Demand for Deforestation in the Amazon*

EDUARDO A. SOUZA-RODRIGUES†
Department of Economics, Yale University

[JOB MARKET PAPER]

November 14, 2011

Abstract

This paper estimates the demand for deforestation on private properties in the Brazilian Amazon. The estimated demand function can be used to study multiple policy interventions with the ultimate goal of preventing deforestation. To recover the demand curve, I use a revealed preference approach and exploit the fact that regional variation in transportation costs can be used to infer variation in the value of forested land relative to agricultural land. By rescaling these costs, I am able to value the difference between forested versus agricultural land in dollars per hectare. I then estimate the model using both parametric and semi-parametric quantile IV estimators. The results sugget that a perfectly enforced Pigouvian tax of US$ 100/ha/year on agricultural land would have maintained 70% coverage of the forested areas on private properties as opposed to 40% coverage observed in the data for 2006. In addition, this would have resulted in US$ 2.1 billion in revenues. Similarly, a payment for ecological services (PES) program paying private landholders at the same rate to prevent deforestation would have achieved the same levels of protection, but would cost roughly US$ 5.33 billion per year. The results also indicate that large landholders are the most responsive to PES programs, which, together with the unequal distribution of land in the Amazon, suggests that these programs are unlikely to reduce local poverty and deforestation simultaneously. Finally, a "back-of-the-envelope" calculation of the

*I am very thankful to Phil Haile, Steven Berry, Don Andrews, Xiaohong Chen, Joseph Altonji, Matthew Kotchen, William Nordhaus, Robert Mendelsohn, Kenneth Gillingham, Chris Udry, Ed Vytlacil, Peter Phillips, John Asker, Chris Knittel, and participants of Yale Workshop of Industrial Organization and Yale Environmental Economics Seminar. I am also greateful to Eustáquio Reis, Newton de Castro, Abraham Parrish and Stacey Maples of the Map Department at Yale University, Jorge Balat, Ted Rosenbaum, Boudhayan Sen, Vitor Farinha Luz, Olga Timoshenko, Pryianka Anand, Daniel Furlan Amaral, Pedro Ferraz Cruz and Renilson Rodrigues da Silva. Finally, I thank Marfisa Queiroz for her support. Financial support from Charles V. Hickox Fellowship and Yale University are gratefully acknowledged. All errors are mine responsibility.

†Yale University, Department of Economics. Email: eduardo.souzarodrigues@yale.edu. Phone: 1 (203)-768-1375.

supply of carbon stock in the Amazon based on the estimated demand function indicates that a REDD+ program fixing the price of carbon at US$1/tC/year would have increased the carbon stock from 4 billion tons of carbon in the privately owned forests to approximately 7 billion tons.

JEL Classifications: Q2, Q57, Q58, L73, L78

KEYWORDS: Amazon, Brazil, deforestation, land use, quantile, instrumental variable, semi-parametric.

1 Introduction

The Amazon is the largest intact piece of contiguous tropical rainforest. It has an unusually rich amount of biodiversity and provides extensive carbon storage and water recycling services. For these reasons, its deforestation has attracted considerable attention over the last two decades. According to satellite images, approximately 15% of the Brazilian Amazon was deforested by 2010, and the average amount of deforestation between 1991 and 2010 was 16.6 thousands square kilometers annually [INPE (2010)].

In this paper, I estimate the demand for deforestation on private properties in the Brazilian Amazon. This demand is defined as the amount of cleared area as a function of the difference between the private value of the agricultural and forested land. The estimated demand function can be used to study multiple policy interventions with the ultimate goal of preventing deforestation. Here I consider three possible policies: (a) payments for ecological services (PES), (b) Pigouvian taxes on agricultural land, and (c) quantitative limits on deforestation allowed on private properties. PES programs are incentive-based mechanisms involving direct payments to suppliers conditional on providing an environmental service [Wunder (2007) and Engel et al. (2008)]. For example, a program may pay a fixed amount of money per hectare to farmers to preserve a given fraction of their land in natural vegetation. To the best of my knowledge, no empirical study addressing these policies for the Brazilian Amazon in a unified and coherent framework currently exists.

The reasons why I study these policies are the following. First, PES programs have been seriously considered in recent years as a viable option to preserve the environment, especially when considering the payments for reduced emissions from deforestation and degradation (REDD+) agreements.[1] According to IPCC (2007), deforestation and forest degradation are responsible for 12–20% of global anthropogenic greenhouse gas (GHG) emissions in the 1990s and early 2000s. Despite these facts, PES programs have not yet been adopted in the Brazilian Amazon. Therefore, an evaluation of both the potential effectiveness and the potential costs of such programs are in order. Second, Pigouvian taxes on agricultural land have also not been adopted in the Brazilian Amazon. They should have similar impacts to PES programs, except for the fact that farmers would have to bear the costs of preserving the rainforest. Third, the Brazilian government, on the other hand, has implemented quantitative limits for land use. By law, landowners are obligated to keep 80% of their land in native forest in the Amazon. In spite of the evidence that this rule has

[1] REDD+ is a carbon credit regime under negotiation in the United Nations Framework Convention on Climate Change (UNFCCC). Under this regime, countries with high emissions can pay to protect forests in developing nations, primarily tropical countries, and count the storage of carbon in protected forests in their overall carbon output.

not been fully enforced (see the discussion about legislation and penalties in Subsection 2.2), one might wonder how costly this policy would have been if it were perfectly enforced. Because all these policies try to influence what farmers are doing with their land, this paper focus on landowners' choices on private properties. Deforestation of public land is also an important problem, but one that I must ignore here.[2]

To estimate the demand curve, I use a revealed preference approach and exploit the fact that regional variation in transportation costs can be used to infer variation in the value of forested land relative to agricultural land. By rescaling these costs using yields, I am able to value the difference between forested versus agricultural land in dollars per hectare. As a result, from the impact of a change in transportation costs on landowners' decisions, it is possible to infer by how much the value of forested land would have to be increased relative to the value of agricultural land to avoid deforestation. The strategy I propose, therefore, is divided into two steps: first, I estimate the effects of transportation costs on deforestation, and second, I rescale these costs using yields to recover the demand function.

One might wonder why one cannot use the price of land instead of transportation costs to estimate this demand. Unfortunately, data on land prices available for the Brazilian Amazon do not distinguish the price of forested land from the prices of agricultural land. As a result, variation in the observed *average* land price cannot be used to infer variation in the relative value of forested versus agricultural land. Another possibility would be to explore data on penalties for illegal deforestation. However, data on punishments are scarce. In addition, significant evidence that the legislation has not been fully enforced in the Amazon exists (see Subsection 2.2).[3]

I collected data from several sources. I combined data from the Brazilian Agricultural Census of 2006 with data on the network of roads, railroads, and navigable rivers in Brazil, freight values, and covariates, such as soil quality and agro climatic conditions. To allow for diminishing (or increasing) returns to agricultural land that may affect farmer's private valuations, I split the sample into different farm sizes and ran the analysis separately for each sub-group. Separating these groups is also important because policy-makers may be interested in differences in the opportunity costs of

[2] Affecting decisions within farmland is an important way to promote conservation, considering that these properties occupy about 18% of the Amazon, according to the Brazilian Agricultural Census of 2006. More importantly, the deforestation has been more intense in the states of the South Amazon (the states of Rondônia and Mato Grosso, see Section 2) where private properties occupy about 45% of their total area.

[3] A third approach would measure the value of alternative land uses by means of "engineering/costing models." This values are based on the revenues and cost information of different alternatives of a representative type of farm [Vosti et al. (2003)]. Although these procedures provide valuable information, they may be potentially limited in recovering the actual preferences of farmers, because there may be private benefits or costs (some possibly non-pecuniary) to alternative land uses that the researcher is unaware of.

these groups. For example, to the extent that they may view PES programs as a way to reduce poverty, they may want to adjust the payments to smallholders.

The first step of my strategy is related to a growing literature that estimates the impact of roads on deforestation.[4] When using cross-sectional data, as I do here, the typical exercise in this literature assumes a logit model for landowners' decisions to deforest; it aggregates their choices at the municipality level and runs an ordinary least squares (OLS) regression of the deforested area on municipality-level variables [Pfaff (1999)]. My procedure improves upon the typical procedure in three aspects. First, transportation costs are instrumented with straight-line distances to the main destinations, which addresses the potential endogeneity of roads and measurement errors in transportation costs. As far as I am aware, no previous study for the Amazon used an instrumental variable approach, despite the fact that the endogeneity of transportation costs is normally acknowledged as an important problem. I find that instrumental variable regressions increase the impact of transportation costs when compared to non-instrumented regressions, which suggests that an attenuation bias from measurement errors may exist.[5]

Second, I use a quantile regression instead of a mean regression, because locations with different levels of deforestation may respond differently to changes in transportation costs. Through the use of the instrumental variable quantile regression estimator (IVQR) proposed by Chernozhukov and Hansen (2008), I obtain estimates suggesting the existence of heterogeneous impacts across quantiles.

Third, I relax the functional form restrictions on the distribution of farmer's unobservable idiosyncratic shocks considering that there is no prior knowledge that justifies a particular shape for this distribution. More specifically, I drop the logit assumption and estimate a semi-parametric quantile IV model using the penalized sieve minimum distance estimator (PSMD) proposed by Chen and Pouzo (2009, 2011). Interestingly, the results establish the suitability of the logit model once there are no significant differences between the logit and the semi-parametric models.

Given the estimates in the first step, I rescale the transportation costs using yields to convert their units from *dollars per ton* of output transported into *dollars per hectare*. The rescaling exercise has two potential problems that have to be addressed. First, because there are hundreds of products being produced in the Amazon, some care is needed in defining the rescaling factor. I select the

[4] See Reis and Guzman (1992), Pfaff (1999), Pfaff et al. (2007), Chomitz and Gray (1996), Chomitz and Thomas (2003), Andersen et al. (2003), Weinhold and Reis (2008), and Igliori (2008). For a review of the literature see Kaimowitz and Angelsen (1998), Bell and Irwin (2002) and Nelson and Geoghegan (2002).

[5] The use of straight-line distances as instruments for transportation costs is similar to the approach adopted by Chomitz and Gray (1996), who studied the determinants of deforestation in Costa Rica, and Banerjee, Duflo and Qian (2009), who studied the impact of transportation costs on local GDP in China.

most representative products in the Amazon to construct an index under the assumption that all these products have the same transportation cost. Second, even if there were only a single product in the Amazon, the quantity of output sold per hectare may be affected by changes in transportation costs. By ignoring this potential effect and assuming that the productivity is independent of the transportation costs, the rescaling exercise might bias the estimates of farmers' private valuations. To handle this problem, I estimate the impact of transportation costs on this index. In the rescaling procedure, I allow the index to respond to these costs.

After rescaling transportation costs, I obtain estimates of the demand for deforestation for different farm sizes in the Brazilian Amazon. The results suggest that a perfectly enforced Pigouvian tax of US$ 100/ha/year on agricultural land would have maintained 70% coverage of the forested areas on private properties as opposed to 40% coverage observed in the data. In addition, it would have resulted in US$ 2.1 billion in revenues. Similarly, a PES program paying private landholders at the same rate to prevent deforestation would have achieved the same levels of protection, but would roughly cost US$ 5.33 billion per year. If the program were able to perfectly target the payments only to farmers who were going to deforest, the cost would be reduced to approximately US$ 2.1 billion per year. The results also indicate that large landholders are the most responsive to PES programs, which, together with the unequal distribution of land in the Amazon, suggests that these programs are unlikely to reduce local poverty and deforestation simultaneously.

In addition to these results, a "back-of-the-envelope" calculation of the supply of carbon stock in the Amazon based on the estimated demand function indicates that a REDD+ program fixing the price of carbon at US$ 1 per ton of carbon (tC) per year would have increased the carbon stock from 4 billion tons of carbon in the privately owned forests to approximately 7 billion tons. The total cost of this program would be roughly US$ 7 billion per year, and the cost per ton of *reduced emissions* of carbon would have been US$ 2.33/tC/year.

Finally, with respect to the quantitative limits in land-use, the required share of 80% of forest cover on private land specified in the Brazilian law would be so expensive for farmers if it were fully enforced that farmers would be willing to pay at least US$ 5.38 billion per year to avoid the enforcement of this rule.

The rest of this paper is organized as follows. Section 2 presents some background information about the occupation of the Brazilian Amazon, the legislation and the local economy. Section 3 provides the theoretical model that guides the empirical application, discusses the identification strategy and elaborates on the econometric procedure. Section 4 presents the data. Section 5 exposes the results of the impacts of transportation costs on deforestation. Section 6 presents the counterfactual share of agricultural land and the total demand for deforestation. Final considerations are provided in Section 7. An Appendix complements the main text with a detailed explanation of the construction of the variables used in the paper.

2 Background

2.1 Brief History of the Occupation of the Amazon

Before the 1960s, the Amazon was barely occupied. The local economy was based on subsistence and a few extraction activities (e.g., the extraction of rubber). During the 1960s and 1970s, however, the military dictatorship promoted the occupation of the region. They constructed hydroelectric facilities, mining, ports, railways, and around 60,000 km of roads [Andersen and Reis (1997)]. Moreover, they promoted land concessions, colonization and titling projects (mostly along the roads); and offered subsidized credits and fiscal incentives for investments in the region. Since this expansion, the population in the Amazon increased from 7.27 million people in 1970 to 22.75 million people in 2007. At the same time, real income per capita increased by 258%.[6]

During the 1980s, the economic recession and hyperinflation led the government to cut investments. After the 1990s, ecological concerns shaped the policies in the Amazon. IBAMA (Brazilian Environment Protection Agency) was created in 1989 to monitor and enforce environmental policies. Investments to monitor deforestation has increased since then, especially in satellite images used to detect fires and illegal deforestation. The government also created large areas of Conservation Units and Indigenous Reserves, which accounted for 44% of the Legal Amazon's (AML) area by 2010 [INPE (2010)]. In 1996, the required share of forest cover on private land increased from 50% to 80% in the Amazon and the penalties for environmental crimes also increased.[7]

2.2 Legislation and Penalties

If a farmer wants to clear a fraction of his/her land, he/she needs to hold many licenses and authorizations, including a detailed plan of management that must be approved by IBAMA. These requirements are costly, time consuming and may take several months to be approved [Hirakuri

[6] Source: http://www.ipeadata.gov.br

[7] In Brazil, there exists two concepts of the Amazon: the Amazon Biome and the Legal Amazon. The Amazon Biome extends over nine countries of South America and occupies an area of 6.4 million square kilometers. The Brazilian Amazon holds 63% of this area (4 million km^2), which corresponds to 49% of the Brazilian territory. The Legal Amazon (AML), on the other hand, is an administrative area in the northern part of Brazil that includes 9 states and around 5 million km^2 of land (about 60% of the Brazilian national territory). It consisted of 771 municipalities in 2006 and included other types of vegetation, in particular, a savannah type of vegetation called the *cerrado*.

(2003)]. Sanctions for forest-related violations include fines ranging from US$ 2,300 to US$ 23,000 per hectare, the seizure of animals, forest products and equipment, and the suspension of activities. The fines are extremely costly to farmers in view of their average (median) gross revenue per hectare, which was, according to Agricultural Census of 2006, US$ 387/ha (US$ 154/ha).

However, there is evidence that the legislation has not been fully enforced. For example, between 2005 and 2009, IBAMA applied 24,161 fines totalling about US$ 7.34 million, but the revenues collected from these fines were only 0.6% of the total value [TCU (2009)]. Moreover, Brito and Barreto (2006) analyzed a sample of 55 court cases against environmental violations in the forest sector in the Pará state between 2000 and 2003 and found that only 2% of the offenders were criminally liable. Therefore, given the apparent small expected cost of punishment, one might expect farmers to slash-and-burn to clear the land without authorization.[8]

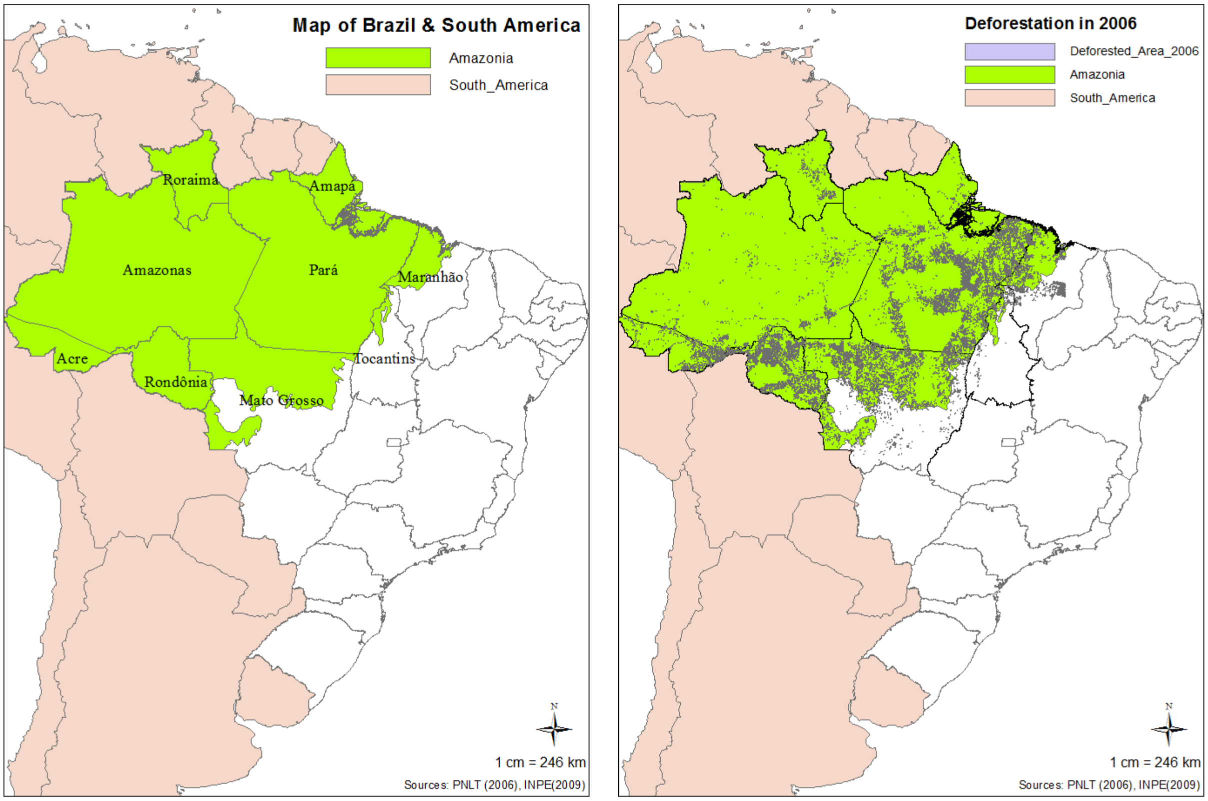
2.3 Modes of Transportation and Area Occupied

Figure 1 shows, in the left panel, the map of Brazil with the location of the Amazon rainforest, the political division and the name of the states in the Legal Amazon. The right panel in Figure 1 includes the area deforested in 2006, according to satellite images. Most of the deforested area is concentrated in the southern and eastern parts of the Amazon Biome, which is normally called the "Arc of Deforestation".

The Arc of Deforestation is characterized by an intense network of roads. Figure 2 presents the transportation network in Brazil. The top left panel presents the navigable rivers, the main ports with their names indicated in the figure and the Amazonian state capitals. The top right panel omits the navigable rivers and instead presents the railroads and the name of the Amazonian state capitals. The bottom left panel shows the location of roads distinguishing paved from unpaved roads. Finally, the bottom right panel puts together the navigable rivers, the roads and the deforested area in 2006.

Figure 1. Deforestation in 2006

[8] Interestingly, according to Brito and Barreto (2006), most of the fines (72%) were due to the illegal storage or transportation of wood. The number of fines for deforestation has increased considerably over time. Prior to 2006, the proportion of deforestation (according to the satellite images) that received fines was about 0.15% in 2003 and 7.9% in 2006. In 2009, this proportion increased to 51% (based on information provided by Pedro Ferraz Cruz, from IBAMA, in a personal message sent on April 7, 2011).

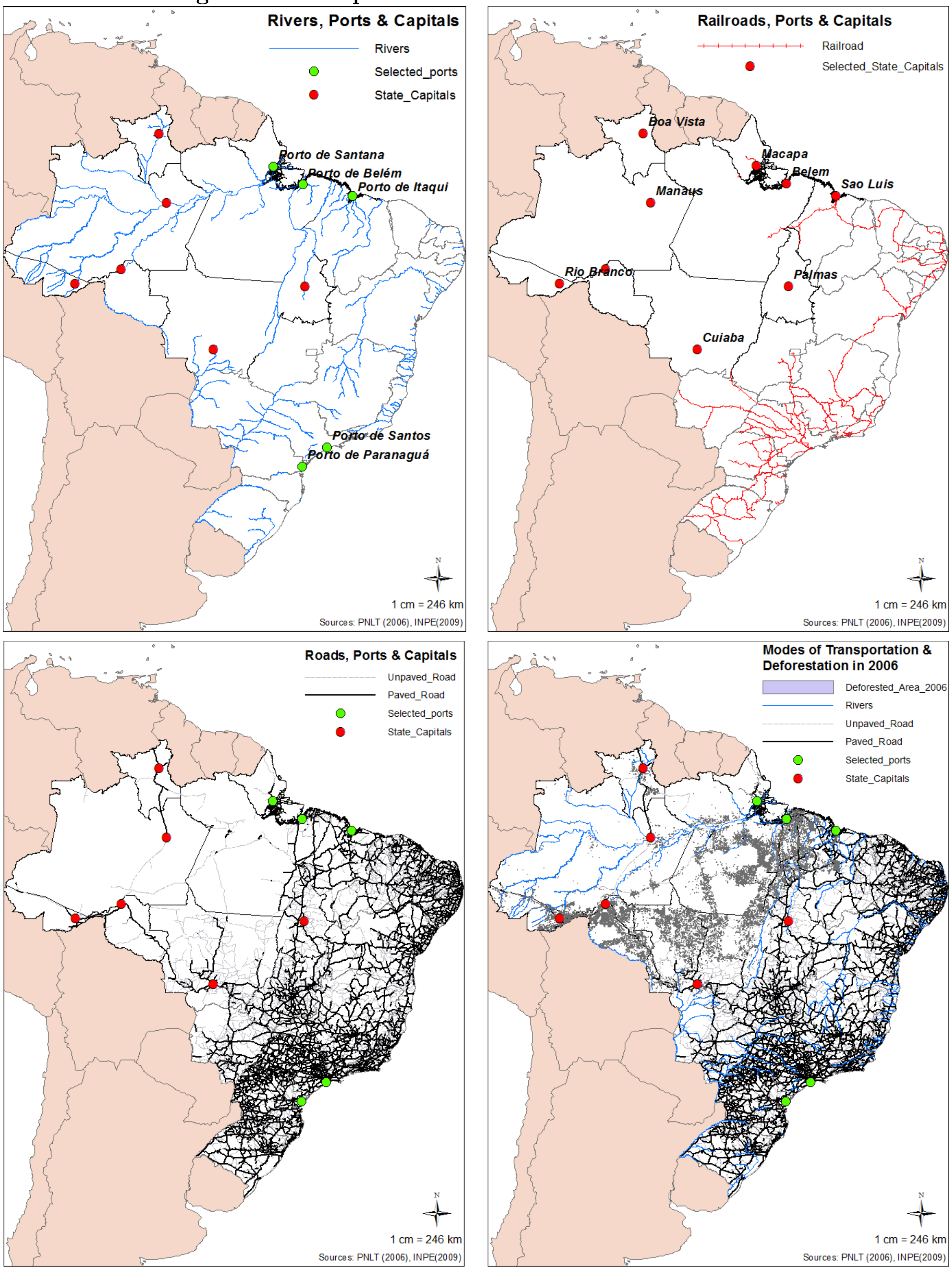


Navigable rivers have always been important to transport products in the Legal Amazon, especially in areas of dense forest cover. In some places, rivers are the only option of transportation for the local population. According to the Ministry of Transport, there are approximately 23 thousand km of rivers in the Amazon basin, of which 16 thousand are navigable. Railroads, on the other hand, are not very prevalent in the Brazilian territory. There are almost 30 thousand km of railroads in the country. These railroads are concentrated in the southeast and are mainly directed to ports. The main ports in the country are also located in the southeast; the most important ports being the Port of Santos and the Port of Paranaguá. Not only is the infrastructure of these ports better, the roads linked to them are also of better quality than in the rest of the country, making them a better option than the ports in the north for exports.

With respect to roads, it is clear from the bottom left panel that paved roads are prevalent in the south area of the country. Overall, most of the roads in the Amazon are unpaved. Indeed, according to the Ministry of Transport, there were 212,098 km of roads in 2006 in the Legal Amazon and 89% of them were unpaved. The few paved roads in the Amazon are federal and tend to connect the main state capitals. The spatial correlation between the location of roads and deforestation can easily be seen in the bottom right panel in Figure 2. Nepstad et al. (2001) and Alves (2002) documented that approximately two-thirds of the total Amazon deforestation between 1978 and

1994 occurred within 50 km of major paved highways.

Figure 2. Transportation Network and Deforestation



Area Occupied. Next, I briefly discuss the most representative goods in the Amazon to provide a better picture of the local economy. Although the Agricultural Census provides detailed information of what is being produced and where, unfortunately, there exists no information available about the exact destinations of these products for each municipality. For this reason, I also provide some information about the proportion of exports for these products.

In terms of production, the Amazon can roughly be divided into three sections. The *Eastern Amazon* (which comprises the states of Tocantins, Pará, Amapá and part of Maranhão, Figure 1) has an economy based on mining, logging, the extraction of açaí (a typical Amazonian fruit), and agriculture based on cattle, rice, and, to a lesser extent, soybeans. The *Western Amazon* (which comprises the states of Amazonas, Acre and Roraima) is based on the extraction of rubber, timber, açaí, and on the mining industry. The agriculture is concentrated in the production of manioc and, to a lesser extent, cattle. Finally, the third part is the *South Amazon* (comprising the states of Rondônia and Mato Grosso). Its economy is strongly based on the production of grains, especially soybeans, corn and rice.

Private farmland occupies about 18% of the Amazon, but this proportion varies depending on the region: it occupies 45% of the South Amazon; 19% of the Eastern Amazon; and 4.5% of the Western Amazon. Most of the private farmland is used for pasture: about 49% according to the 2006 Agricultural Census. Most of the cattle is used to produce beef and a good fraction of the production in the states in the South Amazon are exported, but the other states do not export a significative portion of their production.[9]

The area occupied by crops is a small fraction of the total: 10% of private land in 2006. Its participation, however, has increased lately, especially in the Arc of Deforestation. In terms of area occupied in 2006, soybeans are the most important product (it occupies about 22% of the crop area), followed by corn (11%), manioc (11%), rice (8.4%) and beans (4%).[10] While soybeans and corn are located mostly in the South of the Amazon and are directed to international markets, manioc, rice and beans are consumed domestically, with manioc being more concentrated in pristine areas, possibly for subsistence.[11]

[9] Brazil has the largest number of cattle in the world (about 35% of which were in the Amazon in 2006), is the second largest producer of beef and is the largest beef exporter in the world (USDA, www.fas.usda.gov/psdonline/). Among the states in the Amazon that export beef, Rondônia exported about 25% and Mato Grosso exported about 17% of their production in 2007 [Schlesinger (2009) and Amigos da Terra (2009)].

[10] Source: http://www.sidra.ibge.gov.br/bda/agric

[11] About 42% of the soybeans produced in the Brazilian territory are exported directly (and mostly to China). The other fraction is sold domestically to the crushing industry that produces soybean oil and meal. The oil is used domestically either for biodiesel (14%) or for human consumption. The soybean meal is normally sold to the animal food industry (47%) and exported [Amaral (2010)].

Finally, forests occupy about 37% of the private land. Among the extraction of forest products, the most important in terms of the value of production in the Amazon in 2006 was açaí (41%), timber (39%), Brazilian nuts (5%), hearts of palm (8%), and rubber (1%).[12] Açaí is primarily produced for domestic markets. The logging industry, is located along the Arc of Deforestation, drew 14.2 million cubic meters of timber in 2009 (equivalent to about 3.5 million trees), and directed 36% of its production (after processing) to international markets. The timber consumed domestically is mostly used in the building sector [Pereira et al. (2010)].

3 Model and Estimation

Next, I present a simple stylized model to guide the empirical application. Before presenting the model, a couple of remarks are in order. First, deforestation is defined as the share of agricultural land on private properties. I assume the land was originally forested, so that clearing it for agriculture is equivalent to deforesting. The remaining area can be used for managed forest. The focus is on private properties, because the policies that motivate the present paper try to influence what farmers are doing with their land. Furthermore, the opportunity costs of clearing a plot of private land is presumably different from the cost of clearing public land. Although these policies may affect the total private area, I carry out the analysis conditional on the total farm size. As will be clear in Section 4, I use data from the Brazilian Agricultural Census of 2006 because it provides information on land use within agricultural establishments.[13]

Second, to allow for diminishing (or increasing) returns to agricultural land that may affect farmer's valuations, I split the sample into different farm sizes and run the analysis separately for each sub-group. Third, the available data is aggregated at the municipality level. For this reason, it is not possible to distinguish between a model in which farmers choose the share of agricultural land and a model in which there is a continuum of farmers making discrete choices between clearing their plot of land or not. The typical exercise in the literature that estimates the impact of roads on deforestation assumes a binary choice model for landowners' decisions and aggregates their choices at the municipality level [Pfaff (1999)]. I follow this literature to make my procedure comparable to the existing papers and because the binary choice model is extremely convenient. For any given farm size, one can think of farmers making a sequence of independent choices for each parcel of land based on the difference between the value of agricultural and forested land.

[12] Source: http://www.sidra.ibge.gov.br/bda/agric. Unfortunately I do not have information about the area occupied by these products.

[13] Another option is to use satellite data. However, this data cannot not distinguish between deforestation on private and public land.

Finally, the model I present is static because I have access to cross-sectional data in 2006. The previous Agricultural Census is for 1996, and, so, even if the farmer had deforested his/her plot of land in 1996, forest regrowth may be sufficiently fast so that farmers may have to face the same decision in 2006. Furthermore, if transportation costs tend to decrease over time, the pressure to deforest may have increased over the years, so that if a farmer had cleared his/her land in the past, he/she would probably clear it in 2006 if he/she were to decide in that year.

Next, I proceed with the details of the model.

3.1 Model

Take a parcel of land i located at municipality m and that belongs to a farm of size s. Assume there is a continuum of such parcels, and for each one, the farmer is deciding whether or not to clear the land for agriculture. Let P_{ims} be a vector with output and input farmgate prices and Z_{ims} be the vector of productivity shocks. Define $\Pi^a(P_{ims}, Z_{ims})$ as the expected discounted present value of future profits obtained by using the parcel for agriculture, including the conversion costs, and $\Pi^f(P_{ims}, Z_{ims})$ as the corresponding value obtained leaving the plot as managed forest. Let Y_{ims} equal one if the plot i is cleared and zero otherwise. Then:

$$Y_{ims} = 1\left\{\Pi^a(P_{ims}, Z_{ims}) > \Pi^f(P_{ims}, Z_{ims})\right\},$$

where $1\{.\}$ is the indicator function.

I assume both output and input markets are competitive and all production is sold in nearby markets or exported directly. A no-arbitrage condition implies that local prices are determined by the international price minus the transportation costs to the nearest port. The output local price for agriculture P^a_{ims} is

$$P^a_{ims} = \overline{P}^a - \left(TC_m + \varepsilon^t_{ims}\right)$$

where $\overline{P}^a$ is the international price of the output. The output transportation cost to the nearest port is decomposed into TC_m and ε^t_{ims}. The cost to transport the output from the municipal seat to the nearest port is denoted by TC_m; a proxy for this variable is observed in the data. The deviation of the farm's transportation cost to TC_m is denoted by ε^t_{ims}, is unobserved by the econometrician (but observed by the farmer) and may reflect (i) different locations of parcels within the municipality, and (ii) different quality of the transportation modes available in the municipality. I adopt similar specification for all output and input local prices and assume that all products have the same transportation costs.

The productivity shock is assumed to be the vector $Z_{ims} = (Z_m, U_m(s), \varepsilon^z_{ims})$, where Z_m is a municipality-level vector of observed productivity shifters, such as soil quality and other agro climatic conditions; $U_m(s)$ is a municipality-level unmeasured/unobserved productivity shock; and ε^z_{ims} captures the farmer's unobserved abilities and deviations from the municipality-level variables. Because the empirical analysis is conditionally done on the farm size, it is possible to allow the unobservable $U_m(s)$ to be indexed by the farm size. Interestingly, indexing $U_m(s)$ by farm size permits a richer model than the usual municipality fixed-effect model, because there may be systematic unobservable differences across municipalities in, say, access to good soils and productive technologies that depend on the size of the farm. In particular, there is no restriction in how $U_m(s)$ correlates with $U_m(s')$, for different farm sizes $s \neq s'$. Therefore, while one municipality may be good for agriculture for large farms, it may not be as good for agriculture for smallholders.

Denote the vector of the municipality-level observables by $X_m = (TC_m, Z_m)$. The existing literature typically imposes a single-index structure on the difference between Π^a and Π^f and collapses all individual heterogeneity into a single scalar ε_{ims}. In the present case, these assumptions reduce the model to:

$$\begin{aligned} Y_{ims} &= 1\left\{\Pi^a(P_{ims}, Z_{ims}) - \Pi^f(P_{ims}, Z_{ims}) > 0\right\} \\ &= 1\left\{X'_m\beta_s + U_m(s) - \varepsilon_{ims} > 0\right\}. \end{aligned}$$

In addition, an extreme value distribution for ε_{ims} is typically imposed, which implies the logit model

$$\begin{aligned} Y_m(s) &= \Pr\left(X'_m\beta_s + U_m(s) > \varepsilon_{ims}\right) \\ &= \frac{\exp\left(X'_m\beta_s + U_m(s)\right)}{\left[1 + \exp\left(X'_m\beta_s + U_m(s)\right)\right]}, \end{aligned}$$

where $Y_m(s)$ is defined as the share of agricultural land within farms of size s in municipality m. This logit model can be easily estimated after taking the differences of log shares as:

$$\log\left(\frac{Y_m(s)}{1 - Y_m(s)}\right) = X'_m\beta_s + U_m(s). \tag{1}$$

The typical exercise estimates equation (1) using OLS [Pfaff (1999)]. My procedure improves upon the typical exercise in three aspects. First, transportation costs are instrumented with straight-line distances to the main destinations, D_m, which addresses the potential endogeneity of roads and measurement errors in transportation costs. Second, I use a quantile regression instead of a mean regression, because locations with different levels of deforestation may respond differently to changes in transportation costs. Third, I relax functional form restrictions by dropping the logit assumption to check whether this restriction may drive the results. In the next set of paragraphs, I expose the more flexible model I adopt. However, I leave a discussion of the reasons why transportation costs to the nearest port should be instrumented and under what conditions straight-line distances to the main destinations are expected to be valid instruments for Subsection 3.2.

To allow transportation costs to impact the entire distribution of deforestation, and not only the mean, I turn to a quantile model. The functional form now allows the coefficients to vary with the quantiles of the conditional distribution of $Y_m(s)$. Therefore, instead of estimating the equation (1), I estimate:

$$\log\left(\frac{Y_m(s)}{1-Y_m(s)}\right) = X_m'\beta\left(U_m(s)\right). \tag{2}$$

where $U_m(s)$ is assumed to have a uniform distribution on $[0,1]$ given the instruments. The function $u \mapsto X_m'\beta_s(u)$ is assumed to be strictly increasing and continuous in u. Note that the single-index structure in the quantile regression is not as restrictive as it may appear at first, considering that the coefficients can depend arbitrarily on both the farm size s and the quantile u. This flexibility relaxes the role of the logit structure in determining the shape of the demand for deforestation. I estimate equation (2) using the instrumental variable quantile regression estimator (IVQR) proposed by Chernozhukov and Hansen (2008).

Next, I drop the logit assumption and estimate, for each farm size s and for each quantile $u \in (0,1)$, the transformation model:

$$G_s\left(Y_m(s), u\right) = Z_m'\beta_{su} - TC_m, \tag{3}$$

where the link function $G_s(., u)$ is unknown. Because a normalization for the single-index is required for this semiparametric model, the coefficient on transportation cost is normalized to be minus one and the constant, to be zero. The semiparametric quantile IV model (SPQIV) given in equation (3) is estimated using the penalized sieve minimum distance (PSMD) estimator proposed by Chen and Pouzo (2009, 2011).

Note that $G_s(., u)$ is the inverse of the conditional distribution of ε_{ims}. Denote this distribution by $F_s = G_s^{-1}$. Then, the share of agricultural land in municipality m for farms of size s is given by:

$$Y_m(s) = F_s\left(Z_m'\beta_{su} - TC_m, u\right). \tag{4}$$

where the distribution of unobservable idiosyncratic shocks, ε_{ims}, is allowed to depend on both the farm size and on the municipality fixed effect.

An even more general model can be obtained by dropping the single-index restriction and estimating:

$$Y_m(s) = F_s(X_m, TC_m, U_m(s)). \tag{5}$$

This model (5) is a special case of the generalized regression model with group effects proposed by Berry and Haile (2009). Conditional on the farm size s, it corresponds to the nonparametric quantile IV model developed by Chernozukhov and Hansen (2005). In principle, I could nonparametrically estimate (5) using a two-step nonparametric estimator developed in a companion paper [Souza-Rodrigues (2011)]. Unfortunately, the curse of dimensionality and data limitations prevent me from running a completely nonparametric estimator.[14]

Rescaling Transportation Costs. After estimating equations (2) using IVQR and (3) using PSMD, I rescale the transportation costs to identify the demand for deforestation. In the following paragraphs, I focus the discussion on the SPQIV model given in (3), but the same reasoning can be applied to the logit model in (2).

Because the coefficient of TC_m is normalized to be minus one, the difference in the value of agricultural and forested land is measured in US\$ per ton of output transported. If farms of size s in municipality m sell $q_m(s)$ tons of output per hectare, then the effect of raising the value of the forested area by US\$ t/ha is equivalent to changing the transportation costs from the actual level, TC_m, to a cost $\overline{TC}_m$ that satisfies the equation:

$$\overline{TC}_m - TC_m = \frac{t}{q_s(s)}. \tag{6}$$

Denote by $Y_m^t(s)$ the share of agricultural land when the relative value of forested land increases by US\$ t/ha. The counterfactual fraction of agricultural land is, therefore, given by:

$$Y_m^t(s) = F_s\left(Z_m'\beta_{su} - TC_m - \frac{t}{q_m(s)}, u\right). \tag{7}$$

[14] The estimation procedure proposed in Souza-Rodrigues (2011) requires access to micro-data on landowners' decisions. The procedure is to estimate the expected fraction of agricultural land, $Y_m(s)$, for each municipality, in the first step taking into account the presence of the common shocks affecting all farms in the municipality, i.e., $(X_m, U_m(.))$. The fact that $U_m(.)$ is indexed by s complicates the proofs of the asymptotic results, because $U_m(.)$ has to be treated as a random function. For each municipality, I show that the estimated $\widehat{Y}_m(s)$ converges in probability to the random variable $Y_m(s)$, despite the general nature of the common shocks. In the second step, conditional on farm size s, the procedure runs a NPQIV regression of the predicted $\widehat{Y}_m(s)$ on X_m across municipalities to separate the effects of X_m and $U_m(s)$. In practice, the second step requires a penalized sieve minimum distance estimator for each s that takes into account the preliminary estimator $\widehat{Y}_m(s)$. The preliminary estimator breaks a Lipschitz condition exploited by Chen and Pouzo (2011), which also complicates the proof of the asymptotic results. This procedure requires both a large number of municipalities and a large number of farmers in each municipality to obtain consistency. As might be expected, the larger the number of farmers in each municipality, the faster the rate of convergence.

As mentioned in the Introduction, the rescaling exercise has two potential problems. First, because there are hundreds of products being produced in the Amazon, some form of aggregation is required in defining $q_m(s)$. I select the most representative products discussed in Subsection 2.3. The crucial assumption here is that all these products have the same transportation cost. Second, even if there were only a single product in the Amazon, the quantity of output sold per hectare may be affected by changes in transportation costs. For example, an increase in transportation costs may reduce both the area utilized for agriculture and the quantity produced. To the extent that farmers reduce the use of the worse area first, the quantity produced may not decrease by as much as the area occupied, and, so, the quantity of the output per hectare may increase. By ignoring this potential effect, the estimated counterfactual share of agricultural land would be smaller than the true counterfactual share.

To address this problem, define the function:

$$q_m(s) = q_s(Z_m, TC_m, \varepsilon^q_{ms}),$$

where ε^q_{ms} are unobservable factors affecting local productivity. In this case, the impact of raising the value of forested land by US\$ t/ha is equivalent to changing transportation costs from the actual level, TC_m, to a cost $\overline{TC}_m$ that satisfies the equation:

$$\overline{TC}_m - TC_m = \frac{t}{q_s\left(Z_m, \overline{TC}_m, \varepsilon^q_{ms}\right)}. \tag{8}$$

I estimate $q_s(Z_m, TC_m, \varepsilon^q_{ms})$ using the same set of regressors and instrumental variables as before. Then, the corresponding $\overline{TC}_m$ satisfying (8) is computed for each municipality, for each farm size and for each value t. Finally, in order to compute $Y^t_m(s)$, the term $\left[\frac{t}{q_m(s)}\right]$ in equation (7) is replaced by $\left[\frac{t}{q_s\left(X_m, \overline{TC}_m, \varepsilon^q_{ms}\right)}\right]$. Therefore, the counterfactual share of agricultural land is given by:

$$Y^t_m(s) = F_s\left(Z'_m\beta_{su} - TC_m - \frac{t}{q_s\left(Z_m, \overline{TC}_m, \varepsilon^q_{ms}\right)}, u\right). \tag{9}$$

3.2 Identification Strategy

There are several reasons why one needs to instrument transportation costs. First, roads may be built in response to profitable situations. Unobservable (to the econometrician) soil quality in a given location, for example, may have induced both deforestation and the presence of roads to access this location. Second, previously deforested regions may have a higher demand for improvements in local infrastructure conditions, including better/more roads, which leads to reverse causality in cross-sectional data. In the presence of either omitted variables or simultaneity, the OLS regression of deforestation on transportation costs is expected to overstate the impact of these costs.

In addition, transportation costs are likely to be measured with an error. A common proxy for these costs is the most cost-effective route to valuable markets, which is computed based on the existing network of roads and (sometimes) the freight rate data. This proxy may not be an accurate measurement of the real costs that farmers incur and, so, is potentially mismeasured. Different from the previous cases, the classical measurement error may induce an attenuation bias in the OLS estimates.[15]

In this paper, the proxy for transportation costs is defined as the minimum unit cost (US$/ton) to transport one ton of goods to the nearest port using the most cost-effective route. This proxy is instrumented with straight-line distances to the nearest port and to the nearest state capital. In the following paragraphs, I discuss (i) why one should expect straight-line distances to be strong instruments, and (ii) under what conditions one should expect these instruments to satisfy an exclusion restriction condition.

First of all, it is evident that distances to the nearest port should correlate with the costs to the ports. Furthermore, to the extent that state capitals are connected with better transportation infrastructure, a location close to a state capital should have smaller costs (ceteris paribus) to reach the ports. Therefore, the distance to the nearest capital should also be positively correlated with transportation costs.[16]

Second, because farmers' decisions to deforest depend on productivity factors and on farmgate output and input prices, straight-line distances should not influence their choice once these factors are taken into account. I control for differences in productivity using a measurement of soil quality and agro climatic conditions, such as rain, temperature and altitude, as discussed in Section 4. Variation in local prices is explained by variation in transportation costs to the nearest port, at least for tradable goods.

The instruments may be invalid if there are outputs and/or inputs whose prices are not fixed in the international market. In this case, local market conditions may affect local prices and correlate

[15] Another proxy normally used is the extent of roads per municipality. However, the extent of roads cannot capture the improvements of roads outside the county and fails to distinguish between the roads connected to valuable markets and the roads that run in circles. The most cost-effective route is therefore a preferable proxy.

[16] Another way to look at this problem is to follow the discussion presented by Chomitz and Gray (1996). Because locations of major towns (in the present case, ports and state capitals) were determined by geography and historical reasons long before the expansion of the roads in the 1970s, I could construct an exogenous network of roads by linking the major centers with straight-lines. The distances computed using this exogenous network should be correlated with transportation costs, because the location of the towns creates links between the major centers, but not the precise routing. By noting that using this virtual network to compute distances to main destinations and computing straight-line distances directly to the main destinations provides the same information, I opted for the simpler solution.

with straight-line distances to the main destinations. An important example is local labor markets. For instance, wages may have to increase as the municipalities locate further away from the nearest capital, all else being constant, to compensate workers for working away from desired places. In this case, municipalities further away from the capital may deforest less than a location close to the capital because of wage differences. If these wage differences are not controlled for in the regression and correlate with the instruments, then the instruments are invalid. A similar problem may occur if there are other non-tradable inputs as well as non-tradable outputs.

To minimize this problem, I include in the regressions factors that shift local demand and supply for non-tradable outputs and inputs that may be correlated with straight-line distances. I included the local population, the presence of power plants (mainly hydroelectric facilities) and local mining. While the local population shifts the supply of labor and increases the demand for non-tradables, power plants and mining shifts both the demand for labor and non-tradables.[17]

In case the instruments are invalid even after controlling for these factors, point identification is lost, but partial identification is still possible. For example, the monotone instrumental variable approach of Manski and Pepper (2000) can be used to partially identify the parameters of interest. It would be interesting to obtain results under weaker assumptions, but I leave this extension for a future work.

3.3 Implementation of the PSMD Estimator

To estimate the transformation model:

$$G_s\left(Y_m\left(s\right), u\right) = Z_m' \beta_{su} - TC_m, \tag{10}$$

I use the PSMD estimator proposed by Chen and Pouzo (2009, 2011). I assume (D_m, Z_m) is independent of $U_m\left(s\right)$ for any s, where $D_m = \left(D_m^p, D_m^c\right)$ denotes the straight-line distances to the nearest port and to the nearest capital. For each farm size s and quantile $u \in (0, 1)$, the moment restriction implied by the model is:

$$E\left[\rho_u\left(Y_m\left(s\right), X_m; G_{su}, \beta_{su}\right) \mid D_m, Z_m\right] = 0 \tag{11}$$

where the residual function is:

$$\rho_u\left(Y_m\left(s\right), X_m; G_{su}, \beta_{su}\right) \equiv 1\left\{G\left(Y_m\left(s\right), u\right) \leq Z_m' \beta_{su} - TC_m\right\} - u, \tag{12}$$

[17] Indeed, when regressing the distance to ports and the distance to capitals on these three factors and on other covariates, I find that: (i) the local population predicts the distance to capitals (they are negatively correlated, as expected), but the other factors have insignificant coefficients; and (ii) the presence of power plants predicts the distance to ports (they are positively correlated), but other factors are not significant. More detailed results are available upon request.

and the conditional moment function is:

$$m\left(D_m, Z_m; G_{su}, \beta_{su}\right) \equiv E\left[\rho_u\left(Y_m\left(s\right), X_m; G_{su}, \beta_{su}\right) \mid D_m, Z_m\right]. \tag{13}$$

I approximate G_{su} using an artificial neural network (ANN) sieve approximation, because this non-linear sieve is often, in practice, better able than alternatives to allow for non-linearities in the unknown function [Chen (2007)]. More specifically, I use a sigmoid ANN defined by:

$$sANN\left(k_m\right) = \left\{\sum_{j=1}^{k_m} \alpha_j S\left(\gamma_j Y_m + \gamma_{0,j}\right) : \alpha_j, \gamma_j, \gamma_{0,j} \in \mathbb{R}\right\} \tag{14}$$

where $S : \mathbb{R} \to \mathbb{R}$ is a sigmoid activation function.

For each s and u, the function G_{su} in $sANN\left(k_m\right)$ and the finite dimensional parameter β_{su} are chosen to minimize the criterion function:

$$Q\left(G_{su}, \beta_{su}\right) = \left\{\frac{1}{M}\sum_{m=1}^{M} \widehat{m}\left(D_m, Z_m; G_{su}, \beta_{su}\right)' \widehat{m}\left(D_m, Z_m; G_{su}, \beta_{su}\right) + \lambda_m \widehat{M}\left(G_{su}, \beta_{su}\right)\right\}$$

where M is the number of municipalities; $\widehat{m}\left(.\right)$ is an estimator for $m\left(.\right)$; the penalization parameter, $\lambda_m \geq 0$, converges to zero as $M \to \infty$; and $\widehat{M}\left(G_{su}, \beta_{su}\right)$ is the penalization function.

I take $\widehat{m}\left(.\right)$ to be a series least square estimator of $E\left[\rho_u\left(Y_m\left(s\right), X_m; G_{su}, \beta_{su}\right) \mid D_m, Z_m\right]$. Let $\left\{p_1\left(D_m, Z_m\right), p_2(D_m, Z_m), ...\right\}$ be a sequence of known basis functions that approximate any square integrable real-valued function. Denote $p^{J_M}\left(D_m, Z_m\right) = \left(p_1\left(D_m, Z_m\right), ..., p_{J_M}\left(D_m, Z_m\right)\right)'$ a $(1 \times J_M)$-vector and $P = \left(p^{J_M}\left(d_1, z_1\right), ..., p^{J_M}\left(d_M, z_M\right)\right)'$ an $(M \times J_M)$-matrix. The series LS estimator is given by:

$$\widehat{m}\left(d, z; G_{su}, \beta_{su}\right) = p^{J_M}(d, z)'(P'P)^{-} \sum_{m=1}^{M} p^{J_M}(D_m, Z_m)\rho_u\left(Y_m\left(s\right), X_m; G_{su}, \beta_{su}\right)$$

where $(P'P)^{-}$ is the pseudo-inverse matrix of $P'P$. The penalization function used is:

$$\widehat{M}\left(G_{su}, \beta_{su}\right) = \left\|\nabla_y^2 G_{su}\right\|_{L^2\left(\widehat{f}_{Y(s)}\right)}$$

where ∇_y^2 denotes the second derivative with respect to y and $\|.\|_{L^2\left(\widehat{f}_{Y(s)}\right)}$ denotes the L^2-norm with the empirical measure $\widehat{f}_{Y(s)}$ of $Y_m\left(s\right)$.

4 Data

Next, I describe the set of variables used in this paper. Then, I present some summary statistics.

4.1 Dependent Variable: Deforestation

The Brazilian Agricultural Census of 2006, produced by the IBGE (Instituto Brasileiro de Geografia e Estatística) is the richer dataset available of the agricultural sector.[18] It provides information on land use for different farm sizes; quantity, value and area occupied of major agricultural outputs; herd size; value of land and of other main assets, among other information. The unit of analysis is the "agricultural establishment", be it a household or firm producing any animal or plant output.

The land-use in the Agricultural Census is divided into several categories which were aggregated in two: agricultural and forested land. Agricultural land includes pasture and crops, while forested land aggregates managed forests and forests that are not currently being exploited.[19] The groups of farm sizes considered in this paper are: (i) small farms (those with less than 5 hectares); (ii) small to medium farms (those with an area between 5 and 50 hectares); (iii) medium to large farms (those with an area between 50 and 500 hectares); and (iv) large farms (those with more than 500 hectares).[20]

4.2 Endogenous Regressor: Transportation Costs

The proxy for transportation costs is defined as the minimum unit cost (US$/ton) to transport one ton of goods to the nearest port. It requires definitions about (i) which products are going to be considered; (ii) which ports are included in the calculations, and (iii) how to combine the freight rate data available with the network of roads, railroads and navigable rivers.

The Agricultural Census provides detailed information of what is being produced and where. As discussed in Subsection 2.3, the main products in the Amazon are (i) soybeans, corn, manioc, rice and beans, among crops, (ii) cattle for production of beef and (iii) açaí and timber, among the extraction of forest products. When directed to international markets, these products normally use either the ports in the Amazon (Port of Santana, Port of Belém and Port of Itaqui) or the ports

[18] Available at http://www.ibge.gov.br/home/estatistica/economia/agropecuaria/censoagro/default.shtm and http://www.sidra.ibge.gov.br/bda/agric

[19] Private land-use in the census is divided into the following categories: annual cropland, perennial cropland, pasture (planted and natural), short and long-term fallow, forest (planted and natural), ponds and lakes, constructions, degraded land and unusable land (for economic activities). I define agricultural land as the sum of annual or perennial cropland, pasture, short and long-term fallow, constructions and degraded land. Forested area is the sum of the remaining land uses: natural and planted forest, ponds and lakes and unusable land.

[20] Although the estimation procedure could treat the farm size as a continuous variable, the online information available on the 2006 Agricultural Census provides land-use data separately for 18 classes of farm sizes. For each class, the land-use data is aggregated at the municipality level. For confidentiality reasons, the land-use data is missing whenever the municipality has only one or two farmers in a given class. Yet, by aggregating different classes of farm sizes so that the final aggregated class has more than two farmers in the municipality, it is possible to recover the total land-use information. I therefore aggregated the 18 classes into 4 in trying to achieve a good balance between: (i) not losing too much information from the missing data, (ii) obtaining more or less homogeneous classes and (iii) ease of exposing results.

in the South (Port of Santos and Port of Paranaguá), see Figure 2 in Subsection 2.3. Therefore, I selected these five ports to be the main destinations for the proxy for the transportation costs.

Given the selected destinations, I use the network of modes of transportation produced by the Ministry of Transport for the National Highway Plan (Plano Nacional de Viação).[21] The least cost path to the nearest port is computed in ArcGIS. The calculation divides the entire country into cells corresponding to 1 km^2 and requires the cost to travel over it for each cell. The travel cost depends on if the cell contains a segment of road (paved or unpaved), railroad, navigable river or if it does not contain any transportation mode. The optimization routine in ArcGIS determines the least accumulative cost path to the nearest destination for each cell in the grid. I extract these total costs to the corresponding municipal seats located in those cells. The total costs from the municipal seat to the nearest port is the proxy for transportation costs used in this paper.

Evidently, the procedure described above relies on the rules used to assign costs to the cells in the grid. I used the Vehicle Cost Module of the World Bank's Highway Design Model (HDM-VOC-4) together with the freight rate data collected by SIFRECA (Sistema de Informações de Fretes) to define the unit cost per km to travel over the cells with different modes of transportation. Whenever the SIFRECA's data was incomplete or limited, I complemented it by calling the companies directly and asking them for their freight costs. Because almost all the information I obtained from SIFRECA about the freight values in the Amazon corresponded to costs of transporting soybeans, I assume that all products have the same transportation costs as soybeans, which seems to be correct, at least for bulk products and sacks [Castro (2003)].[22]

Table 1 summarizes the cost weights used to compute the least accumulative cost in ArcGIS. The first column discriminates between the possible modes of transportation considered in the calculations; the second column reports the unit costs used in Brazilian currency (R$), and the third column converts these unit costs into US$.[23] Note that differences in unit costs are in the expected direction, as waterways with good infrastructure are the cheapest mode of transportation, followed by railroads and paved roads. Unpaved roads inside the Amazon are the worst mode of transportation followed by navigable rivers with poor infrastructure and unpaved roads outside the Amazon. Finally, because ArcGIS allows for travelling by land, I imposed high costs to transport goods by land with no mode of transportation so that ArcGIS would avoid computing travelling

[21] Available at http://www.transportes.gov.br/index/conteudo/id/36604 and accessed on 11/24/2010.

[22] SIFRECA provides average road, railroad and waterway freight values for a variety of agricultural products, routes and periods collected from companies in the transportation industry. More details are available at http://sifreca.esalq.usp.br

[23] The average exchange rate in 2006 was US$ 1 = R$ 2.17, where BRL refers to the *real*, the Brazilian currency.

costs using these cells. For brevity, I relegate a detailed description explaining how I obtained these numbers to the Appendix.

TABLE 1. Cost Weights

Cost Weight	R$/ton.km	US$/ton.km
Paved Road	0.07678	0.0353
Unpaved Road - Outside Rain Forest	0.0992	0.0457
Unpaved Road - Within Rain Forest	0.15	0.069
Navigable River - Good Infrastructure	0.0444	0.0204
Navigable River - Poor Infrastructure	0.1139	0.0525
Railroad	0.0608	0.028
Land - Outside Rain Forest	1.5	0.6912
Land - Inside Rain Forest	3	1.3824

4.3 Other Variables

Next, I briefly describe the set of covariates, instrumental variables, and the productivity index.[24]

Covariates. *Temperature and Precipitation.* The Climate Research Unit (CRU) computed the average temperature and precipitation based on the average climate from 1961-1990. Given the high levels of temperature and rain in the region, I expect these variables to be negatively correlated with deforestation. According to Chomitz and Thomas (2003) and the references cited in their work, high levels of rainfall make agriculture unattractive because, among other reasons, cattle are more susceptible to parasites and insect pests; crops are more subject to rotting; yields are depressed by light-limiting cloud cover; mechanization is difficult; and forest burning is incomplete.[25]

Altitude. This data comes from the IBGE. Although high altitudes may be good for the production of rice, I do not have a priori information of whether this variable should be positively or negatively correlated with deforestation.[26]

[24] Some covariates were available at the municipal level for years other than 2006. Because municipal boundaries changed over time, I had to match the previous boundaries to convert the values of these covariates to the boundaries observed in 2006. I summed or averaged the values over the municipalities in 2006, taking a weighted average where appropriate.

[25] The data is available at http://www.ipeadata.gov.br/

[26] The data is available at http://www.ipeadata.gov.br/

Soil quality. This data is produced by IBGE and EMBRAPA (Empresa Brasileira de Pesquisa Agropecuária) and was kindly made available by Professor Eustáquio Reis. It consists of the proportion of the municipal area on each of five aptitude classes of soils. The classes of soil aptitude are: high, medium-to-high, low-to-medium, low, and unsuitable. Factors that could lead to a low ranking include high metal content, poor drainage, high flood risk, uneven ground, low nutrients, and steep slope [Anderson and Reis (2007)].[27]

Local Population. This data comes from the Demographic Census produced by IBGE. In principle, the local population may increase local demand for non-tradables and decrease local wages. Land-use on private land data, however, is affected by the local population depending upon whether agriculture is more or less labor intensive than extractive activities.[28]

Local Mining and Power Plants. This data is available from the National Highway Plan. These variables should shift both the demand for labor and non-tradables.

Instrumental Variables. I computed straight-line distances from the municipal seats to the nearest port and to the nearest capital using ArcGIS.[29]

Productivity Index. From the Agricultural Census, I obtained the quantity sold and area occupied of major agricultural outputs for each municipality for farms of different sizes. I calculated two productivity indices: the first considers the production of crops and the second includes crops and pasture.

The productivity index for crops is based on the main products discussed in Subsection 2.3: soybeans, corn, manioc, rice and beans. For each municipality and for each farm size, I calculated the quantity of output sold per hectare for each product. The productivity index is the weighted average of the output per hectare for these crops, where the weights are the proportion of the area utilized for each crop. By including pasture in this weighted average, I obtain my second measure of productivity. For these indices to make sense in the rescaling exercise, it is crucial to assume that all products have the same transportation costs to the port. More details are presented in Appendix.[30]

[27] The five aptitude classes were aggregated from the 13 soil types that exist in the Brazilian territory. Using the 13 soil types in conjunction with the data on local topography, data from ground surveys, and general familiarity with the land, EMBRAPA soil scientists created the digital map of soil aptitude for agriculture that is used here.

[28] Available at
http://www.ibge.gov.br/home/estatistica/populacao/default_censo_2000.shtm

[29] I excluded Palmas, the capital of the Tocantins state, from the destinations for the straight-line distances because it is a planned city that was built in 1989 with the objective of helping to develop the region.

[30] One may be concerned with the units of the productivity index that averages the quantities of the different products. Access to micro-data on landowners' decisions would avoid this problem, because I could use the yields

4.4 Summary Statistics

There are 528 municipalities in the dataset. I select the 771 municipalities in the Legal Amazon, but excluded those located in the region covered by the *cerrado* (the savanna vegetation). All municipalities with a positive fraction of their area within the Amazon Biome were included in the sample. Table 2 presents some summary statistics. For completeness, I present some information about deforestation based on the satellite data. The average proportion of deforestation across the municipalities from this data is 41%. According to the Agricultural Census data, farms occupy, on average, about 39% of the municipal area. The average number of farms in a county is 1209 (the median is 902) and the fraction of private land used for agriculture is 65% on average. The municipalities are 868 km, on average, (in a straight-line distance) away from the nearest port. The average cost to transport one ton of soybeans is US$ 41, according to my calculations. The average distance to the nearest capital is 316 km.

TABLE 2. Summary Statistics

Statistics	Mean	Stdv	Min	Max
# of Municipalities	528	-	-	-
# of Farms	1222	1165	37	11544
Prop. Deforested - Satellite	41%	31%	0	100%
Prop. Deforested - Census	65%	20%	12%	100%
Share of Farms	39%	27%	0%	98%
Cost to Port (US$/ton)	41.5	33.5	0	163
Distance to Port (km)	868.6	717	0	2627
Distance to Capital (km)	316.2	219.3	0	900

Table 3 provides information about the differences across classes of farm sizes. The numbers in the cells are sample averages across municipalities. It is clear that the land distribution in the Amazon is highly unequal. Despite the fact that large farms (over 500 hectares of land) are a small proportion of the total number of farms (5.4%), they occupy about 50% of the private farmland; while small farms (less than 5 hectares) account for 21% of the farms and occupy only 1% of the private land. Smallholders tend to deforest a larger proportion of their land and concentrate their production more on crops. As farm size increases, the proportion of deforestation decreases and the

of the main product reported by each farmer. Unfortunately, this is not feasible with the aggregated data. A conservative approach would only use the product with the highest yield, since the higher the yield, the more valuable the agricultural land is and, consequently, the higher the value of the forested land has to be to avoid deforestation. Conversely, one might use the least productive good to compute the counterfactual share of agricultural land. The choice of the weighted average for the productivity index is one way to obtain a representative picture of the local production.

production tends to shift from cropland to pastureland. The productivity index for crops suggests increasing returns to agricultural land as farm size increase from less than 5 hectares up to 500 hectares and decreasing returns for large farms. Once I include pasture in the productivity index, the index drops considerably, as might be expected.[31]

TABLE 3. Summary Statistics by Farm sizes (Sample Averages)

Statistics	≤ 5 has	5-50 has	50-500 has	≥ 500 has
# of Farms	302	413.5	353	46.3
Proportion of Farms	21%	33%	31%	5.4%
Share of Area	1.1%	11.6%	38%	50%
Prop. Deforested	90%	71%	69%	62%
Prop. of Crop Area	55%	21%	17%	11%
Prop. of Pasture	28%	48%	50.5%	50.8%
$q^{c}(s)$ - Crops (ton/ha)	1.03	1.04	1.16	0.86
$q^{cp}(s)$ - Past. (ton/ha)	0.68	0.38	0.38	0.25

Finally, Table 4 reports summary statistics for the other regressors. With respect to the agro-climatic conditions, there is little variation in temperature, a high average level of precipitation per year and most of the soil being of poor quality or unsuitable for agriculture. The average number of exploited mineral deposits is 0.8, although some places can have up to 22 deposits. Only 4% of the municipalities in the Amazon have local power plants (most of them are hydroelectric facilities). On average, municipalities have 31 thousand people, but this number varies from places with as few as one thousand to places with one million (the state capital Belém is the municipality with the largest population). There are 27 people per km^2, on average, but the median density of inhabitants per km^2 is 5.9, which provides an idea of how sparsely distributed the population in the Amazon is.

TABLE 4. Summary Statistics - Exogenous Regressors

[31] Although not presented in this table, small farms (less than 5 hectares) tend to be concentrated in perennial crops, mainly manioc, and are primarily located in the Western Amazon. Small to medium (from 5 to 50 hectares) and medium to large farms (from 50 to 500 hectares) have higher fractions of their land in corn, rice and beans and large farms (greater than 500 hectares) concentrate more on corn and soybeans. Large farms are primarily located in the South Amazon, while medium sized farms are more frequently located in the East Amazon and in the Central regions.

Statistics	Mean	Stdv	Min	Max
Temperature (°C)	26.5	0.56	25.1	27.5
Rain (mm/year)	183.9	32.2	111	272.3
Altitude (meters)	116.6	125.9	0	920
Prop. of Soil - Good	0.05	0.17	0	0.99
Prop. of Soil - Good/Medium	0.08	0.22	0	1
Prop. of Soil - Medium	0.04	0.18	0	0.99
Prop. of Soil - Medium/Low	0.51	0.38	0	1
Prop. of Soil - Unsuitable	0.31	0.35	0	1
Number of Mineral Deposits	0.81	2.08	0	22
Local Power Plants	0.04	0.19	0	1
Population (thousands)	31.3	96.4	1.35	1405
Dens. Hab. (pop per km^2)	27.3	147	0.09	2617

5 Effects of Transportation Cost on Share of Agricultural Land

This Section presents the estimated impact of transportation costs on deforestation. It begins by reporting the first stage regression to check for the presence of weak instruments, followed by the results of the logit models and the semiparametric quantile IV model. Finally, it presents the estimates of the productivity index regressions.

5.1 First Stage Regression

Table 5 exposes the results of regressing transportation costs to the nearest port on straight-line distances. For brevity, I omitted the estimated coefficients of the other covariates in this table. It is clear that both straight-line distances to ports and to the nearest capital are strong predictors of costs to ports and that there is no problem with weak instruments in this dataset.

Table 5. First Stage Regression

	(1)
	Cost Port
Distance to Port	**0.0871*****
	(31.35)
Distance to Capital	**0.0273*****
	(5.67)
Observations	**528**
F-statistic	**726.4**
$\mathbf{R^2}$	**0.915**

***t* statistics in parentheses**

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5.2 Results for Logit Models

Next, I present results for the logit models. Table 6 reports the coefficients for costs to the nearest port and the associated t-statistics in parenthesis for the quantile regression (QR), instrumental variable quantile regression (IVQR), OLS and 2SLS for each farm size. For brevity, the coefficients of the other regressors are omitted in this table, but they are reported in Table 8, when I compare the results for the logit models and the SPQIV.

Table 6. Results for Logit Models by Farm Size:[32]

(Coefficients for Costs to Ports)

	Quantiles						
	10	25	50	75	90	OLS	2SLS
Small							
QR	0.0038*	0.0025	0.0028	0.0026	0.0105	0.0030	-
	(2.1999)	(1.314)	(1.342)	(0.948)	(1.472)	(0.945)	-
IVQR	0.0029	0.0018	-0.0002	-0.0013	-0.0027	-	-0.0029
	(1.665)	(0.889)	(-0.068)	(-0.443)	(-0.165)	-	(-0.841)
Small-Medium							
QR	-0.0000	-0.0008	-0.0013	-0.0012	0.0014	0.0001	-
	(-0.037)	(-0.637)	(-1.025)	(-0.691)	(0.492)	(0.035)	-
IVQR	-0.0005	-0.0017	-0.0034*	-0.0050**	-0.0008	-	-0.0034
	(-0.425)	(-1.340)	(-2.506)	(-3.360)	(-0.264)	-	(-1.593)
Medium-Large							
QR	-0.0048**	-0.0036**	-0.0033*	0.0003	0.0040	0.0002	-
	(-3.401)	(-2.510)	(-2.151)	(0.206)	(1.596)	(0.116)	-
IVQR	-0.0069**	-0.0054**	-0.0051**	-0.0038*	-0.0010	-	-0.0041*
	(-4.752)	(-3.698)	(-3.280)	(-2.099)	(0.385)	-	(-2.244)
Large							
QR	-0.0062**	-0.0042**	-0.0051**	-0.0040*	-0.0039	-0.0018	-
	(-3.960)	(-2.643)	(-3.297)	(-2.093)	(-1.712)	(-1.008)	-
IVQR	-0.0062**	-0.0059**	-0.0063**	-0.0065**	-0.0068**	-	-0.0042
	(-3.515)	(-3.668)	(-4.068)	(-3.591)	(-3.124)	-	(-1.910)

t statistics in parentheses

$^{*}p < 0.05$, $^{**}p < 0.01$

I begin the discussion by comparing OLS and 2SLS estimates. Remember that the typical exercise would run an OLS regression on the covariates. In the present dataset this would predict positive impacts of costs to ports on the share of agricultural land for small and medium sized

[32] The number of observations (municipalities) for small farms is 505; for small to medium farms is 528; for medium to large, 526; and, finally, for large farms, 461. For all farm sizes, Hansen's J test of overidentification accepts the null of the validity of the instruments.

farms. When transportation costs are instrumented with straight-line distances, the signs of the coefficients become negative for all farm sizes. Furthermore, OLS coefficients tend to be small in magnitude and not significantly different from zero, while 2SLS coefficients tend to be greater in magnitude (except for smallholders). Interestingly, an omitted variable bias story would suggest that OLS estimates should be upwardly biased in absolute values. The fact that these estimates are closer to zero than the instrumental variable estimates suggests that an attenuation bias from the measurement errors in transportation costs may exist.

Next, I discuss the results for quantile regressions. I begin focusing on small farms (those with less than 5 hectares). All coefficients estimated by IVQR are not significantly different from zero. This seems reasonable, because small farms tend to be concentrated in isolated regions in the Western Amazon and tend to produce manioc, which is consumed domestically and does not require a significant amount of inputs. They are most likely producing for subsistence and not engaged in the market. The shadow value of food must be driving their decision to deforest and not the costs to the nearest port. As a result, the model does not seem to be well suited for them and, so, my strategy most likely fails to identify their demand for deforestation from the variation in transportation costs. Despite these problems, because smallholders occupy only 1% of the private land, their demand for deforestation does not play a major role in environmental policies. Consequently, I proceed by discussing the results for medium and large farms. However, it would be interesting to investigate smallholders' decisions further.

For small to medium sized farms (those with an area between 5 and 50 hectares), IVQR estimates, even though not always significant, tend to be greater in absolute values than the QR estimates, suggesting again that an attenuation bias may be in force. Interestingly, the IVQR coefficients are not stable across quantiles. This pattern suggests that, even after controlling for observable municipality-level variables, farms with different levels of deforested area respond differently to changes in transportation costs to the nearest port.

For medium to large farms (those with an area between 50 and 500 hectares), once more IVQR estimates tend to be greater in absolute value than the QR estimates. Different from small to medium farms, though, the IVQR coefficients decrease in absolute value with the quantiles. Finally, for large farms (with more than 500 hectares), IVQR estimates are all negative, significant, greater in magnitude than the QR coefficients, and fairly stable across the quantiles. An interesting pattern that emerges from the results is that, for each quantile, the estimated coefficients tend to increase in absolute value as the farm size increases.

To have a better sense of magnitude of the estimates, Table 7 reports the implied elasticities of the share of agricultural land to the costs to ports from the IVQR estimates. It is clear that as the farm size increases, the average elasticity increases in absolute value.[33]

Table 7. Elasticities of Transportation Costs to Ports by Farm Size

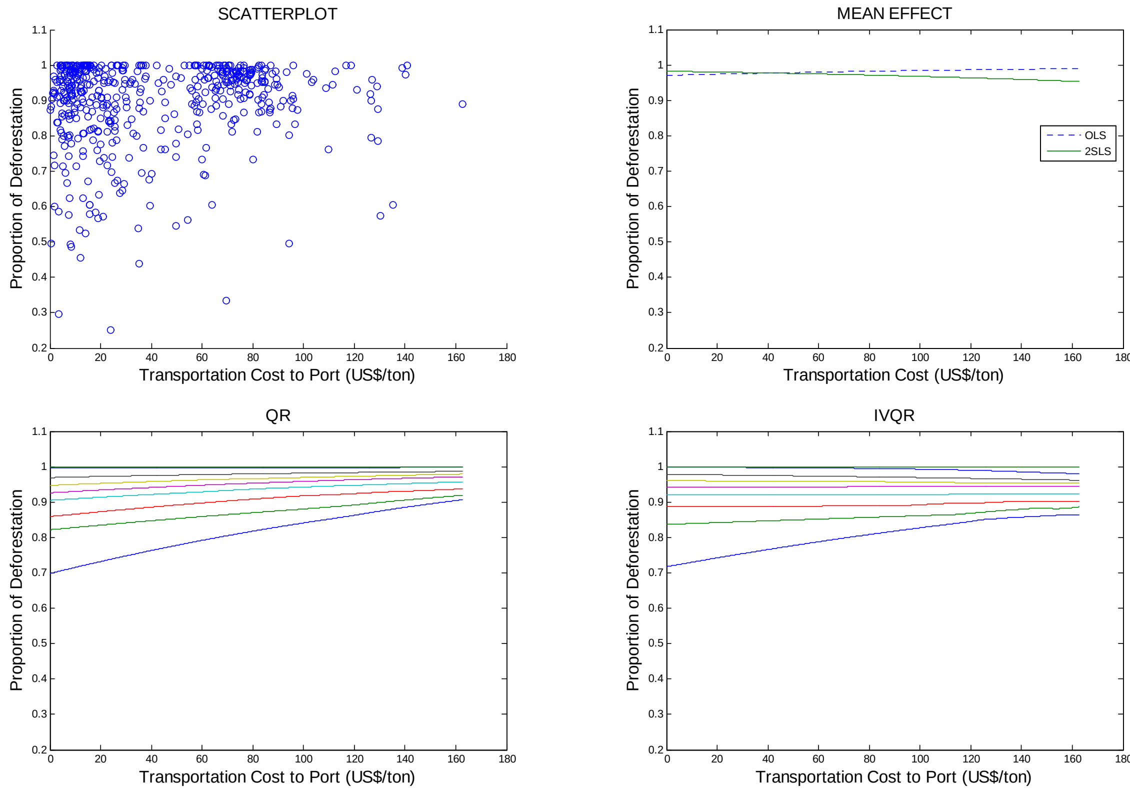
Elasticities	Mean	Stdv	Min	Max
Small Farms	0.009	0.043	-0.15	0.46
Small-Medium Farms	-0.062	0.079	-0.53	0.08
Medium-Large Farms	-0.151	0.205	-1.56	0.03
Large Farms	-0.234	0.264	-1.90	0.00

The differences between QR and IVQR estimates, as well as the heterogeneity in responses across quantiles for transportation costs to the nearest ports, are illustrated graphically in Figures 4, 5, 6, and 7. These figures correspond to, respectively, small, small to medium, medium to large and large farms. For each figure, the top left panel presents a scatter plot with the observations; the top right panel presents the OLS and 2SLS results; the bottom left panel reports the QR results; and the bottom right panel reports the IVQR results. The regressors, other than costs to the nearest port, are fixed at the sample average and costs to ports vary over the observed range in the data. I present curves for the quantiles $u = 0.1, 0.2, ..., 0.9$.[34]

Figure 4. Small Farms - Effect of Costs to Ports on the Share of Agriculture

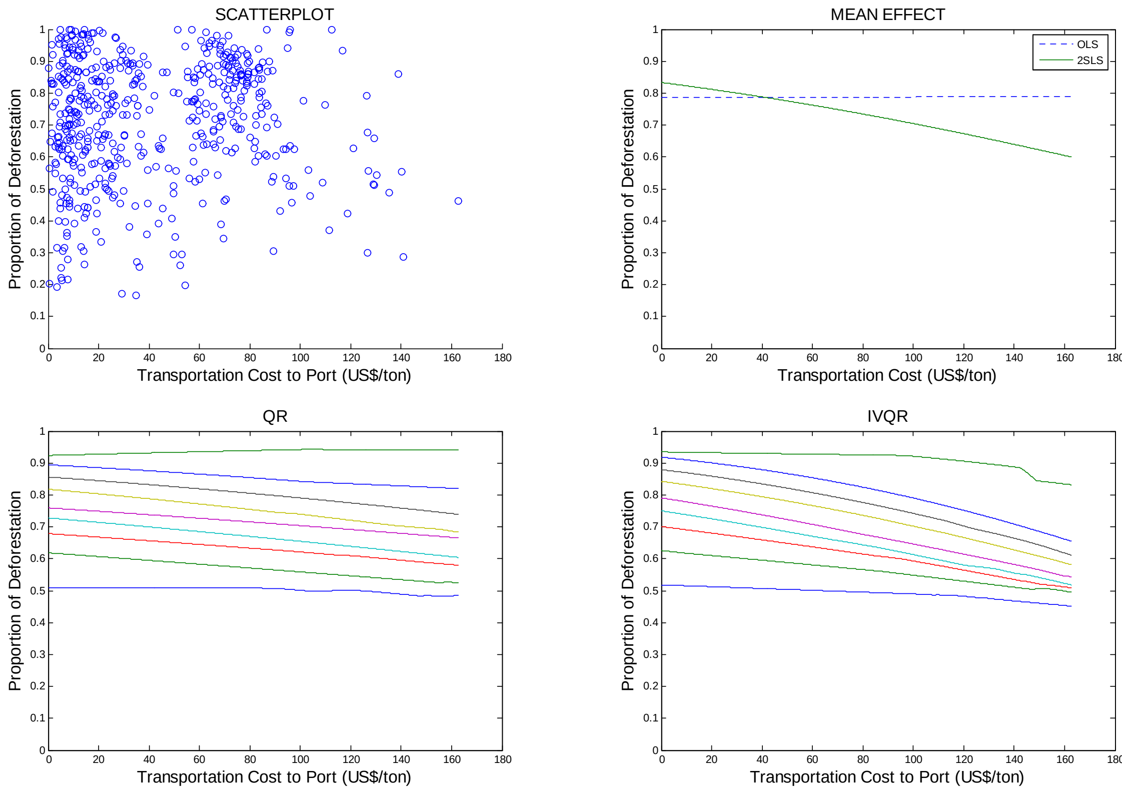
[33] To compute the elasticities, I first rearrange the quantiles for each observation in the data following the procedure proposed by Chernozhukov et al. (2010) to avoid quantile crossing. The model is estimated for quantiles ranging on $\{0.01, 0.02, ..., 0.99\}$. Then, using the observed level of deforestation for each municipality and the rearranged estimates, I recover the associated quantile for any given observation by inverting the equation (2). It is possible, therefore, to compute the elasticity for each point in the dataset assuming a rank invariance property holds.

[34] To compute these figures, I rearranged the quantiles for each evaluating point following the procedure proposed in Chernozhukov et al. (2010).



In Figure 4, it is clear that smallholders tend to use an extremely large fraction of their land for agriculture. The insensibility of their land-use decision to changes in transportation costs to the nearest port is also clear. All estimated curves are quite flat and even increasing for the lower tail quantiles.

Figure 5. Small to Medium Farms - Effect of Costs to Ports on the Share of Agriculture



Figures 5 and 6 present the results for medium sized farms. In both cases, land-use functions estimated by the OLS and QR are flatter (if not increasing) than those estimated by the 2SLS and IVQR, respectively. Moreover, according to the IVQR estimates, the upper tail quantiles do not respond significantly to changes in costs to the nearest port, which suggests that municipalities at those quantiles are so good for agriculture in terms of unobservables that transportation costs would have to increase considerably to start reducing the share of agricultural land. These curves tend to be concave, while the lower tail quantile curves tend to be convex. Similar conclusions hold for large farms, presented in Figure 7, but their sensibility to costs to ports are greater for any quantile.

Figure 6. Medium to Large Farms - Effect of Costs to Ports on the Share of Agriculture

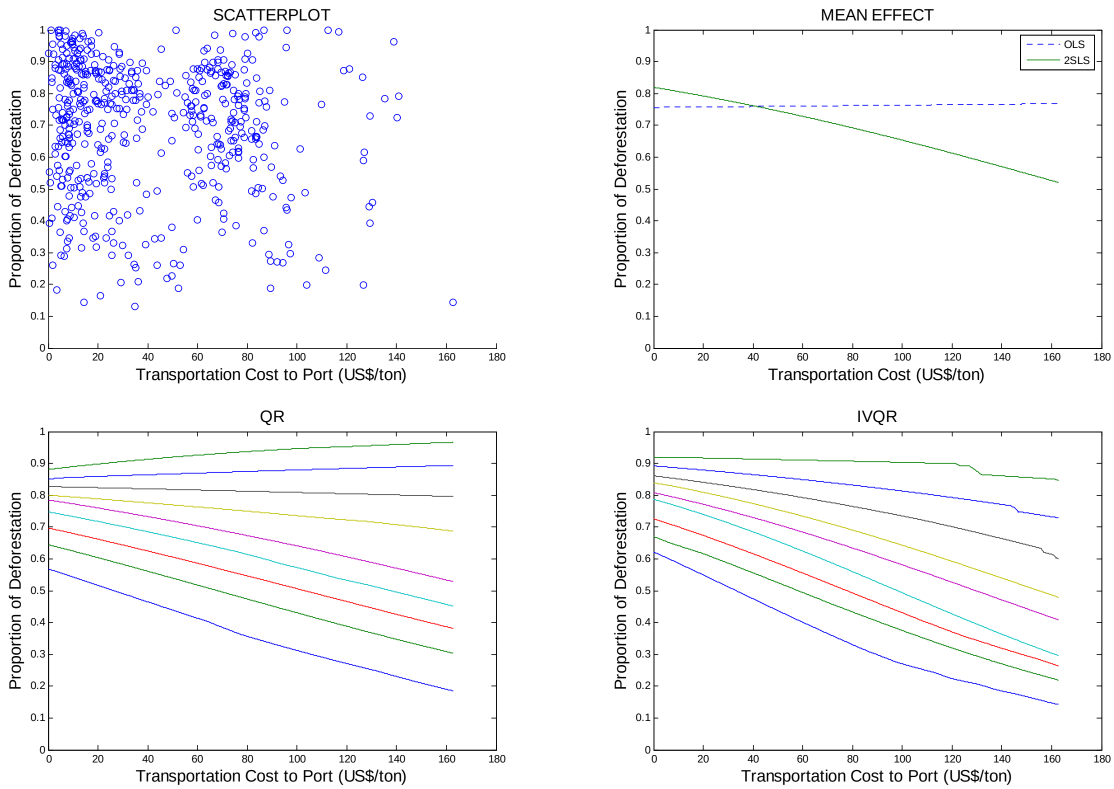
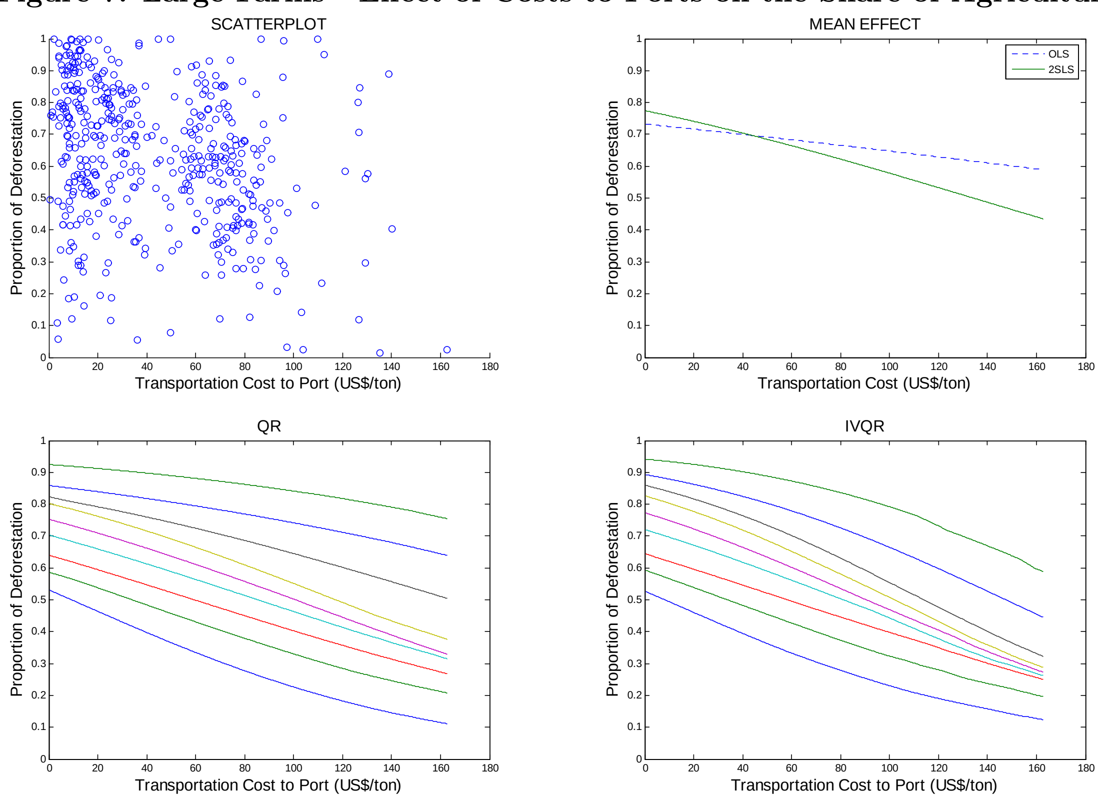


Figure 7. Large Farms - Effect of Costs to Ports on the Share of Agriculture



In short, the findings are: (i) smallholders do not respond significantly to changes in costs to the nearest port; (ii) effects of transportation costs tend to increase with farm sizes; (iii) non-instrumented regressions tend to underestimate the effects of costs to ports when compared to the instrumented estimates; and (iv) there is heterogeneity in responses across the quantiles, depending on the class of farm size.

A natural question to ask is how much of these results are driven by the logit assumption imposed on the regressions. To investigate this restriction, I now turn to the semi-parametric quantile IV model.

5.3 Results for the SPQIV Model

In this subsection, I relax the logit assumption and estimate the transformation model:

$$G_s\left(Y_m\left(s\right), u\right) = Z_m'\beta_{su} - TC_m,$$

using the PSMD estimator proposed by Chen and Pouzo (2009, 2011). I use the sigmoid ANN defined in (14) and I opted for a Gaussian activation function, i.e., I take $S\left(.\right)$ to be a normal cdf. With respect to the penalization parameter, I used $\lambda_m = 10^{-6}$.

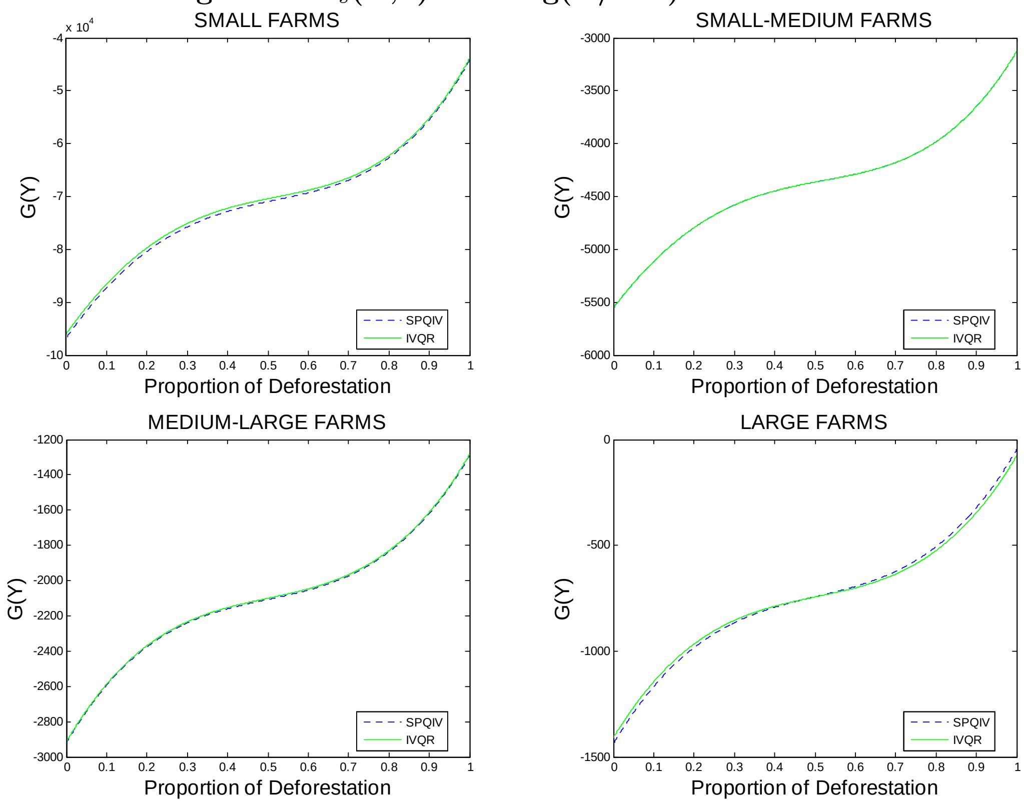
Because the asymptotic theory provides guidance to the rate at which k_m must increase with the data, but not the specific value for k_m, I choose k_m so that the number of coefficients estimated does not exceed the number of moment restrictions. The number of finite dimensional parameters is $\dim\left(\beta\right) = 10$ and the number of parameters of $sANN\left(k_m\right)$ is $3k_m$. So, by choosing $k_m = 3$, the dimension of the parameter vector becomes 19. To approximate the conditional moment $m\left(D_m, Z_m\right)$, I used a basis function $p^{J_M}\left(D_m, Z_m\right)$ with dimension $J_M = 21$.[35]

Although the single-index transformation model reduces the curse of dimensionality to estimate the link function G_{su}, it does not avoid the curse of dimensionality in approximating the function $m\left(D_m, Z_m\right)$. For this reason, I had to be very parsimonious in choosing the number of terms for $p^{J_M}\left(D_m, Z_m\right)$. I opted to use P-Splines(2,4) for both D_m^p and D_m^c, i.e., quadratic splines with 4 knots (chosen at the respective 0.01, 0.25, 0.75 and 0.99 quantiles of D_m^p and D_m^c), and an interaction term $D_m^p \times D_m^c$. The remaining terms, Z_m, entered linearly in $p^{J_M}\left(D_m, Z_m\right)$, yielding the dimension $J_M = 21$. When I experimented by approximating G_{su} with higher dimensions, say, $k_m = 4$, the dimension of the splines for $D_m = \left(D_m^p, D_m^c\right)$ were also increased to cubic splines, i.e., P-Splines(3,4). Including many more higher order terms for any of the instruments in $p^{J_M}\left(D_m, Z_m\right)$

[35] Chen and Pouzo (2011) provide very general results for the rates of convergence for NPQIV models. However, to date, there exists no explicit rate of convergence in terms of the number of observations in the data for nonlinear sieves such as the sANN used in the present paper.

made the estimation routine more difficult, because the matrix $(P'P)$ becomes singular very quickly. Note that to restrict the function $m(D_m, Z_m)$, it would be necessary to specify the conditional distribution of the endogenous variables $(Y_m(s), TC_m)$ given the exogenous (D_m, Z_m). Such a specification is not imposed by the model presented in Section 3.

Figure 8. $\widehat{\mathrm{G}}_s$(Y,u) and Log(Y/1-Y) at the Median



I only present the results for the parsimonious case $k_m = 3$, because I obtained very similar results by setting $k_m = 4$. Figure 8 compares the estimates of the approximating function for G_{su} using $sANN(3)$ and the logistic function for the median. The top left panel presents the results for small farms (less than 5 hectares); the top right panel, for small to medium farms (5-50 hectares); the bottom left panel reports estimates for medium to large farms (50-500 hectares); and the bottom right panel, for large farms (more than 500 hectares). To make the comparisons fair, I rescaled the logistic function from the IVQR estimates to make the coefficient of TC_m equal -1 and the constant term in the single-index equal zero.

The $sANN(k_m)$ approximations for G_{su} are surprisingly similar to the rescaled logistic functions (and similar results hold for almost all the quantiles). The only case where the results differ, and that is not presented in the figure, corresponds to the lower quantiles for smallholders. Interestingly, these correspond to cases where the IVQR estimates for the coefficients of transportation costs are positive.[36]

Not only is the link function similar for the SPQIV and the IVQR estimates, the estimated finite dimensional coefficients, β, are also similar. Table 8 compares the estimated coefficients using the PSMD and the normalized IVQR estimates. I only illustrate the results for the median regression, but this pattern is similar for other quantiles. It is clear how close the estimates are to each other. Although I do not run any formal tests between the IVQR and SPQIV estimates, the results illustrated in Figure 8 and Table 8 suggest that the logit assumption is well suited for this data.[37]

For completeness, I briefly discuss the coefficients of the other covariates. First, as expected, higher levels of temperature and precipitation in the Amazon are worse for agriculture and, so, reduce the share of agricultural land. Coefficients for altitude are small in magnitude and not significant. The proportion of good soils is omitted in the regressions and soils with a worse quality induce less deforestation when compared to good soils, as expected, although these effects are not monotonic in the rank of soil quality. Mining and the presence of power plants are negatively correlated with the share of agricultural land, although they are not significant in most cases. Finally, the local population does not have significant coefficients, except for small to medium farms. Interestingly, omitting the population in these regressions does not affect the coefficients of the other regressors significantly, even for small to medium farms.

Medium sized and large farms are likely to have production functions that are intensive in land and capital, but not in labor. As presented in Section 4, the share of pasture (which does not require a large number of workers) is larger for them than for smaller farms. Furthermore, the population in the Amazon is sparsely distributed and, so, labor is likely a scarce factor in the region. Therefore, the share of wages on costs is probably small, which may help explain why factors that

[36] Although not presented in the figure, when I approximate G_{su} using P-Splines or Hermite polynomials instead of $sANN\,(k_m)$, the estimated functions are close to the rescaled logistic function, but only when Y is around 0.5. The approximations in these cases are too flat and fail to capture any curvature of G_{su} when Y is close to either 0 or 1. As a result, P-Splines or Hermite approximations might predict the share of agricultural land outside the unit interval. Despite the fact that the instruments are strong in a linear sense, as shown in Subsection 5.1, and able to identify β, they may be weak in a "non-linear" sense, i.e., may fail to identify curvatures of G_{su} in a finite dataset. Using the $sANN\,(k_m)$ approximation helps impose more restrictions on G_{su} in the data.

[37] In the estimation routine for the SPQIV model, I used the estimates from IVQR as an initial guess for β_{su} and the best $sANN\,(k_m)$ approximation for the rescaled logistic function as an initial guess for G_{su} (where the rescaling used IVQR estimates). Therefore, as the minimization routine proceeds, it indicates whether the estimates of the IVQR are good estimates for the SPQIV. The results indicate they are indeed good estimates. Even when I started the minimization routine using different initial guesses, I could not obtain better results for the criterion function.

shift the demand and supply of labor are not significant in the regressions presented in Table 8. As a consequence, even though one may argue that the population is endogenous, instrumenting it may not change the primary results.[38]

Table 8. Comparison between the IVQR and SPQIV at the Median[39]

	Small		Small to	Medium	Medium to	Large	Large	
	IVQR	SPQIV	IVQR	SPQIV	IVQR	SPQIV	IVQR	SPQIV
Costs to Ports	-1	-1	-1	-1	-1	-1	-1	-1
Altitude	7.023	7.023	0.037	0.037	-0.009	-0.009	-0.078	-0.078
	(0.069)		(0.234)		(-0.081)		(-0.682)	
Temp	-1614.15	-1634.32	-119.36**	-119.64	-51.40	-51.40	-5.99	-6.15
	(-0.072)		(-2.788)		(-1.848)		(-0.209)	
Rain	-21.52	-21.52	-2.13*	-2.15	-1.36**	-1.40	-1.16*	-1.16
	(-0.069)		(-2.307)		(-2.570)		(-2.223)	
Soil 2	-12375.23	-12839.30	-594.26*	-594.26	-282.78*	-282.78	-222.73	-228.28
	(-0.068)		(-2.025)		(-2.165)		(-1.943)	
Soil 3	-8057.22	-8057.22	-423.02*	-423.02	-152.84*	-152.84	-168.19*	-168.19
	(-0.069)		(-2.327)		(-2.058)		(-2.550)	
Soil 4	-10101.55	-10101.54	-490.54*	-508.93	-260.54**	-267.05	-226.16**	-226.16
	(-0.068)		(-2.382)		(-2.853)		(-3.139)	
Soil 5	-667.35	-667.35	-288.36*	-288.59	-66.39	-66.39	-91.72	-91.72
	(-0.066)		(-2.239)		(-1.235)		(-1.647)	
Mining	-165.91	-165.91	-10.53	-10.53	-14.07*	-14.07	-1.71	-1.71
	(-0.068)		(-1.306)		(-2.167)		(-0.418)	
Power Plants	-304.81	-304.81	-51.83	-51.83	-70.67	-70.67	-23.56	-23.56
	(-0.058)		(-0.682)		(-1.185)		(-0.512)	
Population	-5.67	-5.67	-0.36*	-0.36	-0.14	-0.14	0.004	0.004
	(-0.069)		(-2.026)		(-1.482)		(0.049)	

t-stat in parentheses

$^{*}p < 0.05, ^{**}p < 0.01$

5.4 Endogenous Productivity Index

To estimate farmers' demand for deforestation, it is necessary to rescale the transportation costs using a productivity index. Because this index may respond to transportation costs, Table 9 reports the estimated coefficients of the 2SLS regressions of the log of the productivity index for crops on the same set of regressors and uses the same set of instrumental variables as in the land-use regressions.

[38] One may expect local labor markets to be imperfect, maybe as a result of the sparsely distributed population. A symptom of this imperfection is that farmers rely more on family labor than on hired labor. Indeed, the proportion of family workers in the total number of workers averaged 83% in the Amazon, ranging from a low of 10% in more developed regions in the South Amazon to a high of 99%-100% in the isolated areas in the Western Amazon.

[39] The t-statistics for the normalized IVQR estimates were computed using the delta method.

Formally, the model for the productivity index is:

$$\ln\left[q_m\left(s\right)\right] = X_m'\gamma_s + \varepsilon_m^s.$$

Table 9 reports the estimated coefficients of the costs to ports for each farm size. Once more, for brevity, I omitted the other regressors in the table. I also computed the elasticity of the index with respect to costs to ports at the sample mean (i.e., at $TC_m = 41.5$). As might be expected, the costs to ports do not have a significant impact on the productivity of small farms. This result reinforces the interpretation that small farmers are not engaged in the market.

For small to medium farms, the estimated coefficient is negative. This suggests increasing returns in agricultural land for these farms. More specifically, as farm size increases (all else being constant), landowners may use fertilizers and tractors more intensively, since both fertilizers and mechanization require larger areas to work on than the average small farm. Because a large fraction of inputs used for agriculture in Brazil is imported, one might expect a more intensive use of fertilizers and other inputs as costs to ports decreases.[40] As a result, both total output and total area used for agriculture should increase, but the presence of increasing returns for these farm sizes results in a larger ratio of outputs per hectare.[41] Note that although it is significant, the impact is small in magnitude: at the sample mean, if costs to ports increase by ten percent, the productivity index is reduced by only 0.58%.

Table 9. Results for the Productivity Index (Crops) by Farm Size

	Small	**Small-Medium**	**Medium-Large**	**Large**
Costs to Ports	0.0004	-0.0014*	-0.0007	0.0028*
	(0.449)	(-2.054)	(-1.266)	(2.095)
Elasticity at the Sample Mean	0.017	-0.058	-0.029	0.116
Observations	505	528	526	461

t statistics in parentheses

$^{*}p < 0.05$, $^{**}p < 0.01$

A similar story seems to hold for medium to large farms, despite the fact that the coefficient is not significantly different from zero. For large farms, however, the estimated coefficient of costs to ports is positive and significant. Decreasing returns of agricultural land seems to be in force in this

[40] According to Nogueira (2008), about 57% of the fertilizers used in Brazil in 2006 were imported.

[41] Indeed, these farms tend to concentrate more of their production on corn, rice and beans, which requires more investments in fertilizers and mechanization.

case, which seems reasonable, because diminishing returns to agricultural land should eventually be the dominant force in the agricultural production function.[42]

6 Demand for Deforestation

In this section, I first present the effects of raising the relative value of forested area on the share of agricultural land for each farm size. Then I present the estimated farmers' demand for deforestation and discuss some implications for the three policies considered in this paper.

6.1 Expected Share of Agricultural Land

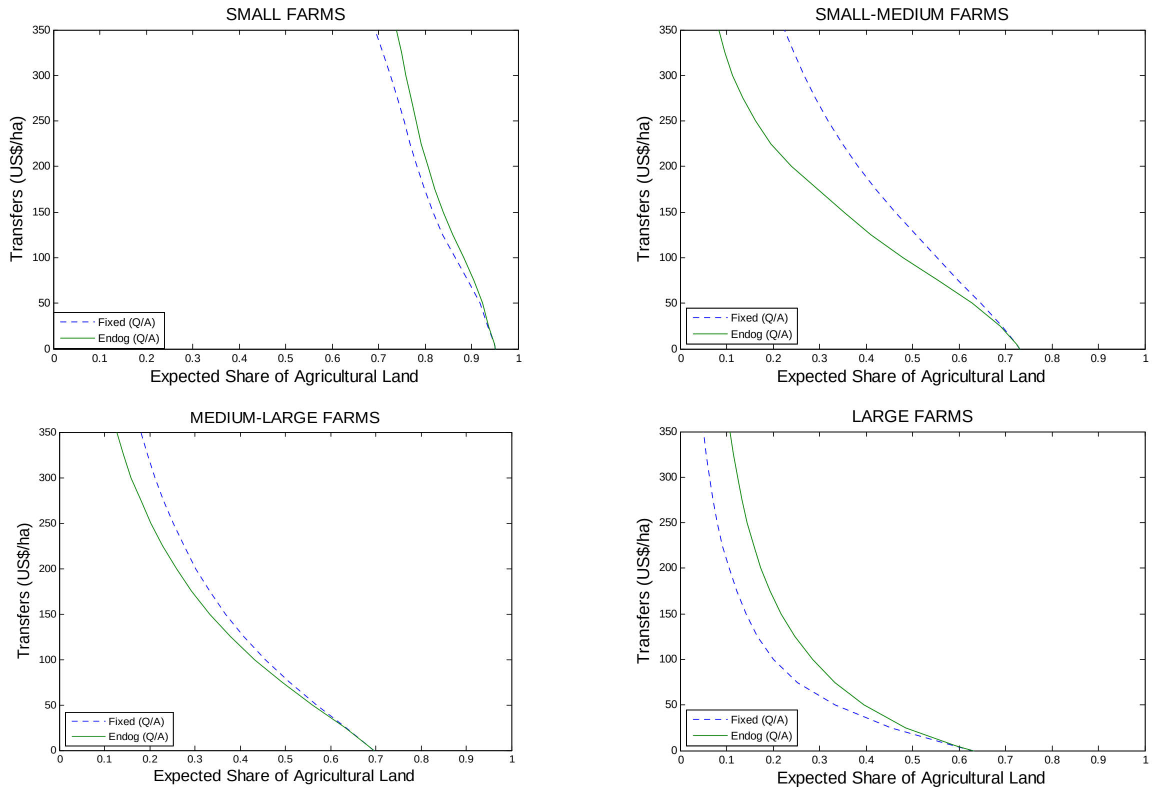
The effects of raising the relative value of forested area on the share of agricultural land are illustrated in Figures 9 and 10. Formally, I use (i) the IVQR estimates of the logit model for their land-use decisions together with (ii) the 2SLS estimated productivity index response to the costs to ports and (iii) equation (9) in Section 3 to predict the fraction of agricultural land on private properties for each municipality in the dataset. The top left panel in Figure 9 presents the results for small farms (less than 5 hectares); the top right panel, for small to medium farms (5-50 hectares); the bottom left panel reports estimates for medium to large farms (50-500 hectares); and the bottom right panel, for large farms (more than 500 hectares). For each panel, I present both results holding the productivity index constant and letting it respond to the transportation costs.

One may interpret the figures in the following way: if the government had increased the relative value of the forested land by, say, paying to avoid deforestation or taxing agricultural land, the expected share of agricultural land on private properties would have been smaller (by some magnitude) than the share observed in the data. The payments/taxes in the figure range from zero to US$ 350/ha/year. To provide a sense of magnitude for the range of payments selected, the average (median) gross revenue per hectare in the Amazon, according to the Agricultural Census of 2006, was US$ 387/ha (US$ 154/ha), while the average (median) price of the land was US$ 640/ha (US$ 490/ha).

In the top left panel, it is clear that smallholders do not reduce the proportion of agriculture significantly with the payments/taxes per hectare. As discussed previously, my strategy is not able to identify their response to payments/taxes. As farm size increases, the actual fraction of agricultural land decreases and the effect of the transfers/taxes increases.

[42] A similar pattern is observed when pasture is included in the productivity index, i.e., positive coefficients for small and large holders and negative coefficients for medium sized farms. However, none of these coefficients are significant.

Figure 9. Taxes/Transfers vs. Expected Share of Agricultural Land - by Farm Size



Small to medium (5-50 hectares) and medium to large (50-500 hectares) farms have similar responses to transfers/taxes. Because of the apparent increasing returns to agricultural land for these farms (estimated in the previous section), it is clear that treating the productivity index as fixed underestimates farmers' response to payments/taxes. The bias is more pronounced for small to medium farms and increases as the amount transferred/taxed augments. Medium sized farms use about 70%, on average, of their private land for agriculture, but they might have used only 24% if they had received payments of US$ 200/ha/year to avoid deforestation. This effect seems strong, but is not surprising given the relatively small gains of the agricultural sector in the Amazon.

Large holders (more than 500 hectares) are the most responsive to hypothetical payments/taxes. They use 63%, on average, of their private land for agriculture, but they would have used only 28% and 17% if they had received payments of, respectively, US$ 100/ha/year and US$ 200/ha/year to avoid deforestation. They respond more according to the estimates, because they are more sensible to changes in transportation costs and because their productivity indices are smaller, on average, when compared to other farm sizes (Table 3). The smaller the productivity, the less valuable the agricultural land, and, so, the smaller the amount of transfers/taxes that are necessary to preserve the forest. Note that, in contrast to medium sized farms, treating the productivity index as fixed overestimates their responses and, so, by correcting for this endogeneity, the impact of transfers/taxes are reduced. I believe that diminishing returns to agricultural land is the main economic force behind the more significant results for large farms.[43]

Table 11 presents some numbers corresponding to the curves exposed in Figure 9. It reports the average effects across municipalities for each farm size for transfers/taxes of (i) US$ 50/ha/year; (i) US$ 100/ha/year; and (i) US$ 200/ha/year.

Table 11. Expected Share of Agriculture vs. Transfers/Taxes - by Farm Size[44]

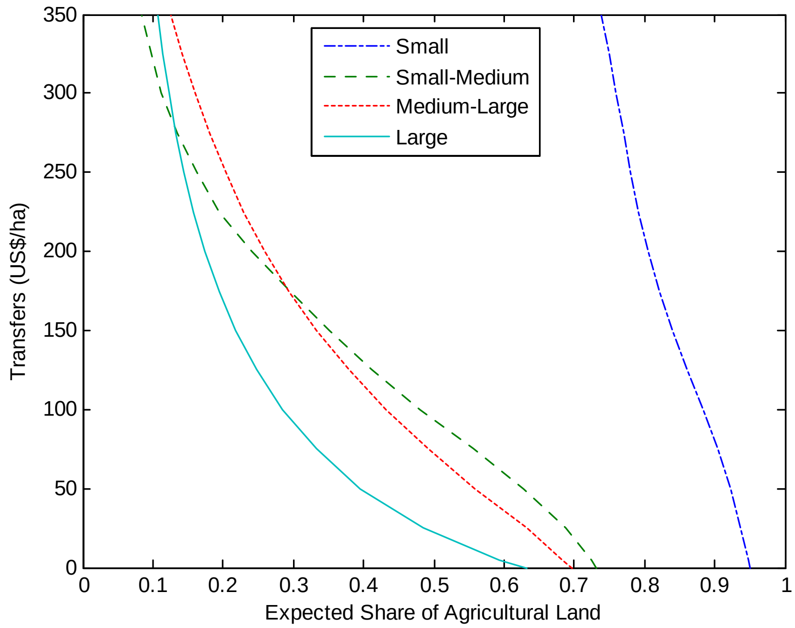
Prop. Deforested	No Transfer	US$ 50/ha/year	US$ 100/ha/year	US$ 200/ha/year
Small				
Fixed $q(s)$	95.1%	91.7%	86.4%	78.3%
Endog. $q(s)$	95.1%	92.3%	88.3%	80.6%
Small-Medium				
Fixed $q(s)$	73.1%	64.7%	55.2%	38.5%
Endog. $q(s)$	73.1%	62.8%	47.9%	23.9%
Medium-Large				
Fixed $q(s)$	69.6%	56.8%	45.5%	30.2%
Endog. $q(s)$	69.6%	55.9%	43.1%	25.9%
Large				
Fixed $q(s)$	63.2%	33.3%	20.1%	10.5%
Endog. $q(s)$	63.2%	39.6%	28.5%	17.4%

Figure 10 puts all curves presented in Figure 9 together for the case where the productivity index responds to transportation costs. It is interesting to see that for medium and large farms, payments/taxes greater than US$ 300/ha/year induce their fraction of agricultural land to approach 10%. The curves tend to be steeper, around 10%, suggesting it becomes increasingly expensive to reduce agricultural share further.

[43] The results are somewhat conservative because I presented the case in which the productivity index is only based on crops. The effects on the share of agricultural land are even more dramatic when the productivity index includes pasture, because it reduces the level of the index considerably.

[44] The reason why the numbers in the first column (with no transfers) differs from the numbers in Table 4 in Subsection 4.4, is that, according to the IVQR estimates, some fraction of the municipalities are associated with positive coefficients for the costs to ports, as presented in Subsection 5.2. To compute farmers' response to transfers, I only considered the observations with negative coefficients for costs to ports. There are 45.5% municipalities associated with negative coefficients for the smallholders; 95.8% for small to medium farms; 98.6% for medium to large; and there are no cases with positive coefficients for large farms.

Figure 10. Taxes/Transfers vs. Expected Share of Agricultural Land - All Farm Sizes

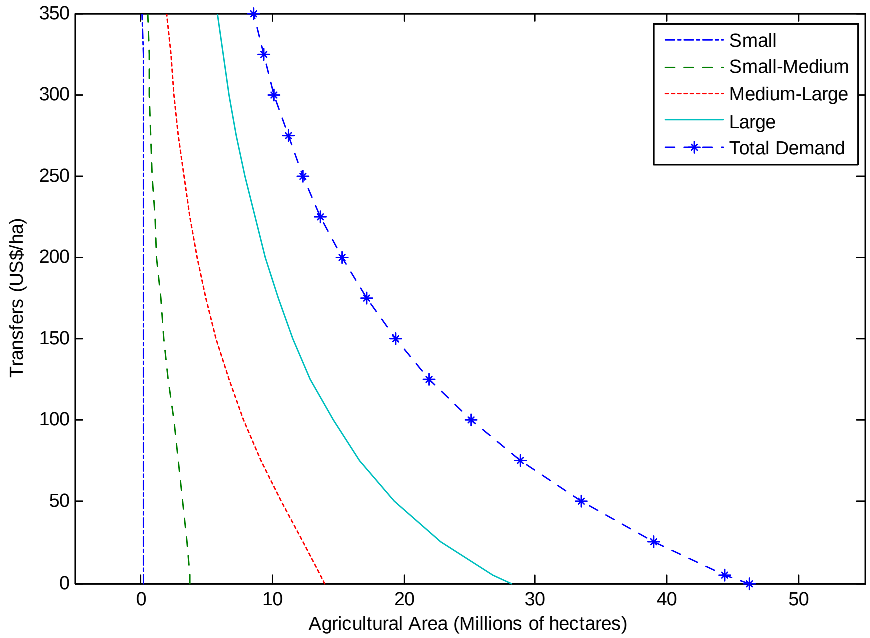


An important note is that the counterfactual exercise refers to what might have happened if the payments/taxes were implemented in the past to avoid deforestation. Now that the land is opened, one should expect the costs to regrow forests to be larger than the payments presented in the counterfactual. The costs are larger, not only because it may be more costly to invest in replanting the forest, but also because it takes time for plants to regrow and that the farmers' private value of young vegetation may be smaller than the value of established forests.

6.2 Total Demand for Deforestation

Next, I present the total demand for deforestation on private properties. For each hypothetical payment/tax, for each farm size and for each municipality, I compute the total deforestation from the predicted share of agricultural land. By summing over the municipalities, I obtain the corresponding demand for each farm size. Finally, the total demand in the Amazon is obtained by summing the farm sizes. Figure 11 shows the demand functions for each farm size and the total demand (allowing for the endogenous productivity index).

Figure 11. Demand for Deforestation



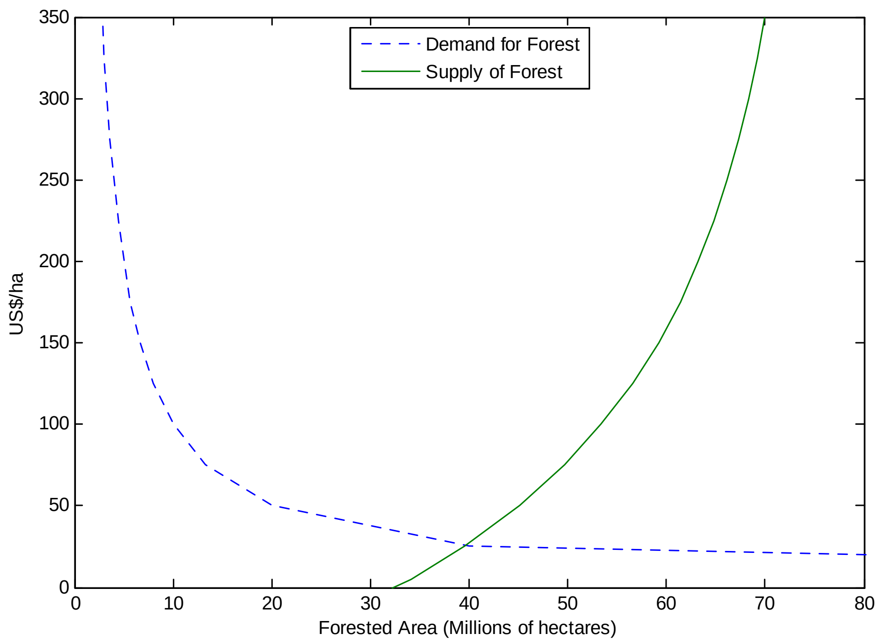
It is clear from the figure that the shape of the total demand function mainly comes from the demand of large farms. The demand function of smallholders, on the other hand, seems almost vertical because of the small area they occupy. In fact, as discussed in Subsection 4.4, smallholders occupy only 1%, on average, of the private land in the Amazon, despite the fact there are 150 thousand in number in the sample; while large holders occupy about 50%, on average, of the private land and are only about 20 thousand in number in the sample. In addition, smallholders are primarily located in isolated regions in the Western Amazon that are not threatened of being deforested, while large farms tend to be located in the South Amazon in the Arc of Deforestation. These facts suggest that policies targeting medium and large farms may be more effective in promoting conservation than policies targeting smallholders. As a consequence, PES programs paying smallholders are unlikely to significantly reduce deforestation and poverty simultaneously in Brazil. Pigouvian taxes targeting large landholders may be more effective in preserving the rainforest and may be preferred from a distributional point of view.

In the next paragraphs, I briefly discuss some implications for each of the three policies I consider in this paper.

PES Programs. It is interesting to see that even small amounts of payments can induce significant changes in the forested area. If PES programs paying US$ 5/ha/year and US$ 25/ha/year to avoid deforestation were implemented, they could have decreased the agricultural area from 46.2 million hectares to 44.3 million and 38.9 million hectares, respectively. Payment of US$ 100/ha/year, by its turn, could have reduced the agricultural area to 25.1 million hectares, and, so, could have preserved 21 million hectares of forest. To have a sense of magnitude, the total private land occupied 78 million hectares in 2006, according to the Agricultural Census. Therefore, a payment of US$ 100/ha/year would have maintained 70% coverage of the forested areas on private properties as opposed to 40% coverage observed in the data.

Despite the strong impacts, preserving the Amazon may be extremely expensive. For instance, the total cost of a program paying US$ 100/ha/year would be roughly US$ 5.3 billion per year. To put these costs into perspective, Norway pledged to donate US$ 1 billion to Brazil's Amazon protection fund in 2008 through 2015 to help fight deforestation. Suppose that Norway's donation were used entirely to pay farmers to avoid deforestation and let the donators and farmers set the price of preserved vegetation freely. Ignoring issues of monitoring and transaction costs, the equilibrium payment would have been US$ 25/ha/year and the total forested area would have been 40 million hectares, instead of the 32 million hectares observed in the data. This simple exercise is illustrated in Figure 12, where the dotted line is Norway's demand for forested area and the full line is the supply of forested land (which is the inverse of farmers' demand for deforestation).

Figure 12. Demand and Supply of Forested Area - Norway's Donation



The primary reason why these programs might have been expensive is the vast area that the Amazon covers. So far, I only considered PES programs with payments directed to every hectare covered by forest on private properties. One way to reduce these costs is to target small regions that are threatened of being deforested. Another possibility is to target transfers to farmers who were going to deforest, and *not* paying those who were *not* going to deforest. For example, if a program paying US$ 100/ha/year to avoid deforestation targets farmers who were going to deforest perfectly, the total cost would be reduced to approximately US$ 2.1 billion per year instead of US$ 5.3 billion. Perfect targeting, however, is unlikely due to problems of asymmetric information. Landowners have better information about their opportunity costs of agricultural land and, so, they may obtain informational rents by pretending that their opportunity costs are higher than their true costs.[45]

Pigouvian Taxes. Instead of payments to avoid deforestation, the same results could have been obtained with Pigouvian taxes on agricultural land, except in that the government would collect revenues from farmers. For instance, perfectly enforced taxes charging US$ 100/ha/year of agricultural land would have the same impact as a PES program, but the tax revenue would have been

[45] For a discussion about the problems of asymmetric information in PES programs, see Ferraro (2008).

US$ 2.1 billion. It is interesting to see that the fines specified in the legislation, ranging from US$ 2,300/ha to US$ 23,000/ha (Subsection 2.2), are not only too high, but are also quite ineffective. Due to the small gains in the Amazon, such high values are not necessary to avoid deforestation. The Brazilian government could be more effective in protecting the rainforest by reducing the value of these fines and increasing enforcement simultaneously.

Quantitative Limit Rule. Limits in land use on private properties is a "command-and-control" instrument that the Brazilian government has adopted. Since 1996, landowners are obligated to keep 80% of their property in native forest in the Amazon. Even though there is evidence that the legislation has not been enforced, as discussed in Subsection 2.2, and as a simple inspection of Table 4 in Subsection 4.4 suggests, it is natural to ask how costly to farmers this policy would have been if it were perfectly enforced. A lower bound on this cost may be obtained by assuming a perfectly enforced Pigouvian tax that induces farmer to only use 20% of their land for agriculture. According to the estimates, the tax would have to charge US$ 195/ha/year of agricultural land. The farmers' lost surplus from this tax is the trapezoid area below the demand curve in Figure 11 between zero and US$ 195/ha. The lost surplus would have been US$ 5.38 billion per year. Therefore, farmers may be willing to pay US$ 5.38 billion to avoid this tax. The amount of money they may be willing to pay to avoid the enforcement of the quantitative rule must be even larger for three reasons: (i) this "command-and-control" policy imposes the same limit on farmers' land-use regardless of the differences in opportunity costs of agricultural land, while Pigouvian tax is a cost-effective price instrument; (ii) the legislation does not allow for managed forests in preservation areas, except under very stringent conditions, but the forested area in the data includes managed forests; and (iii) now that the land is opened, the costs to replant the vegetation adds to the farmers' total costs, since, by law, they must recover the forest at their own expense. Not surprisingly, farmers have systematically tried to alter this law since its implementation [Alston and Miller (2008)].

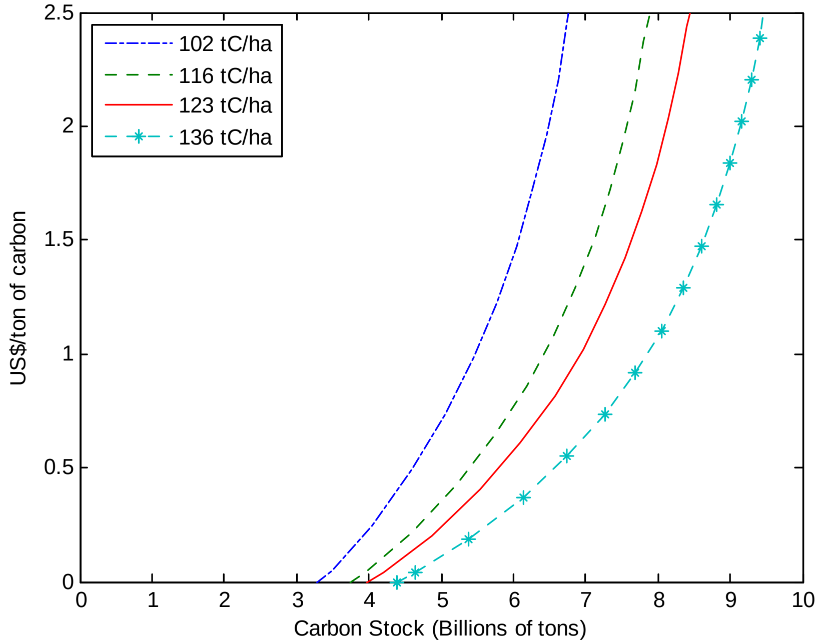
Preserving the Amazon is, therefore, expensive. Some might argue that putting the costs only on the shoulders of the local farmers may be unfair, especially because the benefits of the rainforest may exceed the Brazilian frontiers. A potential source of funds may come from carbon markets and/or REDD+ agreements. For this reason, I discuss next how effective a large scale REDD+ program might have been if it were implemented in the Brazilian Amazon.

Reduced Emissions from Deforestation. [TO BE COMPLETED]

In order to etimate the reduced emissions from deforestation, it is necessary to estimate both the forest loss and the corresponding carbon stock of the land that is cleared. Saatchi et al. (2011) recently estimated a “benchmark” map of biomass carbon stocks over 2.5 billion hectares of forests for the early 2000s, including the Amazon rainforest. They used satellite images to estimate the total carbon stock for forests above 10%, 25%, and 30% tree cover. They obtained point estimates for the carbon density (tons of carbon per hectare) of 102 tC/ha, 116 tC/ha, and 123 tC/ha, respectively, for Brazilian forests. These numbers are somewhat close to the average density of 136 tC/ha reported in Houghton et al. (2001) for the 44 sites in the Amazon where total aboveground biomass (including live and dead) and belowground biomass were directly measured. These sites only measure carbon stocks in mature forests. Because Houghton et al. (2001) documented reasonably close estimates of total carbon stock for several different studies, but few agreements with respect to the spatial distribution, I opted to use these densities assuming each hectare of forest cover would have the same carbon stock independent of its location. As a simplification, I also assume that all the carbon stock would be released into the atmosphere once the land is cleared.

Figure 12 reports the reduced emissions from avoided deforestation using the estimated demand for deforestation and the different carbon densities. Assuming that there are 123 tC/ha in the Amazon rainforest, if REDD+ were implemented and payed US$ 1/tC/year, the carbon stock in the forest would have increased from 4 billion tons of carbon in the privately owned forests to approximately 7 billion tons. To put this quantity into perspective, according to IPCC (2007), the annual emissions of carbon from land-use change in the 1990s was 1.6 billion.

Figure 13. Price of Carbon vs. Carbon Stock



A program paying US$ 1/tC/year to reduce emissions under the assumption that there are 123 tC/ha in the Amazon would have cost roughly US$ 7 billion per year, and a perfectly targeted program only paying the hectares that were going to be deforested would have cost US$ 3 billion per year. Despite the US$ 4 billion loss of a program that pays the same amount for every privately owned hectare in the Amazon, the corresponding cost per ton of *reduced emission* would have been about US$ 2.33/tC/year. Note that this cost is smaller than the price of carbon in the European Union Emissions Trading System. Taking the price as €10/tCO_2 and converting it into dollars per ton of carbon, the result is a price of US$ 44.2/tC.[46]

According to Greenstone et al. (2011), the central value of the social cost of carbon (SCC) for 2010 is US$ 21 per ton of CO_2 emissions. Such a cost implies a marginal damage value of US$ 77/tC. Note that the "back-of-the-envelope" supplies of carbon stock in Figure 13 become steep for small values of carbon prices, primarily because the demand for deforestation becomes steep when the share of agricultural land reaches 10% (Figure 10). A REDD+ program fixing the price of carbon or a Pigouvian tax taxing carbon at the value of US$ 77/tC would have reduced emissions significantly, but not much more than by fixing the price/tax at lower values. For example, impacts would have

[46] The average exchange rate in 2006 was US$ 1 = € 0.829, and one ton of CO_2 corresponds to (12/44) tons of carbon.

been similar to a carbon tax fixed at the value US$ 18/tC (corresponding to US$ 5/tCO_2), which is one of the Greenstone et al. (2011)'s recommended values to conduct cost-benefit sensitivity analyses.

Limitations. There are some limitations in the present paper that I would like to address. First, the counterfactual exercises do not take into account monitoring and transaction costs. PES programs must be more expensive to implement than the estimated here, because they require investments to launch the program, as well expenses in monitoring and enforcement. Taxes and limits on land-use, on the other hand, only require monitoring costs.

Second, there is no general equilibrium effect considered in this exercise. Changing farmers land-use decisions may affect local markets, including local wages, prices of non-tradables or even the international price of soybeans (if Brazil considerably reduces its production of soybeans), that are not taken into account in this exercise.

Third, and similar to the previous point, I have not estimated by how much the total private land would respond to these policies. Although such an exercise is possible, there are some important implications that cannot be addressed with the present dataset. There exists plenty of unprotected public forested land that may be occupied in response to PES programs, for example. These occupations might increase the PES's total costs, as well as the disputes for land and the potential violence associated with these disputes. As such, PES programs, if not carefully designed, may have the unintended consequences of raising local violence. Augmenting the total protected area in the Amazon might reduce these negative effects. Hence, it would be interesting to know how the costs of the PES programs compare to the costs of governmentally protected areas. The results I present in this paper should be viewed, therefore, as only one of the many inputs necessary for a complete evaluation of the costs to preserve the Amazon.

7 Conclusions

In this paper, I estimated farmers' demand for deforestation on private properties in the Brazilian Amazon. This demand is an important input to evaluate several policies that promote conservation. I collected data from several sources for the Brazilian Amazon, estimated both parametric and semiparametric quantile IV models of the impacts of transportation costs on farmers' land use decisions, rescaled these costs and obtained the demand for deforestation. The estimates suggest that large landholders are the most responsive to payments/taxes to preserve the rainforest. This fact, together with the highly unequal distribution of land in the Amazon, suggest that policies targeting medium and large farms may be more effective in promoting conservation than policies targeting smallholders.

The results also indicate that a Pigouvian tax of US$ 100/ha/year on agricultural land would have maintained 70% coverage of the forested areas on private properties as opposed to 40% coverage observed in the data. In addition, it would have resulted in US$ 2.1 billion in revenues. Similarly, a PES program paying private landholders at the same rate to prevent deforestation would have achieved the same levels of protection, but would roughly cost US$ 5.33 billion per year. In addition, a "back-of-the-envelope" calculation of the supply of carbon stock in the Amazon based on the estimated demand function indicates that a REDD+ program fixing the price of carbon at US$ 1/tC/year would have increased the carbon stock from 4 billion tons of carbon in the privately owned forests to approximately 7 billion tons. The total cost of this program would be roughly US$ 7 billion per year, and the cost per ton of *reduced emissions* of carbon would have been US$ 2.33/tC/year.

Finally, with respect to the quantitative limits in land-use, the required share of 80% of forest cover on private land specified in the Brazilian law would be so expensive for farmers if it were fully enforced that farmers would be willing to pay at least US$ 5.38 billion per year to avoid the enforcement of this rule.

There are several directions for future research. First, it would be interesting to access the micro-data on farmers' decisions. The micro-data may provide a richer picture of their opportunity costs for agricultural land and also avoid the potential drawbacks in using an aggregated measure of the productivity index. Furthermore, it may reveal the entire distribution of land use within each municipality, instead of the expected fraction of agricultural land. This distribution may help address issues such as the efficiency loss of PES programs due to asymmetric information about farmers' private valuations and the use of auctions to allocate PES contracts.

Second, it would be interesting to use a panel data model based on satellite images coupled with extra assumptions on the evolution of the private land. This data allows for a dynamic model with irreversible land-use decisions that can be used to study impacts of commodity prices on the rate of deforestation. Furthermore, it can be used to study by how much these prices can affect the effectiveness of the policies considered in this paper.

Third, it would be interesting to complement the results of this paper with a multinomial choice model. Although the results of the multinomial choice model may have to rely on parametric functional forms, since a choice-specific variable required to nonparametrically identify the model is missing in the present dataset [Berry and Haile (2011)], disaggregating agricultural land into pasture and crops may provide a richer description of the agricultural sector. Fourth, it may be possible to obtain results under weaker assumptions than those imposed in this paper. For example, through the use of the monotone instrumental variable approach of Manski and Pepper (2000), I may be able to partially identify the demand for deforestation.

From an econometric point of view, obtaining explicit rates of convergence in terms of the number of observations for the artificial neural networks sieves estimator, as well as the corresponding asymptotic distribution, seems to be in order. Finally, it would be interesting to investigate the extra restrictions in the model that avoid the curse of dimensionality in estimating the conditional moment function used in the PSMD estimator.

References

[1] Amaral, D. F. (2010). "Panorama do Mercado de Oleaginosas: Aproveitamento para Produção de Óleo e Derivados." Presented in: IV Congresso Brasileiro de Mamona e I Simpósio Internacional de Oleaginosas Energéticas.

[2] Amigos da Terra (2009). "A Hora da Conta: Pecuária, Amazonia e Conjuntura". Amigos da Terra - Amazônia Brasileira, São Paulo, Brasil.

[3] Andersen, L.E., C. W. Granger, E. Reis, D. Weinhold and S. Wunder. (2002). *The Dynamics of Deforestation and Economic Growth in the Brazilian Amazon*, Cambridge Univ. Press.

[4] Andersen, L. E. and E. J. Reis (1997) "Deforestation, Development, and Government Policy in the Brazilian Amazon: An Econometric Analysis." In: Hickey, Kevin L. & Demitri Kantarelis (1997) *Our Natural Environment: At a Crossroad*. Worchester, MA: Interdisciplinary Environmental Association.

[5] Anderson, K and E. Reis. (2007). "The Effects of Climate Change on Brazilian Agricultural Profitability and Land Use: Cross-Sectional Model with Census Data." Final report to WHRC/IPAM for LBA project Global Warming, Land Use, and Land Cover Changes in Brazil.

[6] Aston, L. J. and B. Muller (2008). "Legal Reserve Requirements in Brazilian Forests: Path Dependent Evolution of De Facto Legislation." *Revista EconomiA*, Selecta, Brasília (DF), v.8, n.4, 25-53.

[7] Alves, D. (2002). "An Analysis of the Geographical Patterns of Deforestation in the Brazilian Amazon in the Period 1991-1996." In: C. H. Wood and R. Porro (Eds.), *Deforestation and Land Use in the Amazon*. University Press of Florida, Gainesville, FL, 95-106.

[8] Banerjee, Duflo and Qian (2009), "On the Road: Access to Transportation Infrastructure and Economic Growth in China". Mimeo.

[9] Bell, K. and E.G. Irwin (2002). "Spatially explicit micro-level modeling of land-use change at the rural–urban interface". *Agricultural Economics* 27.

[10] Berry, S. T. and P. A. Haile (2009). "Identifcation of a Nonparametric Generalized Regression Model with Group Effects," Discussion paper, Yale University.

[11] Berry, S. T. and P. A. Haile (2010). "Identification in Differentiated Products Markets Using Market Level Data,"Discussion paper, Yale University.

[12] Brito, B. and P. Barreto (2006). "Sugestões para Aumentar a Eficácia da Lei de Crimes Ambientais no Ibama e nos Tribunais de Justiça no Setor Florestal do Pará". 11^o Congresso Brasileiro de Direito Ambiental: Biodiversidade e Direito em São Paulo. São Paulo.

[13] Castro, N. (2003). "Formação de Preços no Transporte de Carga," *Pesquisa e Planejamento Econômico*, v.33, n.1, 167-189.

[14] Chen, X. (2007). "Large Sample Sieve Estimation of Semi-nonparametric Models," chapter 76 in *Handbook of Econometrics*, Vol. 6B, 2007, eds. James J. Heckman and Edward E. Leamer, North-Holland.

[15] Chen, X. and D. Pouzo (2009) "Efficient Estimation of Semiparametric Conditional Moment Models with Possibly Nonsmooth Residuals," *Journal of Econometrics*, vol. 152, pp. 46-60.

[16] Chen, X. and D. Pouzo (2011). "Estimation of Nonparametric Conditional Moment Models With Possibly Nonsmooth Generalized Residuals," forthcoming in *Econometrica*.

[17] Chernozhukov, V. and C. Hansen (2005). "An IV Model of Quantile Treatment Effects," *Econometrica* 73, 245–261.

[18] Chernozhukov,V. and C. Hansen (2008). "Instrumental Variable Quantile Regression: A Robust Inference Approach," *Journal of Econometrics*, Volume 142, Issue 1, 379-398.

[19] Chernozhukov, V., I. Fernández-Val and A. Galichon (2010). "Quantile and Probability Curves Without Crossing," *Econometrica*, 78, Issue 3, 1093-1125.

[20] Chomitz, K. M. and D. A. Gray (1996). "Roads, Land Use and Deforestation: A Spatial Model Applied to Belize". *World Bank Economic Review,* 10(3):487-512.

[21] Chomitz, K.M. and T.S. Thomas (2003). "Determinants of Land Use in Amazonia: a fine-scale spatial analysis". *American Journal of Agricultural Economics*, 85, 1016

[22] Engel,S., S. Pagiola, and S. Wunder (2008). "Designing payments for environmental services in theory and practice: An overview of the issues", *Ecological Economics*, 65, 663-674.

[23] Ferraro, P. J. (2008). "Asymmetric Information and Contract Design for Payments for Environmental Services," *Ecological Economics*, 65, 810-821.

[24] Hirakuri, S. R. (2003). *Can Law Save the Forest? Lessons from Finland and Brazil.* Jacarta, Indonesia, Centre for International Forestry Research.

[25] Houghton, R.A., Lawrence KT., Hackler J.L., Brown S. (2001). "The spatial distribution of forest biomass in the Brazilian Amazon: A comparison of estimates," *Global Change Biology*, 7:731–746.

[26] Kaimowitz, D. and A. Angelsen (1998). *Economic Models of Tropical Deforestation: A Review.* Bogor, Indonesia, CIFOR.

[27] Manski, C. F. and J. V. Pepper (2000). "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica*, 68, Issue 4,997-1010.

[28] Nelson, G.C., Geoghegan, J., (2002). "Modeling deforestation and land use change: sparse data environments." *Agricultural Economics* 27.

[29] Nepstad, D., G. Carvalho, A. C. Barros, A. Alencar, J. P. Capobianco, J. Bishop, P. Moutinho, P. Lefebvre and U. Silva Jr. (2001). "Road Paving, Fire Regime Feedbacks, and the Future of Amazon Forests." *Forest Ecology & Management*, 154. 395-407.

[30] Nogueira, A. C. L. (2008). "Agricultura: o Mercado de Fertilizantes no Brasil," *Informações Fipe, Maio, 5-8.*

[31] IBGE, Instituto Brasileiro de Geografia e Estatística, (2009). *Produto Interno Bruto dos Municípios 2003-2006*. Rio de Janeiro.

[32] Igliori, D. (2008). "Deforestation, Growth and Agglomeration Effects: Evidence from Agriculture in the Brazilian Amazon," *Environmental Economy and Policy Research Working Papers* 29.2008, University of Cambridge, Department of Land Economics.

[33] INPE, Instituto Nacional de Pesquisas Espaciais (2010), *Monitoramento da Floresta Amazônica Brasileira por Satélite – Projeto Prodes.* Available at: <http://www.obt.inpe.br/prodes/>

[34] IPCC (2007). *Climate Change 2007: The Physical Science Basis.* Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

[35] Pereira, D., D. Santos, M. Vedoveto, J. Guimarães and A. Veríssimo (2010). *Fatos Florestais da Amazônia 2010.* Belém, PA: Imazon.

[36] Pfaff, A. (1999). "What Drives Deforestation in the Brazilian Amazon? Evidence from Satellite and Socioeconomic Data". *Journal of Environmental Economics and Management.* 37, 26.

[37] Pfaff A., J.A. Robalino, R. Walker, E. Reis, S. Perz, C. Bohrer, S. Aldrich, E. Arima, M. Caldas, W Laurance and K. Kirby (2007). "Road Investments, Spatial Intensification and Deforestation in the Brazilian Amazon". *Journal of Regional Science,* 47:109-123.

[38] Pfaff, A., J. Robalino and G.A. Sánchez (2009d) "Evaluating the Impacts of Payments for Environmental Services". Duke mimeo after talk at NBER Summer Institute, July 2007.

[39] PNLT, National Highway Plan (2006). Available at http://www.transportes.gov.br/index/conteudo/id/36604 and accessed on 11/24/2010.

[40] Reis, E.J. and R. Guzman (1992). "An Econometric Model of Amazon Deforestation," IPEA, Rio de Janeiro, Brazil.

[41] Saatchia,S.S., Nancy L. Harrisc, Sandra Brownc, Michael Lefskyd, Edward T. A. Mitcharde, William Salasf, Brian R. Zuttaa,b, Wolfgang Buermannb, Simon L. Lewisg, Stephen Hagenf, Silvia Petrovac, Lee Whiteh, Miles Silmani, and Alexandra Morelj (2011). "Benchmark Map of Forest Carbon Stocks in Tropical Regions Across Three Continents," *PNAS*, vol. 108, no. 24, 9899–9904.

[42] Souza-Rodrigues, E. A. (2011). "Nonparametric Estimation of a Generalized Regression Model with Group Effects," Yale University.

[43] Schlesinger, S. (2009). *Onde Pastar? O Gado Bovino no Brasil.* Rio de Janeiro, FASE.

[44] TCU, Tribunal de Contas da União (2009). "Arrecadação de multas administrativas. Versão simplificada das contas do governo da República - Exercício de 2009". Available at: http://portal2.tcu.gov.br/portal/page/portal/TCU/comunidades/contas/contas_governo/contas_09/Textos %20Arrecadacao%20de%20Multas.pdf

[45] Vosti, S. A., E. M. Braz, C. L. Carpentier, M. V. N. d'Oliveira, and J. Witcover (2003). "Small-scale Managed Forestry at the Brazilian Agricultural Frontier: Adoption, Effects and Policy Issues", Presented in: The International Conference on Rural Livelihoods, Forests and Biodiversity, May 2003, Bonn, Germany.

[46] Weinhold, D. and E.J. Reis (2008). "Transportation costs and the spatial distribution of land use in the Brazilian Amazon". *Global Environmental Change* 18:54-68.

[47] Wunder, S. (2007). "The Efficiency of Payments for Environmental Services in Tropical Conservation." *Conservation Biology* 21(1): 48–58.

8 Appendix

This Appendix complements Section 4 about the dataset by providing a detailed explanation about how the transportation costs and the productivity indices were constructed. It also provides a brief explanation about other useful variables not utilized in the regressions, but that are helpful in providing a sense of magnitude for the results.

8.1 Endogenous Regressor: Transportation Costs

The proxy for transportation costs is defined as the minimum unit cost (US$/ton) to transport one ton of goods to the destination. As mentioned in the main text, I used the network of modes of transportation from the National Highway Plan and computed the least cost path to the nearest port in ArcGIS. The calculation divides the entire country into cells corresponding to 1 km^2 and requires the cost to travel over it for each cell. The travel cost depends on if the cell contains a segment of road (paved or unpaved), railroad, navigable river or if it does not contain any transportation mode. The optimization routine in ArcGIS determines for each cell in the grid the least accumulative cost path to the nearest destination. The total costs from the municipal seat to the nearest port is the proxy for the transportation costs.[47]

[47] The network of modes of transport are in a polyline format in ArcGIS while the country is in a polygon. To accomplish this, I first have to transform these polylines and polygons into a raster format (i.e., grid of cells with

The unit cost per km for different modes of transportation is based on the freight rate data collected by SIFRECA and the Vehicle Cost Module of the World Bank's Highway Design Model (HDM-VOC-4). Whenever SIFRECA's data was incomplete or limited, I complemented it by calling the companies directly and asking them for their freight costs.[48] Because almost all the information obtained from SIFRECA about the freight values in the Amazon corresponds to costs of transporting soybeans, I assume that all products have the same transportation costs as soybeans, which seems to be correct, at least for bulk products and sacks [see Castro (2003)].

Table A.1 summarizes the cost weights used to compute the least accumulative cost in ArcGIS. The first column discriminates between the possible modes of transportation considered in the calculations, the second column reports the unit costs used in Brazilian currency (R$), and the third column converts these unit costs into US$.

TABLE A.1. Cost Weights

Cost Weight	R$/ton.km	US$/ton.km
Paved Road	0.07678	0.0353
Unpaved Road - Outside Rain Forest	0.0992	0.0457
Unpaved Road - Within Rain Forest	0.15	0.069
Navigable River - Good Infrastructure	0.0444	0.0204
Navigable River - Poor Infrastructure	0.1139	0.0525
Railroad	0.0608	0.028
Land - Outside Rain Forest	1.5	0.6912
Land - Inside Rain Forest	3	1.3824

Roads. I purchased the freight values of routes from cities in the Legal Amazon to state capitals and ports for the main products in the Amazon. I obtained the costs and distances to transport soybeans by roads for 105 routes and 5 destinations for 2006. The average cost to transport 1 ton

cost values). Then I use the command "cost distance" in ArcGIS to compute the least accumulative cost path to the nearest destination and "cost allocation" to identify which destination is the nearest for each cell. Finally, the command "extract values to points" is used to assign the total costs to the corresponding municipal seats.

[48]The costs obtained directly from the firms were deflated using the National Consumer Price Index (INPC).

of soybeans per km in this data is R$ 0.0767 (US$ 0.0353). That is the value I inputted as the unit cost to travel one cell with a paved road.[49]

To capture the higher costs of travelling on unpaved roads, I increased the unit costs for paved roads by using the World Bank's HDM-VOC-4. This model is designed to calculate unit road user costs for a road section with 1 km length and requires several inputs for the characteristics of the road. I maintained all inputs at the default values, except for changing the road characteristics from paved to unpaved and increasing the roughness of the road to the recommended value for a poor tertiary road. For heavy truck vehicles, the increase in the roughness raised the road user costs by 29%. I adopted the unit cost to travel one km on unpaved road to be 29% higher than the cost to travel on a paved road. Therefore, a cost of R$ 0.0992 (US$ 0.0457) to transport one ton of soybeans per km was assigned to unpaved roads.[50]

Transporting large quantities on unpaved roads within the rainforest is extremely difficult. The poor conditions of the roads, combined with the excess of rain, especially in the rainy season, can make these roads inaccessible. For this reason, I decided to differentiate unpaved roads within the rainforest from the unpaved roads elsewhere. Unfortunately, the freight rate data purchased from SIFRECA does not cover the dense rainforest. To overcome this limitation, I called local companies directly and asked them how much it costs to transport soybeans from Sorriso (in Mato Grosso state), one of the main producers of soybeans, to the Port of Santarém on the Amazon River. The only route available in this case is the "Cuiabá-Santarém" road, which cuts the Amazon almost in the middle from the South to the North and provides a good measure of the difficulties in travelling in the dense jungle. The average cost per km that I obtained is R$ 0.15 (US$ 0.069). I used this unit cost as the cost weight for all unpaved roads within the dense rainforest.[51]

Railroads and Navigable Rivers. Freight values for navigable rivers and railroads are more difficult to obtain because of the reluctance of the firms to disclose their data. The information I was able to obtain from SIFRECA includes the freight values to transport one ton of soybeans in

[49] In SIFRECA's sample, almost all cities were located in the state of Mato Grosso, the main producer of soybeans. The distances range from 277 km to 3,712 km, with an average of 1,863 km. The costs per ton of soybeans transported ranges from US$ 14.2 to US$ 88.47 with an average of US$ 64.

[50] The International Roughness Index (IRI) is an index that measures the deviations of a surface from a true planar surface with characteristic dimensions that affects vehicle dynamics, ride quality, dynamic loads and drainage. It is measured in m/km units. The value recommended for a good primary paved road is 2 m/km. The value I used in order to increase the costs of unpaved roads is 8 m/km, which corresponds to a poor tertiary road. I am grateful to Prof. Newton de Castro, who suggested the use of the World Bank's HDM-VOC to increase the costs of unpaved roads. The HDM-VOC-4 model is available at http://web.worldbank.org/WBSITE/EXTERNAL

[51] I define the dense rainforest as covering the states of Acre, Amazonas, Rondônia, Roraima, Pará and Amapá. The remaining states in the Legal Amazon (Maranhão, Tocantins and Mato Grosso) are not in the dense rainforest.

two routes for a railroad, with an average value of R$ 0.0608 (US$ 0.0279), and one route for one of the most important navigable rivers in the Amazon - the Madeira River waterway. Because of the difficulty in collecting these data, SIFRECA does not provide recent freight rate data for these modes of transportation anymore.[52]

Similar to roads, there are differences in the quality of navigable rivers, depending on the depth of the river, investments in signaling, investments in communications and in the quality of the local ports. Based on conversations with governmental agencies responsible for the administration of the waterways, as well as local companies, I classified the navigable rivers in two types: those with good and those with poor infrastructure. The good rivers include the Madeira River waterway and the Amazon River waterway (linking Manaus to Belém), among a few others. The rivers with poor infrastructure are the remaining navigable rivers.[53]

From SIFRECA's information of the freight cost of the Madeira River waterway combined with the information I obtained directly from the companies for the cost of the Amazon River waterway (from Manaus to Belém), I arrived at a unit cost of R$ 0.0444 (US$ 0.0204) to transport 1 ton of soybeans per km. For the navigable rivers with poor infrastructure, I obtained an average value of R$ 0.1139 (US$ 0.0525) per tons of soybeans per km.[54]

Land. Finally, I imposed high costs to transport products by land with no mode of transportation so that ArcGIS would avoid computing travelling costs using these cells. For the Amazon Biome, I imposed a cost of R$ 3 per km (US$ 1.38) and for land outside the Amazon Biome, a cost of R$ 1.5 (US$ 0.69). The rationale is that moving within the rainforest should be much more costly than for other types of vegetation. There is little guidance to which values should be adopted for transportation in land, but, because a few municipal seats were not connected to any segment of the network and to the extent that these values induce ArcGIS to use cells with some mode of

[52] The routes for railroads are from Cascavel to Ponta Grossa (both in the Paraná state in the South) and from Porto Franco to São Luís (both in the Maranhão state in the Legal Amazon). The Madeira River waterway connects Porto Velho (the capital of the Rondônia state) to Itacoatiara, which is close to Manaus (the capital of the Amazonas state).

[53] The Ministry of Transport classify two types of waterways: *more-navigable* and *less-navigable* rivers. I excluded the *less-navigable* rivers, because they do not seem to be used to transport large quantities of products. Hence, the classification of good- and poor-infrastructure rivers are only restricted to those *more-navigable* rivers.

[54] The unit cost for poor-infrastructure rivers are obtained from the routes Manaus-Tabatinga, Manaus-Barcelos and Manaus-Boca do Acre, all of them in pristine regions in the Western Amazon. The larger costs obtained not only reflect poor signaling and capacity constraints, but also difficulties with the excess of curves, depth of rivers (some of them may have about one meter depth in the dry season, while the Amazon River has, on average, 16 meters depth), as well as the presence of stones and sandbars that shift around over time. To gain a sense of magnitude, it can take about 10 days to go from Tabatinga (located at the border with Venezuela) to Manaus and about 18 days to navigate that stretch of the river in the opposite direction.

transportation, the costs of moving on land should not impact the results significantly.[55]

8.2 Productivity Index

From the Agricultural Census, I obtained the quantity sold and area occupied of major agricultural outputs for each municipality for farms of different sizes. Two productivity indices were computed: the first only considers the production of crops and the second includes crops and pasture.

The productivity index for crops is based on the main products discussed in Subsection 2.3: soybeans, corn, manioc, rice and beans. For each product, j, for each municipality, m, and for each farm size, s, I calculated the productivity per hectare by dividing the total output sold by the total area used for the corresponding crop. I denote this productivity by q_{jm}^{s}, for $j = 1, ..., 5$. Then, I took the weighted average of q_{jm}^{s} across j where the weights are the proportion of the area utilized for each crop, a_{jm}^{s}. The index for the crops is therefore:

$$q_m^c(s) = \sum_{j=1}^{5} \left(\frac{a_{jm}^{s}}{a_m^s} \right) \times \left(q_{jm}^{s} \right)$$

where $a_m^s = \sum_{j=1}^{5} a_{jm}^{s}$. To add pasture in the productivity index, I first assumed that each ox weights half ton and that the entire ox can be used for beef consumption. I therefore multiplied the number of cattle sold by 0.5 to obtain the quantity of beef sold in tons.[56] By dividing this result by the pasture area, I obtain the productivity for cattle, q_m^p. Unfortunately, I do not have information on the number of oxen for different farm sizes, so q_m^p is the same for all farm sizes. Let a_{mp}^{s} be the area occupied by pasture for farms of size s in location m. The second productivity index is a weighted average between q_m^c and q_m^p, where the weights are the proportion of crops and pasture areas:

$$q_m^{cp}(s) = \left(\frac{a_m^s}{a_m^s + a_{mp}^{s}} \right) q_m^c(s) + \left(\frac{a_{mp}^{s}}{a_m^s + a_{mp}^{s}} \right) q_m^p$$

I averaged the indices over the micro-regions to reduce measurement error problems. Micro-regions are administrative areas larger than the municipalities. There are 85 micro-regions in the selected sample for the Amazon.

[55] For those cities not connected to any mode of network in our map, I created straight-lines joining them to the nearest road and assigned the cost of unpaved road (depending on whether they are inside or outside the rainforest) to travel over these straight lines. This procedure is reasonable, because the official map from the National Highway Plan is missing the unofficial roads, and all of them are most likely unpaved.

[56] Informal conversations with farmers suggest that a common rule of thumb is to assume a cow has about 150 kg of "available meat" and a bull has about 270 kg. Because a bull weights around 500 kg, I am overestimating the productivity for pastures.

8.3 Other Useful Variables

Useful variables that are not utilized in the estimation procedure include:

Satellite Data. Satellite images of deforestation in the Brazilian Legal Amazon collected by INPE/PRODES (Brazilian National Institute of Space Research/Programa de Monitoramento da Amazônia Brasileira por Satélite);[57]

Land Prices. This data comes from the Agricultural Census and consists of the farmers' best estimations of the value of their land. Unfortunately, this variable does not come from transaction prices and does not distinguish land prices for different land-uses. The average price per hectare is obtained by dividing the total value of the land in the municipality by the total private land area.

Labor. The Agricultural Census provides the number of workers in the farms (separating family member workers from hired workers), as well as rural wages.

Revenue and Expenditures. The Agricultural Census also contains information on both the revenues and expenditures of the establishment. The revenue is the value from the sale of production and the expenditures include maintenance costs, salaries, rentals of machinery, and other expenses. Similar to land price, however, revenues and costs are not discriminated for each land use.

[57] Available at http://www.obt.inpe.br/prodes/index.html