Roads, Trade and Urbanization in the Tropics: Theory and Evidence from the Brazilian Amazon *

Heitor S. Pellegrina

Brown University

PRELIMINARY VERSION

August 5, 2014

Abstract

The Brazilian Amazon has undergone a great transformation of its economy since the end of the 1990s, when the price of its export-oriented commodities increased in more than 200% in a few years. However, counties in the region experienced the effects of this soaring price differently according to their access to a federal road system constructed in the early 1970s. I take advantage of this institutional setting, and a rich dataset with annual information on agriculture production and labor market employment, to study how trade affects urbanization in developing countries. I find that counties with access to roads expanded more their export of capital intensive commodities and their employment in urban activities after the price shock. Furthermore, structural estimates of a model indicate that an urbanization process would not have occurred in the absence of a capital intensive activity. This article provides evidence that the lack of good trade opportunities in terms of price and products may undermine the benefits of transportation projects in developing countries.

Keywords: Urbanization, Transportation, Trade, Commodities

JEL Classification: F16, O13, O18, O54, N56, Q11, R23

*Email: heitor_pellegrina@brown.edu. I am grateful to my advisers Andrew Foster, Jonathan Eaton, Sriniketh Nagavarapu, and Nathaniel Baum-Snow. I am also thankful to Vernon Henderson, Marcos Rangel, Lucas Scottini, Alex Eble, Ken Chay, David Weil, Leah VanWey, Alexei Abrahams, and participants at presentations held at the Econ department at Brown, the PSTC, the University of São Paulo, the Watson Institute for International Studies, and the Readings and Research group on Trade. I am specially thankful to Peter Richards for many helpful comments and for sharing his vast knowledge of the Amazon. Data for commodity prices in Mato Grosso was provided by IMEA. This research was supported by the National Science Foundation's IGERT and by the Institute for the Study of Environment and Society.

1 Introduction

Urbanization is one of the most established symptoms of development (Kuznets, 1973; Herrendorf et al., 2013). As countries experience economic growth, there is a reallocation of labor from rural activities to urban ones. Theoretically, most of the studies explain the process of urbanization using closed economy models, where a combination of technological improvements in agriculture and a inelastic demand for food decreases the need for labor in the rural sector. Empirically, this literature provides stylized facts at the country level: either comparing the level of urbanization of different countries at a given period, or the same country over time. This body of research, however, still provides very limited microeconomic evidence of the theoretical mechanisms that explain when and why urbanization occurs [1], and, also, how international trade may affect this process (Foster and Rosenzweig, 2007; Gollin et al., 2013; Henderson et al., 2013).

In this article, I work on these two limitations of the current literature by investigating the microeconomic evidence on the expansion of trade and urban activities in the Brazilian Amazon. This region has undergone an intense expansion of its trade activities since the end of the 1990s, when the price of commodities in terms of the local currency increased more than 200% due to variations in the Brazilian exchange rate and the global increase in commodity prices. However, *municípios* (counties) experienced the effects of these macro economic changes differently according to their access to federal roads. The federal highway system in the Amazon was constructed in the early 1970s, and ended up providing the main arterial routes for exporting goods when the commodity prices soared in the end of the 1990s. I explore this institutional setting to study how access to transportation leads to differences in the expansion of trade and urban activities when macro-economic conditions for export improve.

To guide the empirical work, I build a small open economy model with a set of counties with heterogeneous distance to federal roads. The model leads to an expansion of urban activities through the consumption of non-tradeable goods, which are assumed to be produced by the urban centers in each county. Counties that engage in trade activities increase their income, which translates into a higher demand for both tradeable and non-tradeable goods (a Balassa-Samuelson effect). Because non-tradeable goods can only be provided by local labor, the increase in demand pushes wages up in the county, which brings migrants from other regions of Brazil to work in the non-tradeable (urban) sector. This in-migration

[1] See Fajgelbaum and Redding (2014); Storeygard (2013); Jedwab (2013) for recent articles exploring the microeconomic evidence on urbanization.

then reduces wages in the county, leading the local economy to an equilibrium. Based on the characteristics of the Amazon economy, I include the presence of a specific activity with large fixed costs for machine acquisition. Farmers with better access to federal roads face higher output prices because they pay lower transportation costs, and, as a consequence, they are more likely to have profits that are large enough to compensate the fixed costs of venturing in the production of varieties with large fixed costs. This possibility of producing highly mechanized crops provides additional profits for the farmer, which is further translated into a higher demand for non-tradeable goods, and a larger expansion of the urban center.

To analyze the empirical implications of the model, I combine two sources of exogenous variations that affected the export oriented production in the region. First, the construction by the Military government of a large federal road system in Brazilian Amazon in the end of the 1960s . The project consisted of a nearly 20,000 km highway to offer the first set of road access between the Amazon and the rest of Brazil. Before the construction of the federal roads, most of the transportation of goods was carried through waterways, and there was very limited alternative routes in-between counties. The main goal of the federal government was to connect State capitals in the region and with the federal Capital. Given this policy design, I follow a growing literature evaluating the impact of transportation infrastructure (Redding and Turner, 2014), and I compare counties in-between nodal cities that happened to be closer to the transportation infrastructure with counties farther away [2].

Second, I explore the increase in commodity prices in the end of the 1990s, which was directly related to variations in the Brazilian exchange rate and the increase in commodity prices in 2007. Variations in the exchange rate were largely determined by the federal government's efforts to reduce inflation, which reached a level of 2000% in 1993. In 1994, a fixed exchange rate pegged to the dollar was adopted, and, in February of 1999, the Central Bank left this policy to adopt a inflation target one with floating exchange rate. As a consequence, the Brazilian currency devaluated in 300% from 1999 to 2003. In 2007, when the Brazilian exchange rate was back to lower levels, the price of commodities in the international markets soared due to China's economic growth. Combined, these shocks on price sustained a decade long period of high profitability in the production of export oriented commodities in the Amazon.

For the empirical section, first, I proceed with an analysis of the reduced form implications of the model. I use decadal census data to analyze differential trends in population and

[2]See (Redding and Turner, 2014) for a review of the literature on the evaluation of the impact of transportation projects.

income between counties according to their access to federal roads. I show that there is no statistical difference between counties close and far from roads before the 1970s in terms of their population, and a limited one until the end of the 1980s. We observe, however, a large divergence between counties close and far from roads in terms of their population and income after the 1990s. Using annual information on commodity production and employment in the urban and rural sector, I show that the expansion of economic activities in terms of commodity production and employment follows closely the annual increase in commodity prices. During this expansion, there is a widespread expansion in commodities with low fixed costs of production (cattle), but only counties close to federal roads specialized in the production of commodities with large fixed costs for capital (soybeans, corn and cotton). Furthermore, I show that, in spite of a large expansion in the rural production of the Amazon, we observe a limited expansion of employment in the rural sector, but a large one in activities typically associated to urban centers (such as commerce, construction, and services).

In a second empirical section, I estimate the structural parameters of the model using a version of the simulated method of moments. The estimated model has a good fit with the data and is able to tie the relationship between roads and the key variables of the reduced form analysis into an unique theoretical framework. Furthermore, in the absence of a source of exogenous variation that provides comparable counties with different production structures, the estimated model allows me to investigate what would have happened in the absence of a capital intensive sector in the region. In the structural model counterfactual, I find that almost all the job generation in the urban centers can be attributed to the presence of a highly profitable activity with large fixed costs.

The results from this article provides three main contributions for the literature in trade and development. First, this is one of the first studies to provide microeconomic evidence of the theoretical mechanisms relating commodity trade to urbanization.[3] Second, the results from this article suggest that trade and the adoption of large-scale techniques for agricultural production may be central elements to understand urbanization in developing countries (Foster and Rosenzweig, 2007, 2004). Third, there is a recent and already large literature evaluating the impact of transportation projects (Redding and Turner, 2014). However, to the best of my knowledge, this study is the first to explore additional conditions that may foster the realization of the benefits of such investments. In particular, I show that the lack of good macro-economic conditions for trade prevented the urbanization of counties near the

[3]There are two recent articles on a similar vein. See Jedwab (2013) for evidence of the impact of commodity production in Ivory Coast and Ghana, and (Fajgelbaum and Redding, 2014) for an analysis of commodity trade and structural transformation in Argentina.

transportation network.

The remainder of this article is organized as following. In the next section, I discuss the evolution of the Brazilian Amazon economy and the main stylized facts related to population, employment, and agricultural production in the region. In section 3, I develop a model to guide the empirical analysis. In section 4, I study the reduced form implications of the model. In section 5, I estimate the model structurally and I follow with the counterfactual analysis. Finally, in section 6, I conclude the article.

2 The Brazilian Amazon Economy

2.1 Earlier Occupation and The National Plan of Integration (1969)

In 2010, around 10% of the Brazilian population lived in the Amazon region. For most of the 21th century, however, the region was largely unnocupied and sparcely populated. Before the 1970s, most of the transportation in the region was carried through waterways, and there was very limited land connection in-between cities (Pfaff et al., 2007). Using data organized by the *Instituto de Pesquisa Econômica e Aplicada* (IPEA) on population from all the census available in Brazil, we can draw the population's trend from the very beginning of the 21th century and visualize the importance of the federal programs in the Amazon initiated during the 1970s.

The dotted line in figure 1 shows the trend in the Amazon's share of the Brazilian Population. Before the 1960s, the Amazon region represented around 4% to 6% of the Brazilian population across different census years. Furthermore, the solid line shows that, before the 1960s, at most 3 million people lived in the Amazon region, which has more than 4 million km^2 (roughly half the area of the United States). Interestingly, the population data captures the period of economic expansion in the region around 1910, when there was an intense economic expansion associated with the export of rubber for the automobilistic industry in the United States. During this period, the region established several central cities that would later on constitute the bases upon which the federal government designed a transportation project for the region.

In the end of the 1960s, Brazil was going through a period of unprecedented economic growth (with an average of more than 10% a year between 1968 and 1973). During these years, the Military government was running a dictatorship and financed a series of projects to integrate the Amazon region with the rest of the country. The central one was the National Plan of Integration (1969). The goal of the federal government with this plan was twofold: first, to occupy the Amazon region and reduce the likelihood of its occupation by neighbor countries; second, to reduce the flux of migrants to the industrializing centers of the southeast. The central piece of the plan consisted in the construction of a nearly 20,000 km highway system to connect State Capitals in the Amazon between themselves, with the Federal Capital, and with other regions of Brazil [4].

Figure 4 clearly shows that the federal intervention during the 1970s was a watershed for the occupation of the Amazon. From 1970 to 2010, the total population of the region increased in roughly 500% (from 3 to 18 million people). While part of this population's expansion may be attributed to a demographic transition in the region, the large magnitude of the expansion indicates that there was an intense in-migration to the Amazon. Furthermore, other regions in Brazil were also going through a demographic transition during the period and, still, the Amazon's share of the Brazilian population increased from 4% in 1960 to 10% by the end of 2010, when all the regions had already completed their demographic transition.

2.2 Shocks on Commodity Price

After the 1970s, and until the end of the Military's era in the middle of the 1980s, the federal government made several attempts to promote the agricultural production in the Amazon. Farmers who migrated to the region were able to acquire large plots of land for extremely low prices [5]. However, the poor quality of the soil and the climatic conditions prevented them from producing staples that were typical in their native regions (such as coffee and wheat). For most of the initial years of occupation after the 1970s, the wood and the mineral industries remained as the main sources of income for the recently migrated population.

During the 1990s, the conditions for the establishment of a large agricultural industry in the Amazon were met due to a series of technological improvements and macro-economic changes. First, new seed varieties developed by the National Agropecuary Research Insitution (EMBRAPA) allowed the production of soybeans in low latitude areas (Vera-Diaz et al., 2008). Second, mouth vaccines controlling disease enabled ranchers to attend the southeast and the international requirements for meat consumption (Walker et al., 2009). Third, there was an increase of roughly 200% in the price of export oriented commodities in terms of the

[4] See figures A1 and A2 in the appendix for the original maps of the plan.

[5] There are several histories of colonization companies that actually gave huge plots of land for free in order to establish an initial community in the region. Interestingly, the slogan of the government to attract people was *"Terra sem gente para gente sem terra"*, which translated becomes "land without people for people without land".

local currency which made the production of these commodities a highly profitable activity (Richards et al., 2012).

Variations in commodity prices faced by farmers in the region, which is a central source of exogenous variation for this study, is directly related to variations in the Brazilian exchange rate and the international price of commodities. This claim is largely supported by data. The dotted line in figure 2 shows the evolution of the international soybean price given by the World Bank time series multiplied by the exchange rate in Brazil, obtained from the *Banco Central Brasileiro* (BACEN). The solid line presents the price offered by local retailers to farmers in the county of Primavera do Leste (one of the main rural producing counties in the region) measured by the *Instituto Mato-Grossense de Economia Agropecuária* (IMEA). In spite of the level differences between these prices, the variation follows each other closely.

Using annual data on agricultural production from *Produção Agrícola Municipal* (PAM), we observe that, between 1999 and 2004, when the price of commodities soared, there was a large expansion in the production of export oriented commodities. Figure 4 shows that both the total production of cattle and the sum of the land area producing soybeans, cotton and corn follows the increase in commodity prices. Furthermore, in spite of occupying a similar area in the beginning of the 1990s, we do not observe any effect on the production of all the other 15 crops combined [6]. This large expansion in agricultural production did not come without environmental costs. Between 1999 and 2010, there was an intense deforestation in the Amazon (Richards et al., 2012). Between 2001 and 2004, when the annual deforestation rates were at its highest rates in the past 20 years, almost 100.000 km^2 of the Amazon forest was cleared for agricultural production (Morton et al., 2006). Note that specific technological improvements may have provided important conditions for the expansion of these commodities, but they are less likely to explain the timing of expansion in both the cattle and in the soybeans production (besides the fact that they might be endogenous to the profitability of production). Furthermore, aside the expansion in these commodities, several other export oriented crops were affected by the exchange rate shock between 1999 and 2004 in other regions of Brazil. In the appendix, I show a more detailed analysis of the expansion of export-oriented crops between these years.

In spite of a drastic expansion in the agricultural production of the region, annual administrative data from the *Registro Anual de Informações Sociais* (RAIS) on all the formal employment in the region indicates a larger expansion in the generation of jobs in sectors

[6] Other crops in figure 3 includes: manioc, wheat, rice, peanuts, bananas, potatoes, cocoa, coffee, sugarcane, onion, beans, tobacco, orange, tomatoes, and black pepper.

typically associated with urban centers. The solid line in figure 4 combines commerce, construction, and service sectors, which I define as the urban sector for the remainder of this article. The dotted line presents the amount of jobs in the agropecuary, defined as the rural sector. Between 1990 and 1998, employment in the urban and the rural sectors follow a parallel trend. However, after 1998, we observe a divergence between their trends [7].

To summarize these stylized facts, I present in figure 5 all the relevant variables from this section normalized according to their mean and deviation across the years. In 1998, before the commodity price increase, most of the variables were in levels comparable to those in the beginning of the 1990s. In the first five years following the devaluation of the currency in 1999, all the variables achieved unprecedented levels that were sustained until the end of 2010.

2.3 Technological Characteristics of the Agricultural Production

Before moving to the next section, I present central characteristics of the agricultural production in the Amazon which will be used to define the theoretical model.

First, as indicated in figure 3, a few rural activities dominated the dynamics of the agricultural production in the Amazon after the 1990s: soybeans, corn, cotton, and cattle. One important characteristic of production is the complementarity between soybeans, corn, and cotton, but the substitutability between soybeans and cattle. While soybeans work as the leading activity, farmers also produce corn in a second harvest season within the year to generate additional profits [8]. Furthermore, the same machines used to harvest soybeans can be applied on the production of corn. Cotton is usually produced as a additional activity and to diversify the risk and increase profits according to the expected prices.[9] In figure E1 in the appendix, I show a clear positive correlation between soybeans, corn and cotton

[7] The industrial production in the region outside the city of Manaus is very incipient. Manaus is an exception, because the federal government created a free tax zone in the city to give incentives for people to migrate. As a capital of one of the States, this city was excluded from the sample for the following analysis.

[8] Harvesting twice within a year is typically called as double cropping. However, according to the large majority of farmers interviewed in the region, the second season is much less productive and works at most as a limited complement to their production. As such, this technique seems not to be doubling the endowment of land in practice. In general, farmers would have to spend resources covering the land with some type of grass vegetation to protect the soil nutrients from the sun between the months of June until September, when the cultivating season restarts. They take this opportunity to use the land to cultivate something that can actually be sold and generate some additional profits given that labor and capital was already hired for the year. In general, a full typical corn production is not adopted, but a less expensive variety of seed called *Milheto*.

[9] The production of cotton, however, is very risk, which prevents farmers from producing cotton as a leading activity.

at the county level, but a negative one with cattle production. The data also shows that a few counties have no production of corn and cotton in the presence of a substantial soybean production. There are several counties, however, with a small production of soybeans but a large production of cattle, which is consistent with a substitutability between these activities. To simplify the following investigation, I combine the land area use for soybeans, cotton, and corn into one.

Second, there is a key difference between the production of cattle and soybeans: the latter requires large amount of investments in machines and infrastructure such as harvesters and seeders, which are generally imported from abroad. Also, the production of soybeans requires large amounts of imported fertilizers and limestone to neutralize the acidity of the soil and enrich it with nutrients for the crops in the beginning of the season. Cattle production, however, requires much lower investments in machines, and a very limited preparation of the soil. In fact, in table D9 in the appendix, I use data from the Agricultural Census in 2007 to show that there is a stronger correlation between large tractors and soybeans with respect to cattle, but the inverse is true for small tractors (where large tractors are defined as those with more than 100 horsepower). The production of cattle, however, generally precedes the production of soybeans after clearing the land. Therefore, one could argue that cattle production takes some extra costs for clearing the land. The available data does not allow a complete assessment of all the fixed costs involved in the production of soybeans and cattle. Still, it seems plausible to assume that the fixed costs of producing soybeans are larger than the ones related to cattle production. In the structural estimate of the model, where I allow these fixed costs to be arbitrarily different, I find that the production of soybeans have a larger fixed costs than cattle ranching.

3 A Theory for the Amazon Economy

The goal of developing a theoretical model for the region is twofold: first, to study the relationship between the endogenous variables and the sources of exogenous variations in a set of regressions; second, to estimate the structural parameters of the model to explore counterfactuals that would not be feasible under a reduced form approach. I take advantage of the specific characteristics of the Amazon economy to impose assumptions consistent with the data that will keep it tractable for empirical estimation. The main dynamic force in the model is generated by variations in the level of commodity prices common to all counties, and a spatial variation that is given by distance to federal roads. These two elements constitute the main sources of exogenous variation for the empirical analysis in the following sections. Here, I present the main equations of the model. In the appendix, I provide its full derivation.

3.1 The model

The production is divided into a rural and an urban sector. In the rural one, each county i has one immobile representative farmer who decides between producing two export oriented activities: soybeans (S) or cattle (C).[10] Soybeans represent the modern large scale production system that also includes corn and cotton. Because other crops are less important for the economic dynamics of the region and to the patterns of land use (as shown in figure 3), I abstract from them to simplify the model. The land use decisions are based on the Span-of-Control model from Lucas (1978). The farmer decides what to produce in each piece of land ℓ according to prices, technology, the land quality $z_{(\ell)}$, and the fixed cost associated with each commodity (f_S and f_C). Each county has a measure of land T_i such that $\ell \in (0, T_i]$. The urban sector produces a non-tradeable good that only takes labor as input. This sector represent restaurants, education, housing, and other nontradeable activities. Workers in both the rural and the urban sector can migrate between activities and between cities as in typical open city models (Roback, 1982). This is consistent with patterns of migration in Brazil, where 34.5% of the population lives in a city different from the one where they were borned (Census, 2010).

Consumers

Each county i is home to N_i mobile workers and a representative immobile farmer. Both of them have cobb-douglas preferences for a manufacturing good (c_M) produced elsewhere and non-tradeable goods (c_U) produced in the urban center of each county. I assume that soybeans and cattle are not consumed directly by these counties and that, for final consumption, they have to be processed elsewhere. The utility function is given by:

$$U = \delta c_M^{\alpha_M} c_U^{\alpha_U} \tag{1}$$

where I set $\delta = \frac{1}{\alpha_M^{\alpha_M} \alpha_U^{\alpha_U}}$ for algebraic convenience and without loss of generality.

The price faced by consumers depend on the distance a county is from a federal highway.

[10]The results of the model are not dependent on this assumption. I could assume, instead, that there is an arbitrary number of farmers, each owing a share of the county area. As long as the preferences are homothetic, the consumption of each farmer can be aggregated into one single farmer. What is important for the results, however, is that farmers are immobile.

Manufacturing goods have to pay an iceberg cost according to an iceberg decay parameter γ to reach the county. Therefore, manufacturing goods are more expensive farther away from roads. I assume the transportation costs to follow an exponential decay.

$$p_{Mi} = p_M^* e^{\gamma Droads_i}$$

Where p_M^* is given exogenously and $Droad_i$ represents the distance a county i is from a federal road. The price of non-tradables, p_{Ui}, is given endogenously by supply and demand within the city.

Workers may leave the county and migrate to other regions of Brazil if a minimum utility is not achieved. They must be indifferent between living in county i, and an exogenous outside option providing utility V. Manipulating the indirect utility function provides the following non-migration condition.

$$V = \frac{w_i}{e^{\alpha_M Droads_i} P_i} \tag{2}$$

Where $P_i = (p_M^{*\alpha_M} p_{Ui}^{\alpha_U})$.

Rural Sector Production

Farmers may produce two export oriented activities: cattle or soybeans. The latter represents the modern activity with large fixed costs, and the former the traditional one with low capital requirements. The technologies for each are given by the following equations:

$$\begin{aligned} y_{S(\ell)} &= a_S z_{(\ell)} l_{(\ell)}^{\phi} \\ y_{C(\ell)} &= a_C z_{(\ell)} l_{(\ell)}^{\phi} \end{aligned} \tag{3}$$

$y_{S(\ell)}$ and $y_{C(\ell)}$ are the outputs for producing in the unit of land ℓ, $l_{(\ell)}$ is the amount of labor employed, a_S and a_C are parameters related to the productivity of each activity, and $z_{(\ell)}$ represents the distribution of land quality within the county - which could be associated with differences in soil quality, slope or distance to the urban center. For simplicity, I assume that both technologies have the same labor intensity factor ϕ. [11] Note that we can think of these technologies as constant returns to scale where land intensity is given by $1 - \phi$ and land employed is equal to 1. The production of each activity requires a fixed cost f_S and f_C

[11] Note that, even though I assume the same ϕ, I do allow different technologies to have different demand for labor given the use of land. Because I am holding the land in piece ℓ to be 1, the parameter a_S generates differences in the demand for labor for each given piece of land.

such that $f_S > f_C$.

The prices faced by farmers, which is a central for the dynamics of the model, take the following functional form:

$$\begin{aligned} p_{Ci} &= P_C^* e^{-\gamma_C Droads_i} \\ p_{Si} &= P_S^* e^{-\gamma_S Droads_i} \end{aligned} \tag{4}$$

P_C^* and P_S^* are the international price adjusted by the exchange rate for soybeans and cattle respectively. Variations in P_C^* and P_S^* generate level variations that are common to all counties in the Amazon region. $e^{-\gamma Droad_i}$ represents the spatial variation in the level of prices and does not change over time. These two elements combined capture the two sources of exogenous variations that I explore in the empirical section of the article to evaluate the effect on the endogenous variables of the model.

The profit function for each piece of land ℓ is given by:

$$\pi_{(\ell)} = \begin{cases} \max\limits_{l_{(\ell)}} \{p_S e^{-\gamma_S Droad_i} a_S z_{(\ell)} l_{(\ell)}^{\phi} - w_i l_{(\ell)} - f_S\} & \text{if } y_{S(\ell)} > 0 \\ \\ \max\limits_{l_{(\ell)}} \{p_C e^{-\gamma_C Droad_i} a_C z_{(\ell)} l_{(\ell)}^{\phi} - w_i l_{(\ell)} - f_C\} & \text{if } y_{C(\ell)} > 0 \\ \\ 0 & \text{if } y_{S(\ell)} \leq 0 \text{ and } y_{C(\ell)} \leq 0 \end{cases}$$

The landowner maximizes the profit in each piece of land according to the potential profits of each activity and the possibility of not producing anything:

$$\pi_{(\ell)} = \max\{\pi_{S(\ell)}, \pi_{C(\ell)}, 0\} \tag{5}$$

This framework provides a clear economic motivation for decisions on land clearing and the consequences for deforestation. As profits of agricultural goods increase, the opportunity cost of the land increases and farmers clear the forest to produce agricultural goods. In the model, the fixed cost of setting up agricultural activities prevent farmers from clearing all the land of the county for production.

Urban Sector Production

The urban sector produces non-tradeable goods according to a constant return to scale technology with no land requirement:

$$y_{Ui} = a_U l_{Ui} \tag{6}$$

Where y_{Ui} is the total output, a_U is the productivity of the sector and l_{Ui} is the amount of labor employed.

Competitive Equilibrium

Given a vector of prices $p = (p_S^*, p_C^*, p_M^*)$, a distance from roads $Droad_i$, an exogenously given indirect utility function V, and a land endowment T_i, the competitive equilibrium in each county i is characterized by a population N_i, labor demands $\{l_\ell\}_{\ell \in (0,T_i]}$ and L_{Ui}, and patterns of land use such that:

1. *(Non-migration Condition) Workers maximize utility*

$$V_i \geq V \tag{7}$$

2. *(Rural Sector Production) Landowners maximize profits in each plot of land* ℓ

$$\pi_{(\ell)i} = \max\{\pi_{S(\ell)i}, \pi_{C(\ell)i}, 0\} \tag{8}$$

3. *(Urban Sector Production) Firms maximize profits*

$$\pi_U \leq 0 \tag{9}$$

4. *Labor market clears*

$$\int_{\ell \in T_i^S} l_{S(\ell)} d\ell + \int_{\ell \in T_i^C} l_{C(\ell)} d\ell + L_{Ui} = N_i \tag{10}$$

5. *Trade is balanced*

3.2 Comparative Statics

To study the comparative statics, I define a set of parameters that makes the partition of land in the model consistent with some basic facts of the data. In the structural estimates, however, I allow the parameters to be arbitrarily different.

Proposition 1 *(Partition of Land) Define* z_i^{SC} *as the plot* ℓ *where the farmer would be indifferent between producing cattle and soybeans, and* z_i^{CF} *for the plot* ℓ *where the farmer would be indifferent between producing cattle and clearing the forest. If the conditions* $p_S > p_C$, $f_S > f_C$, *and* $z_i^{CF} < z_i^{SC}$ *are satisfied, then the area of the county can be partitioned in three where soybeans is produced in* ℓ *such that* $z_i^{CF} < z(\ell) < z_i^{CS}$, *cattle is produced in* $z_i^{CF} < z(\ell) < z_i^{CS}$, *and forest is left uncleared for* $0 < z(\ell) < z_i^{CF}$.

Proof In appendix.

The parameters required for the proposition to hold are plausible: soybeans must be relatively more profitable than cattle ranching ($p_S^* a_S > p_C^* a_C$) and the fixed costs of producing soybeans must be larger $f_S > f_C$ such that there is a tradeoff between the higher *gross profits* derived from soybeans and its fixed costs. Further, I need cattle to be sufficiently profitable with respect to soybeans, otherwise, we could still have full specialization of the county on soybeans (see appendix for details on other possibilities of partition of land). Condition $z_i^{CF} < z_i^{SC}$ is sufficient for the county to have a positive production of soybeans and cattle.

The partition of land into z_i^{CS} and z_i^{CF} is generated by a combination of differences in fixed costs and the complementarity between land quality and prices. One can show that the cross derivative between prices with respect to land quality is positive. Therefore, as we go from a plot with low quality to higher quality ones, this movement has a larger effect on the *gross profit* of soybeans than on the one for cattle. However, the difference in fixed costs prevents farmers from producing soybeans in every plot. There will be a range of land quality between z_i^{CF} and z_i^{CS} where farmers prefer to produce cattle in spite of the smaller *gross profits*, because *profits net of fixed costs* are larger for cattle production (given that $f_S > f_C$).

Following the parameters in proposition 1, I first show a set of results for the model that should hold given the level of prices p_C^* and p_S^*. Second, I study how variations in p_C^* and p_S^* affect the production of counties given the distance from roads $e^{\gamma Droads_i}$. Given the general equilibrium effects within each county, the model does not allow for closed form partial derivatives. In the appendix, I show a set of figures with simulations where I study the comparative statics of the model.

Variation in Distance to Roads given International Price

For a given international price level p_C^* and p_S^*, decreasing distance to roads increase the land area dedicated to soybeans, the total income of the county, the population, and the employment in the urban center. However, the effect is ambiguous for the land used for pasture and may follow a U-shape depending on the parameters of the model.

The intuition for the possibility of a U-shape relationship between cattle production and distance to roads is the following. When we increase distance to roads, the price faced by farmers in the county for commodities is smaller due to the transportation costs. At the marginal land producing soybeans ($\bar{z}_i$), cattle becomes more profitable because of its lower fixed costs, and there is a reduction in the area producing soybeans but an expansion in the one producing cattle. However, at the marginal land where the farmer is indifferent between clearing the forest and producing cattle ($\underline{z}_i$), an increase in distance to roads makes the production of cattle unprofitable, decreasing its use of land area. Therefore, at one margin, there is an increase in the amount of land producing cattle while, at the other, there is a decrease. The effect of distance on the production of cattle is thus ambiguous and depends on the magnitude of the effect on each margin

Variation in International Prices given Distance to Roads

When international prices p_C^* and p_S^* increase, there is a positive impact on the land area used to soybean production, the total income in the city, the population, and the generation of jobs in the urban centers. The effect of the price increase is magnified by the access to federal highways, i.e., counties closer to roads have an even larger expansion on these endogenous variables. The impact of the increase in commodity prices on cattle production, on the other hand, is heterogeneous and depends on the initial production before the price shock.

Here, the main intuition behind the heterogeneity is that cattle ranching precedes the production of soybeans. If a county has an already large production of cattle before the positive shock on commodity prices, there will be an expansion of soybeans over land used for pasture, with limited expansion of pasture over forest areas (which are the remaining areas of the county where the productivity of land is already too low to justify even the production of a low fixed cost activity). However, in counties that have a small production of both soybeans and cattle production in an initial period before the shock, there will be first an expansion of the activity with lower fixed costs over forest areas, i.e., cattle. Note that the model is able to rationalize one of the main stylized facts discussed in the literature about deforestation in the Amazon (Barona et al., 2010): a weak correlation between soybean production and deforestation, but a strong correlation between cattle ranching and deforestation.

4 Reduced Form Evidence and the Dataset

4.1 The Dataset

For the empirical investigation, I combined information from a series of different sources to organize a dataset at the county level with economic production variables and geographic characteristics related to agricultural productivity. A thorough description of the source of information for each variable is given in the appendix. Figure 6 presents the main structure of the dataset, where I present the county boundaries in 2010, the federal highway system constructed, and the project lines according to the documentation in the National Plan of Integration of 1969.

The information available allowed me to construct two panels at the county level. One that goes back to the 1950s with decadal information on population and income. And another that starts in 1990, which includes annual information on commodity production and jobs in the formal sector. Because several counties were created between 1970 and 1990, I used information available from the Brazilian census bureau (IBGE) on the date of creation of each county to improve consistency of the boundaries over time [12].

I included in the dataset several controls that are potential determinants of agricultural productivity. First, I brought information on rainfall and temperature in September obtained from the WorldClim (Global Climate Data). This month is a crucial one for the beginning of the harvest season, when the rainfall patterns determine the decisions for seeding the plot. Furthermore, I added information on corn and soybeans suitability from the FAO Global Agro-Ecological Zones, which is calculated based on several geographic characteristics of the region and has been used in recent articles (Bustos et al., 2013) [13]. Finally, I also calculated the average slope of the county using information on altitude provided by the NASA'a Shuttle Radar Topography.

The central explanatory variable in the analysis is distance of the centroid of each county to a federal highway constructed during the 1970s. One caveat about this measure is that I am implicitly assuming that being farther away from a highway does not make a county closer to another transportation system. To avoid this problem, I excluded all the counties

[12] I used the information provided by the *Instituto Brasileiro de Geografia e Estatística* to create an algorithm that aggregates information of each year according to the set of counties that were already created in the initial year of the dataset (i.e., 1950 and 1990). The dataset provided by IBGE contains information on the date of creation of each city as well as the cities from which each city was derived.

[13] The FAO Global Agro-Ecological Zones data provide suitability for different levels of input use. Because of the large-scale pattern of production in the region, with several farmers producing in farms with over 500 ha, I included information on high input.

further than 300 *km* from a highway from the analysis (around 1% of the sample) [14]. Another caveat is that federal highways in the Amazon have different qualities. Actually, the quality *on federal highways* is far more heterogeneous than the quality *off federal highways*. Furthermore, in the sample, most of the roads were paved several years after the implementation of the National Plan of Integration. Therefore, the quality of the road is more likely to be correlated with further economic expansion of the counties than its initial condition. In spite of that, to address this concern, I control in the empirical section for region fixed effects, which makes counties under comparison more likely to face the same quality of the transportation network.

4.2 Empirical Strategy

For an initial evaluation of the differential impact of the commodity price shock on the counties of the region according to their access to the federal highway, I use the following specification:

$$Y_{it} - Y_{i(t-1)} = \theta Close_i + X_i\beta + \alpha_j + \epsilon_i \tag{11}$$

Where the subscript t denotes a year, j a region fixed effect, and i a county. $Close_i$ is a dummy equal to 1 if the county is closer than 100 *km* from a federal highway, X_i is a set of control variables, and ϵ_i is the unobserved component. θ is the parameter of interest.

Equation 11 is estimated for different time periods t, for example, $t = 1991$ and $(t-1) = 1970$. In this case, I estimate the impact of the federal highways on the economic activities before the shock in commodity prices. For $t = 2010$ and $(t-1) = 1991$, I estimate the impact of being close to the federal highways after the commodity price shock. For the dependent variables with annual frequency, I am able to explore with more precision the timing of the price shock. For these variables, I set a regression for $t = 1999$ and $(t-1) = 1994$, and another for $t = 2004$ and $(t-1) = 1999$. This time period structure, with periods before and after the price shock, leads to the final specification to estimate θ:

$$\overbrace{(Y_{it} - Y_{i(t-1)})}^{\text{After Price Shock}} - \overbrace{(Y_{i(t-1)} - Y_{i(t-2)})}^{\text{Before Price Shock}} = \theta Close_i + X_i\beta + \alpha_j + \epsilon_i \tag{12}$$

To identify the parameter θ on equation 12, I need to assume the differential impact

[14] Another problem is that there are only counties in the far northwest of the Amazon that are farther than 300 *km* from a highway network. To some extent, these counties are in very remote areas, with almost no meaningful economic production, and provide very little variation for the analysis of this study.

of prices on counties that are close to highways to be uncorrelated with unobserved shocks that coincides with both the timing of the price shock, and the location of counties. Note that year specific technological improvements do not violate the identification assumption, as long as they are sufficiently accessible for farmers who are both close and far from highways. However, the assumption might be violated if being close to a road is correlated with the suitability for production. For example, if the route of highways were constructed to avoid steep areas, counties close to them would be flatter and more suitable for a mechanized production. Once we have the price shock, these counties would potentially increase more their agricultural output with respect to counties farther from the highways because of their better geographic characteristics for production.

An analysis of the control variables can shed light on the plausibility of the identification assumption. Table 1 presents the geographic variables included in X_i used as controls in the estimation of equations 11 and 12. In columns 2-3 and 5-6, I present the relationship between the control variables and distance to roads in elasticity to facilitate the interpretation and the comparison of the coefficients. In general, panel A of table 1 presents a weak correlation between distance to roads and the geographic variables in both statistical and economic terms. The area of the county, however, is an exception and has a large and significant correlation with distance to federal highways. In part, this is due to the mechanical correlation between distance to federal highways and State capitals generated by the design of project. Counties closer to the State Capitals are generally smaller, and counties close to the federal highway system ended up being those closer to State Capitals. In spite of this, the combination of a weak correlation with geographic variables and a strong one with distance to State Capitals reinforces the fact that the main determinant of roads was the government‘s goal of connecting State Capitals. Given this framework, in all the estimates of equation 11 and 12, I control for a cubic polynomial on distance to State Capitals.

Another evidence that supports the validity of the identification assumption comes from an investigation of the population before the construction of the federal highways. In table 1 panel B, while the full sample in column 2 presents a moderate and statistically significant correlation between distance to federal highways and population levels before the implementation of the transportation project, once we drop State capitals from the sample in column 3, the magnitude of the correlation reduces to almost zero in 1950 and by half in 1960. This provides evidence that there was limited connection in-between counties, and that the route of the federal roads were not constructed to benefit specific counties with larger economic production. For the following analysis, I drop State Capitals from the sample and compare only counties in-between.

Several recent articles on the evaluation of transportation projects use the least-cost spanning tree network as an instrument for actual distance to federal roads (Faber, 2014). The main goal of this instrument is to obtain a source of variation that is uncorrelated with the demand for transportation, which could lead to a simultaneity bias with the expansion in economic activities. In the present context, however, it is unlikely that the federal government planned the route of the roads according to an economic boom that occurred around 30 years later. Furthermore, this instrument would mechanically increase the correlation between distance to roads and geographic characteristics. Given that agricultural production is at the core of the Amazon's economy, this would raise concerns about omitted variable bias.

Another option of instrument would be to use the project lines in the documentation of the plan (Baum-Snow, 2007). Table 1 panel C presents a strong correlation between actually constructed highways and distance to project lines, which leads to a strong first stage. However, investigations of the correlation between project lines and the covariates in table 1 does not provide evidence that the instrument is less correlated with potential omitted variables. Actually, when highways are instrumented by the project lines, the magnitude of the coefficients are very similar, but statistically weaker, which led me to opt for the ordinary least square approach.

4.3 Results

In this section, I present the results from estimates of equation 11 and 12 using the main endogenous variables from the theoretical model as dependent variables.

Population and Income (1950-2010)

As initial evidence of the impact of the federal highway system in the Brazilian Amazon and the commodity price shock, figure 7 presents the difference between counties close and far from federal roads on population and income for each census year. From 1970 to 1980, in the first 10 years after the beginning of the implementation of the National Plan of Integration in the region, there is a divergence in trends between counties close and far from federal roads. This divergence process slows down during the 1980s. In the following 20 years after the 1990s, however, we observe an intense new process of divergence between counties, which corresponds to the period when the economic production of commodities for external markets expanded in the region.

Table 2 presents the results from the specification 11 for the time lag that captures the first 20 years after the implementation of the project, and another that captures the following divergence after the 1990s. Column 1 shows a large impact of the construction of roads on population. Given the baseline level (reported in the table) the magnitude of the coefficient reinforces the fact that the impact is large and that the initial level of occupation in the Amazon was low. When we control for a series of geographic variables and initial conditions in column 2, the coefficient remains relatively stable. Note that a F-test of the joint statistical significance rejects the null-hypothesis of no joint effect of the control variables at the 1% level (*p-values* are reported in the table). This is a systematic pattern across most of the specifications in this reduced form section, which shows that the controls have a strong predictive power on the dependent variable. In column 3, where I present the results for the period between 2010 and 1991, we can see a strong divergence process. The impact of being close to roads is much larger during this period. In column 5, I estimate equation 12 to assess the impact of commodity prices on the trends between 1991-1970 and 2010-1991. The effect of being close to roads confirms an intensification of the divergence after the 1990s. Also, note that the controls for geographic characteristics have a limited effect on the magnitude of the coefficients across all the specifications, which reinforces the fact that distance to roads is orthogonal to characteristics related to agricultural productivity.

The divergence between counties close and far from roads after the 1990s is clearer when we look at income, which is presented in panel B of table 2. There is a very limited divergence between 1991 and 1970. However, a strong one after the 1990s, when the export oriented agricultural industry emerged in the Amazon. In general, all the qualitative conclusions from population can be extended for the analysis of income.[15] Coefficients are large in magnitude, stable across specifications, and they suggest a large divergence after the 1990s, more than 20 years after the construction of the federal road system. However, the fact that total income did not follow closely the population expansion between 1970 and 1990 is puzzling. In part, this could be due to the fact that in the first 20 years after the construction of roads, there was still a large proportion of the population living from a subsistence production, which is poorly captured by a monetary measure such as income. Unfortunately, good measures of subsistence production is not available for earlier years. Still, for the scope of this article,

[15]Here, I use total income because it is more closely related to total population than income per capita. Furthermore, the model does not provide any theoretical prediction on the impact on income per capita. Mainly, because of the high migration rates across counties, the gap in income per capita should be small. In fact, in the next section, I show evidence that wages are very similar across different counties independent of their distance to federal highways.

it seems sufficient to document that using both population and income variables to analyze trends in the economy of the Amazon leads to the same conclusion: that an additional shock occurred after the 1990s that intensified a divergence in economic production between counties close and far from federal roads.

Commodity Production (1990-2010)

The data on population and income does not allow me to explore with more precision the timing of the increase in commodity prices. Here, I use annual data on commodity production to evaluate the divergence in trends between the periods of 1999-1994 and 2004-1999.

As an initial assessment of the differential trends, figure 8 shows the difference in the average production of commodities between counties close and far from federal highways. The solid line in the figure presents a clear divergence between 1999 and 2004, when the commodity prices reached a historical peak. In spite of the aggregate cattle production following the trend in soybeans during this period (see figure 3), we do not observe any differential trend here. According to the model, this could be attributed to the highly heterogeneous impact of the price shock on cattle production. To investigate this point, figure 9 shows the trends in cattle production in counties that reached 2010 being large producers of soybeans (above the 95th percentile), medium producers (between the 95th and the 90th percentile), and the rest of the counties in the Amazon (below the 90th percentile). Between 1994 and 1999, trends in the production of cattle follow a relatively similar path across different counties. However, after 1999, counties that reached 2010 as large soybean producers have a clear decrease in their production of cattle, while counties with a small production have an expansion in their cattle production of more than 60%. Interestingly, the rank also changed during this period. Counties that became large producers of soybeans in 2010 went from being large to small producers of cattle before and after the commodity price shock. Counties that reached 2010 as small producers of cattle, however, went from being small to large producers of cattle.

Table 3 shows the estimates of equation 11 and 12 using commodity production as the dependent variables. As suggested in figure 8, there is a large impact of prices in the production of soybeans between 2004 and 1999 as shown in column 3 and 4 of panel A. Between 1994 and 1999, however, the coefficients suggest a process of convergence between the production of counties close and far from highways. Taken together, the estimated effect for the period after and before the shock suggest a divergence in the soybean production of counties close and far from roads that coincides with the period when the price of export oriented commodities were increasing. The difference in trends, presented in columns 5 and

6, confirms a divergence process in soybean production. In general, the statistical properties of the coefficients are good: they are statistically significant at the 5 % level, and they are not affected by the inclusion of controls. The magnitude of the coefficient is also substantial when we look at counties far from roads in the initial period (given by the baseline).

When we look at all the 15 remaining crops in panel B, the results are much smaller in economic terms, and statistically not different from zero. This suggests that during this period, there was a very limited variation in the production of other crops besides soybeans and cattle. This result support the modeling decision to abstract from these crops. The results for cattle production, which are presented in panel C, are unstable. [16]. As discussed above, this can be attributed to the highly heterogeneous impact of prices on cattle production.

Employment in the Urban and Rural sectors (1990-2010)

So far, the reduced form section has shown results that are consistent with the model: the shock on commodity prices has a positive impact on production which is magnified by the proximity to federal roads. In other words, when conditions for exporting commodities improved, counties closer to the transportation system were disproportionately benefited by the facility to export their products. Here, I evaluate how much of this effect is translated into an expansion of the urban centers in the Amazon.

As in the previous analysis, I begin by showing the difference between counties close and far from federal highways over time in figure 10. Again, the timing of expansion in the generation of jobs follows the increase in commodity prices in the region. Figure 1A indicates that the difference between counties close and far from federal highways rose after 1999. However, when we look at both of them in the same axis in panel B, we can clearly see that the magnitude of the expansion in the urban sector was much larger than in the rural one. This result supports the fact that the commodity expansion had an effect on the generation of jobs in the urban centers of the Amazon. In spite of the agricultural production being at the core of the economic dynamics of the region, the impact of the commodity price shock had a far larger impact on the generation of jobs out of the rural sector.

Table 4 confirms the evidence from figure 10. Columns 1 and 2 present the results for the period before the price shock. Between 1994 and 1999, there was a very limited expansion of jobs in both the urban and the rural sector. On the other hand, the estimated effect for

[16] Unfortunately, I can not proceed with a similar analysis as the one in table 3 because a dummy equal to one if the county ended up producing soybeans is highly endogenous. When I proceeded with such empirical exercise, the coefficients were highly unstable across specifications.

the period between 2004 and 1999 is much larger, with a greater magnitude for the urban sector. The estimated effect of the commodity prices on the change in trends in column 5 and 6 show the same patterns.

4.4 Robustness

In this section, I proceed with additional exercises to address potential concerns on the interpretation of the results.

Non-linear effects

To increase statistical power and simplify the interpretation of the results in the previous section, I adopted the simpler approach of defining a dummy variable that indicates whether a county is close or far from the federal highway system. Here, I show that non-parametric estimates confirm the reduced form effect. Figure 11 shows non-parametric regressions adjusted by a cubic polynomial on the distance from State Capital using the procedure suggested in Robinson (1988) [17]. The non-parametric estimates reveal that part of the results could be driven by a set of counties that are very close to roads. In fact, some counties that are just on the federal highways became important service centers for the region. Next, I study the robustness of the results when I drop these counties from the sample.

Dropping counties specialized in the service sector

The theoretical model in this article does not account for the possibility of some counties becoming service centers for the region. In part, most of the public institutions that provides documentation and environmental permits for production are located in the State Capitals of the region, which are already excluded from the sample. However, some counties along federal highways became centers where farmers obtain services such as machine repairs and bank loans. This is potentially an important issue for the interpretation of the results through the lens of the proposed theoretical model. To deal with this concern, I present in table 5 the results from the main endogenous variables in the model dropping every county with the centroid closer than 25 km from a federal highway. Because I lose around 20% of the sample, I estimate the most parsimonious specifications where I control for a polynomial of

[17] The procedure resembles a Frisch-Waugh-Lovell theorem. First, I non-parametrically regressed the dependent variables and the controls against distance to roads and saved the residuals. Then, I regressed the residuals of the dependent variables against the residuals of the controls ones to estimate a set of parametric coefficients, as in the Frisch-Waugh-Lovell theorem. Finally, I run the non-parametric regression of the dependent variables subtracted from the predictions of the control variables using the estimated parameters.

distance to State Capitals. Table 5 panel A shows the results from the baseline estimates, and in panel B the results when we drop counties close to the federal highways. In spite of excluding 20% of the sample, which are counties that according to the non-parametric regression could affect the results, the point-estimate results are statistically the same and the qualitative interpretation is largely unaffected. While some counties may have become important centers for the Amazon, this does not seem to be a first order issue driving the results.

5 Structural Estimates

[18] What would have happened to urbanization in the Brazilian Amazon in the absence of a capital intensive crop? To answer this question with a quasi-experimental design, one would need a source of exogenous variation to generate comparable counties with and without the production of soybeans. This is clearly unfeasible in the present setting, because soybean production is highly endogenous to suitability conditions, access to transportation, and commodity prices. However, under stronger assumptions, a structural estimate of the model can provide us with counterfactuals to answer this question. Furthermore, a good fit of the estimated model with the data would reinforce the plausibility of the mechanisms explored in this article. In this section, first, I discuss the variations in the data that identify the structural parameters of the model. Then, I present the estimation procedure, the results from the estimation, and the fit of the model with a series of stylized facts from the data. Finally, with the estimated parameters, I proceed with counterfactuals of the model.

5.1 Identification of the Parameters

Unfortunately, not all the parameters of the model are separately identified. First, p_S and a_S, as well as p_C and a_C, are always combined in the equations defining the equilibrium of the model. Therefore, I assume a_S and a_C to be equal to one. Also, I am not able to disentangle a level effect of the productivity parameters from the level of the fixed costs. In other words, the parameters of the model related to the productivity of soy and cattle could all be multiplied by a factor, and, still, I would be able to fit the data as long as I increase the level of the fixed costs. As such, I focused on the estimation of the difference between the productivity of soybeans with respect to cattle, and I assume $p_C = f_C = 10$

[18] This is a very preliminary exercise. Comments are very welcomed.

and $\gamma_C = 0.0001$. The same argument follows for the estimation of the parameters of $z(\ell)$. I am not able to estimate both parameters from the extreme value distribution. Therefore, I assume T, the parameter governing the location of $z(\ell)$, to be equal to 1 [19].

Also, with respect to the parameters related to consumption, a_U and p_M appear together with the indirect utility function in the non-migration condition:

$$w_i = (\frac{Ve^{\gamma\alpha_M Droad_i} p_M^{\alpha_M}}{a_U^{\alpha_U}})^{\frac{1}{\alpha_M}}$$

There are no other equations defining the equilibrium which allows V, a_U and p_M to be separately identified. Therefore, I assume a_U and p_M to be equal to one.

Even though I am not able to estimate many parameters of the model, I can still explore a series of counterfactuals with the ones I can identify. The following vector of 8 exogenous parameters in the model are estimated:

$$\Sigma = (\theta, \phi, p_S, f_S, \gamma_S, \gamma_M, V, \alpha_M)$$

Below, I discuss the identification of each of these parameters:

- θ: this parameter governs the dispersion of the extreme value distribution. The movement of z^{SC},z^{SF}, and z^{CF} (i.e., the marginal land where the representative farmer is indifferent between soybeans versus cattle, soybeans versus forest, and cattle versus forest) identifies this parameter.

- p_S: the price of soybeans provides information on the level of production, land, and labor employment. The difference between the employment in labor and land with respect to cattle generates the differences in the demand for land and labor we observe in the data.

- γ_S and γ_M: the parameters on the transportation costs generate differences in production between counties close and far from highways. In other words, if there were no transportation costs, the model would predict all the counties to have the same level of production.

- f_S: the absence of production in the model where soybeans have a better price than cattle is only justified by the impossibility of obtaining positive profits in areas with

[19] The identification of the scale and the dispersion parameters of the Fréchet is a common issue in the Trade literature. See (Eaton and Kortum, 2002).

lower land quality, which is determined by the size of the fixed costs. Furthermore, note that if f_S is sufficiently low, all the counties would specialize on the production of soybeans (as long as $p_S > p_C$). The absence of a widespread production of soybeans, therefore, also provides variation to infer f_S from the data.

- ϕ: the parameter on labor intensity factor will be associated with the ratio between labor to land. For low levels of ϕ, production would require a small ratio between land and labor. Therefore, the amount of labor and land in the county together provides the identification of ϕ.

- V: this parameter captures several pieces of the migration condition. Here, the level of wages help me identify this parameter.

5.2 Estimation procedure

The method used to estimate the parameters of the model resembles the simulated method of moments. First, I use the actual data for distance to roads and the area of the county to generate a simulated dataset using the theoretical model. Then, I calculate a set of moments from this artificial dataset with a given set of parameters, and I compare them to the actual moments in the data. Finally, an algorithm procedure follows with several comparisons between the moments generated by the model and the ones from the data to minimize their difference.

Note that, in each interaction of the model, I have to calculate the general equilibrium in each county, which makes this process computationally intensive. Therefore, to simplify the estimation, I do not assume a specific structure for the noise generated by the data as in a typical simulated methods of moments estimation - which are generally modeled as productivity shocks or idiosyncratic utility preferences for living in a specific county [20]. Instead, I focus on a parsimonious number of moments that are not dependent on the structure of noise in the data. More specifically, the following 10 moments are chosen:

$$
\begin{aligned}
M^x_{avg}(\Sigma) &= \bar{y}^x - \bar{y}^x_s(\Sigma) \\
M^x_{\beta}(\Sigma) &= \hat{\beta}^x - \hat{\beta}^x_s(\Sigma)
\end{aligned} \tag{13}
$$

[20] This can also be interpreted as a calibration where I use actual data to infer the moments from the model.

Where $x \in \{L_a, L_u, T_S, T_C, w\}$, $\hat{\beta}$ is the coefficient of an ordinary least square regression with a constant of the dependent variable x on distance to roads $Droads$, and s stands for the moments coming from the simulated dataset. There are two sets of moments. First, the average value of the dependent variables, which provides the identification for the parameters determining the level of production in the region. Note that, by focusing on the average of the endogenous variables of the model instead of their variance, I am allowing the actual moment in the data to be matched by a deterministic model that does not incorporate the noise in the data. The second set of moments are the coefficients from an ordinary least square coefficients. They provide the variation to estimate the transportation costs in the model. By using the ordinary least square coefficients as a set of moments, I attempt to unify into a single theoretical framework the results that would be obtained by running 5 independent regressions on the relationship between the dependent variables and distance to federal highways [21].

Defining $M(\Sigma) = [M^x_{avg}(\Sigma)\ M^x_{\beta}(\Sigma)]$ as the vector of moments. Σ_0 as the true values of Σ. The estimation is based on the condition that

$$E[M(\Sigma_0)] = 0$$

The minimization algorithm searches for $\hat{\Sigma}$ to achieve:

$$\hat{\Sigma} = \arg\min_{\Sigma}\{M(\Sigma)WM(\Sigma)'\}$$

Where W is a weighting matrix. I choose the weighting matrix that equalize the order of magnitude of the different moments used for the estimation.

For the reduced form section, I used data on formal employment in the Amazon region, which is available on annual basis and allowed me to explore the timing of expansion in the production of commodities. However, while using formal employment may provide the same qualitative results as using informal one (as long as they are proportional to each other in every year), for the estimation of the structural parameters of the model this would be a very concerning limitation. For example, we would understate the value of ϕ, which is related to the proportion between labor and land. To avoid this problem, I estimated the structural model on data on land use available from the Agricultural Census of 2007, and the whole

[21] Note that, because I use the actual data on distance to roads for the simulated moments, using the ordinary least square coefficients would provide the same results as using the covariance. However, I preferred to use the information on the coefficients as they provide a more intuitive way of relating the structural model estimates to the reduced form one.

universe of employment (formal and informal) provided by the Census of 2010.[22]

5.3 Results

Table 6 presents the results for the fit of the model. In spite of a very parsimonious number of parameters (8 parameters), the model is able to fit the data well. Panel A indicates a very good fit of the simulated dataset with the actual data in terms of the average in the endogenous variables of the model L_a, L_u, T_S, T_C, w. Furthermore, panel B shows that the model is also able to reproduce broadly the qualitative patterns of the relationship between roads and the endogenous variables. In particular, the coefficient for wages, soybeans and urban employment have a point-estimate which is close in magnitude to the actual ones. It is reassuring that a unique theoretical framework with 8 parameters is able to capture all the qualitative interpretations one would get by running 5 different regressions, and also most of the magnitudes.

The model does not necessarily have to fit the data, however, part of the fit can be attributed to the minimization algorithm. Here, I provide two additional elements of the predictive power of the model that are not directly related to the minimization procedure. First, figure 12 shows the results from estimating the non-parametric relationship between the dependent variables and distance to roads. The model does a reasonable job fitting the non-linear relationship between the endogenous variables and distance to roads. To some extent, this is an out-of-the-sample fit of the model. Even though the minimization algorithm uses the coefficient of the regressions on distance to federal roads, this is not sufficient to produce the non-linearity we observe in the figure. In a simple regression analysis, for example, we would need to transform the explanatory variable (by taking the square or the log) to generate a non-linear relationship. Second, note that not only the non-linear relationship of the simulated dataset resembles the actual one, but also the noise in the relationship between production variables T_S and T_C also follows each other. In the model, this is generated only by the variation in the size of counties.

The best fit of the model was achieved at the parameters given in table 7. The parameter ϕ affects the productivity of both cattle and soybeans. Different from previous estimation in the Trade literature, it does not have any specific interpretation here. p_S is estimated to be 9 times higher than p_L. This shows that soybeans are in fact more profitable than cattle. However, the large fixed costs of soybeans with respect to cattle prevents farmer

[22]Part of the data also seemed to have absurd high levels of soybean production and urban activities. Drop these counties from the sample (around 7 counties are dropped out of 424).

from specializing in soybeans. The transportation costs for soybeans is also larger, which is an additional force that pushes soybean production closer to roads. γ_M is estimated to be very low. This is in accordance to figure 12, where we do not observe a large variation in wages across different counties. In other words, there is little need for compensation on wages. $\bar{v}$ is the parameter related to V (such that $V = exp(\bar{v})$. The parameter α_M indicates that roughly 50% of income goes to imported goods. Finally, $\phi = 0.55$ provides the labor intensity factor. Mundlak et al. (1999) provides a review of the literature on the estimates of intensity factors in agricultural production functions. Previous estimates of the labor intensity varies widely, from 0.11 to 0.66, with the median estimate across articles being at 0.42. Therefore, it seems that the estimate of this parameter depends on the context, but it is reassuring that the level found is in-between the range of parameters previously found in the literature.

5.4 Counterfactual Analysis

Given the parameters estimated in the previous section, I follow with four counterfactuals to learn about the mechanisms of the model. First, I increase the fixed costs of soybeans production to make it inviable. Second, I proceed with a similar analysis for cattle. Third, I increase the proximity of every county to federal roads in 10%, and, fourth, I decrease the distance of counties to federal highways. The results for the counterfactual analysis are presented in table 8.

When we exclude the possibility of soybean production, there is an expansion of cattle. This is to be expected. Land that was being used for soybeans will now be used for cattle production. Note, however, that the effect is not a large one, mainly because soybeans were occupying a small share of the total land in each county. In spite of that, both urban and rural employment almost disappear in the absence of the soybean sector. This result shows the importance of the possibility of producing a highly profitable activity for the urbanization process in the Amazon, and it also indicates why the Amazon did not engage in the production of these crops before. In the absence of the commodity price shock, farmers would not be able to pay for the large fixed costs of producing these activities.

Finally, the effect of varying distance to roads have a moderate effect on the production of the region. This indicates that marginal changes in the infrastructure system may not bring drastic changes in the production of the region.

6 Conclusion

Once the Brazilian federal government constructed a large highway system for the Amazon in the early 1970s, millions of workers migrated to the region with the hope of producing agricultural varieties that were common in their native cities. However, for almost 30 years following the initial efforts of the government to populate the Amazon, the economy of the region would be characterized by frustrated attempts to establish an agricultural industry. In the middle of the 1990s, technological advances in seed made the production of grains possible in the low latitude region of the Amazon, and the drastic increase in commodity prices made the production of these commodities a highly profitable activity. Once these conditions for the establishment of an export oriented agricultural industry came together, farmers who were close to the federal highway system were disproportionately benefited by the facility to export their products. As a consequence, a large urban sector rose to support the increasing demand for goods generated by this export oriented sector.

The results from this paper provide three main contributions to the literature on development and trade. First, this is one of the first studies to provide microeconomic evidence on the mechanisms that may foster urbanization. According to the results from the article, the absence of access to transportation, good prices for exporting goods, and the possibility of producing mechanized crops may prevent the establishment of urban centers in rural settings of developing countries. Second, recent articles highlight the role of mechanization and scale for the low agricultural productivity in developing countries (Foster and Rosenzweig, 2011; Restuccia and Adamopoulos, 2014; Adamopoulos and Restuccia, 2014). The present results suggest that access to transportation infrastructure and the profitability or production may constitute essential elements for the establishment of a large scale rural production. Third, this article has shed light on the additional elements that may foster the benefits from investments in transportation infrastructure. In the Brazilian Amazon, it took more than 30 years for the federal highway network to substantially affect its economic activities. The evidence in the article suggests that the lack of the right macro-economic conditions for trade have undermined the impact of roads in the region.

It is important to recognized that the Brazilian economy provides favorable conditions for agricultural production that limits the external validity of this study. First, some regions may not have access to seeds suited for the production of large-scale mechanized crops. In Brazil, the availability of seeds suitable for low latitude areas took several years of research financed by the federal government. Second, Brazil has a huge population that has systematically migrated to take advantage of economic opportunities in several historical moments of economic growth. This population represented an important source of labor for the expansion of economic activities in the Amazon.

In spite of these advantages given by the Brazilian institutional setting, it is still interesting how farmers in the Amazon region managed to expand their production of export-oriented commodities by more than 300% in roughly 15 years. The production of grains is a large-scale enterprise and its production requires technology, large investments in machinery, and access to international markets, which makes this region a specially good case for studying the emergence of large-scale agricultural enterprises in developing countries. Here, I focused on the role of transportation infrastructure for the development of agriculture. In future research, I intend to explore additional conditions, such as access to credit and technology, and the role they might play in the expansion of large scale agricultural production.

References

Tasso Adamopoulos and Diego Restuccia. The size distribution of farms and international productivity differences. *American Economic Review*, 104(6):1667–97, 2014.

Elizabeth Barona, Navin Ramankutty, Glenn Hyman, and Oliver T Coomes. The role of pasture and soybean in deforestation of the brazilian amazon. *Environmental Research Letters*, 5(2):024002, 2010.

Nathaniel Baum-Snow. Did highways cause suburbanization? *The Quarterly Journal of Economics*, 122(2):775–805, 2007.

Paula Bustos, Bruno Caprettini, and Jacopo Ponticelli. Agricultural productivity and structural transformation: Evidence from brazil. *Working Paper*, 2013.

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

Benjamin Faber. Trade integration, market size, and industrialization: Evidence from china's national trunk highway system. *The Review of Economics Studies*, 2014.

Pablo Fajgelbaum and Stephen Redding. External integration, structural transformation, and economic development: Evdence from argentina 1870-1914. *NBER Working Paper*, 20217, 2014.

Andrew D Foster and Mark R Rosenzweig. Agricultural productivity growth, rural economic diversity, and economic reforms: India, 1970–2000*. *Economic Development and Cultural Change*, 52(3):509–542, 2004.

Andrew D Foster and Mark R Rosenzweig. Economic development and the decline of agricultural employment. *Handbook of development economics*, 4:3051–3083, 2007.

Andrew D Foster and Mark R Rosenzweig. Are indian farms too small? mechanization, agency costs, and farm efficiency. *Econojmic Growth Center, Yale University New Haven CT*, 2011.

Douglas Gollin, Remi Jedwab, and Dietrich Vollrath. Urbanization without structural transformation: Evidence from sub-saharan africa. 2013.

J Vernon Henderson, Mark Roberts, and Adam Storeygard. Is urbanization in sub-saharan africa different? 2013.

Berthold Herrendorf, Richard Rogerson, and Akos Valentinyi. Growth and structural transformation. 2013.

Rémi Jedwab. Urbanization without structural transformation: Evidence from consumption cities in africa. *Processed, George Washington University*, 2013.

Simon Kuznets. Modern economic growth: findings and reflections. *The American Economic Review*, pages 247–258, 1973.

Robert E Lucas. On the size distribution of business firms. *The Bell Journal of Economics*, pages 508–523, 1978.

Douglas C Morton, Ruth S DeFries, Yosio E Shimabukuro, Liana O Anderson, Egidio Arai, Fernando del Bon Espirito-Santo, Ramon Freitas, and Jeff Morisette. Cropland expansion changes deforestation dynamics in the southern brazilian amazon. *Proceedings of the National Academy of Sciences*, 103(39):14637–14641, 2006.

Yair Mundlak, Donald F Larson, and Rita Butzer. Rethinking within and between regressions: The case of agricultural production functions. *Annales d'Economie et de Statistique*, pages 475–501, 1999.

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

Stephen J Redding and Matthew A Turner. Transportation costs and the spatial organization of economic activity. *Chapter in progress for the Handbook of Regional and Urban Economics*, 2014.

Diego Restuccia and Tasso Adamopoulos. Land reform and productivity: A quantitative analysis with micro data. *Working Paper*, 2014.

Peter D Richards, Robert J Myers, Scott M Swinton, and Robert T Walker. Exchange rates, soybean supply response, and deforestation in south america. *Global Environmental Change*, 22(2):454–462, 2012.

Jennifer Roback. Wages, rents, and the quality of life. *Journal of Political Economy*, 1982.

Peter M Robinson. Root-n-consistent semiparametric regression. *Econometrica: Journal of the Econometric Society*, pages 931–954, 1988.

Adam Storeygard. Farther on down the road: transport costs, trade and urban growth in sub-saharan africa. 2013.

Maria del Carmen Vera-Diaz, Robert K Kaufmann, Daniel C Nepstad, and Peter Schlesinger. An interdisciplinary model of soybean yield in the amazon basin: the climatic, edaphic, and economic determinants. *Ecological Economics*, 65(2):420–431, 2008.

Robert Walker, John Browder, Eugenio Arima, Cynthia Simmons, Ritaumaria Pereira, Marcellus Caldas, Ricardo Shirota, and Sergio de Zen. Ranching and the new global range: Amazônia in the 21st century. *Geoforum*, 40(5):732–745, 2009.

Figure 1: Historical Population of the Brazilian Amazon.

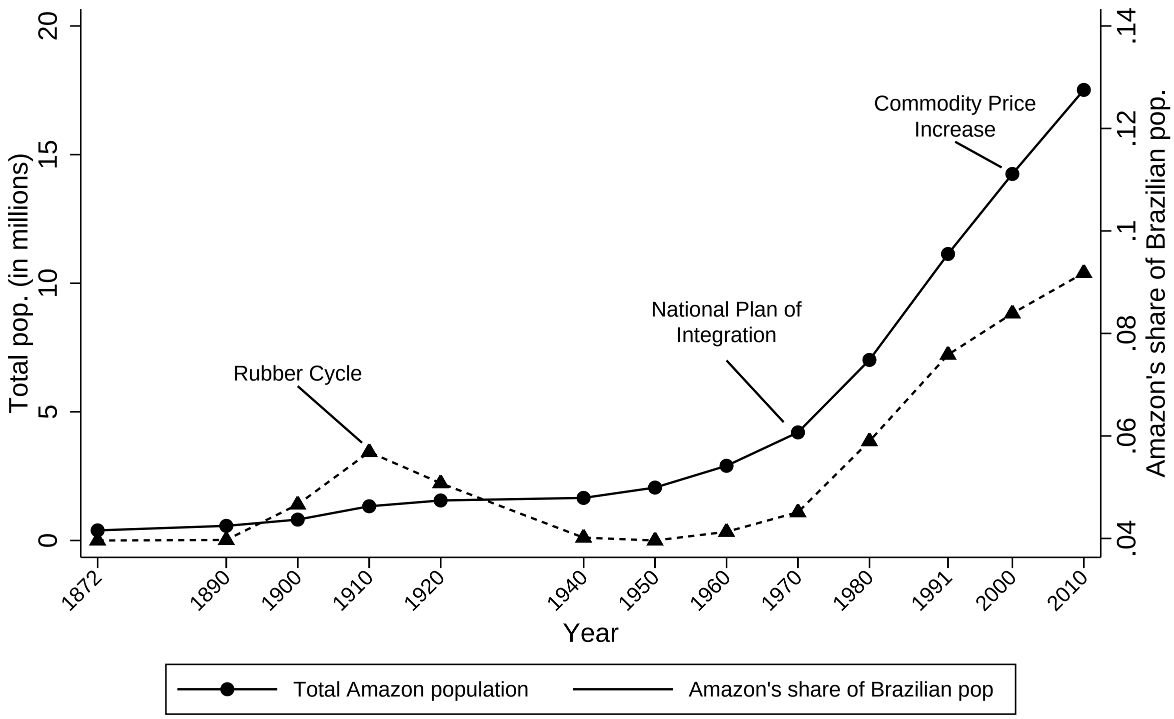


Figure 2: Soybean price at local retailers and the international soybean price adjusted by exchange rate.

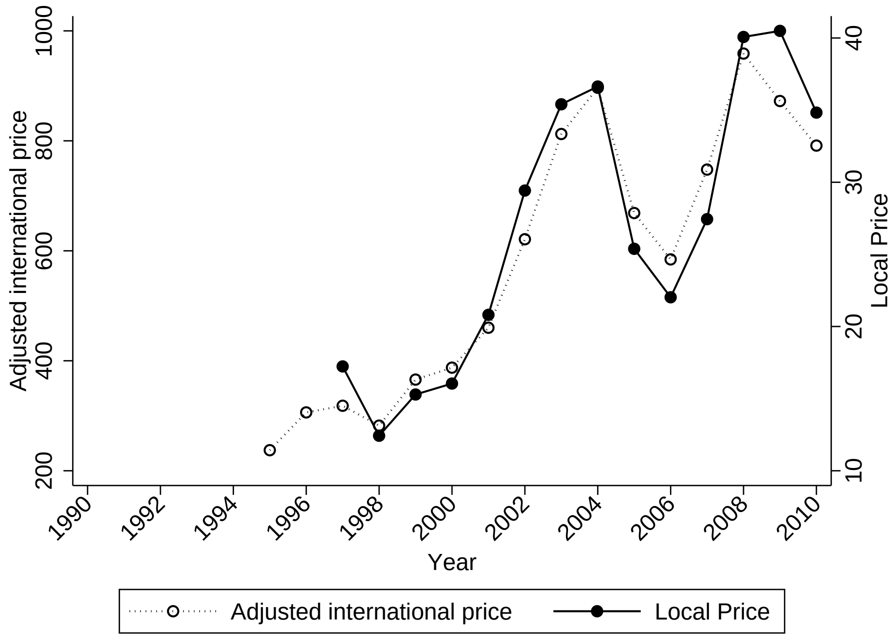


Figure 3: Trends in the agricultural production in the Brazilian Amazon.

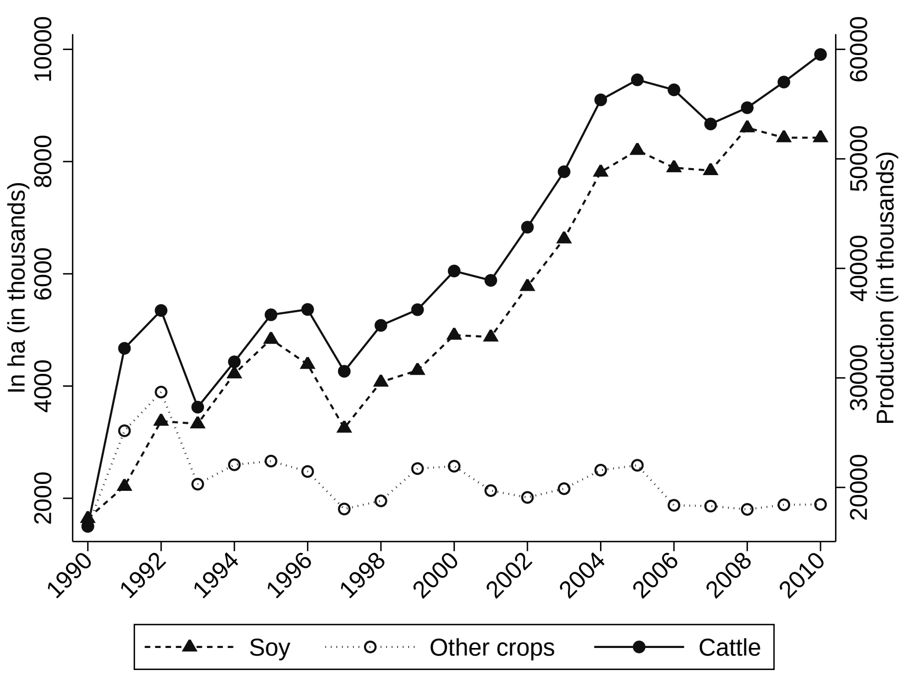


Figure 4: Trends in employment by sector in the Brazilian Amazon (in thousands).

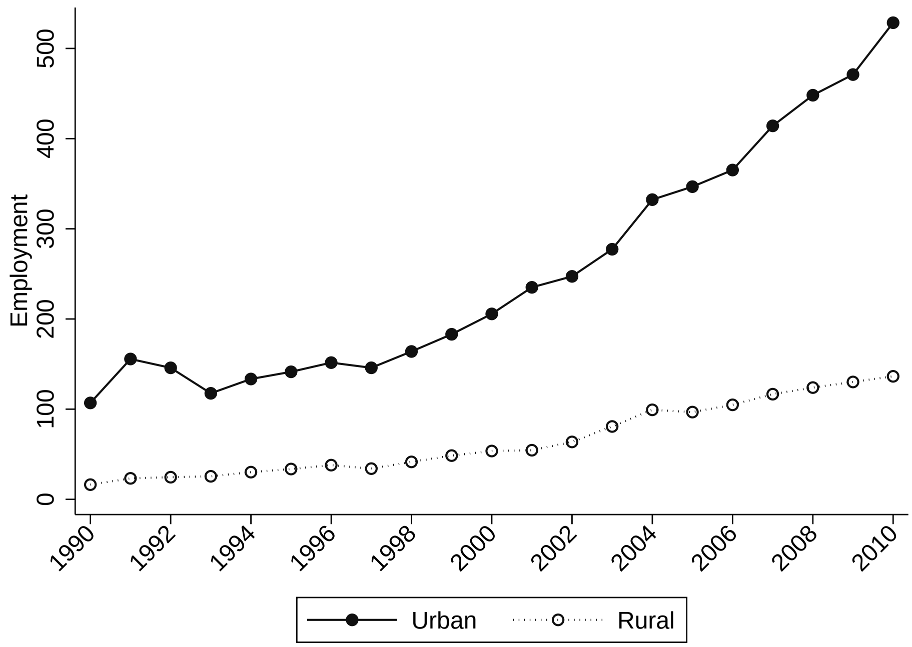
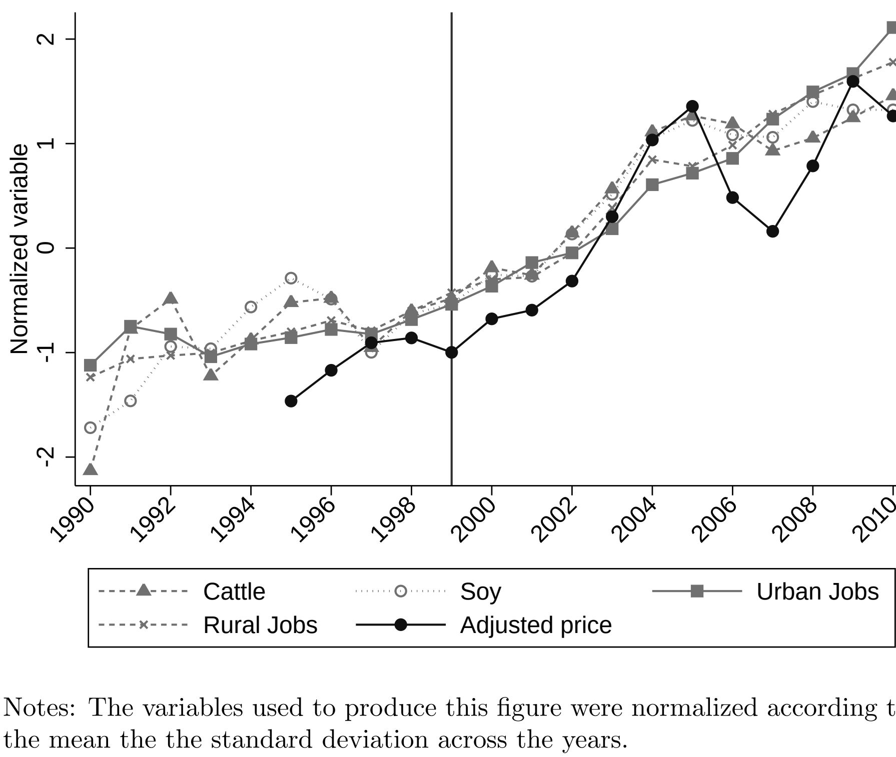


Figure 5: Trends in main variables over time.



Notes: The variables used to produce this figure were normalized according to the mean the the standard deviation across the years.

Figure 6: County Boundaries, Project Lines, and Constructed Roads.

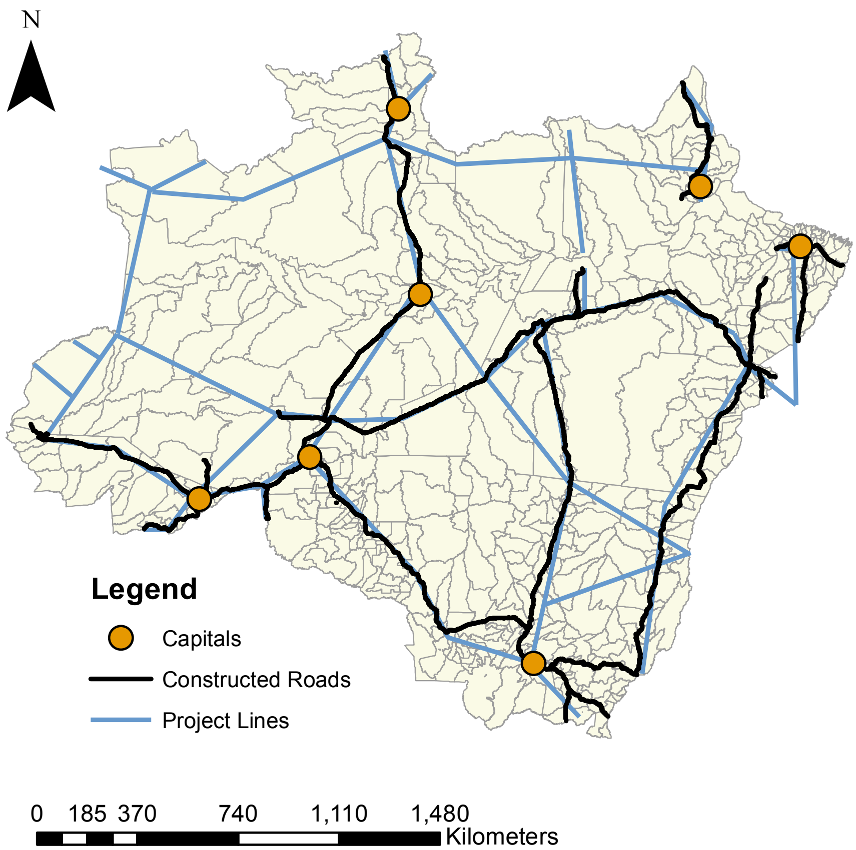
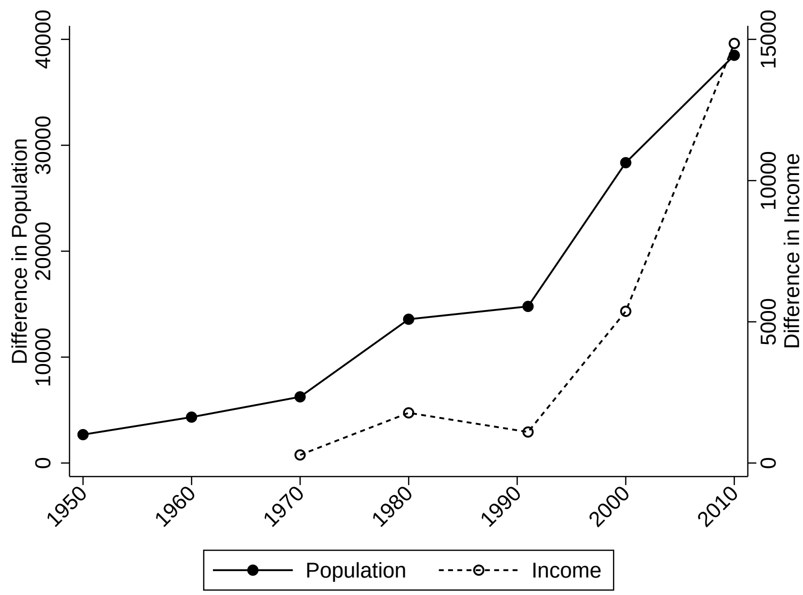


Figure 7: Difference between counties close and far from federal roads in terms of population and income.



Notes: This figure reports the difference in the average population and total income between counties closer than 100 km from a constructed federal road and those farther away.

Figure 8: Difference between counties close and far from federal roads in terms of commodity production.

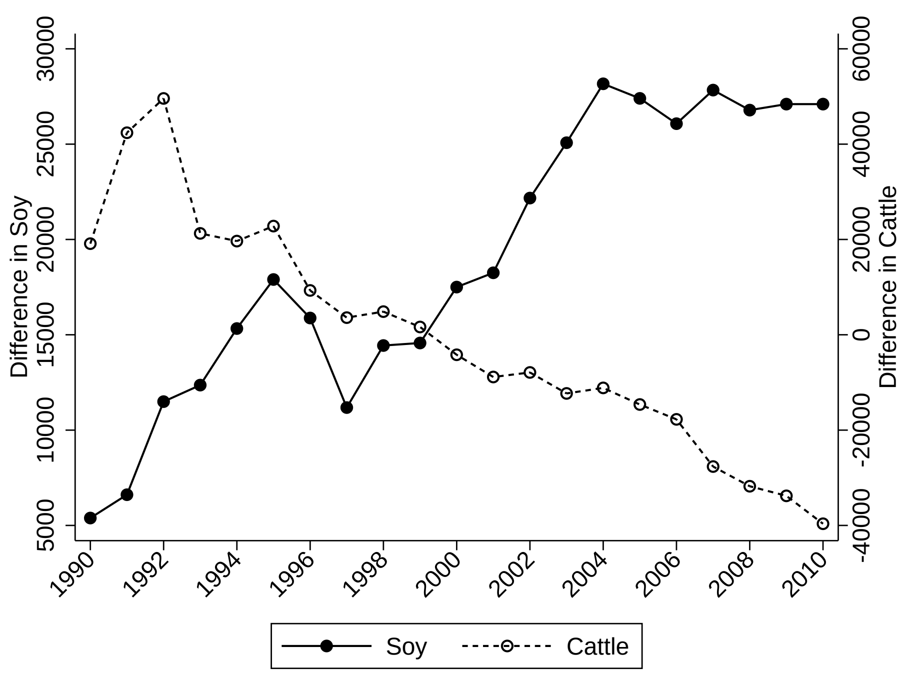
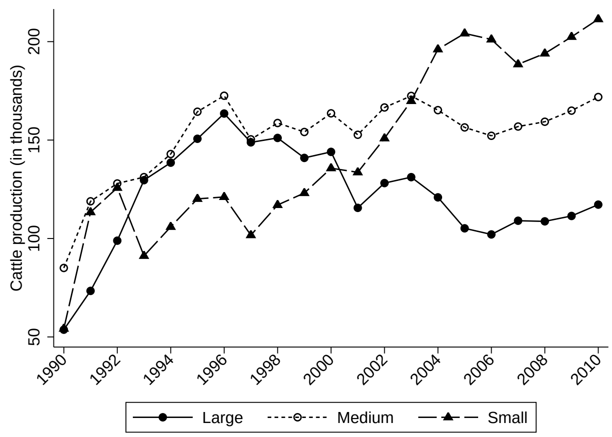


Figure 9: Heterogeneity in cattle production: difference between counties close and far from federal roads in terms of their production of soybeans in 2010.



Notes: Large producers are the counties that reached 2010 above the 95th percentile in terms of soybean production. Medium are those between the 95th and the 90th percentile. And small producers are the rest of the counties.

Figure 10: Difference between counties close and far from federal roads in terms of employment in the urban and rural sectors.

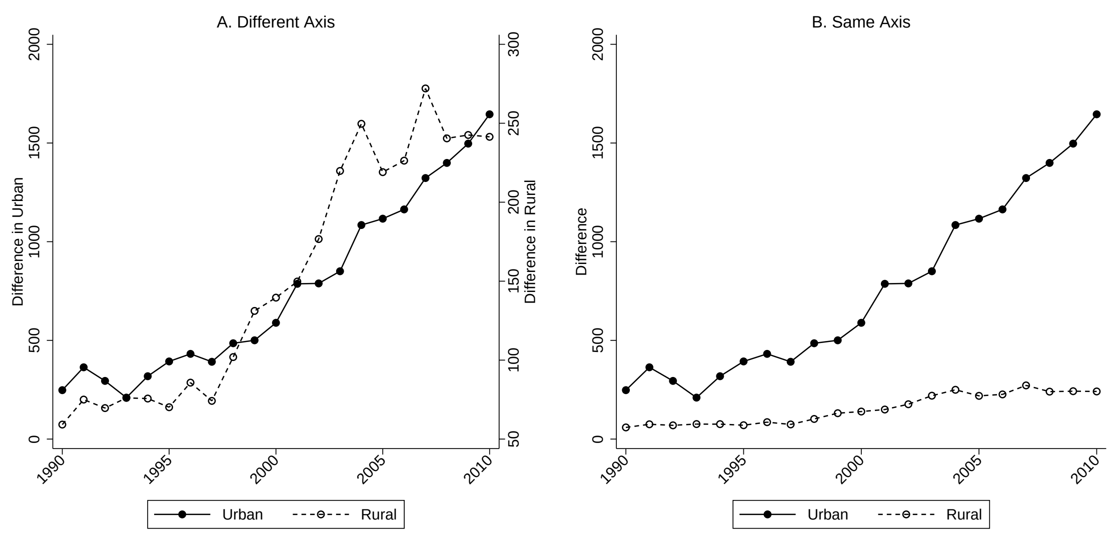
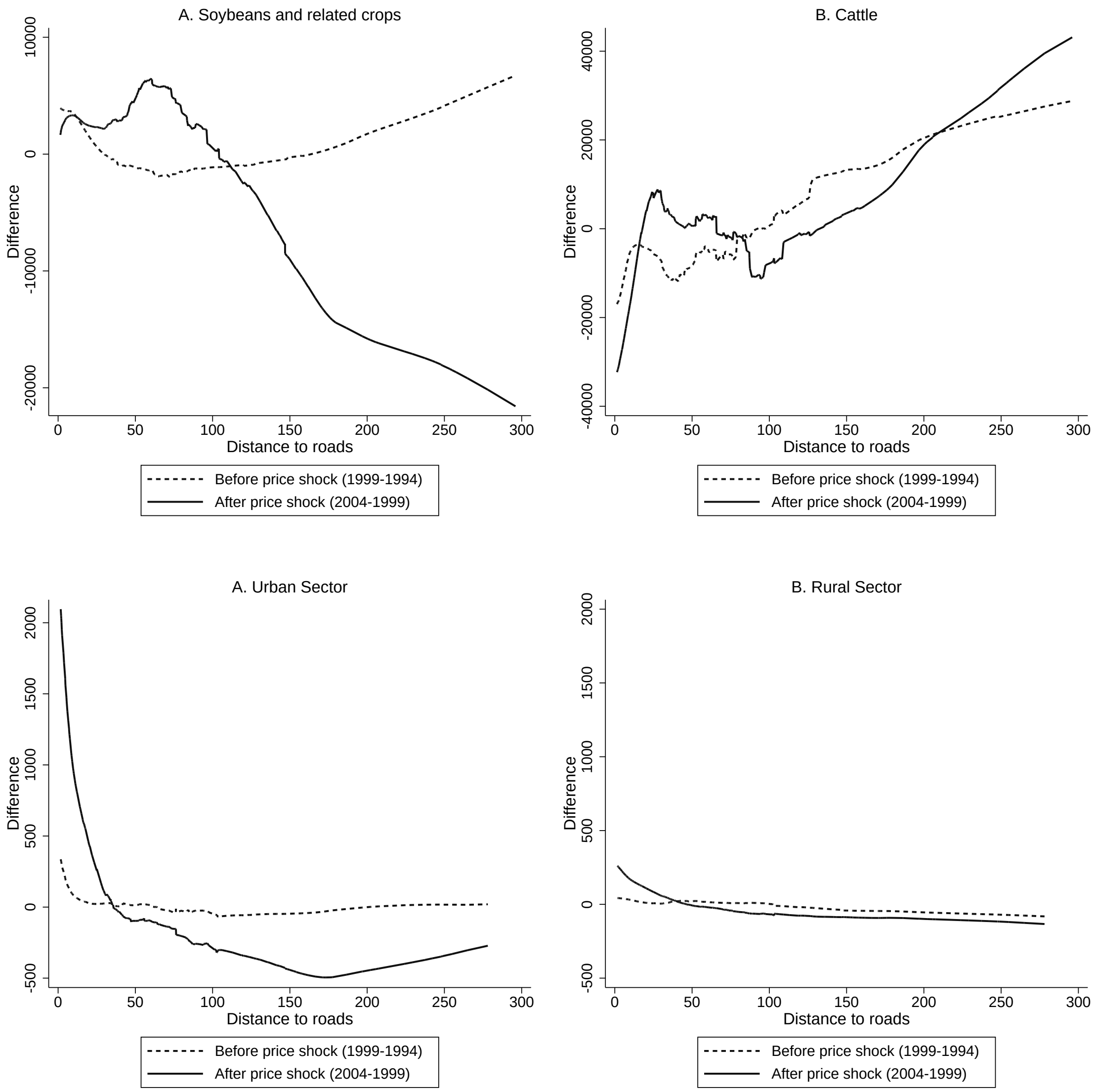


Figure 11: Non-parametric regressions of the differential trend in the main endogenous variables of the model.



Notes: The non-parametric regressions were adjusted by a cubic polynomial on distance to roads following the procedure suggested in Robinson (1988).

Figure 12: Non-parametric regressions of the endogenous variables in the model on distance to roads (actual and simulated dataset).

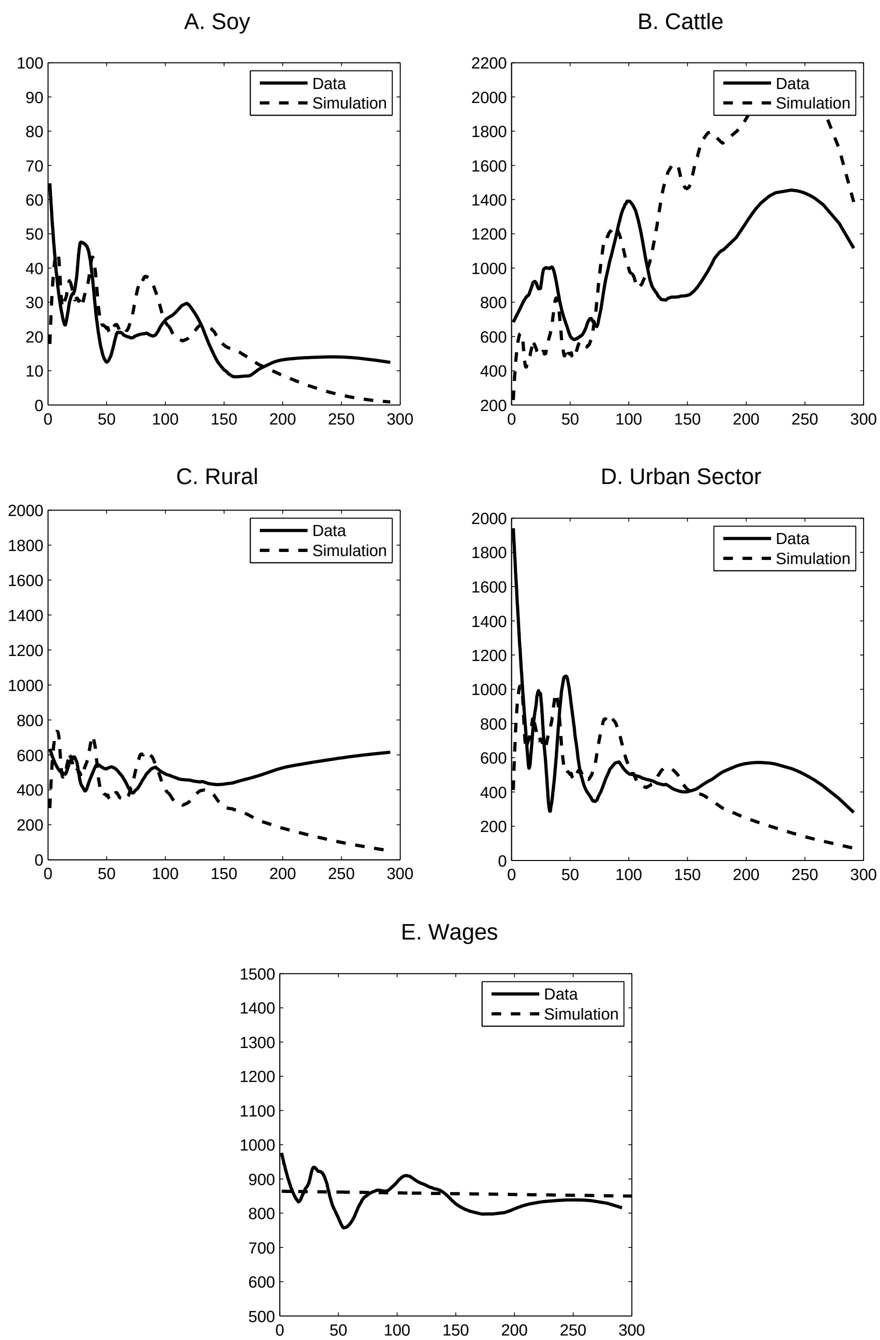


Table 1: Descriptive Statistics

	1950-2010 dataset			1990-2010 dataset		
		Elasticity wrt Roads			Elasticity wrt Roads	
		Full Sample	No State Capitals		Full Sample	No State Capitals
	Mean			Mean		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. *Geographic Characteristics*						
Area	35206.90	0.80*	0.80*	12825.72	0.46*	0.46*
	(51082.49)	(0.03)	(0.02)	(19321.80)	(0.09)	(0.09)
Corn suitability	5.30	0.01	0.01	5.20	-0.01	-0.01
	(0.53)	(0.01)	(0.01)	(0.56)	(0.01)	(0.01)
Soybeans suitability	4.20	-0.03	-0.03	4.37	-0.04	-0.04
	(0.84)	(0.02)	(0.02)	(0.98)	(0.02)	(0.02)
Slope	4.37	0.23	0.25	2.70	0.07	0.07
	(7.70)	(0.31)	(0.34)	(5.91)	(0.19)	(0.20)
Distance to rivers	34.53	0.07	0.06	34.95	0.06	0.06
	(22.30)	(0.08)	(0.08)	(25.64)	(0.06)	(0.06)
Temperature in september	267.60	0.01*	0.01*	263.07	0.01	0.01
	(8.78)	(0.01)	(0.01)	(11.63)	(0.01)	(0.01)
Rainfall in september	71.83	0.06	0.07	72.62	0.06	0.06
	(26.04)	(0.05)	(0.06)	(26.72)	(0.03)	(0.03)
Panel B. *Initial Conditions*						
Population in 1950	18504.99	-0.07	-0.02	-	-	-
	(28226.78)	(0.05)	(0.05)			
Population in 1960	26126.17	-0.10*	-0.05	-	-	-
	(43720.05)	(0.04)	(0.04)			
Panel C. *Determinants of Road*						
Distance to capitals	248.43	0.53*	0.53*	300.98	0.27*	0.26*
	(169.93)	(0.05)	(0.05)	(202.29)	(0.03)	(0.03)
Distance to project lines	86.18	0.64*	0.63*	82.09	0.50*	0.50*
	(56.44)	(0.11)	(0.12)	(6.01)	(0.07)	(0.07)
Obs						

Notes: * p<0.5. Robust standard errors at the State level in parentheses. Columns 2 to 5 presents coefficients of regressions of the log of each variable against the log of distance to roads. Column 3 drops 7 State capitals which where nodal points in the Federal government project.

Table 2: Effect of access to federal roads on population and income by period.

	Period					
	2004-1994		2004-1999		(2004-1999)-(1999-1994)	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. *Population*						
Close	14,922	11,785	26,593***	30,420**	11,671	18,635*
	(9,402)	(9,748)	(9,942)	(13,301)	(7,192)	(10,765)
Baseline	22,134	-	52,536	-	-	-
P-value	-	0.001	-	0.068	-	0.26
Panel C. *Income*						
Close	1,509	461.1	18,446**	15,560*	15,046***	14,602**
	(1,287)	(1,141)	(7,650)	(8,126)	(5,369)	(6,261)
Baseline	889	-	4,773	-	- -	
P-value	-	0.001	-	0.001	- 0.001	
Dist to Capitals	Y	Y	Y	Y	Y	Y
Region FE	N	Y	N	Y	N	Y
Geography	N	Y	N	Y	N	Y
Initial Condition	N	Y	N	Y	N	Y

Notes: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors at the county level in parentheses. Every regression controls for county and year fixed effect. Dist to Capitals is a cubic polynomial on the distance of each county to the State Capital. Geography represent controls for: area of the county, corn suitability, soybeans suitability, slope, temperature in September, and rainfall in september. Initial condition controls for the population in 1950 and 1960. Region FE stands for a dummy wether the county is the southern part of the Amazon (the States of Mato Grosso or Rondonia). Sample includes 99 counties. *Baseline* stands for the mean for the counties far from roads in the initial period. *P-value* is the result from a joint test of the statistical significance of the control variables.

Table 3: Effect of access to federal roads on agricultural production by period.

	Period					
	2004-1994		2004-1999		(2004-1999)-(1999-1994)	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. *Soybeans and related crops (in ha)*						
Close	-1,517	-2,662**	10,979*	11,541*	12,496***	14,202**
	(2,070)	(971.2)	(5,161)	(4,923)	(3,284)	(5,251)
Baseline	3,620	-	4,366	-	-	-
P-value	-	0.001	-	0.001	-	0.001
Panel B. *Other crops (in ha)*						
Close	-3,845	-3,523	1,218	2,425	1,218	2,425
	(3,108)	(3,053)	(2,401)	(2,787)	(2,401)	(2,787)
Baseline	5,199	-	6,801	-	-	-
P-value	-	0.001	-	0.001	-	0.001
Panel C. *Cattle (output)*						
Close	-17,383	-14,765	-6,276	-21,195	11,108	-6,430
	(11,207)	(7,740)	(36,091)	(21,256)	(30,308)	(25,964)
Baseline	111303	-	138071	-	-	-
P-value	-	0.001	-	0.053	-	0.053

Notes: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors at the state level in parentheses. Controls are explained in table 2. Sample includes 259 counties.

Table 4: Effect of access to federal roads on employment in the urban and rural sector by period.

	Period					
	2004-1994		2004-1999		(2004-1999)-(1999-1994)	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. *Urban Sector*						
Close	111.0	154.0	648.6***	507.3***	537.6***	353.3*
	(134.7)	(152.1)	(153.7)	(161.3)	(206.4)	(213.0)
Baseline	272	-	321	-	-	-
P-value	-	0.082	-	0.080	-	0.167
Panel B. *Rural Sector*						
Close	54.56**	21.48	147.7***	93.87*	93.16*	72.39
	(26.16)	(28.40)	(43.08)	(53.04)	(52.72)	(67.09)
Baseline	51	-	75	-	-	-
P-value	-	0.001	-	0.001	-	0.016

Notes: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors at the state level in parentheses. Controls are explained in table 2. Sample includes 259 counties.

Table 5: Robustness check.

	Dependent Variables and Period				
	(2010-1991)-(1991-1970)		(2004-1999)-(1999-1994)		
	Pop	Inc	Soy	Urban	Rural
	(1)	(2)	(3)	(4)	(5)
Panel A. *Baseline Results*					
Close	11,671	15,046**	12,496***	573.6***	93.16*
	(7,192)	(5,369)	(3,284)	(206.4)	(52.72)
Obs	99	99	259	259	259
Panel B. *Dropping Service Counties*					
Close	15,374**	13,243**	11,917*	389.3***	33.81
	(7,274)	(5,796)	(5,658)	(93.63)	(85.93)
Obs	80	80	205	205	205

Notes: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors at the state level in columns 3-5 and at the county level in columns 1 and 2 in parentheses. All regressions control for a cubic polynomial of the distance a county is from a State Capital.

Table 6: Fit of the model

Panel A. *Endogenous Variables* (M_{avg})					
	T^S	T^C	L^A	L^U	w
Data	26	914	495	673	861
Model	29	912	490	675	860
Panel B. *Regressions on Roads* (M_β)					
	β_{T_S}	β_{T_c}	β_{L_a}	β_{L_u}	β_w
Data	-0.12	2.10	-0.08	-2.66	-0.24
Model	-0.10	7.61	-1.62	-2.28	-0.04

Table 7: Estimated parameters

θ	p_S/p_C	f_S/f_C	γ_S/γ_C	γ_M	$\bar{v}$	α_M	ϕ
31.57	9.62	679.2	1.289	-0.0001	3.28	0.48	0.55
(2.04)	(1.12)	(208)	(0.12)	(0.00001)	(0.11)	(0.01)	(0.01)

Notes: Bootstrapped standard errors in parentheses (50 samples). P_C and f_C are equal to 10.

Table 8: Counterfactuals

	T^S	T^C	L^A	L^U	w
Model	29	912	490	675	860
No soy	0	942	31	48	860
Closer to Roads (10%)	32	929	532	732	860
Farther from Roads (10%)	27	895	454	625	860

A The Plans of Integration of the Brazilian Amazon

Figure A1: The National Plan of Integration (1969).

Figure A2: The National Transportation Plan (1973).

B Datasets

B.1 NASA's Shuttle Radar Topography Mission

I use topographic data produced as part of NASA's Shutte Radar Topographic Missions. I use the 3arc second scale (around 90 m^2) to calculate the standard deviation in altitude in each city as our measure of slope. For the altitude we used the average altitude in each municipality.

B.2 FAO-GAEZ

To obtain information on soy and corn suitability, I use data from the Global Agro-Ecological Zones database produced by the FAO. The suitability is measured as the potential yields attainable for a crop in a certain geographical area. They use information on climatic conditions, slope, and the level of technology available to produce a raster file of potential yields for different levels of technology. I included the high input type of technology, which corresponds to the large-scale mechanized techniques that are used in the region. I then measure the averate suitability in each municipality.

B.3 RAIS

For employment I used data from the *Relação Anual de Informações Sociais* (RAIS). This dataset is provided by the Ministry of Labor from their official recors on all the formal job employments in Brazil. To obtain the dataset, I had to require access to a website that compiles the information at the county level in each year since the 1980s. However, the way the website is organized required a webscrapping technique to automatize the download of the information. I used the internet language *iMacro* that has been recently developed specially for this type of procedure.

B.4 World Climate

For information in temperature and precipitation we used the WorldClim dataset available at http://www.worldclim.org/. The WorldClim data is a set of global climate layers (climate grids) with a spatial resolution of 1 square kilometer. The dataset provides montly precipitation and temperature averages using information from 1950 to 2000. For each municipality, calculated the average monthly precipitation and temperature in each month of the year.

B.5 IMEA

Data for local retailer's price of soybeans and cattle was obtained from the *Instituto Mato Grossense de Economia Agropecuária*. This a public research institute that is part of the State government of Mato Grosso. They provide daily price offered by retailer's in several cities for soybeans from 1996 to 2013.

Table B1: Summary of Datasets

Dataset (1)	Institution (2)	Information (3)	Frequency/Years (4)
		Dataset 1: 1950-2010	
CENSUS	IBGE (IPEA)	Population	Decadal
CENSUS	IBGE (IPEA)	Income	Decadal (since 1970)
		Dataset 2: 1991-2010	
RAIS	Ministry of Labor	Labor per sector	Annual
PAM	IBGE (IPEA)	Land use per activity	Annual
PPM	IBGE (IPEA)	Cattle production	Annual
		Both datasets: geographic characteristics, shapefiles and others	
GAEZ	FAO	Soy and corn suitability	
SRTM	NASA	Slope and elevation	
WorldClim	WorldClim	Rainfall and temperature	
Cities map	CENSUS	Area and political boundaries	
Road map	IBAMA	Road network	
Rivers map	IBAMA	Principal rivers	
Agro CENSUS	IBGE	Number of tractors	2007

Notes: (IPEA) indicates that, even though the primary source of information is collected by another institution, the information for this article was organized and published in the websited of the *Insituto de Pesquisa e Economic Aplicada* (IPEA), a public research institute of the federal government.

C Comparative Statics

Figure C1 summarizes the main relationships between the endogenous variables in the model (population, income, land use and labor employment) to the exogenous variables. Panel A shows the production characteristics before the shock on commodity prices. Panel B shows the production after a positive shock. As in section 3 of the paper, I divide the discussion in two.

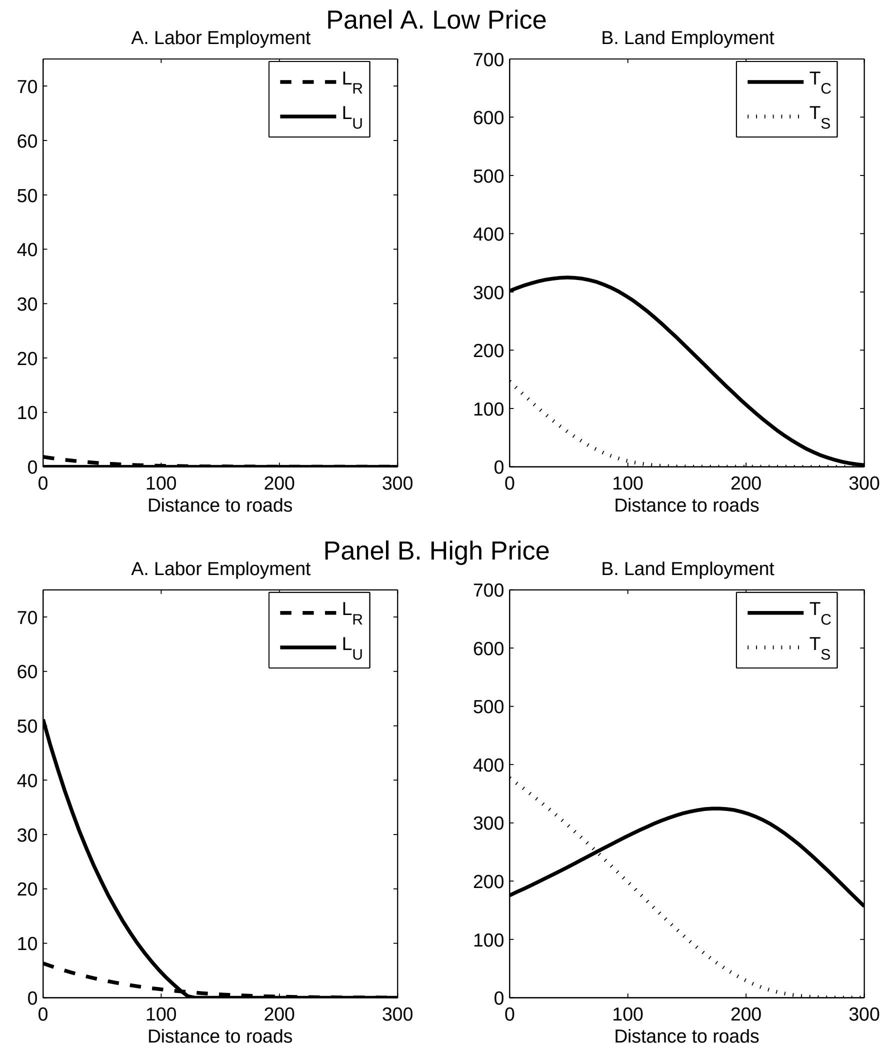
Variation in Distance to Roads given International Price

Panel A figure B shows a monotonic relationship between the land area used for soybean and distance to roads. However, cattle follows a U-shape relationship with distance to roads. In the first few kilometers, there is a positive relationship, but after 100 km the relationship is clearly negative. As discussed in the main body of the article, this can be attributed to two margins of substitutions: the one between cattle and soybeans, and the one between cattle and forests. Given the parameters of the model, employment is very low and there is an almost inexistent urban center.

Variation in International Prices given Distance to Roads

Panel B shows the results from simulations when I increase commodity prices by 100%. First, the simulation predicts one main stylized fact from the literature: that soybean production precedes cattle ranching. Counties close to federal roads who were initially larger producer of cattle expand their production of soybeans over pasture. They become smaller producers when compared to counties at a medium distance to roads (100-200 km). The expansion of cattle in counties farther from roads leads to a large deforestation. Note, for example, that a county at 150 km from a road has an expansion of roughly 200 land units in the production of agricultural goods (from 200 to 400), while counties at 0 km from a road have an expansion of around 150. Depending on the parameters of the model, this could be magnified.

Figure C1: Relationship between the endogenous variables in the model and distance to roads.



Notes: Parameters used for the simulation are: $p_M = 10$, $\phi = 0.1$, $f_S = 15$, f_C, $\alpha_U = 0.5$, $V = 5$, $a_S = a_C = a_U = 1$, $\theta = 4$, $\mu = 1$, $\gamma = 0.005$. μ is the level parameter of the Fréchet. For low prices, in panel A, we have $p_S = 7.5$ and $p_C = 1.5$. In panel B prices are doubles.

D Additional Evidence of the Impact of Commodity Prices on Production

As discussed in section 2 of this article. I argue that the devaluation of the exchange rate between 1999 and 2004 had a large impact on the price of export oriented product faced by farmers in the Amazon. As a consequence, this had a large impact on the profitability of the rural activities which led to a large expansion in production. One natural question is whether the increase in agricultural production was specific to the products exported in the amazon region. If that were the case, this would suggest that, besides the price shock, there were additional technological improvements in production that may raise questions on the interpretation of the results. In particular, it would cast doubts on the established association between commodity production and commodity prices.

Figure D1 presents the results from an analysis of several commodities that are recognized as important for the Brazilian export basket. Each variable in figure D1 is the total amount produced in Brazil normalized by its mean and standard deviation across the years. To facilitate the visualization, I divide the set of commodities in two with no specific rule. Panel A in figure D1 presents the first set and shows that both soybeans and cattle follows the international commodity price increase adjusted by the exchange rate. However, tobacco, which is primarily produced in the south of Brazil, has a dramatic increase during this period.

Panel B registers the expansion of two additional products during this period: wheat and grapes. Both of them are primarily produced in the south of Brazil. It is interesting to note that sugarcane, which is largely produced in the southeast, did not follow the variation in the exchange rate during the period. One of the main final goods from sugarcane between 1999 and 2003 was the production of ethanol. During this period, there was a massive adoption of bi-fuel technologies in Brazilian cars and, as a strategy to reduce the effect of the exchange rate devaluation on prices, the government controlled the price of fuel and ethanol during the period (this has been a common strategy from the government to influence inflation in the past two decades). When the international price of oil increased after 2005, however, the government allowed the price of ethanol and fuel increase, which could have motivated the expansion in sugarcane production. To some extent, the different behavior in the production of sugarcane is an exception that confirms the rule.

Finally, I obtained aggregate information on certain exporting agricultural products in Brazil from 1995 to 2004. Figure D2 presents the trend in the normalized exporting in a set of commodities with available information. Data is given in FOB dollars. Note that, between 1995 and 2004, the price of commodities were relatively stable in the international markets. Therefore, the expansion in trends can be attributed to an increase in the total amount of exported commodities. Trends in all the commodities in the figure are very similar and follow the international price of soybeans adjusted by the exchange rate. This supports the price of soybeans as a good proxy for the price of commodity in general.

Figure D1: Trends in the production of a set of relevant export oriented commodities in Brazil.

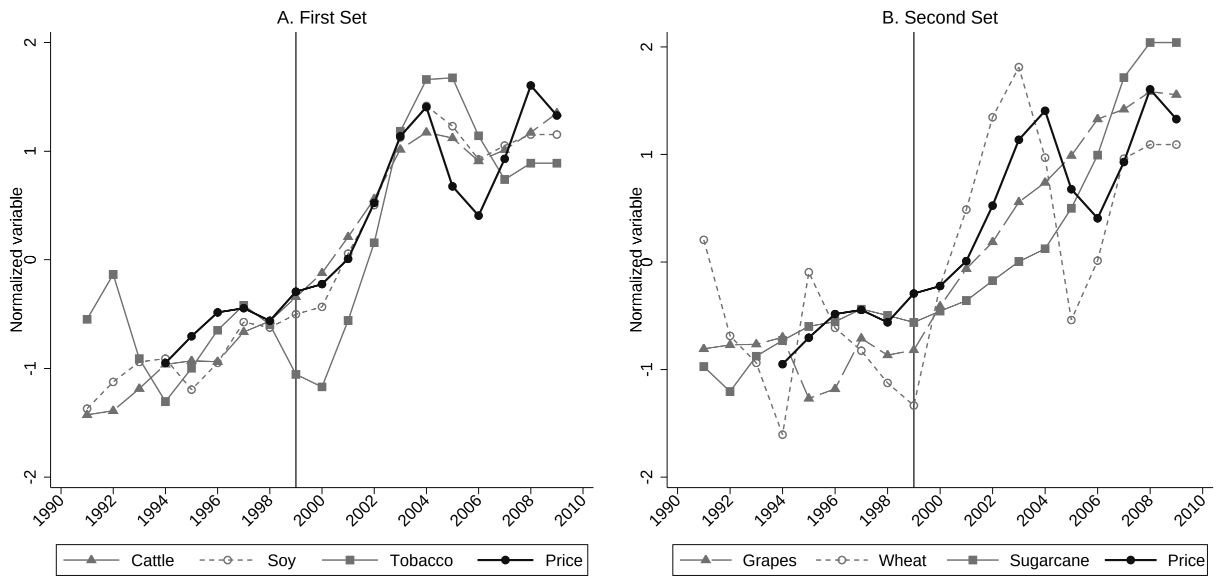
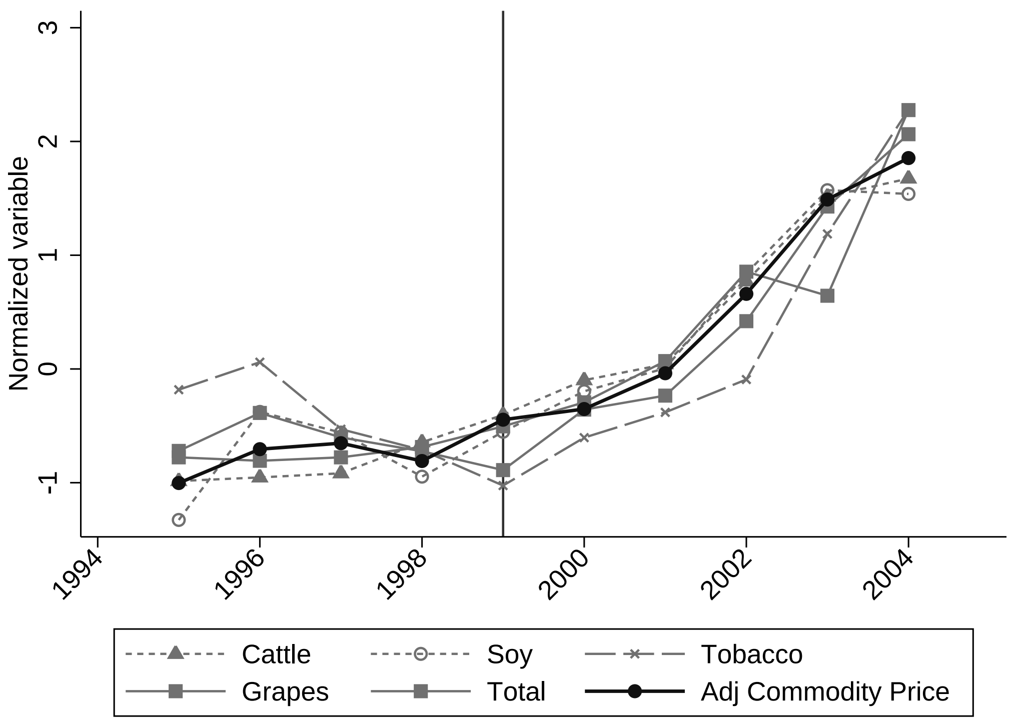


Figure D2: Trends in the exporting of a set of relevant commodities in Brazil (FOB data).



E Technological Characteristics of the Agricultural Production

In section 2 of this article I discuss the complementarity and substitutability between the main agricultural varieties produced in the Amazon region. Here, I provide data evidence to complement the discussion.

Figure E1 presents evidence of complementarity between soybeans, corn, and cotton, but a substitutability between soybeans and cattle. Further discussion is included in the main article.

Figure E1: Complementarity and Substitutability between the Main Agricultural Activities.

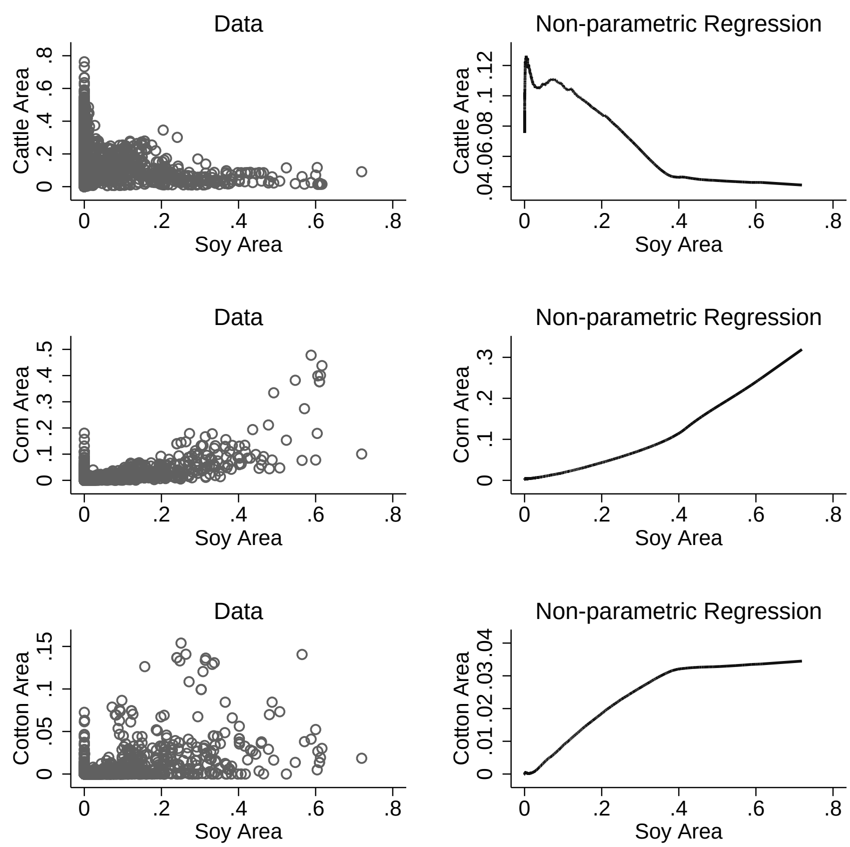


Table E1 presents evidence of larger fixed costs in the production of cattle with respect to soybeans.

Table E1: Mechanization of Soybeans and Cattle

	Dependent variable	
	Small Tractors	Big Tractors
	(1)	(2)
Log of cattle	9.473***	4.332**
	(1.613)	(1.747)
Log of soybeans	5.601***	14.58***
	(1.196)	(2.065)
Obs	225	225
R^2	0.348	0.491

Notes: *** p<0.01, ** p<0.05, * p<0.1. Robust standard in parentheses. Data from Agricultural Census of 2007. Big tractors have more than 100 horsepower.

F Model Details

F.1 Equations defining the equilibrium

Consumers

Assume consumers have a cobb-douglas preference:

$$U = \delta c_M^{\alpha_M} c_U^{\alpha_U} \tag{14}$$

Where $\delta = \dfrac{1}{\alpha_M^{\alpha_M}\alpha_U^{\alpha_U}}$. c_M are the goods imported from other regions:
Demand is given by:

$$\begin{aligned} c_{Mi} &= \alpha_M \frac{w_i}{p_M} \\ c_{Ui} &= \alpha_U \frac{w_i}{p_U} \end{aligned} \tag{15}$$

Where i indexes the city. Workers satisfy an indirect utility that provides the non-migration condition.

$$V = \frac{w_i}{e^{\gamma \alpha_M Droad_i} P_i} \tag{16}$$

Where the price index is equal to $P = (p_M^{\alpha_M} p_{Ui}^{\alpha_U})$. We will see that, given the firm profit maximizing conditions, $p_{Ui} = w_i / a_U$. $Droads$ is the distance from roads. The indirect utility function can be rewritten as:

$$\begin{aligned} V &= \frac{w_i}{e^{\alpha_1 \gamma Droad_i} p_M^{\alpha_M} p_U^{\alpha_U}} \\ &= \frac{w_i^{1-\alpha_U} a_U^{\alpha_U}}{e^{\gamma \alpha_M Droad_i} p_M^{\alpha_M}} \end{aligned} \tag{17}$$

Firms

For production, there will be 2 technologies for agricultural production:

$$y(\ell)_S = z(\ell) l(\ell)_S^{\phi} \tag{18}$$

$$y(\ell)_C = z(\ell) l(\ell)_C^{\phi} \tag{19}$$

Where ℓ defines the plot which are distributed according to a measure of land T_i. Therefore, $\ell \in (0, T_i]$. Also, note that the production will depend on the relative price of soybeans and cattle. Here, I abstract from potential differences in total factor productivity as I am not able to disentangle these from the price (in other words, $p_S a_S$ and $p_C a_C$ are combined into one variable). Each plot has a realization of an extreme value distribution (gumbel type II) [23].

The maximization problem, if producing the modern crop, is:

$$\max p_S e^{-\gamma_S Droad_i} z(\ell) l(\ell)^{\phi} - w_i l(\ell) - f_S \tag{20}$$

For cattle, the maximization problem is:

$$\max p_C e^{-\gamma_C Droad_i} z(\ell) l(\ell)^{\phi} - w_i l(\ell) - f_C \tag{21}$$

And we have the following maximization conditions.

$$\begin{gathered} w_i = p_U a_U \\ w_i = p_C e^{-\gamma_S Droad_i} z(\ell) \phi l(\ell)^{\phi-1} \\ w_i = p_S e^{-\gamma_C Droad_i} z(\ell) \phi l(\ell)^{\phi-1} \\ w_i > 0 \end{gathered} \tag{22}$$

The last condition is given by an inequality for the plots with no production.

Potential profits in each plot is given by:

$$\begin{gathered} \pi(\ell)_S = (\frac{p_S e^{-\gamma Droad_i} z(\ell)}{w_i^{\phi}})^{\frac{1}{1-\phi}} \Phi - f_S \\ \pi(\ell)_C = (\frac{p_C e^{-\gamma Droad_i} z(\ell)}{w_i^{\phi}})^{\frac{1}{1-\phi}} \Phi - f_C \\ \pi(\ell) = 0 \end{gathered} \tag{23}$$

Where $\Phi = (\phi^{\frac{\phi}{1-\phi}} - \phi^{\frac{1}{1-\phi}})$ The landowner can choose to not use the land. If he decides

[23] I chose this distribution because it is readily available in Matlab.

not to clear the land, profits are equal to zero.

The production is given by the following equations:

$$Y_{Si} = \left(\frac{p_S \phi}{w_i}\right)^{\frac{\phi}{1-\theta}} \int_{\bar{z}_i}^{\infty} z^{\frac{1}{1-\phi}} dz$$

$$Y_{Ci} = \left(\frac{p_C \phi}{w_i}\right)^{\frac{\phi}{1-\theta}} \int_{\underline{z}_i}^{\bar{z}_i} z^{\frac{1}{1-\phi}} dz$$

Labor Market Clearing

Note that, because this is an open city model, N_i is endogenous.

$$L_i^S + L_i^C + L_i^U = N_i \tag{24}$$

Where $l(\ell)_C$ and $l(\ell)_S$ are given by the following in each plot of land with positive production (labor demand is zero for uncleared land):

$$\begin{aligned} l(\ell)_S &= (\frac{p_S e^{-\gamma_S Droads_i} \phi z(\ell)}{w_i})^{\frac{1}{1-\phi}} \\ l(\ell)_C &= (\frac{p_C e^{-\gamma_C Droads_i} \phi z(\ell)}{w_i})^{\frac{1}{1-\phi}} \end{aligned} \tag{25}$$

Total labor in the soybean sector is given by the following (where T_i^S is the total area producing soybeans and T_i^C the one producing cattle) equation. Also, assume that each city has a measure T_i of land:

$$\begin{aligned} \int_{\ell \in T_S} l(\ell)_S d\ell &= (\frac{p_S e^{-\gamma_S Droads_i} \phi}{w_i})^{\frac{1}{1-\phi}} \int_{\ell \in T_i^S} z(\ell)^{\frac{1}{1-\phi}} d\ell \\ L_i^S &= (\frac{p_S e^{-\gamma_C Droads_i} \phi}{w_i})^{\frac{1}{1-\phi}} \int_{\ell \in T_i^S} z(\ell)^{\frac{1}{1-\phi}} d\ell \end{aligned} \tag{26}$$

And cattle is obtained analogously:

$$L_i^C = (\frac{p_C e^{-\gamma_C Droads_i} \phi}{w_i})^{\frac{1}{1-\phi}} \int_{\ell \in T_i^C} z(\ell)^{\frac{1}{1-\phi}} d\ell \tag{27}$$

In the service sector, labor will be defined by the non-tradeable sector, which means that the demand for labor from the landowner and the workers will have to meet the supply of goods.

Non-tradeable condition

$$\alpha_U I_i = p_{Ui} y_{Ui}$$
$$\alpha_U I_i = w_i L_{Ui} \tag{28}$$

And the total income is equal to:

$$I_i = Profits + Wages$$
$$I_i = \int_{\ell \in T_S} \pi_S(z(\ell))d\ell + \int_{\ell \in T_C} \pi_C(z(\ell))d\ell + w_i N_i \tag{29}$$

Tradeable condition

For the tradable balance we need the following:

$$\alpha_M I_i = Soybeans_i + Cattle_i$$

$$\alpha_M I_i = \int_{\ell \in T_i^S} p_S e^{-\gamma_S Droads_i} y_S(z(\ell))d\ell - \int_{\ell \in T_i^S} f_S d\ell$$
$$+ \int_{\ell \in T_i^C} p_C e^{-\gamma_C Droads_i} y_C(z(\ell))d\ell - \int_{\ell \in T_i^C} f_C f(z) dz(\ell)$$
$$\alpha_M I_i = \int_{\ell \in T_i^S} \frac{w_i l(\ell)_S}{\phi} d\ell - f_S T_i^S \tag{30}$$
$$+ \int_{\ell \in T_i^C} \frac{w_i l(\ell)_C}{\phi} d\ell - f_C T_i^C$$
$$\alpha_M I_i = \frac{w_i}{\phi}(L_{Si} + L_{Ci}) - T_i^S f_S - T_i^C f_C$$

Note that we use the following transformation here.

$$w_i = p_x e^{-\gamma_x Droads_i} z(\ell) l(\ell)^{\phi-1} \phi$$

$$\frac{l_i(\ell) w_i}{\phi} = p_x e^{-\gamma_x Droads_i} z(\ell) l(\ell)^{\phi} \tag{31}$$

$$\frac{l_i(\ell) w_i}{\phi} = p_x e^{-\gamma_x Droads_i} y_x(\ell)$$

Where $x \in \{S, C\}$. Combining both the tradeable and the non-tradeable sector (and aggregating the labor demand in the agricultural sector), we have:

$$L_i^S = \frac{\alpha_U}{\alpha_M} \left(\frac{L_{Ai}}{\phi} - T_i^S \frac{f_S}{w_i} - T_i^C \frac{f_C}{w_i} \right) \tag{32}$$

L_{Ai} is the total employment in agriculture. Note that, in the absence of roads, we would have the same proportion of people employed in the service sector and people employed in the agricultural one in every city. However, because of the fixed costs, we do not have the exact relationship.

Equilibrium

With the following set of exogenous parameters

$$\Sigma = (\gamma_S, \gamma_C, \theta, p_S, p_C, p_U, a_U, \phi, V, f_S, f_C, \alpha_U, \alpha_S)$$

, the equilibrium is then given by:

- *Landowners maximize profits*
- *Firms maximize profits*
- *Agents maximize utility*
- *Agents are indifferent between different cities*
- *Labor market clears*
- *Balance in tradeable and non-tradeable sector*

The equilibrium provides endogenously $X_i = (L_i^S, L_i^C, L_i^U, T_i^S, T_i^C, w_i, p_i^U, N_i)$

Equations defining the equilibrium

With the following set of exogenous parameters:

$$\Sigma = (\gamma, \theta, p_S, p_C, p_U, a_U, \phi, V, f_S, f_C, \alpha_M, \alpha_U)$$

, the equilibrium is then given by following equations:

$$L_i^S = (\frac{p_S e^{-\gamma_S Droads_i}\phi}{w_i})^{\frac{1}{1-\phi}} \int_{\ell \in T_i^S} z(\ell)^{\frac{1}{1-\phi}} d\ell \tag{33}$$

$$L_i^S = (\frac{p_S e^{-\gamma_S Droads_i}\phi}{w_i})^{\frac{1}{1-\phi}} \int_{\ell \in T_i^S} z(\ell)^{\frac{1}{1-\phi}} d\ell \tag{34}$$

$$L_i^C = (\frac{p_C e^{-\gamma_C Droads_i}\phi}{w_i})^{\frac{1}{1-\phi}} \int_{\ell \in T_i^C} z(\ell)^{\frac{1}{1-\phi}} d\ell \tag{35}$$

$$L_i^U = \frac{\alpha_U}{\alpha_M} \left(\frac{L_{Ai}}{\phi} - T_i^S \frac{f_S}{w_i} - T_i^C \frac{f_C}{w_i} \right) \tag{36}$$

$$w_i = (\frac{V e^{\gamma \alpha_M Droad_i} p_M^{\alpha_M}}{a_S^{\alpha_U}})^{\frac{1}{1-\alpha_U}} \tag{37}$$

$$p_{Ui} = \frac{w_i}{a_U} \tag{38}$$

$$Y_{Si} = \left(\frac{p_S e^{\gamma_S Droads_i}\phi}{w_i} \right)^{\frac{\phi}{1-\phi}} T_i \int_{\ell \in T_i^S} z^{\frac{1}{1-\phi}} d\ell \tag{39}$$

$$Y_{Ci} = \left(\frac{p_C e^{\gamma_C Droads_i}\phi}{w_i} \right)^{\frac{\phi}{1-\phi}} T_i \int_{\ell \in T_i^C} z^{\frac{1}{1-\phi}} d\ell \tag{40}$$

The equilibrium provides endogenously $X_i = (L_i^S, L_i^C, L_i^U, T_i^S, T_i^C, w_i, p_i^U, N_i, Y_{Si}, Y_{Ci})$

F.2 Proof of Proposition (Partition of Land)

So far, I have worked with an arbitrary partition of land T_i^S and T_i^C. Here, I define the plots of land in the county with cattle, soybeans, or forest according to quality of the plot $z(\ell)$.

In the main body of the article, I stated the following proposition, which I prove below.

Proposition 2 ***(Partition of Land 1)*** *Define z_i^{SC} as the plot (ℓ) where the farmer would be indifferent between producing cattle and soybeans, and z_i^{CF} for the plot (ℓ) where the* *farmer would be indifferent between producing cattle and clearing the forest. If the conditions $p_S > p_C$, $f_S > f_C$, and $z_i^{CF} < z_i^{SC}$ are satisfied, then the area of the county can be partitioned in three where soybeans is produced in ℓ such that $z_i^{CF} < z(\ell) < z_i^{CS}$, cattle is produced in $z_i^{CF} < z(\ell) < z_i^{CS}$, and forest is left uncleared for $0 < z(\ell) < z_i^{CF}$.*

Proof First, let's define z_i^{CF}, z_i^{CS}, and z_i^{SF}. From the profit function, we can establish the condition such that $\pi_{S(\ell)} \geq \pi_{C(\ell)}$ that will provide an equation for z_i^{CS}. Manipulation of this inequality leads to:

$$z_{(\ell)} \geq (\frac{w_i^{\phi}}{p_{Si} - p_{Ci}})(f_S - f_C)^{1-\phi}\Phi^{\phi-1} \tag{41}$$

Where $\Phi = (\phi^{\frac{\phi}{1-\phi}} - \phi^{\frac{1}{1-\phi}})$. An increase in $z_{(\ell)}$ always makes this inequality more likely to be satisfied, which defines z_i^{CS} as an unique value of $z_{(\ell)}$ that satisfies the above equation as an equality. We can derive z_i^{CF} from $\pi_{C(\ell)} \geq 0$, which leads to the following:

$$z_{(\ell)} \geq e^{\gamma Droads_i} \frac{w_i^{\phi}}{p_{Ci}^{*}} (\frac{f_C}{\Phi})^{1-\phi} \tag{42}$$

Again, there is only one z_i^{CF} that makes the above equation an equality [24]. We can also define z_i^{SF} in the same manner.

First, note, given $p_S > p_C$, better land will always be used for soybeans, because of the complementarity between the quality of land z_i and the price of the activity p_S and p_C. In other words, the cross-derivative between z_i and p_S or p_C is positive. As a consequence, we know that for $z_i > z_i^{CS}$, the land will be more profitable producing soybeans. Note also that, because $p_S > p_C$ and $f_S > f_C$, then $z_i^{CS} > 0$ in equation 41. Similar logic can be used to show that $z_i^{CF} > 0$ and $z_i^{SF} > 0$.

Now, assume that $z_i^{CF} < z_i^{SC}$. If $z_i^{SF} < z_i^{CF}$, then it means that at the plot z_i^{SF} producing cattle leads to negative profits, while producing soybeans leads to a non-negative one. However, if this is true, for the plot of land z_i^{SC}, the farmer would not be indifferent between producing cattle and soybeans, because of $z_i^{SF} < z_i^{SC}$ and the positive cross-derivative between z_i and p_S and p_C, at plot z_i^{SC} the farmer would actually have a larger profit producing soybeans, which contradicts the definition of z_i^{SC}. If $z_i^{SF} > z_i^{CS}$, it would mean that at the plot z_i^{CS} the farmer has a negative profit (again, I use the positive cross-derivative), which contradicts the fact that the farmer already has a positive profit with cattle at that point.

[24] Note that here the assumption that both cattle and soy have the same ϕ plays an important role in facilitating the solution of the problem. With different ϕ, we would have multiple solutions. However, the same patterns we will look at the data would be possible to be generated by using different ϕ.

Finally, analysing $z_i^{CF} < z_i^{SC}$ provides the parameter condition for the partition:

$$(\frac{f_S - f_C}{f_C})^{1-\phi} > \frac{p_{Si} - p_{Ci}}{p_{Ci}} \tag{43}$$

This last inequality concludes the proof. □

Proposition 3 ***(Partition of Land 2)*** *If the conditions $p_S > p_C$, $f_S > f_C$, and $z_i^{CF} \leq z_i^{SC}$ are satisfied, then the area of the county can be partitioned in two where soybeans is produced in ℓ such that $z_i^{SF} < z(\ell)$, and forest is left uncleared for $0 < z(\ell) < z_i^{SF}$. No cattle is produced.*

Proof Above z_i^{SC}, it is always more profitable to produce with soybeans. If $z_i^{CF} \leq z_i^{SC}$, it means that at the point when cattle becomes profitable for production, soybeans are already a more profitable activity. Therefore, no cattle is produced. □

Proposition 4 ***(Partition of Land 3)*** *If the conditions $p_S > p_C$ and $f_S \leq f_C$ are satisfied, then the area of the county can be partitioned in two where soybeans is produced in ℓ such that $z_i^{SF} < z(\ell)$, and forest is left uncleared for $0 < z(\ell) < z_i^{SF}$. No cattle is produced.*

Proof If the price of soybeans is better than the one for cattle, and on the top of that its fixed costs are smaller, then soybeans is always more profitable than cattle production. In this case, the county specialized in soybean production. □

If we follow analogously with $p_C > p_S$, then we have the cases with cattle being produced in the better plots within the county and another 3 partitions of land are obtained.

F.3 Revisiting equations defining the equilibrium

With the partition of $z(\ell)$. We can use the following transformation for the land used for soy (given proposition 1), which is very useful for the structural estimation:

$$\frac{\int_{\ell \in T_i^S} z(\ell)^{\frac{1}{1-\phi}} d\ell}{T_i^S} = \int_{z^{SC}}^{\infty} z(\ell)^{\frac{1}{1-\phi}} f(z|z > \bar{z}_i) dz(\ell)$$

$$\begin{aligned}
\int_{\ell \in T_i^S} z(\ell)^{\frac{1}{1-\phi}} d\ell &= T_i^S \int_{z^{SC}}^{\infty} z(\ell)^{\frac{1}{1-\phi}} f(z|z > z_i^{SC}) dz(\ell) \\
&= P(z > z_i^{SC}) T_i \int_{z^{SC}}^{\infty} z(\ell)^{\frac{1}{1-\phi}} \frac{f(z)}{P(z > z_i^{SC})} dz(\ell) \\
&= T_i \int_{z^{SC}}^{\infty} z(\ell)^{\frac{1}{1-\phi}} f(z) dz(\ell)
\end{aligned}$$

Using these transformations, we can restate the equilibrium conditions as:

$$T_i^S = T_i \int_{z^{SC}}^{\infty} z(\ell) f(z) dz(\ell) \tag{44}$$

$$T_i^C = T_i \int_{z^{CF}}^{SC} z(\ell) f(z) dz(\ell) \tag{45}$$

$$L_i^S = (\frac{p_S e^{-\gamma_S Droads_i} \phi}{w_i})^{\frac{1}{1-\phi}} \int_{z^{SC}}^{\infty} z(\ell)^{\frac{1}{1-\phi}} dz(\ell) \tag{46}$$

$$L_i^C = (\frac{p_C e^{-\gamma_C Droads_i} \phi}{w_i})^{\frac{1}{1-\phi}} \int_{z^{CF}}^{SC} z(\ell)^{\frac{1}{1-\phi}} dz(\ell) \tag{47}$$

$$L_i^U = \frac{\alpha_U}{\alpha_M} \left(\frac{L_{Ai}}{\phi} - T_i^S \frac{f_S}{w_i} - T_i^C \frac{f_C}{w_i} \right) \tag{48}$$

$$w_i = (\frac{V e^{\gamma \alpha_M Droad_i} p_M^{\alpha_M}}{a_S^{\alpha_U}})^{\frac{1}{1-\alpha_U}} \tag{49}$$

$$p_{Ui} = \frac{w_i}{a_U} \tag{50}$$

$$Y_{Si} = \left(\frac{p_S e^{\gamma_S Droads_i} \phi}{w_i} \right)^{\frac{\phi}{1-\phi}} T_i \int_{\ell \in T_i^S} z^{\frac{1}{1-\phi}} d\ell \tag{51}$$

$$Y_{Ci} = \left(\frac{p_C e^{\gamma_C Droads_i} \phi}{w_i} \right)^{\frac{\phi}{1-\phi}} T_i \int_{\ell \in T_i^C} z^{\frac{1}{1-\phi}} d\ell \tag{52}$$