, *626* (7999), 555–564.

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

Gollin, Douglas and Julien Wolfersberger, "Agricultural Trade and Deforestation: the Role of New Roads," 2023.

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.

Hochard, Jacob and Edward Barbier, "Market accessibility and economic growth: Insights from a new dimension of inequality," *World Development*, 2017, *97*, 279–297.

Jayachandran, Seema, "How economic development influences the environment," *Annual Review of Economics*, 2022, *14*, 229–252.

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.

Kline, Patrick and Enrico Moretti, "People, places, and public policy: Some simple welfare economics of local economic development programs," *Annu. Rev. Econ.*, 2014, *6* (1), 629–662.

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.

Masuda, Yuta J, Teevrat Garg, Ike Anggraeni, Kristie Ebi, Jennifer Krenz, Edward T Game, Nicholas H Wolff, and June T Spector, "Warming from tropical deforestation reduces worker productivity in rural communities," *Nature communications*, 2021, *12* (1), 1601.

Mitchard, Edward TA, "The tropical forest carbon cycle and climate change," *Nature*, 2018, *559* (7715), 527–534.

Monte, Ferdinando, Stephen J Redding, and Esteban Rossi-Hansberg, "Commuting, migration, and local employment elasticities," *American Economic Review*, 2018, *108* (12), 3855–3890.

Morten, Melanie and Jaqueline Oliveira, "The effects of roads on trade and migration: Evidence from a planned capital city," *American Economic Journal: Applied Economics*, 2024, *16* (2), 389–421.

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

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

Restrepo, Verónica Salazar and Gabriel Leite Mariante, "Does Conservation Work in General Equilibrium?," 2023.

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.

Sant'Anna, Marcelo, "How green is sugarcane ethanol?," *Review of Economics and Statistics*, 2021, pp. 1–45.

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.

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

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

Stickler, Claudia M, Michael T Coe, Marcos H Costa, Daniel C Nepstad, David G McGrath, Livia CP Dias, Hermann O Rodrigues, and Britaldo S Soares-Filho, "Dependence of hydropower energy generation on forests in the Amazon Basin at local and regional scales," *Proceedings of the National Academy of Sciences*, 2013, *110* (23), 9601–9606.

Tsivanidis, Nick, "Evaluating the Impact of Urban Transit Infrastructure: Evidence from Bogotá's TransMilenio," 2019.

Tsuda, Shunsuke, Yoshito Takasaki, and Mari Tanaka, "Human and nature: economies of density and conservation in the amazon rainforest," *Available at SSRN 4652170*, 2023.

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

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.

FOR ONLINE PUBLICATION

Appendix to "Transportation Infrastructure and Deforestation in the Amazon"

Table of Contents

A Proofs of the Theoretical Model

Lemma 1. *The probability that a farmer will choose cleared 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.

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

■

Lemma 2. *Total income accrued to forest land equals total income accrued to cleared 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 cleared land and forest 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,

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

■

Lemma A.1. *Offered price distribution from region $o \in R$ to region $d \in O$ $(G_{o,d}(p))$ is a uni-*

variate 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 $(G_d(p))$ 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 cleared land ($\bar{G}^C_{o,d}(p)$) 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 $s = \frac{\tau_{od} {q_o^C}^\alpha w_o^\gamma r^{1-\alpha-\gamma}}{p}$

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

■

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 $G^c_{o,d}$ and $G^f_{o,d}$ only by a constant factor, conditioning on cleared or forest land will result in the same above integral above, up to a multiplicative constant. Therefore

$$\pi^C_{o,d} = \bar{p} \frac{\phi_{od}}{\Phi_d}$$

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

■

A.1 Derivation of the model

In this section we present a step by step guide to arrive at our main expression in equation 6. Given the expressions already presented in the paper, solving for the deforestation amounts to isolating the correct terms in a system of equations. We start with (5):

$$\log Y_o = \log x' + \gamma \log CMA_o + \log \left(A_o^C (q_o^C)^{-\theta\alpha} + A_o^F (q_o^F)^{-\theta\alpha} \right) + \log FMA_o, \qquad \text{(A.7)}$$

Substituting the Cobb-Douglas implication for land share $Y_o = \left(q_o^C L_o^C + q_o^F L_o^F \right) / \alpha$, we have:

$$\log \frac{q_o^C L_o^C + q_o^F L_o^F}{\alpha} = \log x' + \gamma \log CMA_o + \log\left(A_o^C (q_o^C)^{-\theta\alpha} + A_o^F (q_o^F)^{-\theta\alpha}\right) + \log FMA_o, \tag{A.8}$$

Substituting for the land supply through deforestation $q_o^F = B_o(L_o^F)^\eta$:

$$\log \frac{q_o^C L_o^C + B_o(L_o^F)^{\eta+1}}{\alpha} = \log x' + \gamma \log CMA_o + \log\left(A_o^C (q_o^C)^{-\theta\alpha} + A_o^F (B_o(L_o^F)^\eta)^{-\theta\alpha}\right) + \log FMA_o, \tag{A.9}$$

From Lemmas 1 and 2 we have:

$$q_o^F = \left[\frac{A_o^C L_o^F}{A_o^F L_o^C}\right]^{\frac{1}{1+\theta\alpha}} B_o(L_o^F)^\eta \tag{A.10}$$

Using this expression and collecting terms we arrive at:

$$\begin{aligned} \log \frac{B_o(L_o^F)^{\eta+1}}{\alpha}\left(\left(\frac{A_o^C L_o^F}{A_o^F L_o^C}\right)^{\frac{1}{1+\theta\alpha}} \frac{L_o^C}{L_o^F} + 1\right) &= \log x' + \gamma \log CMA_o + \\ \log\left[A_o^F (B_o(L_o^F)^\eta)^{-\theta\alpha}\left(\left(\frac{A_o^C L_o^F}{A_o^F L_o^C}\right)^{\frac{1}{1+\theta\alpha}} \frac{L_o^C}{L_o^F} + 1\right)\right] &+ \log FMA_o \end{aligned} \tag{A.11}$$

The expression in parentheses will cancel out. Finally, we use $CMA_o = \rho FMA_o := MA_o$ to arrive at expression (6) in the text.

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 $d \in O$

$$X_{od}^m = x_m M_o (\tau_{o,d}^m q_o{}^{\alpha_m} w_o{}^{\gamma_m})^{-\theta_m} (CMA_d^m)^{-1} X_d^m \tag{B.6}$$

And the condition of equilibrium in the manufacturing sector is

$$Y_o^m = \sum_d X_{od}^m = x_m M_o (q_o{}^{\alpha_m} w_o{}^{\gamma_m})^{-\theta_m} \sum_d \tau_{o,d}^m{}^{-\theta_m} (CMA_d^m)^{-1} X_d^m \tag{B.7}$$

Agriculture

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

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

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

$$X_{od} = x \left(\tau_{od} w_o^{\gamma} \right)^{-\theta} \left(A_o^C (q_o^C)^{-g\theta\alpha} + A_o^F (q_o^F)^{-g\theta\alpha} \right)^{\frac{1}{g}} (CMA_d)^{-1} X_d \tag{B.9}$$

The equilibrium in agricultural markets is given by

$$Y_o = \sum_d X_{od} = x \left(w_o^{\gamma} \right)^{-\theta} \left(A_o^C (q_o^C)^{-g\theta\alpha} + A_o^F (q_o^F)^{-g\theta\alpha} \right)^{\frac{1}{g}} \sum_d \tau_{od}{}^{-\theta} (CMA_d)^{-1} X_d \tag{B.10}$$

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 Labor Mobility

As discussed in the main text, we consider a model with perfect labor mobility as a reasonable approximation for the Amazon as it captures the fact that the population of the localities in the frontier is more or less being determined by the comparison of productivity and cost of living in these localities with real wages prevailing in the rest of country. It is possible to extend our theoretical model to include imperfect labor mobility.

Following Kline and Moretti (2014), Monte et al. (2018) and Morten and Oliveira (2024), we consider a model in which agents have heterogeneous tastes across locations.[2] This implies that location decisions will be driven not only by differences in real wages as in the model with perfect mobility but also by differences in preferences across locations.

Formally, we assume that the indirect utility of agent i living in location d is

$$V_{id} = \frac{w_d}{P_d} \times \xi_d(i), \tag{C.1}$$

in which $\xi_d(i)$ is Fréchet-distributed with shape parameter $1/\epsilon$.

Utility maximization implies that the share of individuals born in o who choose to live in d is

$$\pi_{od} = \frac{(w_d/P_d)^{1/\epsilon}}{\sum_d' (w_d'/P_d')^{1/\epsilon}}, \tag{C.2}$$

Let N_d be the total number of individuals living in each destination d. Summing the expression above across origins, we obtain the following expression connecting local population, nominal wages and cost of living:

$$N_d = \left(\frac{w_d}{P_d}\right)^{1/\epsilon} \times \bar{U}, \tag{C.3}$$

in which $\bar{U} = \sum_o (\bar{N}_o / \sum_d' (w_d'/P_d')^{\epsilon})$ is the mean utility across locations and $\bar{N}_o$ is the number of individuals born in each origin o.

Equation (C.3) is a (inverse) supply curve. The parameter $1/\epsilon$ governs the responsiveness

[2] We refrain from modeling the differences in tastes across locations (e.g., differences in amenities or in the costs of housing and/or other non-tradable goods). However, including these dimensions would not change the conclusions of this section.

of local population to differences in real wages across locations. If $\epsilon \to 0$, population is perfectly elastic and the model will be equal to the model with perfect mobility discussed in the main text.

We solve the model following the same steps used to solve the model presented in the main text. We start from expression for output (eq. (5)) and then use the factor shares and the factor supply curves to express deforestation as a function of parameters. We obtain the following closed form solution connecting deforestation and market access:

$$\begin{aligned}[(1+\epsilon)(1+\eta+\eta\theta\alpha)+\gamma\theta\epsilon]\log L_o^F+\gamma\theta\epsilon\log\left(1+\frac{A_o^C}{A_o^F}\left(\frac{L^F}{L^C}\right)^{\frac{\theta\alpha}{1+\theta\alpha}}\right)= \\ \log C+[(1+\epsilon)(1+\gamma)+\epsilon]\log MA_o,\end{aligned} \tag{C.4}$$

in which C is a constant.

Notice that equation (C.4) converges to equation (6) if $\epsilon \to 0$. However, different from the equation in the main model, it does not enable us to evaluate the effects of changes in transportation costs on deforestation using solely the coefficients from regressing deforestation on market access. To see this, notice that the elasticity of deforestation with respect to market access is

$$\frac{d\log L_o^F}{d\log MA_o}=\frac{(1+\epsilon)(1+\gamma)+\epsilon}{(1+\epsilon)(1+\eta+\eta\theta\alpha)+\gamma\theta\epsilon\left[1-\left(\frac{\theta\alpha}{1+\theta\alpha}\right)\left(\frac{\frac{A_o^C}{A_o^F}\left(\frac{L^C}{L^F}\right)^{\frac{\theta\alpha}{1+\theta\alpha}}}{1+\frac{A_o^C}{A_o^F}\left(\frac{L^C}{L^F}\right)^{\frac{\theta\alpha}{1+\theta\alpha}}}\right)\right]} \tag{C.5}$$

This elasticity depends not only on parameters, as in the model with mobile labor, but also on quantities such as relative productivity between types of land and land use. These quantities (as well as all parameters) would have to be estimated for counterfactuals to be conducted using the model.

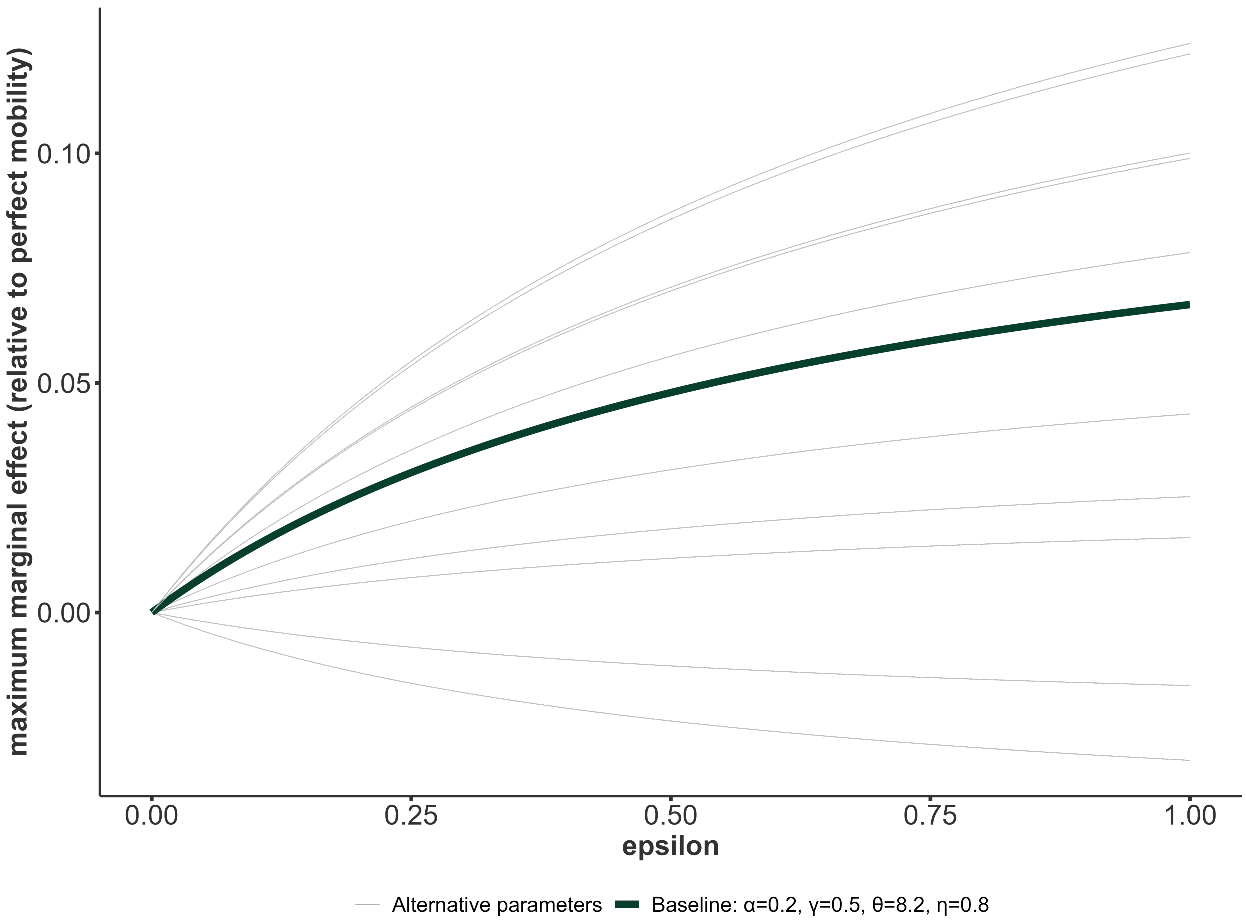
However, the expression above implies that, for reasonable values of the other model parameters, mobility frictions will not influence much the results presented in this paper. To see this, we being by noticing that the elasticity of deforestation with respect to market access is bounded above:

$$\frac{d\log L_o^F}{d\log MA_o} \leq \frac{(1+\epsilon)(1+\gamma)+\epsilon}{(1+\epsilon)(1+\eta+\eta\theta\alpha)+\gamma\theta\epsilon\left[1-\left(\frac{\theta\alpha}{1+\theta\alpha}\right)\right]} \tag{C.6}$$

We then use expression (C.6) to compute – for given values of α, γ, θ and η – an upper bound of the difference between the elasticity of the model with perfect mobility presented in the main text and the model with imperfect mobility presented here. Figure C.1 presents the results for this exercise. It plots the difference for the two models for different values of ϵ. Increasing mobility frictions within the range of parameters found in the empirical literature would increase the elasticity of deforestation with respect to market access.[3] In the baseline simulation (using α, γ, θ and η discussed in the main text), we find that the model with imperfect labor mobility would generate a elasticity of deforestation with respect of market access at most 7% higher than the model with perfect mobility. The simulations with other parameters are quantitatively similar – for most cases the elasticity of deforestation with respect to market access increases as mobility frictions increase but for a few parameters' combinations the effects decrease. However, these changes are always modest, indicating that the model with perfect mobility generates a key elasticity for conducting counterfactuals that is very similar to the one obtained in the model with imperfect mobility for all plausible combinations of parameters. Thus, despite being unrealistic, the hypothesis of perfect mobility does not seem to influence our results much while enabling us to link the model to the data in a quite transparent way.

[3] Using data from the U.S., Diamond (2016) estimates an elasticity of local employment with respect to wages of 2 for high-skill workers and 4 for low-skill workers, Monte et al. (2018) estimates an elasticity of 2.5 with substantial heterogeneity between regions, and Adao et al. (2019) estimates an elasticity of 1.3. Using data from Brazil, Morten and Oliveira (2024) estimates an elasticity of local employment with respect to wages of 6. Together, these estimates suggest that ϵ is between 0 and 1.

Figure C.1: Imperfect Labor Mobility vs. Perfect Labor Mobility



Notes: The figure depicts the maximum difference of the elasticity of deforestation with respect to market access between the model with perfect labor mobility presented in the main text and the model with imperfect labor mobility presented in the appendix. The vertical axis presents the percent difference between the two models whereas the horizontal axis presents different values of the inverse labor supply elasticity (ϵ). The green line depicts the results for the baseline simulation ($\theta = 8.2$, $\eta = 0.8$ as implied by the baseline estimates, see p. 29-30). The light grey lines depict the results for other parameter values ($\theta = 8.2$ and $\eta = 0.6$, $\theta = 8.2$, $\eta = 0.4$, $\theta = 8.2$ and $\eta = 1$, $\theta = 8.2$ and $\eta = 1.2$, $\theta = 6.5$ and $\eta = 0.67$, $\theta = 6.5$ and $\eta = 0.8$, $\theta = 6.5$ and $\eta = 0.55$, $\theta = 4$ and $\eta = 0.75$, $\theta = 4$ and $\eta = 0.90$, $\theta = 4$ and $\eta = 1.08$, $\theta = 8.2$ and $\eta = 0.45$). All simulations assume $\alpha = 0.20$ and $\gamma = 0.50$.

D Additional Results

Table D.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 D.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 D.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 D.4: Trade elasticity in the literature

Paper	Preferred θ
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 (θ) found in the economics literature.

Table D.5: Estimation Results Controlling for Local Infrastructure

	(1)	(2)	(3)	(4)	(5)	(6)
	log (Deforestation)					
log(Market Access)	0.47***	0.5***	0.47***	0.47***	0.49***	0.46***
	(0.13)	(0.13)	(0.14)	(0.13)	(0.13)	(0.13)
$R^2(within)$	0.15	0.16	0.18	0.15	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)	0.95***	0.95***	0.95***	0.95***	0.95***	0.95***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Lat-Long	Yes	Yes	Yes	Yes	Yes	Yes
Distances	No	Yes	Yes	No	Yes	Yes
Soil	No	No	Yes	No	No	Yes
Observations	1,278	1,278	1,278	1,278	1,278	1,278
F Statistic	99,956	103,394	102,706	98,707	102,782	102,116

Notes: This table reports the results for estimating Equation (9) controlling for total infrastructure inside the buffer of 400 km. All specifications include municipality and state-year fixed effects. We continue excluding observations within a buffer of radius $d = 400$km as an instrument for market access as in Table 2. Columns 1-3 controls for the cubic polynomial of total infrastructure inside the buffer and columns 4-6 controls for the total infrastructure inside the buffer interacted with year dummis. Standard errors clustered at the municipality level are reported in parenthesis. *** p<0.01; ** p<0.05; * p<0.10.

Table D.6: 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				22,078
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 D.7: 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

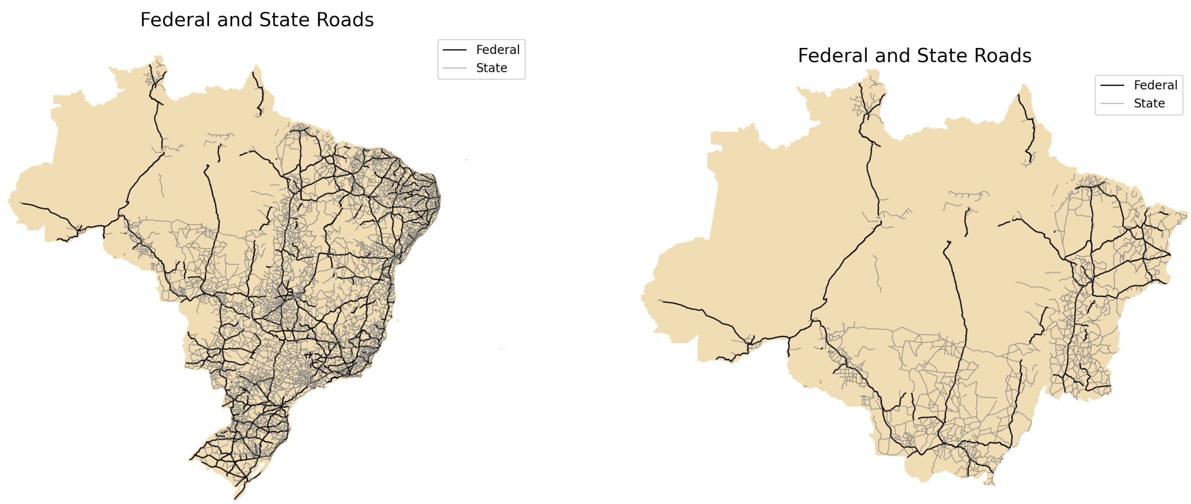
Notes: This table reports the results of estimating Equation (9) using different weighting procedures. All specifications include municipality fixed effects, state-year fixed effects, geographic variables (cubic polynomials on latitude and longitude, the distance to the coast and to *Brasília*, and suitability 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.

Table D.8: Market Access and Deforestation

	(1)	(2)	(3)	(4)	(5)	(6)
	log (Deforestation)					
log(Market Access)	0.17***	0.19***	0.18***	0.17***	0.19***	0.18***
	(0.06)	(0.06)	(0.05)	(0.06)	(0.06)	(0.06)
R^2 (within)	0.21	0.20	0.22	0.21	0.20	0.22
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				128,250	130,266	131,612
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), using as left-hand side variable the cumulative deforestation up to decade t. 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 $d = 400$km 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.

Figure D.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.

Table D.9: Market Access and Deforestation (with Cerrado)

	(1)	(2)	(3)
log(Market Access)	0.56***	0.53***	0.5***
	(0.14)	(0.14)	(0.13)
log(Market Access).Cerrado	-0.45	-0.14	-0.16
	(0.32)	(0.39)	(0.38)
R^2	0.14	0.15	0.17
Observations	1,278	1,278	1,278
Lat-Long	Yes	Yes	Yes
Distance	No	Yes	Yes
Soil	No	No	Yes

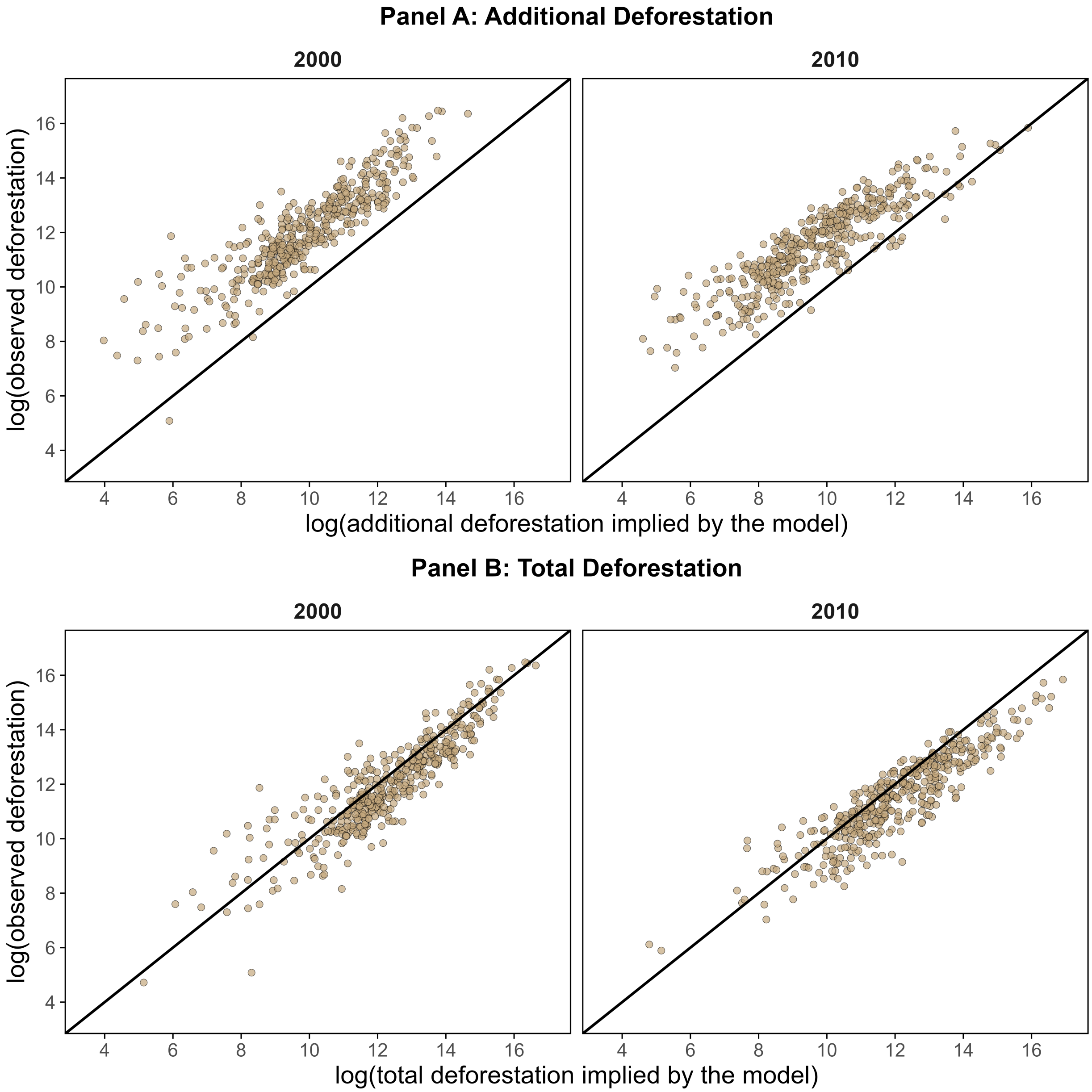
Notes: This table reports the results of estimating Equation (9) allowing for heterogeneous effect by biome. We flag all municipalities in the Legal Amazon that touches the Cerrado biome to create a dummy variable 'Cerrado' that is interacted with the market access variable. All specifications include municipality and state-year fixed effects. Columns 1 includes cubic polynomials of latitude and longitude as controls ('lat-long'). Column 2 includes distance to the coast and distance to *Brasília* as additional controls ('distance'). Column 3 includes 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 Legal Amazon during the period 1990-2019. Minimum comparable areas in the Cerrado totals 188. Standard errors clustered at the municipality level are reported in parenthesis. *** $p<0.01$; ** $p<0.05$; * $p<0.10$.

Table D.10: Market Access and Deforestation (Different Clusters)

	(1)	(2)	(3)	(4)	(5)	(6)
	log (Deforestation)					
log(Market Access)	0.45**	0.51**	0.47**	0.47**	0.52**	0.49**
Cluster municipality (# 426)	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)
Cluster mesoregion (# 30)	(0.21)	(0.20)	(0.20)	(0.21)	(0.21)	(0.21)
Cluster state-year (# 27)	(0.19)	(0.19)	(0.18)	(0.19)	(0.18)	(0.18)
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
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) for different ways of clustering the standard errors. 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 $d = 400$km 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. We assign significance levels based on the least significant result among all the cluster specifications *** p<0.01; ** p<0.05; * p<0.10.

Figure D.2: Measures of model-implied deforestation vs. observed deforestation



Notes: Panel A depicts the relationship between the (log) of additional deforestation implied by the model and the (log) observed deforestation in each decade. Additional deforestation is computed using the formula $\beta \times \Delta \log MA_{o,t+1} \times L^{F}_{o,t}$. Panel B depicts the relationship between the (log) of total deforestation implied by the model and the (log) observed deforestation in each decade. Total deforestation is computed using the formula $(1 + \beta \times \Delta \log MA_{o,t+1}) \times L^{F}_{o,t}$. Each dot represents a municipality. The dark line shows the 45° degree line.