

# The Causes and Consequences of Agricultural Specialization in Brazil

Heitor S. Pellegrina\*

*Brown University*

**JOB MARKET PAPER**

November 17, 2016

[[Latest Version](#)]

## Abstract

I formulate and estimate a spatial model to study causes and consequences of the regional patterns of agricultural specialization in Brazil. Both heterogeneities in natural advantages across counties and input intensity across crops are central causes of agricultural specialization, but differences in transportation costs across crops play a minor role. I evaluate the consequences of trade with China and public research that adapted soybeans to tropical regions. These shocks expanded the production of land-intensive crops in particular regions, releasing labor to economic activities elsewhere, generating gains that are not captured by local measures of the impact of these shocks. Moreover, macroeconomic conditions shape the return to adapting soybeans, for example, the internal rate of return is 30% lower in the absence of trade with China.

**Keywords:** Agriculture, Trade, Spatial Economics, Nature, Agricultural Research

**JEL classification:** J16, N5, N7, O13, O18, Q17

---

\*Ph.d. Candidate at Brown. Email: [heitor\\_pellegrina@brown.edu](mailto:heitor_pellegrina@brown.edu). I am grateful to Jonathan Eaton and Andrew Foster for guidance and support during the course of this project. I am thankful to Jesse Shapiro and Matthew Turner for numerous comments on this paper. Special thanks to Sebastian Sotelo for detailed discussion. I appreciate comments from Guillermo Alves, Juliano Assunção, Joaquin Blaum, Daniel Bjorkegren, Arthur Bragança, Lorenzo Caliendo, Farid Farrokhi, Michael Gechter, Samuel Kortum, Kim Ruhl, Anja Sautmann, Lucas Scottini, Meredith Startz, Daniel Tanis, and participants in presentations held at Brown, Penn State, the Conoco-Phillip Seminar, NEUDC and FREIT-EIIT. Part of this paper was written while I was visiting the Economics department at Penn State University. I am grateful to Institute at Brown for Environment and Society (IBES) and Development Bank of Latin America (CAF) for financial support for this research. All errors are my own.

# 1 Introduction

There are vast differences in the regional patterns of agricultural specialization. For example, in Brazil, in 2006, farms in the county of Cuiabá-MT specialized in the large-scale production of soybeans to foreign markets and employed an average of 85 *ha* of land, whereas farms in Ilhéus-BA specialized in the small-scale production of fruits for the domestic market and employed an average of 2.9 *ha*. These disparities are also observed within activities: cattle ranchers in Cuiabá-MT employed an average of 111 *ha* and, in Ilhéus-BA, 37 *ha*. While part of these differences may be caused by nature, the literature in spatial economics argue that they are also shaped by market conditions related to the location of population, such as the price of land and the access to consumers (Fujita et al., 2001; Samuelson, 1983). This literature often refers to natural advantages as the “First nature” causes of specialization and to market conditions related to the location of population as the “Second nature” causes.

Among policy makers, there is great interest in policies that change the agricultural specialization of specific regions. For example, the Brazilian government funded agricultural research that allowed the production of soybeans in tropical regions (Assuncao and Braganca, 2015), public research in the United States fostered the expansion of hybrid-corn in southern areas of the country (Griliches, 1957,5) and, nowadays, there are a number of research cooperation projects in African countries such as Mozambique and Ghana (Bank, 2009; Scoones et al., 2016). While the goal of these policies is typically to increase exports and agricultural output in specific areas, their effects propagate across regions and activities due to trade, land market adjustments and migration. Therefore, understanding the aggregate consequences of such policies requires a framework that accounts for the economic interactions between regions and activities.

I formulate and estimate a spatial model to investigate the regional patterns of agricultural specialization in Brazil. In doing so, I assemble comprehensive county-level data on agricultural production, trade, commodity prices and transportation. I present facts about Brazilian agriculture for the baseline year of this study, 2006. Motivated by these facts, the model combines elements from comparative advantage theories (Costinot et al., 2016; Donaldson, 2012; Donaldson and Hornbeck, 2016; Sotelo, 2016) with market-access theories of spatial location of population and agriculture (Fujita, 2012; Fujita et al., 2001; Samuelson, 1983). I quantify the role of different causes of specialization and how these causes shaped the consequences of two major shocks to specialization: the adaptation of soybeans to tropical regions and trade with China.

The structure of the model is the following. There is a set of counties connected by transportation network. Counties trade the output of multiple agricultural sectors and one urban sector in an Eaton and Kortum (2002) fashion. Each agricultural sector presents different technologies in terms of transportation costs, trade elasticities, land intensity and intermediate input intensity. To define farm size, I assume decreasing returns to scale in agricultural production and a minimum labor requirement to

set up a farm. Workers are mobile and earn wages. Landowners are immobile and earn land rents. I introduce international trade assuming that the global economy has a fixed expenditure in each sector and an international competitor that competes with domestic producers. With this formulation, the international prices of commodities are still influenced by shocks to the domestic economy.

Six sources of heterogeneity determine agricultural specialization in the model: (i) natural advantages bring classic technological comparative advantage forces of specialization, (ii) low land rents are beneficial for the production of land-intensive crops, (iii) proximity to urban centers induces the production of intermediate input intensive crops, (iv) lower wages increase the competitiveness of crops with stronger decreasing returns, (v) proximity to consumers is beneficial for the production of perishable products, and (vi) counties with a cheaper bundle of inputs are more competitive in activities with high trade elasticities. Despite accounting for many determinants of specialization, the model delivers gravity equations that show how each source of heterogeneity translates into observable patterns of specialization in a tractable way.

There are three sets of parameters in the model: (1) trade costs and trade elasticities, (2) agricultural technologies and consumers' preferences, and (3) natural advantages and amenities. As in [Donaldson \(2012\)](#), I infer trade costs from price differentials between counties. I then use estimates of trade costs to identify trade elasticities from estimates of gravity equations. To estimate parameters related to the agricultural production function, I minimize the distance between statistics from the agricultural census and predicted statistics given by the structural parameters of the model. I calibrate consumption shares to match the absorption of the Brazilian economy in each sector. Following [Allen and Arkolakis \(2014\)](#), I recover the distribution of natural advantages and amenities as “residuals” that make the model perfectly explain the spatial distribution of workers and wages for 2006.

With the model estimated, I simulate the economy shutting down different sources of heterogeneity driving specialization and I verify the ability of the simulated economy to predict different features of the data. While the full model correlates perfectly with this share of farms in each activity, this correlation decreases to 0.05 without differences in natural advantages across counties and to 0.45 without heterogeneity in input intensity across crops. Despite the fact that perishable goods are 4 times more costly to transport than non-perishables ones, differences in transportation costs across crops play a minor role. As an alternative exercise, I shut down different sources of technological heterogeneity and I re-estimate the residuals associated with natural advantages. The variance of these residuals increases by 30% when I restrict the input intensity parameters of the model to be the same, but there is almost no change in this variance when I restrict transportation costs to be the same.

I then study the effects of shocks to agricultural specialization. First, I simulate the model in a counterfactual scenario where soybeans are not produced in tropical regions. Even though the production of soybeans in tropical regions account for 0.5% of Brazilian's GDP, Brazilian's GDP falls by 0.85%. In part, this occurs because in areas where soybeans “disappear”, land is occupied by alternative activities that are less-land intensive and that absorb labor from other regions of the econ-

omy. The migration of workers to the area where soybeans “disappear” minimizes the local loss in GDP, but this comes at a cost for other regions of the economy that as a consequence have less labor to produce their own products. As such, measuring the local GDP loss without accounting for the propagation of shocks across regions understate the aggregate loss for the economy. Looking at welfare, I found an aggregate loss of roughly 1%, but with important distributional impacts: landowners generally lose more than workers, but landowners in temperate climate regions producing soybeans in the counterfactual scenario are better-off without the adaptation since they no longer compete with soybean producers in tropical regions.

With the aggregate gains from tropical soybeans in hand, I bring information on the budget of the Brazilian research department (EMBRAPA)<sup>1</sup> that developed the technology that allowed the production of soybeans in tropical regions to calculate the annualized internal rates of returns to research. Restricting the measure of benefits to changes in the total output of the soybean sector, I find a measured return of 30%. Using changes in total Brazilian GDP, which accounts for the propagation of shocks in the economy, I find a return of 50%. Furthermore, I use the model to analyze the returns to research under different economic and institutional conditions and I show that: (i) if Brazil does not export soybeans to China, returns are 30% lower; (ii) if there is an additional *ad valorem* land tax of 25%, returns are 10% lower.

Also, I evaluate a counterfactual scenario where Brazil does not export to China. Even though exports to China are largely concentrated in agricultural commodities. I find a decrease in the number of workers in urban sectors. This occurs because agricultural exports to China are concentrated in land-intensive activities that release labor for other regions and activities. While exports to China account for 1.6% of the Brazilian output, total output decreases by 4.8% and aggregate welfare by 4.7%.

I close my investigation testing the ability of the model to predict historical data on agricultural specialization that was not used in the estimation. To do so, I use State-level data on the share of the area employed in each activity. I find that simulations of the model without the adaptation of soybeans and the trade with China decrease the ability of the model to predict the share of land use in each State and activity in 2006. Looking at data for the 1965, I find the opposite result: the absence of the studied shocks increases the correlation of the model with the data. In other words, the regional patterns of specialization in the simulated economy become more similar to actual patterns of specialization observed in the data for years that preceded these shocks.

This paper relates to growing research formulating quantitative trade and spatial models (Allen, 2014; Baum-Snow et al., 2016; Costinot and Donaldson, 2014; Costinot et al., 2016; Desmet et al., 2015; Donaldson, 2012; Donaldson and Hornbeck, 2016; Fajgelbaum and Redding, 2014; Monte et al., 2015; Morten and Bryan, 2015; Nagy, 2015; Sotelo, 2016). The main distinguishing feature of this paper is the introduction of both heterogeneity in input intensity and worker mobility into a

---

<sup>1</sup>EMBRAPA stands for *Empresa Brasileira de Pesquisa Agropecuária*.

quantitative trade model. These extensions bring causes of agricultural specialization emphasized by market-access theories of spatial location of agriculture to a recent quantitative literature in spatial economics (Fujita et al., 2001; Samuelson, 1983). I find that these causes are important drivers of regional specialization and that the interaction between heterogeneity in input intensity and worker mobility matters for the aggregate consequences of economic shocks.

This paper speaks to different fronts of research in development economics. First, by formulating a framework that accounts for the general equilibrium effects of agricultural innovations, this paper provides a complementary approach to research evaluating the impact of agricultural innovations using reduced form techniques (Assuncao and Braganca, 2015; Bustos et al., 2016; Foster and Rosenzweig, 2004). Second, this paper relates to studies on farm size distribution and productivity (Adamopoulos and Restuccia, 2014; Eastwood et al., 2010; Foster and Rosenzweig, 2011). Variation in farm size across regions and countries has been typically attributed to imperfections in land and credit markets, or to heterogeneity in natural advantages for large-scale activities. In this paper, spatial forces govern agricultural specialization as well as farm size.

Finally, the cost-benefit analysis of the adaptation of soybeans to tropical areas relates to a large literature evaluating the returns from agricultural research (Alston et al., 2000; Evenson, 2001; Griliches, 1958,6). To the best of my knowledge, this is the first paper to evaluate the returns to agricultural research using a general equilibrium framework.

The rest of this paper is organized as follows. Section 2 presents the data. Section 3 presents facts about Brazilian agriculture. Section 4 develops the model. Section 5 estimates the model. Section 6 investigates causes of agricultural specialization. Section 7 studies consequences of shocks to agricultural specialization. Section 8 concludes.

## 2 Data

The dataset contains information on: (1) agricultural production, (2) road networks, (3) trade between counties and external markets, and (4) wholesale agricultural prices. The unit of observation is a county (*municípios*), which provides a fine level of geographical disaggregation: there are 5564 of them and the median county has roughly 400  $km^2$ . Panel A in Figure 1 shows the territorial division of counties.

First, the primary data for this paper comes from the agricultural census of 2006 organized by the Brazilian Institute of Geography and Statistics (IBGE). Information on wages and labor employment is available at the county level. Data on revenues, land use and the number of farms are disaggregated by county and agricultural activity. I aggregated information into 11 agricultural sectors that I can link with export information in the trade data. I combine this data with the demographic census of 2010 to obtain total population per county.

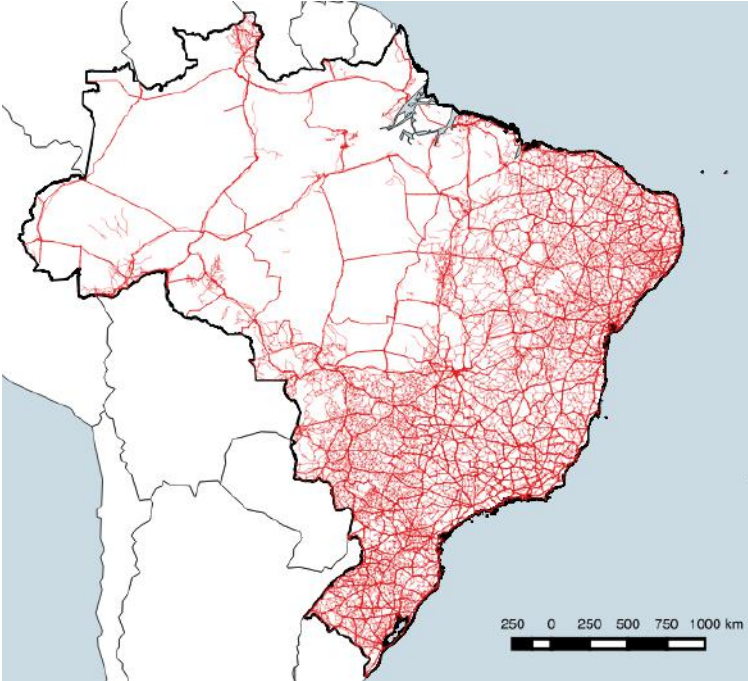
Second, I use electronic maps on the Brazilian road network. I connect this network with counties

Figure 1: Summary of the Data

A. Territorial Division of Counties



B. Road Network



Notes: Panel A highlights the territorial division of counties in the data. Panel B shows the complete road network used to calculate the minimum travel distance between counties.

using the official location of their administrative centers. I then define a matrix of nodes and arcs that allows me to calculate a county-to-county matrix of minimum travel distances. Figure 1 shows the road network used to construct this matrix.<sup>2</sup> To connect the Brazilian economy to the international market in the estimation of the model, I identify 14 major ports in Brazil. I then assume that counties reach markets overseas using the closest major port, that all ports in Brazil are equally distant to markets overseas and that counties reach neighboring markets such as Argentina and Paraguay using the internal road system.

Third, the Brazilian Ministry of Development, Industry and External Commerce provides information on the total value exported (free on board) from every county to different external markets. Information is disaggregated by products and includes annual information at the county level by sector from 2000 to 2015.

Fourth, I collected county-level information on wholesale prices of 31 varieties of agricultural products. Information is available at a monthly frequency for a subset of counties and comes from official statistics of the State Secretaries of Agriculture. The final wholesale price dataset goes from 1998 to 2015.

In addition, I constructed measures of agricultural suitability from pixel level maps from FAO-GAEZ. These measures are not used in the structural estimation of the model, but I use them for auxiliary analyses. I use data for 22 crops and, for each crop, two types of technologies are available: one related to high use of intermediate inputs and another for low use.<sup>3</sup>

### 3 Facts

This section presents three facts about Brazilian agriculture which motivate the building blocks of my model. Fact 1 shows that natural advantages alone cannot explain a strong link between market access and the regional patterns of agricultural specialization. Fact 2 discusses variation in the use of inputs across crops and counties. Fact 3 presents the agricultural export of Brazil. These facts motivate the formulation of a spatial model where natural advantages are not the only cause of specialization, where heterogeneity in input intensity is a potential cause of specialization, and where different activities have various degrees of integration with international markets.

*Fact 1: Natural advantages alone cannot explain why market-access access is associated with the regional patterns of agricultural specialization.*

---

<sup>2</sup>Given the density of the network, I simplified the road network using only interstate highways, and I assume that the administrative center of each county is at the euclidean distance from the nearest interstate highway. I used Dijkstra's algorithm to calculate the travel distance between counties. Note that I do not have information on the railroad system. According to interviews with producers, the bulk of agricultural production is transported using the highway system. In addition, it was mentioned that, because of competition between modes of transportation, differences in freight costs tend to be small. For comparison, in the US, [Duranton et al. \(2013\)](#) find that 70% of the exported value uses the road system.

<sup>3</sup>See [Costinot et al. \(2016\)](#) for a detailed description of the FAO-GAEZ measures of agricultural suitability.

As an initial inspection of the data, I construct a measure of market access that is consistent with previous measures used in the literature (Baum-Snow et al., 2016; Donaldson and Hornbeck, 2016). For each county  $i$ , I weight the population in every county  $n$  ( $Pop_n$ ) according to the inverse of the minimum travel distance between county  $i$  and  $n$  ( $MTD_{in}$ )

$$MA_i \equiv \sum_{n=1}^C \exp(-\lambda MTD_{in}) Pop_n, \quad (1)$$

where  $C$  is the set of counties in the data and I set  $\lambda = 0.01$ , which is consistent with the magnitude of trade costs in Brazil.<sup>4</sup>

Figure 2 shows a particular region of Brazil where land regulation and natural advantages are relatively homogeneous. Counties closer to the Atlantic coast have better market access compared to remote regions in the west. Panels B and C indicate that counties with better market access have a larger proportion of farms producing vegetables and a lower proportion of farms producing cattle. Panel D shows that counties with lower market access also have larger farms, indicating that heterogeneity in the use of land is associated with the regional patterns of specialization.

Figure 3 shows non-parametric regressions of agricultural production variables on market access controlling for measures of agricultural suitability from FAO. To do so, I estimate non-parametric regressions using the residuals of a regression of each variable on the measures of agricultural suitability. I divide each variable by their average in counties with the highest values in terms of market access (above 1%). Controlling for agricultural suitability, market access is still positively associated with specialization in the production of vegetables, and negatively associated with farm size or specialization in the production of cattle.<sup>5</sup> The model will account for the possibility that differences in natural advantages influence both market-access and agricultural specialization, and it will also contain mechanisms that explain why this link holds when we control for differences in natural advantages.

In addition, Figure 3 presents a positive relationship between market access and the average land price in a county, a *proxy* for land rents. This indicates the presence of a spatial equilibrium where land rents are jointly determined with the location of population and agriculture. Importantly, figures 2 and 3 are based on cross-sectional information on the location of population and, in principle, they could be rationalized without worker mobility. However, general statistics on migration indicate substantial mobility in Brazil. According to the Brazilian annual household survey of 2006, PNAD, about 55% of the heads of households lived in a county where they were not born and 40% lived in a State where

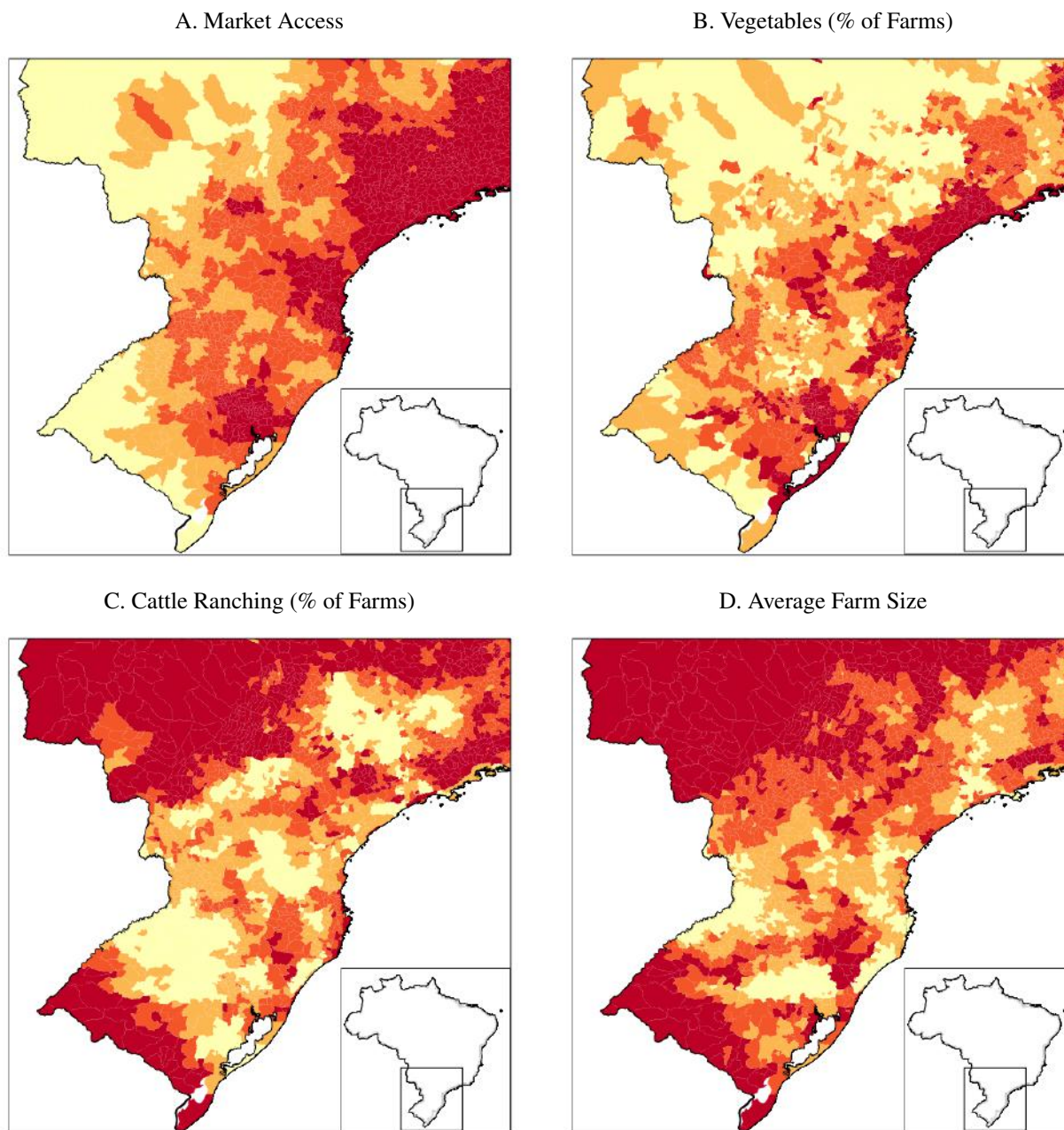
---

<sup>4</sup>An alternative approach would be to estimate  $\lambda$  to maximize the fit of regressions with key variables in the data. For example, setting the decay parameter to 0.016 would minimize the residuals of a regression of log of farm size on log of market access.

<sup>5</sup>In the appendix, I show that the relationships in Figure 3 are also robust to controls for land and credit market characteristics and the educational composition of farmers.

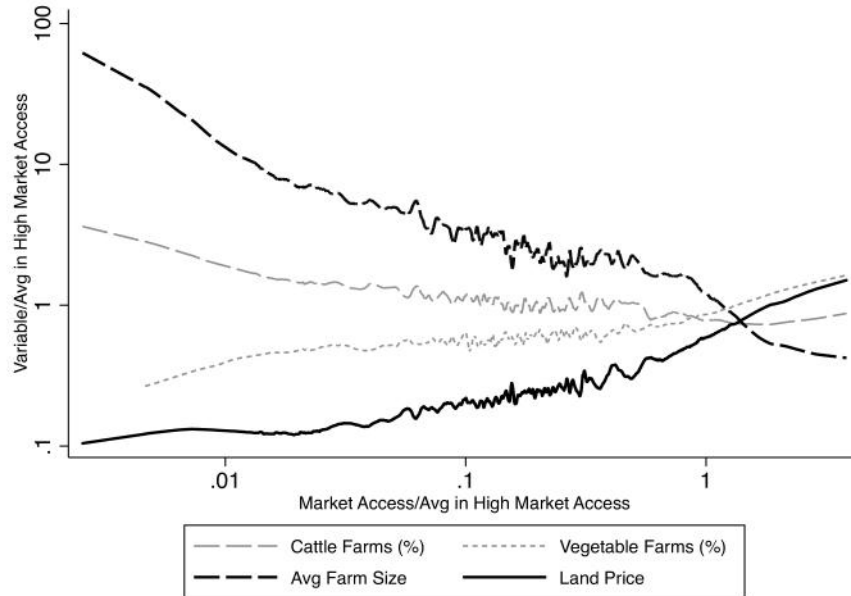


Figure 2: Market Access, Agricultural Specialization and Farm Size (2006)



Notes: The figure plots county-level data on agricultural activity and market access. Colors in each map show different quartiles of the data. Darker colors represent higher quartiles. Market access is a measure of distance of a county to population elsewhere  $MA_i \equiv \sum_{n \in C} \exp(-0.01 \times MTD_{in}) \times Pop_n$ . Average farm size is the total land used in agriculture divided by the total number of farms in each county.

Figure 3: Market Access and Agricultural Specialization controlling for Measures of Agricultural Suitability from FAO



Notes: The figure shows non-parametric regressions of different variables on market access controlling for measures of agricultural suitability from FAO. To do so, I first obtain the residuals from each variable on the full set of agricultural suitability measures. I then divide each variable by its average in counties with the highest market access (1% highest). I use locally weighted scatterplot smoothing with a bandwidth of 0.025. Land price is the total value of land in a county divided by total land use.

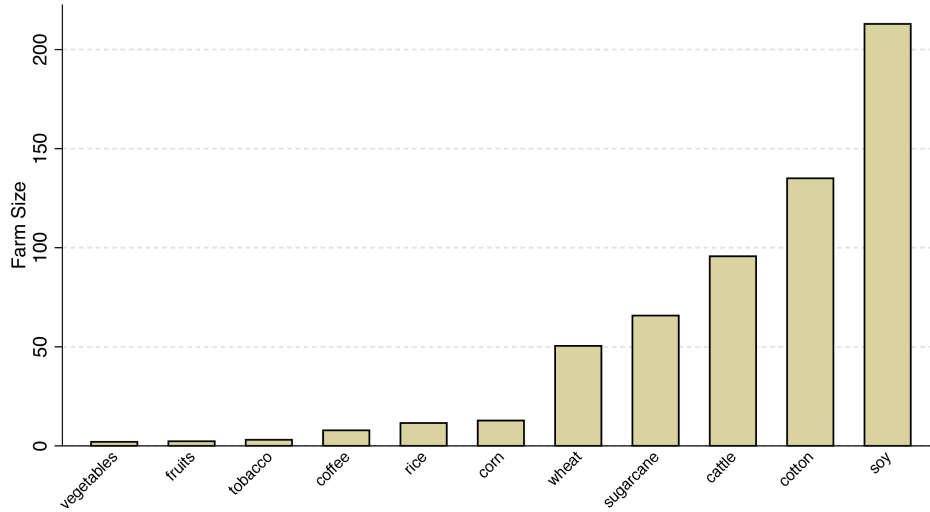
they were not born.<sup>6</sup>

*Fact 2a: There are large differences in farm size across crops. Both county and crop specific factors explain a large part of this variation.*

Using county-level information on farm size disaggregated by activity, Figure 4 shows that the average farm size in each activity is drastically different: soy uses in average 212 *ha*, whereas coffee employs 11 *ha*. Table 1 reports regressions of farm size on different sets of fixed effects to study the sources of variation in farm size. County fixed effects explain 28% of this variation, which indicates that an important part of the variation can be explained by factors such as the price of land that are common to all activities within a county. Column 2 shows that activity fixed effects explain about 35% of the variation, which indicates that a large part of the variation is explained by factors that are common to each agricultural activity. Together, both sets of fixed effects explain 60% of the variation. Therefore, a large part of the variation explained by county-fixed effects is not already absorbed by activity fixed

<sup>6</sup>In the appendix, I construct a panel data with a subset of variables that are measured since the 1960's by the census bureau and I measure how changes in the location of population are associated with changes in agricultural activity. To construct the measure of market access for previous periods, I assume the same transportation infrastructure in terms of minimum travel distance (*MTD*) and the decay parameter across periods so that changes in the measure of market access over time are uniquely driven by changes in the location of the population. With this procedure, I still find a strong link between market access and the patterns of agricultural specialization.

Figure 4: Average Farm Size per Activity



Notes: The figure shows the overall average farm size per activity using information on average farm size per activity in each county.

Table 1: Sources of Variation in Farm Size across Activities and Counties

	Regression of Farm size on		
	County FE (1)	Activity FE (2)	County & Activity FE (3)
$R^2$	0.28	0.35	0.60
Obs	30884	30884	30884

Notes: Regressions for this table use data on average farm size per activity and county. Farm size is constructed by dividing total land use by the total number of farms in each activity and each county.

effects since the explanatory power of the regression increases by 25 percentage points.<sup>7</sup>

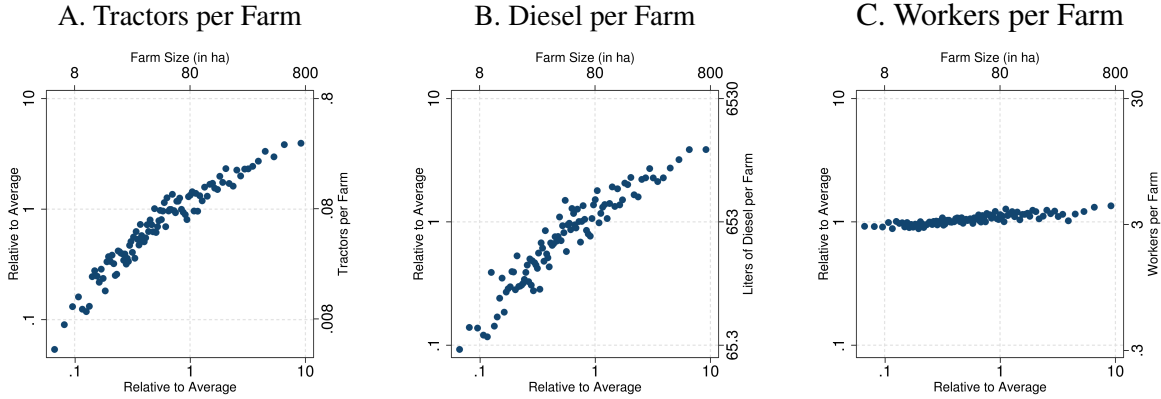
*Fact 2b: There are large differences in the use of land and intermediate inputs per farm across counties, but small differences in the use of workers per farm.*

Based on county-level data, Figure 5 reports the use of land, diesel, and tractors per farm according to the average in each percentile of the distribution of farm size. Each dot in Figure 5 represents 1/100 of counties. There are vast differences in farm size across counties. For example, in lower percentiles, farms are smaller than 10 ha, whereas in higher percentiles, they are larger than 500 ha, a 50-fold difference. Variation in tractors and diesel per farm is large and positively associated with farm size. However, the data shows low variation in workers per farm despite drastic differences in farm size: the 90 and 10 percentile of the average employment of labor is equal to 3.9 and 2.3,<sup>8</sup> whereas for

<sup>7</sup>In the appendix, I investigate factors driving variation in county fixed effects. I show that wages, land rents and measures of agricultural suitability from FAO explain a large part of the variation in county FE.

<sup>8</sup>Information includes both temporary and permanent workers, but temporary workers account for 8% of the labor force in agriculture.

Figure 5: Farm Size and the Use of Inputs across Counties



Notes: The figure shows the total use of inputs divided by the total number of farms in each county. I group the sample in a hundred groups according to the farm size percentile of the county such that each dot represents 1/100 of counties.

farm size, these percentiles are equal to 207 and 12 *ha*. This fact motivates the approach that I take to generate farm size in the model.<sup>9</sup>

*Fact 3: There is large heterogeneity in export activities across crops.*

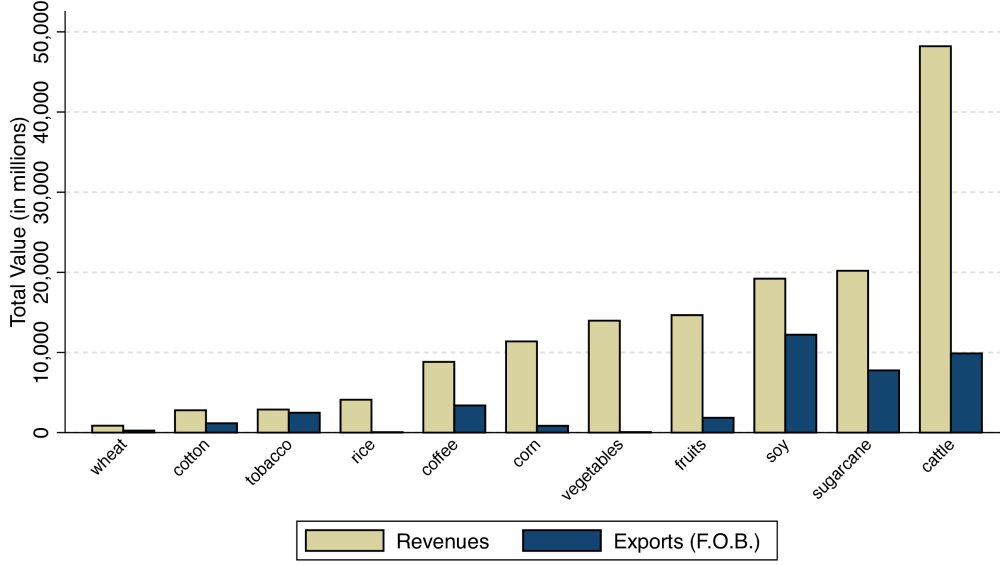
Figure 6 shows the distribution of exports and revenues for different crops. There is large heterogeneity in total revenues. Also, there is large heterogeneity in the export-orientation of different crops: about 60% of Brazilian soybeans are exported, while 30% of fruits are exported. Importantly, Brazil is a large global producer of commodities. For example, it accounts for 16% of the global production of beef, 30% of soybeans, 35% of coffee, and 25% of sugarcane. Therefore, shocks to the Brazilian agriculture likely affect the international price of commodities.

## 4 Model

This section presents a quantifiable spatial model that accounts for different sources of agricultural specialization. I use the framework in [Eaton and Kortum \(2002\)](#) to model trade between counties. I extend this framework in two ways. First, I introduce multiple agricultural sectors, which allows me to define agricultural technologies with different input intensities and transportation costs. This structure implies gravity equations defining trade relationships that I use to study the causes of agricultural specialization analytically. Second, I allow for worker mobility between counties as in [Allen and Arkolakis \(2014\)](#).

<sup>9</sup>In the appendix, I show several additional facts related to farm size in Brazil. In particular, I find that a regression of the log of agricultural workers on the log of the number of farms has a  $R^2$  of 0.881. Adding the log of farm size interacted with the log of the number of farms increases the  $R^2$  to 0.886, indicating that additional information on farm size does help predict the amount of workers in each county. For comparison, I bring information about agriculture in the United States and I found very similar patterns. For example, I also found low variation in workers per farm but drastic differences in farm size. I interpret the fact that Brazil and the United States share similar characteristics in terms of the employment of labor per farm as a property of modern techniques used in agricultural production.

Figure 6: Revenues and Total Exports per Agricultural Activity (2006)



Notes: The figure shows aggregate revenues and exports per activity in the data. I multiply exports in dollars by 2.1, the conversion rate for the Brazillian currency (Reais) in 2006.

## 4.1 Setup

The economy consists of a set of counties ( $C$ ) indexed by  $i$  or  $n$ , a measure of population ( $N$ ) that is free to move between counties and sectors, immobile landowners in each county who obtain rents from land, multiple agricultural sectors indexed by  $k$  and an urban sector indexed by  $u$  representing non-agricultural activities. Both landowners and workers consume where they live. The set of all sectors in the economy, both urban and agricultural ones, is given by  $K$ . Within each sector, there are different varieties of goods. For example, cattle ranching is a broad agricultural activity that produces different varieties of meat and coffee is a broad activity that produces different varieties of coffee beans. In spatial equilibrium, counties may produce goods in both urban and agricultural sectors. As such, counties may differ in terms of the degree of urbanization, that is, the share of workers employed in the urban sector relative to agriculture.

**Preferences** Consumers have Cobb-Douglas preferences over final goods from each agricultural sector  $k$ , and a constant elasticity of substitution between varieties of final goods within each sector

$$U_i \equiv \prod_{k=1}^K (C_{ki})^{\mu_k} a_i, \quad (2)$$

where  $C_{ki} \equiv \left( \int c_{ki}(v) \frac{\sigma_k - 1}{\sigma_k} dv \right)^{\frac{\sigma_k}{\sigma_k - 1}}$ ,  $c_{ki}(v)$  is the consumption of variety  $v$ , the utility of residents of county  $i$  is given by  $U_i$ , the elasticity of substitution between varieties by  $\sigma_k$ , the share of expenditure on goods from sector  $k$  by  $\mu_k$  (and  $\sum_k \mu_k = 1$ ) and the amenity of county  $i$  by  $a_i$ . There is a *continuum*

of varieties such that  $v \in [0, 1]$ .

**Urban Sector** The urban sector produces varieties of urban goods ( $v$ ) with the following technology

$$q_{ui}(v) = z_{ui}(v)n_{ui}(v), \quad (3)$$

where  $z_{ui}(v)$  is the efficiency of technology in each county,  $n_{ui}(v)$  is the labor employed in production and  $q_{ui}(v)$  is the output. Perfect competition ensures that the price of one unit of variety  $v$  produced in the urban sector  $p_{ui}(v)$  is  $w_i/z_{ui}(v)$ .

**Agricultural Sector** A farm can produce a variety of agricultural good ( $v$ ) with the following technology

$$q_{ki}(v) = \begin{cases} z_{ki}(v)l_{ki}(v)^{\gamma_k}C_{ui}(v)^{\alpha_k} & \text{if } n_{ki}(v) \geq f_k \\ 0 & \text{if } n_{ki}(v) < f_k \end{cases}, \quad (4)$$

where  $q_{ki}(v)$  is the quantity of variety  $v$  in sector  $k$ ,  $l_{ki}(v)$  is the use of land,  $C_{ui}(v)$  is a composite good associated with the urban sector,  $n_{ki}(v)$  is the employment of labor,  $z_{ki}(v)$  is an efficiency term,  $f_k$  is the minimum number of permanent workers required to set up a farm,  $\gamma_k$  is the share of land,  $\alpha_k$  is the share of urban inputs and  $\gamma_k + \alpha_k < 1$ .<sup>10</sup>  $C_{ui}(v)$  captures the role intermediate inputs that are easier to be purchased in urbanized counties, for example, tractors, technical assistance and fertilizers. Because the technology is decreasing returns to scale, farm size is well defined. Assuming that farms make zero profits, the farm gate price of variety  $v$  in sector  $k$  ( $p_{ki}(v)$ ) is

$$p_{ki}(v) = \kappa_k \frac{P_{ui}^{\alpha_k} r_i^{\gamma_k} w_i^{1-\gamma_k-\alpha_k}}{z_{ki}(v)}, \quad (5)$$

where  $\kappa_k \equiv \frac{f_k^{1-\alpha_k-\gamma_k}}{(1-\alpha_k-\gamma_k)^{1-\alpha_k-\gamma_k} \gamma_k^{\gamma_k} \alpha_k^{\alpha_k}}$ . Note that the aggregate production function of each sector is still constant returns to scale with a labor share of  $1 - \alpha_k - \gamma_k$ .<sup>11</sup> This formulation reconciles micro data patterns about profits and farm size presented in [Foster and Rosenzweig \(2011\)](#) with the fact that estimates of the aggregate agricultural production function typically do not reject constant returns to scale ([Mundlak et al., 1999](#)).<sup>12</sup>

<sup>10</sup>With this formulation, I am assuming that land is homogeneous within each county. I use this simplification because land heterogeneity within counties is likely small compared to differences between regions and counties, which is the focus of this article. See [Sotelo \(2016\)](#) for an alternative approach that focuses on land heterogeneity within regions.

<sup>11</sup>In the appendix, I show that if we consider  $n_{ki}(v)$  as permanent workers and include temporary workers ( $n_{ki}^T(v)$ ) in the production function with an intensity parameter ( $\phi_k$ ), the total mass of workers in each county is divided into a mass of temporary workers and a mass of permanent ones, but the aggregate production function of the county still has a labor share of  $1 - \alpha_k - \gamma_k$ . As a result, conclusions from the simulation of the model related to changes in the total mass of workers in each county do not change if we consider the use of temporary workers. Given that 8% of workers in the dataset are temporary, I use the more parsimonious formulation presented here.

<sup>12</sup>This production function is consistent with the following facts presented in [Foster and Rosenzweig \(2011\)](#): total

**Trade** Following Eaton and Kortum (2002), I pursue a probabilistic approach for the realization of efficiencies ( $z_{ki}(v)$ ). I assume that the technology available for farms and firms in county  $i$  to produce variety  $v$  in sector  $k$  is drawn from a Fréchet distribution (i.e., a Type-II extreme value distribution)

$$F_{ki}(z) = \exp(-T_{ki}z^{-\theta_k}), \quad (6)$$

where  $T_{ki}$  is a scale parameter that increases the probability of receiving good technological draws for  $z$  ( $T_{ki} \geq 0$ ) and  $\theta_k$  defines the dispersion of efficiency draws.  $T_{ki}$  is specific to county  $i$  and relates to the absolute advantage in the production of varieties for sector  $k$  and  $\theta_k$  is common across counties and captures the intensity of comparative advantage forces *within* sector  $k$ . Here,  $T_{ki}$  represents the natural advantage of a county.

Counties must incur an iceberg trade cost to reach county  $n$  such that delivering 1 unit of good requires  $\tau_{kin} \geq 1$  units, where  $\tau_{kii} = 1$  for all  $i$ . Therefore, the price that county  $i$  can offer for variety  $k$  to county  $n$  is  $p_{kin}(v) = \tau_{kin}p_{ki}(v)$ .

As demonstrated in Eaton and Kortum (2002), under this formulation, total sales in sector  $k$  ( $X_{kin}$ ) from origin county  $i$  to destination county  $n$  in sector  $k$  are given by

$$X_{kin} \equiv \mu_k \pi_{kin} E_n = \mu_k \frac{T_{ki} (P_{ui}^{\alpha_k} r_i^{\gamma_k} w_i^{1-\alpha_k-\gamma_k} \tau_{kin})^{-\theta_k}}{\sum_{n'=1}^C T_{kn'} (P_{un'}^{\alpha_k} r_{n'}^{\gamma_k} w_{n'}^{1-\alpha_k-\gamma_k} \tau_{kn'n})^{-\theta_k}} E_n, \quad (7)$$

where  $E_n$  is the total expenditure from consumers in  $n$  and  $\pi_{kin}$  is the probability that county  $i$  is the lowest cost provider of a variety to destination  $n$  in sector  $k$ . In addition, the price index for sector  $k$  in county  $i$  becomes<sup>13</sup>

$$P_{ki} = \Gamma_k \left( \sum_{n=1}^C T_{kn} (P_{un}^{\alpha_k} r_n^{\gamma_k} w_n^{1-\alpha_k-\gamma_k} \tau_{kin})^{-\theta_k} \right)^{-\frac{1}{\theta_k}}, \quad (8)$$

where  $\Gamma_k \equiv \kappa_k \Gamma \left( \frac{\theta_k + 1 - \sigma_k}{\theta_k} \right)^{1/(1-\sigma_k)}$  and  $\Gamma$  is the gamma function.<sup>14</sup> This price index can be used to simplify total sales in (7)

$$X_{kin} = \mu_k \Gamma_k^{-\theta_k} T_{ki} (P_{ui}^{\alpha_k} r_i^{\gamma_k} w_i^{1-\alpha_k-\gamma_k} \tau_{kin})^{-\theta_k} P_{kn}^{\theta_k} E_n. \quad (9)$$

This equation defines the value exported from county  $i$  to  $n$  as a function of a set of origin county's characteristics ( $T_{ki} (P_{ui}^{\alpha_k} r_i^{\gamma_k} w_i^{1-\alpha_k-\gamma_k} \tau_{kin})^{-\theta_k}$ ), the total expenditure of the destination county on commodities from sector  $k$  ( $\mu_k E_n$ ), bilateral trade costs ( $\tau_{kin}^{-\theta_k}$ ), and the price index of the destination

---

labor costs per acre decreases with farm size, output per acre decreases with farm size and total labor per acre decreases with farm size.

<sup>13</sup>The price index in the consumer maximization problem is given by  $P_{ki} \equiv \left( \int p_{ki}(v)^{1-\sigma_k} dv \right)^{\frac{1}{1-\sigma_k}}$ .

<sup>14</sup>In order to have a well defined price index, I assume that  $\theta_k > \sigma_k - 1$ . In the estimation of the model,  $\Gamma_k$  is not separately identified from the level of  $T_{ki}$ .

$(P_{kn}^{\theta_k})$ , which is a term that captures the competition with other counties in the economy. Note that the dispersion parameter ( $\theta_k$ ) governs the elasticity of trade with respect to trade costs ( $\tau_{kin}$ ).<sup>15</sup>

## 4.2 Spatial Equilibrium

In spatial equilibrium, goods market clear (output is equal to the total value exported in any county  $i$  and sector  $k$ ), trade balances (total expenditure is equal to total sales in any county  $i$ ), and total payments for factors of production (workers, landowners, and urban inputs) equal total sales. The combination of these three conditions gives

$$E_i + E_{ui}^A = \sum_{n=1}^C \sum_{k=1}^K X_{kin} = \sum_{n=1}^C \sum_{k=1}^K \pi_{kin} \left( \mu_k E_n + \mathbf{1}(k=u) E_{un}^A \right), \quad (10)$$

where  $\mathbf{1}(k=u)$  is an indicator function for the additional expenditure for urban goods coming from agriculture ( $E_{un}^A$ ).<sup>16</sup> Because I assume that workers and farmers can move freely between counties, welfare must equalize

$$W \leq \frac{w_i}{P_i} a_i, \quad (11)$$

where the inequality holds as an equality if the population in county  $i$  ( $N_i$ ) is positive ( $N_i > 0$ ) and  $P_i \equiv \prod_k (P_{ki}/\mu_k)^{\mu_k}$  is the price index of county  $i$ . Local land markets must clear

$$L_i = \sum_{k=1}^K L_{ki}, \quad (12)$$

where  $L_i$  is a fixed supply of agricultural land in county  $i$  and  $L_{ki}$  is the demand for land from each sector  $k$  (note that  $L_{ui} = 0$ ). To close the equilibrium, population must be fully employed  $N = \sum_k \sum_i N_{ki}$ , where  $N_{ki}$  is the number of workers in sector  $k$  in county  $i$ . The following definition summarizes the spatial equilibrium, where I define a set of vectors to simplify notation.<sup>17</sup>

**Definition 1.** (*Spatial Equilibrium*) Given a set of counties ( $C$ ), a vector of parameters defining consumers' preferences  $\{\boldsymbol{\sigma}, \boldsymbol{\mu}\}$ , a vector of parameters defining the production technology  $\{\boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{f}\}$ , a vector of natural advantages and amenities  $\{\boldsymbol{T}, \boldsymbol{a}\}$ , a matrix of bilateral trade costs  $\{\boldsymbol{\tau}\}$ , a measure of population  $N$ , and a vector of land supply  $\{\boldsymbol{L}\}$ , the spatial equilibrium is defined by endogenous

<sup>15</sup>The gravity equation for the urban sector is given by setting  $\alpha_k = \gamma_k = 0$ .

<sup>16</sup>More specifically,  $E_i$  is equal to payments for rents and wages ( $E_i \equiv \sum_{k=1}^K \frac{1-\alpha_k}{1-\alpha_k-\gamma_k} w_i N_{ki}$ ) and  $E_{ui}^A$  is the sum of the demand from the agricultural sector for urban goods ( $E_{ui}^A \equiv \sum_{k=1}^K \alpha_k \frac{w_i N_{ki}}{(1-\gamma_k-\alpha_k)}$ ).

<sup>17</sup>Allen and Arkolakis (2014) and Allen et al. (2014) derive sufficient conditions for the existence and uniqueness of equilibrium in a large class of spatial models. However, their proofs do not cover the case of models with multiple sectors with free labor mobility. Therefore, I studied equilibrium properties simulating the model. The main intuitions apply here. The existence and uniqueness of equilibria depend on the magnitude of agglomeration gains and congestion costs in the economy. With the estimated model, I investigated the presence of multiple equilibria using a large set of different initial guesses. I found the same equilibrium independent of the initial guess.



vectors of workers allocation, wages, prices, and welfare  $\{\mathbf{N}, \mathbf{w}, \mathbf{p}, W\}$  such that: (1) land markets clear in each county; (2) farms and firms make zero profits; (3) labor is fully employed; (4) welfare equalizes; (5) goods market clear; (6) trade balances; (7) total payments to factors of production equal total sales.

### 4.3 Absolute and Comparative Advantage

To understand how comparative advantages work in the model, it is useful to define an overall absolute advantage term for sector  $k$  in county  $i$  as

$$\bar{T}_{ki} \equiv \underbrace{T_{ki}}_{\text{Natural Advantages}} \underbrace{\left( P_{ui}^{\alpha_k} r_i^{\gamma_k} w_i^{1-\alpha_k-\gamma_k} \right)^{-\theta_k}}_{\text{Input Costs}}. \quad (13)$$

This term shows that the absolute advantage of a county is given by natural advantages and endogenous input prices. Note that the absolute advantage ( $\bar{T}_{ki}$ ) is *unobservable* in the data, but it can be recovered using the structure of the model. This *unobservable* term determines *observable* patterns of sectoral specialization. To see this, take two counties  $i$  and  $i'$  and two sectors  $k$  and  $k'$ . Then we can combine (9) and (13) to obtain the following equation governing the patterns of specialization

$$\frac{X_{kin} / X_{k'in}}{X_{k'in} / X_{k'i'n}} = \frac{\bar{T}_{ki} \tau_{kin}^{-\theta_k}}{\bar{T}_{k'i} \tau_{k'in}^{-\theta_k}} \bigg/ \frac{\bar{T}_{k'ik'} \tau_{k'ik'}^{-\theta_{k'}}}{\bar{T}_{k'i'} \tau_{k'i'n}^{-\theta_{k'}}}. \quad (14)$$

This equation shows that, when the absolute advantage ( $\bar{T}_{ki}$ ) of a sector  $k$  increases relative to the absolute advantage in another sector  $k'$  ( $\bar{T}_{k'i}$ ), that is, when there is an increase in its comparative advantage in sector  $k$ , the ratio of exports from county  $i$  to county  $n$  in sector  $k$  relative to  $k'$  increases compared to the same ratio for county  $i'$ . This expression shows that transportation costs ( $\tau_{kin}$ ) and productivity dispersion ( $\theta_k$ ) can also induce specialization. While the dispersion of efficiencies ( $\theta_k$ ) governs the comparative advantage *within* sectors in a single sector version of the model, the comparative advantage *between* sectors in the multiple sector version is characterized by differences in  $\bar{T}_{ki}$  and transportation costs.<sup>18</sup>

Equations (13) and (14) reveal six sources of heterogeneity driving specialization.

First, heterogeneity in natural advantages ( $T_{ki}$ ) drives specialization due to classic forces of specialization related to technological comparative advantage.

Second, heterogeneity in land intensity induces specialization because of variation in land rents

<sup>18</sup>An analogous equation can be obtained for the number of workers in each activity:

$$\frac{N_{kin} / N_{k'in}}{N_{k'in} / N_{k'i'n}} = \frac{\bar{T}_{ki} \tau_{kin}^{-\theta_k}}{\bar{T}_{k'i} \tau_{k'in}^{-\theta_k}} \bigg/ \frac{\bar{T}_{k'ik'} \tau_{k'ik'}^{-\theta_{k'}}}{\bar{T}_{k'i'} \tau_{k'i'n}^{-\theta_{k'}}}.$$

across counties. Even though an increase in land rents ( $r_i$ ) in a county affects all crops, due to technological differences in land intensity ( $\gamma_k$ ), it has a disproportional effect on the absolute advantage ( $\bar{T}_{ki}$ ) of land-intensive activities, generating forces of comparative advantages.<sup>19</sup>

Third, access to intermediate inputs induces specialization in the production of crops that are intensive in intermediate inputs. This force is captured in the model by the price of intermediate inputs ( $P_{ui}$ ).

Fourth, counties with lower wages are more competitive in the production of activities with stronger decreasing returns (lower  $\alpha_k + \gamma_k$ ), which is associated with a higher labor share in the aggregate production function.<sup>20</sup>

Fifth, proximity to consumers is beneficial for perishable activities (larger  $\tau_{kin}$  for the same  $i$  and  $n$ ).

Sixth, counties with lower costs of a bundle of inputs have a competitive advantage in the production of crops with high trade elasticity. To understand this mechanism, note that trade elasticities ( $\theta_k$ ) govern the relative importance of natural advantages with respect to the costs of a bundle of inputs. When  $\theta_k$  is large, variation in the costs of a bundle of inputs (13) translate into larger variation in absolute advantage and reduces the relative importance of variation in natural advantages ( $T_{ki}$ ). Intuitively, the efficiency draws ( $z_{ki}(v)$ ) are more similar across counties for activities with high  $\theta_k$ , which means that counties have less incentives to trade based on differences in their efficiency draws.

## 4.4 International Trade

I introduce international trade as follows. First, I assume that the rest of the world has a fixed expenditure in each sector ( $E_{kF}$ ). Second, I assume that there is an international competitor with an exogenous absolute advantage ( $\bar{T}_{kF}$ ). Third, given that Brazil imports a negligible amount of agricultural commodities, I assume that the domestic economy does not import from the international competitor. Given these assumptions and denoting the foreign market by  $F$ , foreign exports from each county is given by

$$X_{kiF} \equiv \frac{\bar{T}_{ki} \tau_{kiF}^{-\theta_k}}{\bar{T}_{kF} + \sum_{i'=1}^C \bar{T}_{ki'} \tau_{ki'F}^{-\theta_k}} E_{kF}, \quad (15)$$

where  $E_{kF}$  is the total expenditure of the foreign market and  $\tau_{kiF}$  is the distance to the closest port. With this formulation, the domestic economy still affects international prices. An increase in the

---

<sup>19</sup>Solving the land market clearing condition, the land rent is given by  $r_i = w_i s_i (N_i/L_i)$ , where  $s_i = \sum_k \frac{\gamma_k}{1 - \gamma_k - \alpha_k} \frac{N_{ki}}{N_i}$  captures the average land intensity in the county. In this case, regions with higher density, that is, regions with higher  $N_i/L_i$  tend to have higher rental prices. This would generate the land rent gradient observed in Fact 2.

<sup>20</sup>We can substitute the welfare equalization condition ( $w_i = W(P_i/a_i)$  if  $N_i > 0$ ) to obtain an absolute advantage equation that depends on amenities ( $a_i$ ) and the price index of a county ( $P_i$ ). This shows that, because a lower price index leads to lower wages, the connection of a county with the rest of the economy has an indirect effect on the absolute advantage term due to the effect that the price index ( $P_i$ ) has on the price of labor ( $w_i$ ).

absolute advantage of the domestic economy increases total sales to the world ( $E_{kF}$ ) relative to the international competitor.<sup>21</sup> Finally, to close the equilibrium of the model with trade, I assume that the government takes all trade surplus with a lump-sum tax on income and uses it to accumulate foreign reserves.

## 4.5 Example: The Economy on a Line

Figure 7 shows the distribution of workers across activities and counties in a simple economy with the spatial structure of a line, two agricultural sectors and no heterogeneity in natural advantages, amenities or land endowment. In this economy, consumers spend half of their income on each sector.

Panel A shows the allocation of workers when both agricultural sectors have the same land intensity ( $\gamma_1 = \gamma_2 = 0$ ), intermediate input intensity ( $\alpha_1 = \alpha_2 = 0$ ), labor requirement ( $f_1 = f_2 = 1$ ), trade elasticity ( $\theta_1 = \theta_2 = 4$ ) and trade cost ( $\tau_{kin} = \tau_{kin}$  for any  $i, n$  and  $k$ ). In this case, there is no source of heterogeneity between the two sectors and, therefore, every county has the same proportion of workers employed in each sector. Note that workers tend to live in the center of the line, which is the location that minimizes trade costs with the rest of the economy. Interestingly, even though there is no congestion in land markets since I assume no use of land, workers do not live in a unique location. This occurs because counties obtain efficiency draws ( $z_{ki}(v)$ ) for an infinite number of varieties and the Fréchet distribution ensures that every location is sufficiently competitive in a positive measure of varieties to have a positive mass of workers in equilibrium.

Panel B shows the allocation of workers when I introduce different sources of comparative advantage revealed by equations (13) and (14). Either differences in land-intensity ( $\gamma_1 > \gamma_2$ ), intermediate input intensity ( $\alpha_1 < \alpha_2$ ), labor share ( $1 - \alpha_1 - \gamma_1 < 1 - \alpha_2 - \gamma_2$ ), transportation costs ( $\tau_{1in} > \tau_{2in}$  for any  $i$  and  $n$ ) or trade elasticities ( $\theta_1 < \theta_2$ ) can generate the allocation of workers in Figure 7, where counties closer to the center of the economy specialize in one of the sectors of the economy. This simplified version of the model shows that, even with no diversity in natural advantages, there are at least five additional mechanisms that can explain observed patterns of specialization.

Finally, since there are no differences in natural advantages ( $T_{ki}$ ) or amenities ( $a_i$ ) in this simplified economy, the population density is higher in the center of line due to transportation costs. In the data, heterogeneity in the productivity fundamentals of urban activities ( $T_{ui}$ ) is a major factor explaining

---

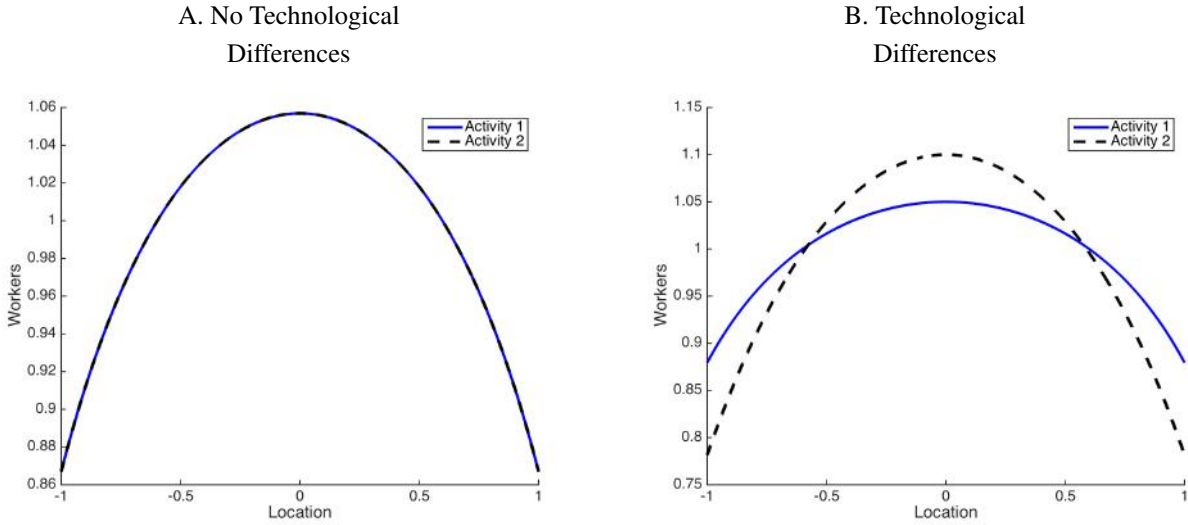
<sup>21</sup>This formulation is analogous to assuming that there is an upward sloping demand for domestic products with an elasticity of  $-\theta_k$ . To see this, first define  $P_{kiF} \equiv \Gamma_k(\bar{T}_{ki}\tau_{kiF}^{-\theta_k})^{-1/\theta_k}$ , which is a term associated with the contribution of the domestic production to the price index of products in the international market,  $P_{kF} \equiv \Gamma_k(\bar{T}_{kF} + \sum_{i=1}^C \bar{T}_{ki}\tau_{kiF}^{-\theta_k})^{-1/\theta_k}$ . Then, we can write total exports as

$$X_{kiF} = P_{kiF}^{-\theta_k} P_{kF}^{\theta_k} E_{kF}.$$

In the case where the contribution of the domestic production for the international price index is small ( $P_{kiF}$  is sufficiently large compared to  $P_{kF}$ ), then  $P_{kF}$  can be considered fixed for small changes in  $P_{kiF}$ . In that case, the elasticity of total exports in sector  $k$  relative to the price index of exported varieties is equal to  $-\theta_k$ .

differences in population density. Also, because some agricultural sectors demand more intermediate inputs from urban activities than others, the patterns of agricultural specialization in a region has also an influence on urban activities. Therefore, the model presents circular causations where agricultural activity fosters urban activity and vice versa.

Figure 7: Economic Activity on a Line



Notes: The figure shows simulations of the model on a set of counties distributed on a homogeneous line with two sectors. Panel A shows the number of workers for two sectors with the same technological properties. Panel B shows production when we introduce any source of technological heterogeneity in terms of land intensity, intermediate input intensity, trade costs and trade elasticities.

## 5 Estimation

I estimate the model in three stages. First, I estimate trade costs and trade elasticities ( $\tau_{kin}$  and  $\theta_k$ ). Second, I estimate the production function ( $\gamma_k$ ,  $\alpha_k$  and  $f_k$ ). Third, I recover expenditure shares ( $\mu_k$ ), natural advantages ( $T_{ki}$ ) and amenities ( $a_i$ ). I close this section with tests of the fit of the model.

### 5.1 Trade Costs ( $\tau_{kin}$ ) and Trade Elasticity ( $\theta_k$ )

**Method** As in Donaldson (2012), I first infer trade costs from price differential between varieties of goods across counties, I then use estimates of trade costs to identify trade elasticities in gravity equations. However, different from Donaldson (2012), I do not observe if a price is related to the origin or the destination of a variety. In this case, a potential concern discussed in Eaton and Kortum (2002) is that trade cost between two locations places only an upper bound on price differentials and, therefore, price differentials provide downward biased estimates of trade costs. To minimize this bias, Eaton and Kortum (2002) use the maximum price differential between countries as their estimate of trade cost. Based on this technique, I follow Allen (2014) and I use the maximum realization of

the monthly price differential within a triple of year, variety in sector  $v$ , and bilateral pair, which I define by  $p_{ymkin}^{Max}(v) \equiv \max\{p_{y1ki}(v)/p_{y1kn}(v), \dots, p_{y12ki}(v)/p_{y12kn}(v)\}$ , where  $m$  stands for month,  $y$  for year,  $k$  for sector,  $v$  for variety, and  $i$  and  $n$  for counties.<sup>22</sup> In what follows, I drop the index for variety  $v$  to simplify notation. I assume that trade costs have the following parametric form:  $\tau_{ymkin} = \exp(\beta_{ym} + \beta_{yk} + \beta_{mk} + \delta_k MTD_{in} + \varepsilon_{ymkin})$ , where  $\beta_{yk}$ ,  $\beta_{mk}$  and  $\beta_{ym}$  capture macroeconomic shocks and fluctuations on prices,  $\delta_k$  the geographic trade costs related to distance between counties and  $\varepsilon_{ymkin}$  the unobserved shocks affecting the price wedge.<sup>23</sup> With these assumptions,<sup>24</sup> I estimate the geographic trade cost ( $\delta_k$ ) with

$$\log p_{ymkin}^{Max} = \beta_{ym} + \beta_{yk} + \beta_{mk} + \delta_k MTD_{in} + \varepsilon_{ymkin}. \quad (16)$$

To estimate  $\theta_k$ , I assume the following parametric form for trade costs:  $\tau_{ykin} = \exp(\delta_k MTD_{in} - 1/\theta_k \varepsilon_{ykin})$ . With this assumption, log-linearize equation (9) and add the subscripts for year to obtain

$$\ln X_{ykin} = \beta_{yki} + \beta_{ykn} - \theta_k \hat{\delta}_k MTD_{in} + \varepsilon_{ykin}, \quad (17)$$

where  $\beta_{yki}$  captures the absolute advantage ( $\bar{T}_{ki}$ ) and  $\beta_{ykn}$  the terms related to the destination  $n$  ( $P_{kn}^{\theta_k} E_n$ ). Using estimates of geographic trade cost between any two pairs  $i$  and  $n$  ( $\hat{\delta}_k MTD_{in}$ ), the gravity equation identifies the dispersion parameter ( $\theta_k$ ). Statistical power allows me to estimate transportation costs and trade elasticities for three major groups: perishable agricultural sectors (vegetables and fruits), cereal (rice, soybeans, corn and wheat) and non-perishable sectors (remaining agricultural sectors). For the urban sector, I have information on trade but not on prices, therefore, I assume that trade costs for the urban sector are the same as trade costs for non-perishable goods.

**Results** Table 2 shows that an increase of 1000 *km* in bilateral distance is associated with 4.5% increase in the price gap for non-perishable goods, 17% for cereals, and 16% for perishable products. Therefore, transportation costs for perishable products is roughly 4 times larger than for non-perishable goods. As in Duranton et al. (2013), I find that these transportation costs are associated with the weight per value (tons per 1000 dollars) in the trade data: for perishable goods this ratio is about 10, for cereals 12, and for non-perishables 1.

---

<sup>22</sup>In the appendix, I present an alternative procedure for the estimation of trade costs that is based on distance from counties to port-counties instead of distance between counties. Port-counties generally have no production of agricultural products and, therefore, they can be assumed to be the destination of varieties. For non-perishable and cereals, I obtain trade costs that are similar to the ones obtained in my baseline estimates of trade costs. However, for perishables, I found much smaller trade costs. This can be attributed to the fact that perishable products are not internationally traded and, therefore, the assumption that port-counties is the destination if perishable goods is not valid.

<sup>23</sup>Figure A.1 in the appendix shows that a semi-log relationship between price wedges and minimum travel distance between counties fits the data well.

<sup>24</sup>Because I use data from different periods to estimate a static model, for the model to be internally consistent with the expressions that I use to estimate trade costs and trade elasticities, I have to assume that each period represents a different static equilibrium.

Table 3 reports results for trade elasticities.<sup>25</sup> For perishables (vegetables and fruits), the elasticity is much smaller than cereals and non-perishables. This is consistent with the hypothesis that variation in the types of fruits is larger than variation in the types of soybeans and wheat, which would translate into larger variation in the efficiencies available to produce varieties of fruits than in the efficiencies available to produce varieties of wheat.

Table 2: Estimates of Trade Costs

Estimates of Trade Costs ( $\delta_k$ )	$\log p_{ymkin}^{Max}$ (1)
Non-Perishables $\times$ MTD/1000	0.045 (0.001)
Cereals $\times$ MTD/1000	0.171 (0.022)
Perishables $\times$ MTD/1000	0.159 (0.013)
Obs	65755
$R^2$	0.68

Notes: Robust standard errors in parenthesis. Distance is the minimum travel distance between two counties. Cereals include varieties of rice, corn, soybeans and wheat. Perishables include varieties of fruits, vegetables and sugarcane. Non-perishables include all the remaining varieties. See appendix for a complete description of the varieties included in each group. Every regression controls for year-variety, month-variety, and month-year fixed effects.

Table 3: Estimates of Trade Elasticity

	Agriculture			
	Urban (1)	Non-Perishable (2)	Cereals (3)	Perishables (4)
Trade Elasticity ( $\theta_k$ )	3.63 (0.953)	7.29 (2.317)	5.71 (1.781)	3.92 (0.840)
Obs	5577	2757	2001	2495

Notes: Robust standard errors in parenthesis. Reported values are the negative of the estimated elasticity using the gravity equation. Every regression controls for origin-product-year and destination-product-year fixed effects.

<sup>25</sup>Results in this paper are comparable with previous estimates of trade elasticity in the literature. Several studies found that the trade elasticity for the manufacturing sector is around 4 (Caliendo and Parro, 2014; Eaton and Kortum, 2002; Simonovska and Waugh, 2014). Caliendo and Parro (2014) found trade elasticities to be larger in the agricultural sector using data on international trade (around 8.11). Similarly, I found the magnitude of the dispersion of efficiencies to be larger for non-perishable and cereals, which have a disproportional representation in international trade data.

## 5.2 Production Function ( $\gamma_k$ , $\alpha_k$ and $f_k$ )

**Method** I use the minimum distance method (Newey and McFadden, 1994) to back out the triple  $\varphi \equiv (\mathbf{f}, \boldsymbol{\gamma}, \boldsymbol{\alpha})$ . This method consists of two steps: first, calculate a set of sample moments  $m$  from the data; second, minimize the distance between these moments and their model counterparts  $\hat{m}(\varphi)$ . Formally, I seek  $\varphi$  that achieves

$$\hat{\varphi} = \arg \min_{\varphi \in \Phi} (m - \hat{m}(\varphi))' \hat{W} (m - \hat{m}(\varphi)), \quad (18)$$

where  $\hat{W}$  is an estimated matrix of weights and  $\Phi$  is the parameter space.<sup>26</sup> I assume that  $E(m - \hat{m}(\varphi_0)) = 0$ , where  $\varphi_0$  is the true value of  $\varphi$ .  $m$  contains direct moments from the data, as well as indirect moments such as regression coefficients.

Using the agricultural census of 2006, I construct four sets of moments. First, I use the coefficients from a county-level regression of total payments for workers on total revenues in each sector. Since the production function is Cobb-Douglas, these coefficients identify labor shares ( $1 - \alpha_k - \gamma_k$ ), because an increase in revenues leads to the same proportional increase in payments for workers in each sector.

Second, I include the coefficients of a county-level regression of the total number of agricultural workers on the total number of farms in each activity. These coefficients identify the labor requirement  $f_k$ , because in the model each sector employs the same number of workers per farm, and an increase in the number of farms leads to the same proportional increase in the number of workers.

Third, the model implies that crops with systematically higher revenues per land are less intensive in land ( $\gamma_k$ ). Therefore, I estimate county-level regressions of revenues per land on a sector on the revenues per land of a base activity. The coefficients from these regressions identify land intensities relative to the base activity.

Fourth, I use the average farm size per activity. Farm size in each county and activity is given by a non-linear function of the supply of land and the number of farms. These moments contribute to the identification of the relative magnitudes of  $f_k$ ,  $\gamma_k$  and  $\alpha_k$ .

Finally, these moments identify the level of  $\alpha_k + \gamma_k$ , but not the level of  $\gamma_k$  and  $\alpha_k$  separately. To pin down their level, I constrain all parameters to match the aggregate share of total payments for workers in each sector.<sup>27</sup>

---

<sup>26</sup>I use the inverse of the estimated covariance matrix of  $m$  for  $\hat{W}$ . To calculate the variance of the estimated parameters, I multiply the inverse of this covariance matrix by the Jacobian of  $m(\hat{\varphi})$ , which I construct numerically, and I invert the result from this product.

<sup>27</sup>Though I do not have a formal proof of the identification, in 1000 repeated estimations of the production function using random initial guesses, I found the same result.

Table 4: Parameters

	Labor	Factor Intensity		Expenditure
	Requirement $\hat{f}_k$	Land $\hat{\gamma}_k$	Urban $\hat{\alpha}_k$	Share $\hat{\mu}_k$
Cattle	2.872 (0.036)	0.422 (0.006)	0.408 (0.012)	0.131 (-)
Coffee	3.658 (0.083)	0.048 (0.132)	0.760 (0.135)	0.018 (-)
Corn	2.407 (0.062)	0.374 (0.006)	0.351 (0.042)	0.036 (-)
Cotton	3.470 (0.631)	0.066 (0.046)	0.921 (0.036)	0.005 (-)
Fruits	2.628 (0.055)	0.056 (0.074)	0.565 (0.075)	0.043 (-)
Rice	3.438 (0.061)	0.174 (0.007)	0.643 (0.035)	0.013 (-)
Soy	2.422 (0.200)	0.383 (0.002)	0.567 (0.017)	0.023 (-)
Sugarcane	3.500 (0.164)	0.119 (0.020)	0.826 (0.022)	0.042 (-)
Tobacco	2.372 (0.098)	0.100 (0.024)	0.832 (0.043)	0.001 (-)
Vegetables	2.799 (0.034)	0.093 (0.127)	0.330 (0.129)	0.047 (-)
Wheat	0.797 (0.910)	0.634 (0.002)	0.326 (0.032)	0.002 (-)
Urban	1 (-)	0 (-)	0 (-)	0.633 (-)

Notes: Standard errors in parenthesis. (-) denotes the calibrated parameters. Columns 1 to 3 show estimates of the production function by minimizing the distance between statistics in the data and statistics predicted by the model. Column 4 shows the expenditure share of the economy in each activity. I calibrate the expenditure share to match the total absorption of the domestic economy in each activity. I assume that the urban sector does not use any input besides labor.

**Results** Table 4 shows rich heterogeneity in agricultural technologies. While the share of land in soybeans is equal to 0.42, for fruits it is equal to 0.06. Furthermore, there are substantial differences in the use intermediate inputs. Sugarcane has a high intermediate input intensity, which is consistent with the fact that sugarcane in Brazil is frequently used as an input in the production of ethanol and farms are often integrated with ethanol plants.<sup>28</sup> Cattle production, on the other hand, has a small share of intermediate inputs and a large share of land. Therefore, better market access provides low

<sup>28</sup>In interviews with agricultural investors and farmers, it was mentioned that a rule of thumb for the sector is that a sugarcane farm must be within 100-150 km from an ethanol plant to be profitable.



benefits for cattle ranchers and high land rents have a large negative effect on their absolute advantage. The combination of these two technological properties explain why the production of cattle is often located in remote regions.<sup>29</sup> Finally, there are small differences in the labor requirement ( $f_k$ ) between activities: most of them employ between 2.5 to 3.5 workers per establishment.

### 5.3 Expenditure Shares ( $\mu_k$ ), Natural Advantages ( $T_{ki}$ ) and Amenities ( $a_i$ )

**Method** I calibrate expenditure shares ( $\mu_k$ ) to match the absorption of the Brazilian economy in each sector.<sup>30</sup> Table 4 reports results for this calibration.

With the preference and technological parameters, one can prove that there is a unique distribution of natural advantages ( $T_{ki}$ ), amenities ( $u_i$ ) and absolute advantages for the international competitor ( $\bar{T}_{kF}$ ) that rationalizes the observed equilibrium in the data in terms of the spatial distribution of wages ( $w_i$ ) and workers ( $N_{ki}$ ).<sup>31</sup> To adapt the strategy presented in Allen and Arkolakis (2014) to prove this result, I have to add two elements that are specific to my framework. First, there are multiple sectors of production. Second, there is a productivity fundamental for an international competitor that I identify using aggregate data on Brazilian exports for each sector.

**Proposition 1.** *(Recovering Natural Advantages and Amenities) Given an observed equilibrium distribution of economic activity given by a vector of wages and workers allocation  $\{\mathbf{N}, \mathbf{w}\}$ , a vector of parameters defining consumers' preferences  $\{\boldsymbol{\sigma}, \boldsymbol{\mu}\}$ , a vector of parameters defining the production technology  $\{\boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\alpha}, \mathbf{f}\}$ , a vector of parameters defining the total world consumption and domestic exports for each activity  $\{\mathbf{E}^F, \mathbf{X}^F\}$ , a matrix of bilateral trade costs  $\{\boldsymbol{\tau}\}$ , a vector of land supply  $\{\mathbf{A}\}$ , there is a unique (up-to-scale by sector) vector of natural advantages and amenities  $\{\mathbf{T}, \mathbf{a}\}$  as well as foreign markets competitiveness terms  $\{\mathbf{T}_F\}$  such that  $\{\mathbf{N}, \mathbf{w}\}$  characterizes a spatial equilibrium.*

*Proof.* See Appendix. □

**Results** Figure 8 presents an initial inspection of the spatial variation in terms of the distribution of natural advantages for cattle and vegetables, as well as the price of land and intermediate inputs. In the

---

<sup>29</sup>The aggregate production function implied by my estimates are consistent with previous literature estimating aggregate agricultural production functions. Mundlak et al. (1999) present a review of this literature. My estimates imply an aggregate land share of 0.25 and a intermediate input share of 0.50. Using country-level data, Mundlak et al. (1999) estimate 0.47 for land and 0.45 for capital and fertilizer. Previous estimates of land share are between 0.1 and 0.3 with a median of 0.24. For capital, which can be considered as intermediate inputs in my model, estimates are between 0.1 and 0.65 with a median of 0.38.

<sup>30</sup>I equalize revenues in each sector with total income adjusting for exports abroad, the additional expenditure of farmers on intermediate goods, and the lump-sum tax from the government. The model implies 37% of expenditure and 15% of GDP in agriculture.

<sup>31</sup>To recover natural advantages, I adopt two additional procedures. First, to simplify the estimation and the counterfactual analysis, I assume in the structural model the existence of only one destination for exports. Note, however, that for the estimation of trade elasticities  $\theta_k$ , I used information for different foreign destinations. Second, farms per sector are observed, but not workers per agricultural sector. Therefore, I predict the number of agricultural workers using the estimates of the production function.

south of Figure 8, in regions closer to Uruguay and Argentina, natural advantages for cattle ranching are better. This region, however, is not specialized in the production of cattle despite favorable natural conditions to do so (see Figure 2 on the location of cattle ranchers). Panel D suggests that this can be explained by market conditions: land rents in the southern part of this region are generally higher.

Panel B indicates large dispersion in the natural advantage for the production of vegetables. Different from the natural advantages for cattle, there is no clear visual pattern, which contrasts with the clear spatial pattern that we observe in Figure 2, where vegetables are concentrated closer to the Atlantic coast where market access is higher. This shows that, for the model to match the data, market conditions related to input prices must play a role.

In summary, Figure 8 suggests that natural advantages are unlikely to fully explain the patterns of specialization in the data. In the next section, I study the causes of specialization in a systematic way. But, before, I present several tests of the fit of the model.

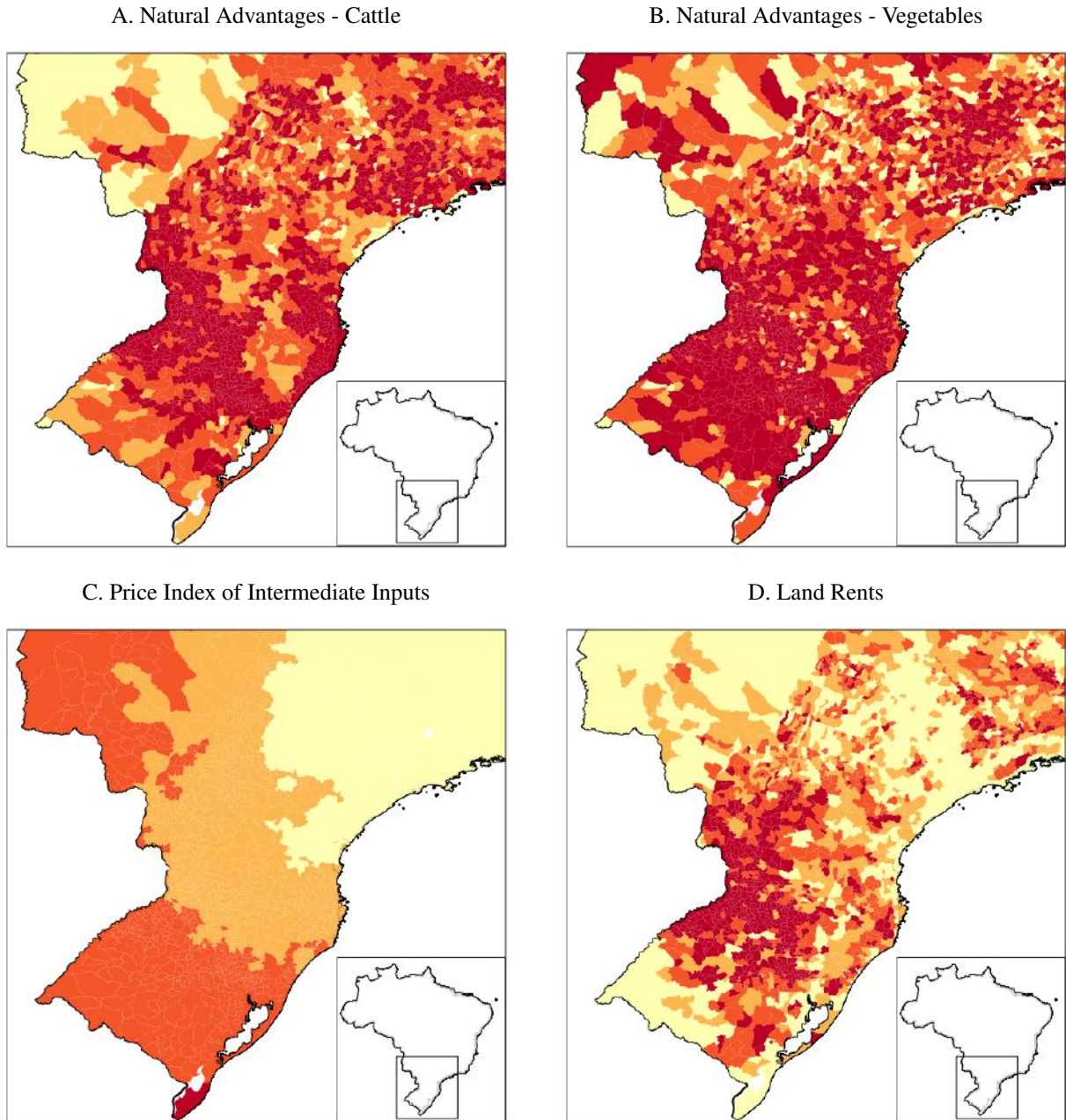
## 5.4 Fit of the Model

Given that I recover natural advantages as residuals from the model, I study whether these residuals are actually correlated with measures of agricultural suitability. I found that a regression with the full set of agricultural suitability variables from FAO interacted with dummies for each sector explains 74% of the variation in natural advantages implied by the model.

The estimation of the model is such that it perfectly explains the estimated patterns of agricultural specialization in terms of workers' allocation. Figure 9 reports the fit of the model with two alternative measures of specialization that were not targeted in the estimation: the share of the area in each activity and the share of revenues. Predictions from the model concentrate in the 45 degree line in both cases. A regression of the log of the share of area in the data on the share of area in the model has a slope of 1.21 with an  $R^2$  of 0.67. For the share of revenues, the elasticity is equal to 1.19 with an  $R^2$  of 0.50. In addition, Figure A.2 in the appendix shows that the model predicts well the farm size distribution: a regression of the model on the data has a slope of 0.76 with an  $R^2$  of 0.50. These results indicate that the assumptions about the production function driving the relationships between area, revenues and farm size provide a reasonable approximation to the data.

Since I do not use data on intermediate inputs in the estimation, I tested if the model can predict county-level information on intermediate input use. Despite the simple approach that I take to introduce intermediate inputs, the model still has a good fit. Panel A in Figure 10 shows that a regression of the expenditure on intermediate inputs per farm in the data on the expenditure per farm in the model has a slope of 0.67 with an  $R^2$  of 0.77. Panel B shows similar results for tractors per farm in the data on expenditure per farm in the model. A regression of the share of intermediate inputs costs to agricultural revenues in the data on its model counterpart has a slope of 0.34 with and a  $R^2$  of 0.22. The lower  $R^2$  here indicates that disaggregating the intermediate input into different types of inputs

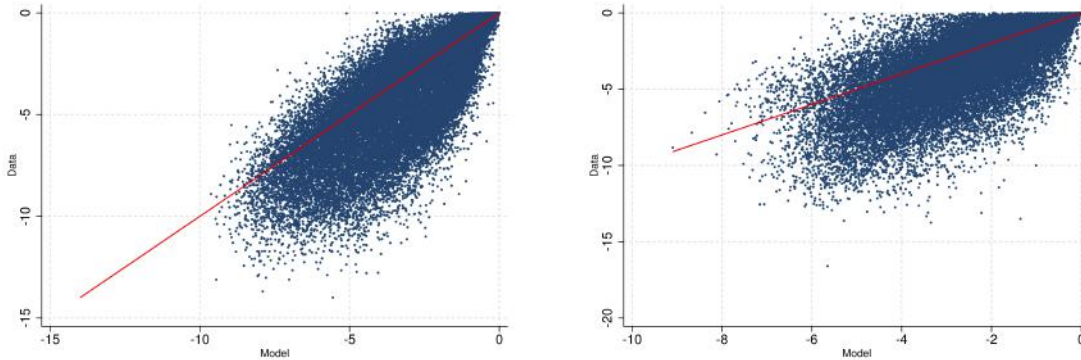
Figure 8: Natural Advantages and the Price of Inputs



Notes: The figure shows natural advantages in cattle and vegetables implied by the model, as well as the price of intermediate inputs and land rents. Colors are divided in quartiles. Darker colors show higher values.

can potentially improve the fit of the model. In addition, a regression of the intermediate input cost per land in the data on the predictions from the model has a slope of 0.70 with an  $R^2$  of 0.61.

Figure 9: Non-Targeted Measures of Specialization  
 A. Share of Area                      B. Share of Revenues



Notes: The figure provides two alternative measure of specialization that are not targeted in the estimation of the model. The red line shows the 45 degree line.

Finally, I test if the model can predict the average price of land in the data, which is also not used in the estimation. I find that a regression of log of land prices in the data on log of land rents in the model has a slope of 0.55 with a large  $R^2$  of 0.54. In part, the fit is not better because property prices is an imperfect *proxy* for land rents.

## 6 Causes of Agricultural Specialization

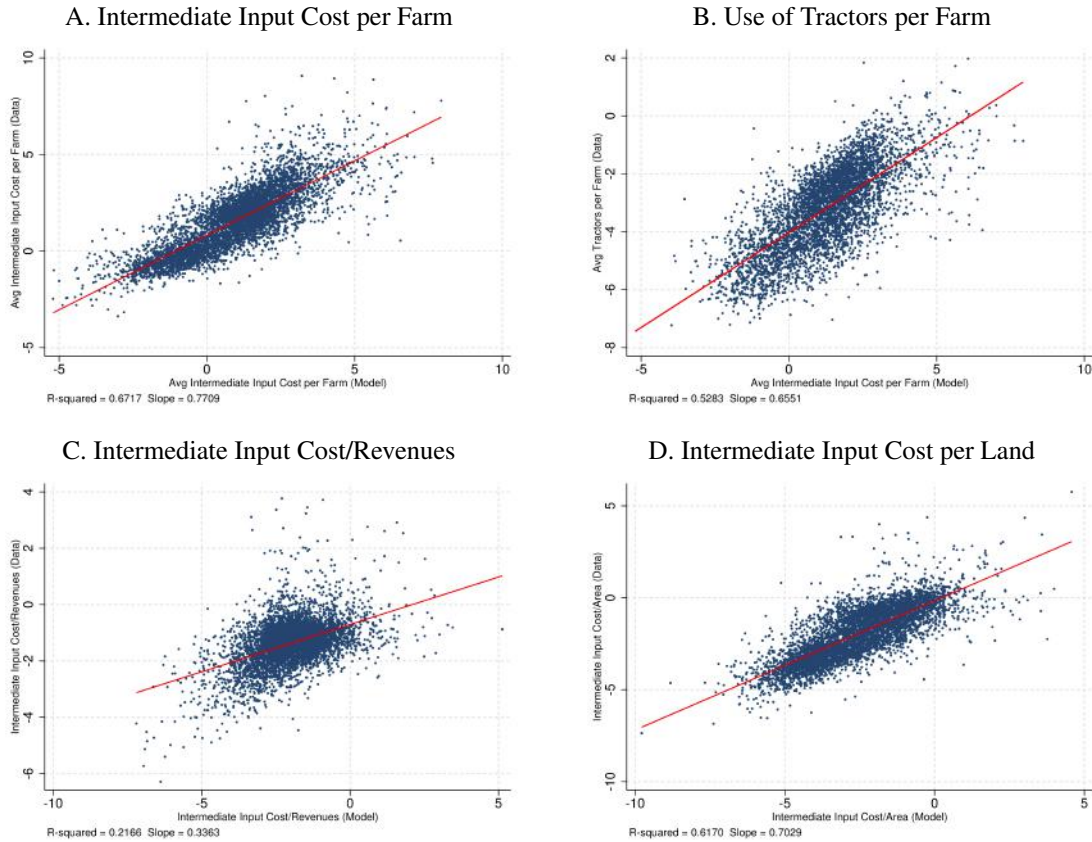
Having estimated the model, I investigate what causes the regional patterns of agricultural specialization. To do so, I shut down different sources of heterogeneity in the model, I simulate the economy without these sources of heterogeneity and I test the ability of the simulated economy to predict different features of the data. First, I study the link between market-access and agricultural specialization. I then analyze the overall patterns of agricultural specialization.

### 6.1 Causes of the link between Market-Access and Agricultural Specialization

Figure 9 presents non-parametric regressions of the share of farmers specialized in cattle and vegetables on market access when I shut down different sources of technological heterogeneity between specific activities. To estimate the non-parametric regressions, I follow the same procedure used to construct Figure 3.

Panel A shows that assuming that the production of cattle has the same trade costs or intermediate input intensity as vegetables have a small impact on the link between market access and the specialization in cattle. However, when I assume that the production of cattle uses land with the same

Figure 10: Correlation between Urban Expenditure in the Model and the Use of Intermediate Inputs in the Data (Non-Targeted)



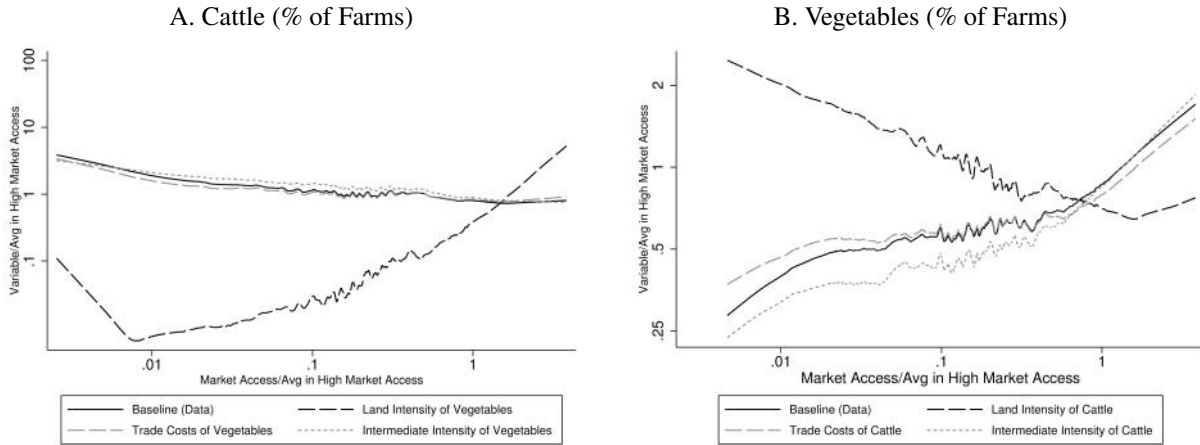
Notes: The figures investigate the correlation between the expenditure on inputs from urban activities postulated in the model and the actual use of intermediate inputs in the data. The red line in the figure shows the slope of the regressions.

intensity as vegetables, there is a large change in the location of production. Panel B shows analogous results for vegetables. Reducing trade costs of vegetables by roughly 75% to match trade costs for non-perishables, the model is still able to replicate the data well. However, when I assume that vegetables use land or intermediate inputs with the same intensity of cattle, there is a large change in the link between market access and the specialization in vegetables. In both cases, differences in transportation costs have a small influence on the location of production.

## 6.2 Causes of Agricultural Specialization

Table 5 reports correlations between the model and the data for the whole agricultural sector when I shut down different sources of technological heterogeneity. For natural advantages ( $T_{ki}$ ), I assume that all counties have the same average natural advantage. For technological parameters ( $\gamma_k$ ,  $\alpha_k$ ,  $\theta_k$  and  $\delta_k$ ), I assume that all activities have the same parameter. I use the average parameter value across

Figure 11: Revisiting the Relationship between Market Access and Agricultural Specialization



Notes: The figure shows the link between agricultural specialization and market access shutting down different sources of technological heterogeneity between crops in the model. To construct this figure I use the same procedure used to construct Figure 3.

farms for this analysis.<sup>32</sup>

Column 2 shows that when all counties have the same natural advantage, the correlation between the model and the data decreases to 0.05. Shutting down all the heterogeneity in input intensity reduces this correlation to 0.45. Columns 3 to 5 show that each source of input intensity is important to explain the patterns of specialization in the data. Shutting down differences in trade costs has a small effect on this correlation. Looking at farm size as a robustness, we get the same conclusions.

As an alternative exercise, I shut down different sources of technological heterogeneity and I re-estimate the residuals associated with natural advantages and amenities. I find that, without heterogeneity in input intensity, the variance of residuals on natural advantages is about 30% higher. This indicates that allowing for heterogeneity in input intensity in the model reduces substantially how much the model relies on residuals to explain the spatial allocation of workers. Shutting down differences in trade costs and elasticities, however, leads to a small change in the ratio of variances. For amenities, there are small changes overall, which indicates that technological heterogeneity in the agricultural sector do not explain variation in real wages.

In summary, results indicate that both differences in natural advantages across counties and heterogeneity in input intensities across crops are central causes of specialization. Next, I show that interactions between these two causes are important to understand the aggregate consequences of policies and economic shocks that change the regional patterns of agricultural specialization.

<sup>32</sup>Note that, for some cases, if I assume an average value for some parameters, the sum of intensities may exceed 1. Therefore, I change the remaining inputs to force the sum of intensities to be equal to 1 while keeping the relative importance of the remaining inputs the same. For example, if land intensity is equal to 0.4 and intermediate input intensity is 0.4, and I want  $1 - \gamma_k - \alpha_k$  to be equal to 0.4, I then assume that land intensity and intermediate input intensity are both equal to 0.3.

Table 5: Evaluating the Causes of Agricultural Specialization

	No Heterogeneity in						
	Nat.	Input	Specific Input Intensity			Trade	Trade
	Adv.	Intensity	Interm.	Land	Labor	Costs	Elasticities
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
First Exercise: Correlation between baseline and counterfactual economy in terms of							
% of farms in Activity	0.048	0.441	0.391	0.292	0.616	0.997	0.949
Farm Size per Activity	0.271	0.459	0.868	0.870	0.863	0.998	0.941
Second Exercise: Re-estimate residuals and measure ratio of variances							
Ratio for Natural Advantages	-	1.270	1.315	1.054	1.239	0.982	0.942
Ratio for Amenities	-	1.063	1.058	0.937	1.049	1.021	0.990

Notes: The table shows, in the first exercise, the correlation between counterfactuals with the baseline model when I shut down different sources of technological heterogeneity. Column 1 assumes that all counties have the same average natural advantage. Columns 2 to 7 show results when I assume that all technological parameters have the same average value where I average according to the number of farms in each activity. Share of farms in activity is the percentage of farms relative to the total number of farms in each county. Farm size per activity is the correlation between the counterfactual model and the baseline. In the second exercise, I re-estimate the residuals from the model shutting down different sources of heterogeneity. I present the ratio of the variance in the counterfactual relative to the baseline model.

## 7 Consequences of Shocks to Agricultural Specialization

In this section, I study the consequences of two shocks to agricultural specialization: the adaptation of soybeans to tropical regions and the rise of exports to China. In addition, I use these shocks to test the ability of the model to predict historical data on agricultural specialization.

### 7.1 Adaptation of Soybeans to Tropical Regions

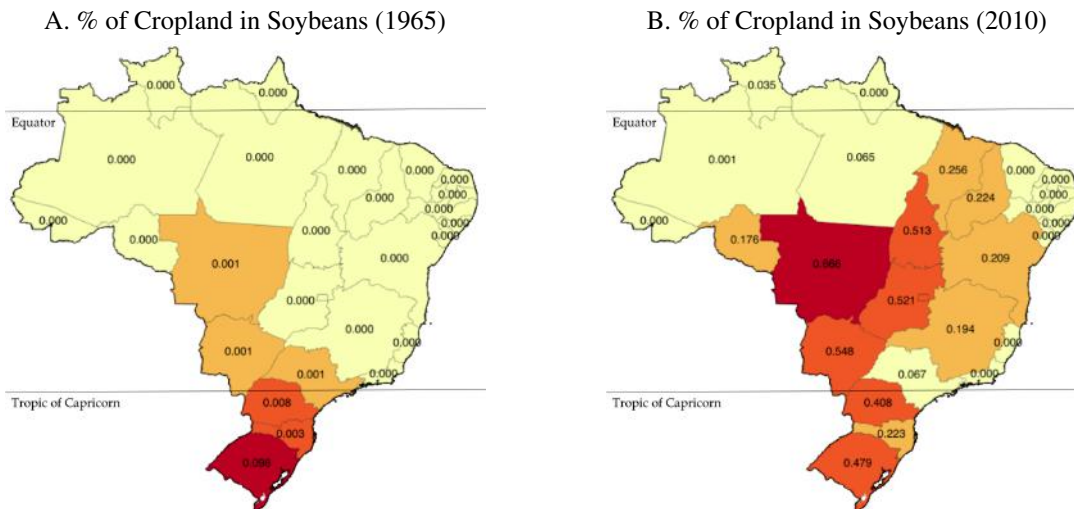
**Context** Soybeans originate from temperate areas. Nowadays, however, about 40% of the production of soybeans in Brazil comes from tropical regions in the Brazilian Savanna region (*Cerrado*) and this region accounts for 12% of the global production of soybeans. The development of the technology that allowed the production of soybeans in tropical regions is largely attributed to investments in agricultural research from the Brazilian government since the 1960's ([Assuncao and Braganca, 2015](#); [Spehar, 1995](#)). There were two technical requirements for the production of soybeans in tropical regions: first, the correction of the high soil acidity in the Brazilian Savanna; second, the development of varieties that were adapted to the fact that, in regions closer to the equator, there is smaller variation in daily sunlight exposure over the year.<sup>33</sup>

Figure 12 presents the expansion of soybeans from 1965 to 2010. While in the mid-1960's the

<sup>33</sup>See [Assuncao and Braganca \(2015\)](#) for a detailed description of the adaptation of soybeans to the tropics.

production of soybeans was still largely concentrated below the tropic of capricorn in the south of Brazil, by 2010, a large share of the cropland area in tropical regions produced soybeans. Next, I simulate the economy in the hypothetical scenario in which soybeans can only be produced in the three States below the tropic of capricorn. In other words, I investigate what would occur if tropical soybeans were to “disappear”.

Figure 12: The Adaptation of Soybeans to Tropical Regions



Notes: The figure shows the share cropland in each State producing soybeans. It indicates a large expansion of production to tropical regions from 1965 to 2010.

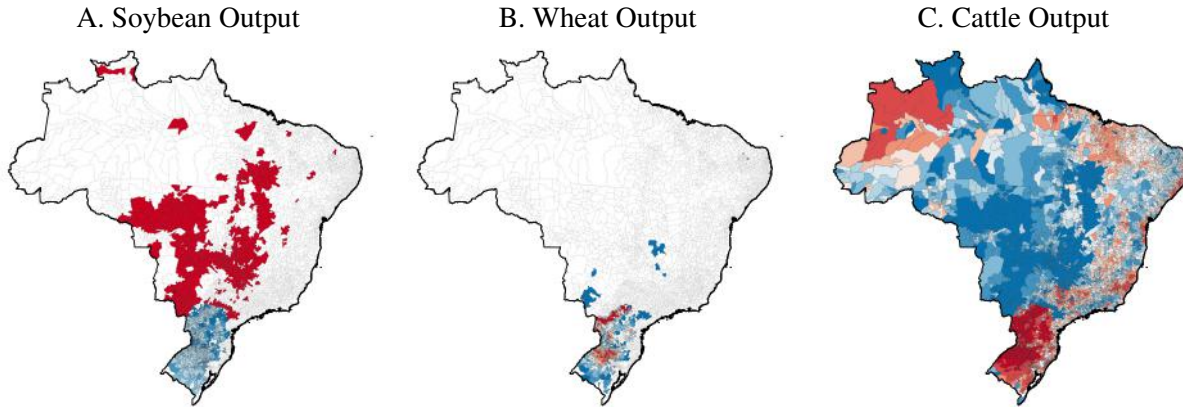
**Effects per Activity** Panel A in Figure 13 shows counties with a relative decrease in the production of soybeans in red and, in blue, those with a relative increase. The relative production of soybeans moves towards the south of Brazil. Table 6 reports details about the effects of these changes. Even though tropical regions account for roughly 33% of the total output of soybeans, Brazilian’s total output of soybeans falls by only 18%. This occurs because general equilibrium effects increase the price of soybeans by 15%, incentivizing an expansion in the production of soybeans in the south.<sup>34</sup> This change in the location of soybeans pressures land markets in the south where the Brazilian production of wheat is concentrated (see Panel B in Figure 13), reducing the production of wheat by 5%. In the tropics where land is released, there is an increase in the production of cattle. Table 6 indicates that, in magnitude, effects are larger for land-intensive activities that compete geographically with soybeans for land: cattle, wheat and cotton.

**Effects by Region** Column 1 in Table 7 presents effects on GDP according to three types of regions:

<sup>34</sup>I take the average price of soybeans across counties weighting the price according to share of population.



Figure 13: Effects of no Adaptation of Soybeans to Tropical Regions on Three Land-Intensive Activities



Notes: The figure shows in warmer colors (red) counties with a relative decrease in output and in cold colors (blue) counties with a relative increase in output when there is not production of soybeans in the tropics.

(1) the area in the tropics where soybeans “disappear”, (2) the temperate climate region below the tropic of capricorn that still produces soybeans in the counterfactual scenario, and (3) the rest of the country that does not produce soybeans. In tropical regions where soybeans “disappear”, total GDP decreases by 1.4%, even though soybeans represent 2.8% of the GDP in this region. In temperate regions in the south where the production of soybeans increases, total GDP is 0.04% larger. Therefore, in these two regions, economic adjustments minimize the aggregate impact of the local negative shock on the production of tropical soybeans. In the rest of the country, however, there is a GDP loss that magnifies the aggregate impact of this shock. When we aggregate the effects across these three regions, the aggregate GDP loss is larger than the local GDP loss associated with tropical soybeans: even though tropical soybeans represent 0.5% of Brazilian’s GDP, Brazilian’s GDP falls by 0.85%.<sup>35</sup>

Columns 2 to 5 in Table 7 and Figure 14 provide details to understand mechanisms driving GDP changes. Column 2 calculates changes in GDP keeping the price index of the baseline economy.<sup>36</sup> It shows that changes in the price index have a small effect on changes in total GDP for tropical regions, indicating that changes in output prices do not explain why GDP decreases in tropical regions without soybeans. Column 3 indicates that the number of agricultural workers increases in tropical regions where soybeans “disappear”, which decreases the average farm size in these region. This occurs because crops that occupy the land being used for soybeans are less land-intensive and absorb workers from the rest of the economy. In the south and in the tropical region where soybeans “disappear” there is an increase of 0.15% and 0.40% in their labor force, but the labor force in the rest of the country falls

<sup>35</sup>This result is consistent with reduced form evidence in Bustos et al. (2016), where the authors find that the expansion of soybeans increase out-migration.

<sup>36</sup>More specifically, I multiply the GDP in the counterfactual scenario in each county  $i$  and activity  $k$  by  $P_{ki}^{SOY} / P_{ki}^{NOSOY}$  where  $P_{ki}^{SOY}$  is the price index in the baseline scenario and  $P_{ki}^{NOSOY}$  is the price index in the counterfactual scenario.

Table 6: Effect of No Adaptation of Soybeans (in percentage change)

	GDP (1)	Export (2)	Price (3)
All Agriculture	4.99	-6.62	0.08
Urban	-0.71	0.01	-0.01
Cattle	1.31	7.29	-1.30
Coffee	-0.12	0.47	-0.08
Corn	0.16	2.36	-0.78
Cotton	0.85	2.88	-0.47
Fruits	-0.00	0.36	-0.14
Rice	0.07	1.49	-0.29
Soy	-18.1	-28.1	14.7
Sugarcane	0.00	0.82	-0.34
Tobacco	-1.36	-0.69	0.09
Vegetables	0.06	0.37	-0.14
Wheat	-5.14	-17.4	3.11

Notes: The table shows the effects of the absence of soybeans in the tropics. Effects are presented in percentage terms times 100. Price is the average price index weighted by the proportion of workers in each county.

by 0.05%. Therefore, the influx of labor to the tropical region where soybeans “disappear” minimizes local losses, but this comes at a cost for the tropical region not producing soybeans since there is a reduction in the labor force available for the production of its own products.

Note that the GDP loss in tropical regions where soybeans “disappear” depend on how much the increase in the value of alternative activities compensate losses associated with soybeans. The magnitude of this compensating mechanism depends on the quality of the land that is released for alternative activities. Non-reported results show that the quality of the land in tropical regions released by soybeans to other commodities is low (low  $T_{ki}$ ). This is consistent with the fact that an important part of the adaptation of soybeans to tropical regions was associated with the correction of the soil acidity.

**Welfare Effects** There is an aggregate welfare loss of 1.04%. I find substantial distributional impacts. The aggregate welfare of landowners decreases by 1.32%, whereas the welfare of workers decreases by 0.98%. This occurs because the reduction in the production of soybeans generates disproportional losses for the factor used intensively in production since it is a land-intensive sector. However, despite the aggregate decrease in the welfare of landowners, the welfare of landowners in the south increase by 2.7%.

Table 7: Effect of No Adaptation of Soybeans per Region (in percentage changes)

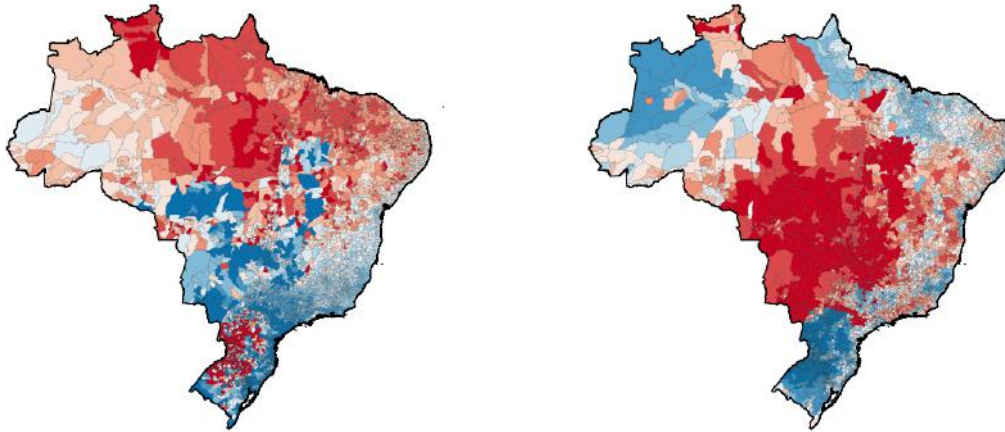
Types of Regions	Economic Adjustments				
	GDP (1)	No Price Adj (2)	Ag Workers (3)	All Workers (4)	Rents (5)
Tropical Regions no Soy	-0.87	-0.78	-0.45	-0.06	-2.41
Tropical Regions with Soy	-1.41	-1.25	3.32	0.37	-11.1
Temperate Regions	0.04	-0.54	-1.71	0.15	3.53
All Regions	-0.84	-0.84	-0.50	4.40	-2.78

Notes: The table shows results of no adaptation of soybeans on total output and different adjustment mechanisms in the economy. Results are disaggregated according to three different types of regions that were affected by general equilibrium effects in different ways. Column 2 shows changes in GDP holding the price fixed at the baseline simulation level. To do so, I divide GDP in the counterfactual in each county and sector by the price index in the counterfactual and I multiply it by the price index in the baseline.

Figure 14: Effects of No Soybeans on Adjustment Mechanisms

A. Workers

B. Rents



Notes: The figure shows in warmer colors (red) counties with a relative decrease in output and in cold colors (blue) counties with a relative increase in output when there is not production of soybeans in the tropics.

**Cost-Benefit Analysis** To evaluate the cost and benefit of public research on soybeans, I bring information on the budget of the Brazilian research department (EMBRAPA) since 1975 when it was created. I follow the procedures adopted in previous evaluations of the return of agricultural research to make my results comparable (Alston et al., 2000; Evenson, 2001; Griliches, 1958,6). I close the books in 2006 and I assume no improvement in productivity after 2006 besides the ones realized until then.

To calculate future costs, as in Griliches (1958), I assume an annual “maintenance” cost of research that keeps productivity equal to the one in 2006. From 1975 to 2006, about 5-10% of the total annual budget of EMBRAPA was allocated to research on soybeans. To account for research costs that are shared with research on other crops, I assume that 20% of its total annual budget went to research on soybeans.

To calculate future gains, I assume that soybeans in the tropics generate an annual gain equal to the

Table 8: Estimates of the Internal Rate of Return of Agricultural Research on the Adaptation of Soybeans to Tropical Regions

	Measures of Benefits			Alternative Scenarios	
	Soy Output (1)	Total Output (2)	Total Welfare (3)	No China (4)	Land Taxes (5)
IRR	29.7%	53.0%	58.5%	35.5%	46.2%

Notes: The table shows results of the annualized internal rate of return from agricultural research on the adaptation of soybeans to tropical regions. Data for costs of research comes from the Brazilian federal department of agricultural research, EMBRAPA. Columns 1 to 3 show results when we use different measures of benefits in the baseline case. Columns 4 to 5 calculate the returns to agricultural research when we assume different economic conditions for the economy. Columns 4 to 5 use changes in total output to measure the benefit cost ratio.

one in 2006. To account for the phase-in of benefits before 2006, I assume that the gains in previous years are given by the gain in 2006 times the ratio of soybeans exports relative to exports in 2006.<sup>37</sup>

Table 8 presents the annualized internal rate of return to research on soybeans in the year of 2006. This measure is the interest rate that would make the net present value of all the costs and benefits of soybeans in 2006 equal to zero. Column 1 shows returns if we only considered the effects on the production of soy, which is a common approach in the literature. In this case, the return would be equal to 35%. This measure, however, understates the aggregate consequences to the economy. Column 2 shows that the return is about 50% larger when we consider the total change in GDP. Column 3 shows that calculating the returns using the welfare gains, which accounts for the role of prices and amenities, provides similar results.<sup>38</sup>

Finally, I take advantage of the model to study the returns to research under different economic and institutional conditions. Table 8 shows that the return is 30% lower if Brazil does not export to China.<sup>39</sup> Also, adding an *ad valorem* tax of 25% on land, which disproportionately reduces the competitiveness of land-intensive activities, decreases the return by roughly 10%.

## 7.2 Trade with China

**Context** The economic growth of China in the past few decades had a dramatic effect on the international demand for commodities. In 2006, for example, China accounted for 40% of the global imports of cotton and soybeans. The global demand from China was crucial for the increase in agricultural exports of regions in Latin American and African Countries in recent decades (Gollin et al., 2016;

<sup>37</sup>A large part of production is exported and data for exports captures the influence of international prices on the benefits from soybeans for previous years. Another approach would be to use a linear phase-in of benefits since the 1980's when soybeans started to expand in the tropical areas. Using this alternative approach leads to larger gains. Therefore, I keep the more conservative value.

<sup>38</sup>Given that the Brazilian GDP in 2006 was about R\$2.4 trillion and preferences are homothetic, a welfare loss of 1.04% is equivalent to the same monetary loss in total output.

<sup>39</sup>To do so, I subtract from the total world expenditure the consumption from China ( $E_{kF} - E_{kCHINA}$ ).

Jenkins et al., 2008). For example, Brazilian exports to China in 1990 was only 4% of exports in 2006. Next, I study the hypothetical scenario where Brazil is not able to export to China. To do so, I assume the total world expenditure does not include imports from China ( $E_{kF} - E_{kCHINA}$ ). For simplicity, I do not consider the full general equilibrium effects that the absence of China would cause to other countries in the world.

**Effects per Activity** Table 6 shows that agricultural exports falls by 15%. Three products lead this effect: soybeans, cotton and tobacco. In most cases, for each sector, exports decrease proportionately less than the share of exports to China. This occurs because general equilibrium effects reallocates part of goods that were flowing to China to other countries. The absence of exports to China reduces the price of factors in Brazil, which makes the domestic economy more competitive relative to the international competitor.

Table 9: Effect of No Exports to China (in percentage changes)

	Changes w.r.t. Baseline			% of Exports to China (5)
	Output (1)	Export (2)	Price (3)	
All Agriculture	-2.14	-14.6	-0.10	16.9
Urban	-5.01	-5.45	-0.00	5.44
Cattle	-1.33	-1.16	-0.16	2.17
Coffee	-2.38	-1.30	0.02	1.10
Corn	-0.66	-1.79	-0.55	4.80
Cotton	-19.0	-42.7	-0.23	43.6
Fruits	-0.96	-1.55	0.00	1.46
Rice	-0.34	-4.81	0.04	4.39
Soy	-27.2	-40.3	-2.25	45.6
Sugarcane	-2.83	-2.52	-0.03	2.57
Tobacco	-7.44	-4.12	-0.24	5.61
Vegetables	-0.24	-1.40	-0.03	1.46
Wheat	6.10	28.0	-4.52	1.60

Notes: The table shows the effects of no exports to China. Effects are presented in percentage terms times 100. Price is the average price index weighted by the proportion of workers in each county. Exports to China is the proportion of exported revenues that are imported by China in the baseline.

**Effects by Region** To investigate how trade with China affects different regions of Brazil, table 10 presents the effects according to each quartile in the distribution of changes in export per county. Interestingly, even though exports to China are concentrated in agricultural activities, the absence of these exports increases the number of agricultural workers in the economy. This occurs because the

production of commodities that are exported to China absorbs a large part of the land supply in the economy, but uses a small proportion of labor, increasing the labor force available for urban activities. Column 3 shows that there is an overall increase in the number of workers in agricultural sector when Brazil does not export to China, and that exporting regions are the ones absorbing labor. This indicates that the rise of exports to China leads to the formation of regions with fewer workers, larger farms and more urbanized areas in terms of the share of agricultural workers relative to urban ones.

Table 10: Effect of No Exports to China per Region (in percentage changes)

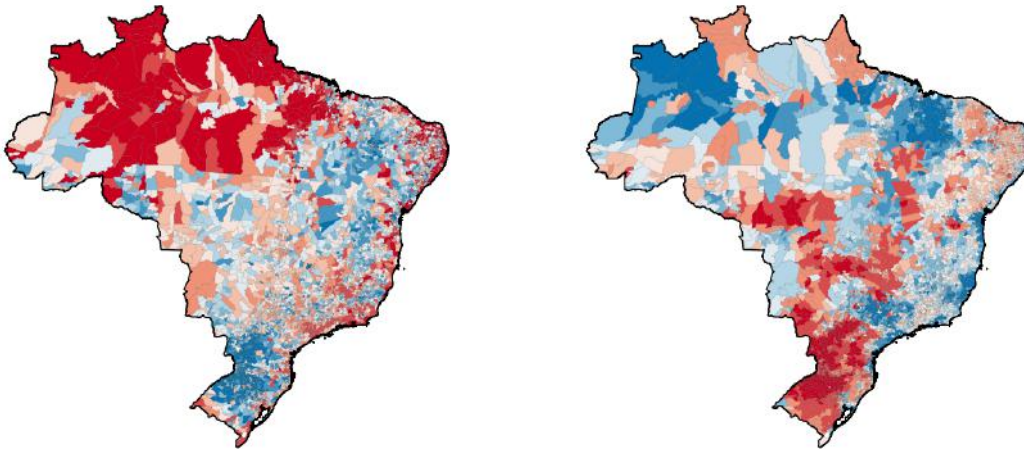
Types of Regions	Total	Total	Economic Adjustments			
	Exports (1)	GDP (2)	No Price Adj (3)	Ag Workers (4)	All Workers (5)	Rents (6)
<b>Changes in Exports</b>						
- 4th Quartile	-4.12	-4.15	-4.11	2.17	0.61	-4.02
- 3th Quartile	-5.02	-4.61	-4.58	2.04	0.17	-4.11
- 2th Quartile	-5.42	-4.94	-4.94	2.30	-0.32	-4.22
- 1th Quartile	-14.3	-5.22	-4.96	3.68	0.92	-8.62
All Regions	-7.17	-4.89	-4.84	2.61	-3.08	-6.20

Notes: The table shows results of no export to China on total output and different adjustment mechanisms in the economy. Results are disaggregated according to four quartiles relative to changes in exports to clarify the adjustment mechanisms in the economy.

Figure 15: Effects of No Export to China on Adjustment Mechanisms

A. Workers

B. Rents



Notes: The figure shows in warmer colors (red) counties with a relative decrease in output and in cold colors (blue) counties with a relative increase in output when there are not exports to China.

**Welfare Effects** The absence of exports to China leads to an aggregate welfare loss of 4.7%. Landowners have a welfare loss of 5.6%, whereas the welfare of workers decrease by 4.6%. Landowners in counties with low increase in exports, however, are better off when Brazil does not export to China.

### 7.3 Using the Shocks to Test the Model

Here, I use historical data not used in the estimation to test if the model can predict patterns of agricultural specialization in the economy before the two studied shocks hit the economy. Unfortunately, a rich county level data on agricultural production is not available for previous years to run this test.<sup>40</sup> Therefore, I aggregate county-level predictions from the model to obtain State-level aggregates. I bring information from PAM (*Produção Agrícola Municipal*) and from the Agricultural Census to construct a State level dataset with information on the share of land used in each activity for 1965.

Table 11 shows that the baseline economy has a correlation of 0.9 with the patterns of specialization in terms of the share of area in each activity for 2006. This can be considered a within sample fit since I estimated the parameters of the model with the agricultural census of 2006. Without the adaptation of soybeans and the exports to China, there is a large drop in the correlation of the model with the data for 2006 from 0.9 to 0.4, indicating that these shocks are important to explain agricultural specialization. For 1965, the model has a weak correlation with the data. However, the absence of the studied shocks increases the correlation between the model and the data from 0.6 to 0.8. Therefore, the absence of the studied shocks make the patterns of agricultural specialization in the simulated economy more similar to the patterns of specialization in the data before these shocks actually hit the economy.

Table 11: Correlations between the Share of the Area in each State and Activity in the Data and in Simulations of the Model

	Correlations with Share of Land	
	After Shocks	Before Shocks
	2006 (1)	1965 (4)
With shocks	0.916	0.597
No shocks	0.471	0.824
Obs	139	139
Number of States	27	27

Notes: The table shows the correlations between the log of the share of the area in each State used for each agricultural activity in the data and in the model. I only use observations for activity and State pairs that are available in every year.

<sup>40</sup>There are two major data limitations here. First, there was a large multiplication of counties over the years. While in the dataset for 2006 there are 5564 counties, in the 1960's there were less than 3000. Furthermore, I do not have sufficiently rich information to separate the production of vegetables from fruits. Therefore, I combine these two activities into one.

## 8 Conclusion

I formulated and estimated a spatial model to study the causes and consequences of the regional patterns of agricultural specialization in Brazil. I found that both heterogeneities in natural advantages across regions and in input intensity across crops are important causes of agricultural specialization. Using the terms coined by spatial economists, I found that both “First nature” and “Second nature” causes of specialization play an important role.

I find that the interaction between these causes is important to understand the aggregate consequences of economic shocks and policies. In particular, the interaction between worker mobility and heterogeneity in input intensity can create situations where negative productivity shocks in a region can lead to an influx of migrants. Looking at welfare, I find that the gains from technological innovations are unequally distributed across regions, workers and landowners, with some regions potentially being worse-off with productivity shocks in other parts of the economy. Further, I show that macroeconomic and institutional contexts of a country shape the returns to agricultural research. Importantly, the effects found in this paper are context specific, but the model is flexible and could be adapted to study agricultural specialization in different countries.

Finally, while several tests indicate that the model formulated in this paper provides a reasonable approximation to the economic forces driving agricultural specialization in the data, it abstracts from critical elements of policy debate on agricultural trade in developing countries. For example, the model does not account for the role of trading firms, the use of natural resources, the tradeoffs between commercial and subsistence agriculture, or the imperfections in insurance markets. Here, some of these elements are captured by the measure of natural advantages in the model, but, as we obtain more and better data on agriculture, they could be modeled to address different questions in the future.

## References

- Adamopoulos, T. and Restuccia, D. (2014). The size distribution of farms and international productivity differences. *The American Economic Review*, 104(6):pp. 1667–1697.
- Allen, T. (2014). Information frictions in trade. *Econometrica*, 82(6):2041–2083.
- Allen, T. and Arkolakis, C. (2014). Trade and the topography of the spatial economy. *The Quarterly Journal of Economics*, 129(3):1085–1140.
- Allen, T., Arkolakis, C., and Takahashi, Y. (2014). Universal gravity. *NBER Working Paper*, (w20787).
- Alston, J. M., Chan-Kang, C., Marra, M. C., Pardey, P. G., and Wyatt, T. J. (2000). *A meta-analysis of rates of return to agricultural R&D: Ex pede Herculem?*, volume 113. Intl Food Policy Res Inst.



- Assuncao, J. and Braganca, A. (2015). Technological change and deforestation: Evidence from the brazilian soybean revolution. *Working Paper*.
- Bank, W. (2009). Bridging the atlantic: Brazil and sub-sharan africa: South-south partnering for growth.
- Baum-Snow, N., Henderson, V., Turner, M. A., Zhang, Q., and Brandt, L. (2016). Highways, market access, and urban growth in china. *Working Paper*.
- Bird, J. and Straub, S. (2014). The brasilia experiment: road access and the spatial pattern of long-term local development in brazil. *World Bank Policy Research Working Paper*, (6964).
- Bustos, P., Caprettini, B., and Ponticelli, J. (2016). Agricultural productivity and structural transformation: Evidence from brazil. *American Economic Review*, 106(6):1320–65.
- Caliendo, L. and Parro, F. (2014). Estimates of the trade and welfare effects of nafta. *The Review of Economic Studies*, page rdu035.
- Costinot, A. and Donaldson, D. (2014). How large are the gains from economic integration: Theory and evidence from u.s. agriculture, 1880-1997. *Working paper*.
- Costinot, A., Donaldson, D., and Smith, C. (2016). Evolving comparative advantage and the impact of climate change in agricultural markets: Evidence from 1.7 million fields around the world. *Journal of Political Economy*, 124(1):205–248.
- Desmet, K., Nagy, K. D., and Rossi-Hansberg, E. (2015). The geography of development: Evaluating migration restrictions and coastal flooding. *Working Paper*.
- Donaldson, D. (2012). Railroads of the raj: Estimating the impact of transportation infrastructure. *Forthcoming The American Economic Review*.
- Donaldson, D. and Hornbeck, R. (2016). Railroads and american economic growth: A market access approach. *The Quarterly Journal of Economics*, 131(2):799–858.
- Duranton, G., Morrow, P., and Turner, M. A. (2013). Roads and trade: Evidence from the us. *The Review of Economics Studies*.
- Eastwood, R., Lipton, M., and Newell, A. (2010). Farm size. *Handbook of agricultural economics*, 4:3323–3397.
- Eaton, J. and Kortum, S. (2002). Technology, geography, and trade. *Econometrica*, 70(5):1741–1779.
- Evenson, R. E. (2001). Economic impacts of agricultural research and extension. *Handbook of agricultural economics*, 1:573–628.
- Fajgelbaum, P. and Redding, S. J. (2014). External integration, structural transformation and economic development: Evidence from argentina 1870-1914. Technical report, National Bureau of Economic Research.
- Foster, A. D. and Rosenzweig, M. R. (2004). Agricultural productivity growth, rural economic diversity, and economic reforms: India, 1970–2000. *Economic Development and Cultural Change*, 52(3):509–542.
- Foster, A. D. and Rosenzweig, M. R. (2011). Are indian farms too small? mechanization, agency

- costs, and farm efficiency. *Economic Growth Center, Yale University New Haven CT*.
- Fujita, M. (2012). Thünen and the new economic geography. *Regional Science and Urban Economics*, 42(6):907–912.
- Fujita, M., Krugman, P. R., and Venables, A. (2001). *The spatial economy: Cities, regions, and international trade*. MIT press.
- Gollin, D., Jedwab, R., and Vollrath, D. (2016). Urbanization with and without industrialization. *Journal of Economic Growth*, 21(1):35–70.
- Griliches, Z. (1957). Hybrid corn: An exploration in the economics of technological change. *Econometrica*, 25(4):501–522.
- Griliches, Z. (1958). Research costs and social returns: Hybrid corn and related innovations. *The Journal of Political Economy*, pages 419–431.
- Griliches, Z. (1964). Research expenditures, education, and the aggregate agricultural production function. *The American Economic Review*, 54(6):961–974.
- Jenkins, R., Peters, E. D., and Moreira, M. M. (2008). The impact of china on latin america and the caribbean. *World Development*, 36(2):235–253.
- Mas-Colell, A., Whinston, M. D., Green, J. R., et al. (1995). *Microeconomic theory*, volume 1. Oxford university press New York.
- Monte, F., Redding, S., and Rossi-Hansberg, E. (2015). Migration, commuting, and local employment elasticities.
- Morten, M. and Bryan, G. (2015). Economic development and the spatial allocation of labor: Evidence from indonesia. *Working Paper*.
- Mundlak, Y., Larson, D. F., and Butzer, R. (1999). Rethinking within and between regressions: The case of agricultural production functions. *Annales d'Economie et de Statistique*, (55-56):475–501.
- Nagy, D. K. (2015). City location and economic development. *Working Paper*.
- Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of econometrics*, 4:2111–2245.
- Oliveira, O. and Morten, M. (2014). Migration, roads and labor market integration: Evidence from a planned capital city. *Working Paper*.
- Samuelson, P. A. (1983). Thünen at two hundred. *Journal of Economic Literature*, pages 1468–1488.
- Scoones, I., Amanor, K., Favareto, A., and Qi, G. (2016). A new politics of development cooperation? chinese and brazilian engagements in african agriculture. *World Development*, 81:1–12.
- Simonovska, I. and Waugh, M. E. (2014). The elasticity of trade: Estimates and evidence. *Journal of international Economics*, 92(1):34–50.
- Sotelo, S. (2016). Domestic trade frictions and agriculture. *Working Paper*.
- Spehar, C. R. (1995). Impact of strategic genes in soybean on agricultural development in the brazilian tropical savannahs. *Field Crops Research*, 41(3):141–146.

## A Details of the Datasets

**Agricultural Production** The agricultural census of 2006 is organized by the Brazilian Statistical Bureau, *Instituto Brasileiro de Geografia e Estatística* (IBGE) and is publicly available online in their website: [www.sidra.org.br](http://www.sidra.org.br). Information for previous years is available under request.

**Trade** The Brazilian government updates on a monthly basis all the information on Brazilian exports and imports on the Aliceweb system ([www.aliceweb.mdic.gov.br](http://www.aliceweb.mdic.gov.br)), which is a webpage with information on the f.o.b. value, the quantity, and the weight of exports by county of origin and country of destination. I developed a webscraping code to download the data. I constructed the code using the Python package *Selenium* to access the website, and *BeautifulSoup* to parse the information. Information is disaggregated according to 4-digit NCM-level categories, where NCM stands for Nomenclature of Mercosur.

**Agricultural Prices** Several States in Brazil have a Secretary that provides services for the agricultural sector. One of the common services is the measurement of agricultural prices for farmers and consumers in different counties. Information comes from four State institutes: *Instituto Matogrossense de Economia Agrícola* (IMEA-MT), *Instituto de Economia Agrícola de São Paulo* (IEA-SP), *Empresa de Pesquisa Agropecuária e Extensão Rural de Santa Catarina* (EPAGRI-SC), *Secretaria de Desenvolvimento do Estado de Goiás* (SED-GO), and *Secretaria de Agricultura do Estado da Bahia* (SEAGRI-BA). Most of the information is available on their websites, but in some cases information was available under request. See table A.7 for details on prices.

**Transportation** For the transportation network, the central dataset comes from the *Departamento Nacional de Infraestrutura e Transporte* (DNIT). The department provides shapefiles with information on the road network that can be used in Geographic System Information (GIS). To use the transportation network, I follow Donaldson (2012) and Sotelo (2016), and I transformed the transportation network into a graph of nodes and edges. I constructed a square matrix where each row represents a node and the entry in the matrix represents the distance from each node to other nodes in the network. With this matrix, I use the Dijkstra algorithm to find the minimum travel distance between nodes in the graph.

**Agricultural Suitability** Data on agricultural suitability comes from the Global Agro-Ecological Zones database produced by the FAO and is publicly available on their website. The suitability is measured as the potential yields attainable for a crop in a certain geographical area. FAO uses 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 and low type of technology, which corresponds to the intensity of inputs such as mechanization and fertilizers that are used in production. Pixels in each map are classified with a value between 0 and 10000, with higher values denoting pixels with better suitability. With these maps, I created a set of variables that captures the average value of pixels that lie within the boundaries of each county. For a thorough description of this dataset, look at Costinot et al. (2016).

**Agricultural Production (Historical Data)** To construct the historical dataset on agricultural production I use information from PAM, *Produção Agrícola Municipal*, and I combine with information on pasture use from the Agricultural Census. PAM is an annual survey on agricultural production for a subset of agricultural products. With this dataset, I am able to construct 10 broad categories of agricultural products. I group fruits with vegetables in the same category. An important limitation of this dataset is that there was a large multiplication of counties over the years. Therefore, I aggregated the data to the State level, which is a geographic unit that can be compared over the years.

**Costs of Research on Soybeans** Information on the budget costs were obtained by contacting EM-BRAPA. It provides information on the total costs of the agency from 1975 to 2015 and the total costs on the Soybeans research department.

## **B Additional Facts**

This section provides additional details about facts on the Brazilian agriculture.

### **B.1 Market Access and Agricultural Specialization**

Table A.1 shows the correlation between market access and agricultural specialization. The second column of each variable shows the effect of including controls for agricultural suitability from FAO. In general, there is a moderate effect on the point estimate of the correlation and a substantial increase in the  $R^2$ . This indicates that agricultural suitability as measured by FAO is correlated with market access and it can explain a substantial part of the variation in the dependent variables. For land and credit market characteristics, the effects on point-estimates are much smaller.

Table A.2 shows the correlations between market access and agricultural activity using panel data. Given that I include year and county fixed effects, the elasticity is identified from changes over time, which controls for unobserved factors that are constant over time. I assume the same infrastructure of 2010 to construct the measure of market access so that its variation is uniquely driven by changes in population. I find links between changes in market access, the scale of production, and the specialization of farmers in the production of cattle that are consistent with previous findings. These links hold when I include variables for differential trends in agricultural suitability or distance to Brasília and State capitals to capture the Federal policies aimed at changing the location of population. Adding changes in the education of the population also does not change the results, which indicates that self-selection on skills does not affect the link between workers' location and the spatial distribution of agriculture. Furthermore, using distance to roads constructed to connect Brasília to State capitals as proposed in Oliveira and Morten (2014) and Bird and Straub (2014), I find similar results.

### **B.2 County FE and Market Access**

In the facts section of the paper, I present regressions of farm size on county and activity fixed effects. Here, I investigate factors driving the variation in county fixed effects. Table A.3 shows that an increase in 1% in the measure of market access decreases the county fixed effect by 0.13%, including agricultural suitability measures decreases. This indicates that an increase of 1% in market access decreases average farm size by 0.13%, controlling for crop fixed effects.

### **B.3 Use of Labor and Farm Size in Brazil and the United States**

Table A.4 shows additional evidence on the relationship between farm size, the number of farms

and the number of agricultural workers. Column 1 shows that the log of the number of farms in a county, by itself, explains 0.881 of the variation in the log of the total number of agricultural workers with a slope of 0.93. When we add interactions for farm size, which would capture the fact that counties with larger farms also have in average more workers, the explanatory power of the regressions increases by 0.005. This indicates that additional information on farm size contributes little to explain the variation in agricultural workers. Note that similar patterns can be observed in the United States. Increasing the number of farms by 1% increases the number of workers by 1.06%, and the regression has an  $R^2$  of 0.519. Using additional information on farm size increases the explanatory power of the regression by 0.07.

Table A.5 shows the distribution of workers per farm and farm size according to deciles of farm size. The ratio between the bottom and upper decile in Brazil for workers per farm is equal to 2 and for farm size 56. In the United States, the ratio for workers per farm is equal to 3 and for farm size 1221. The fact that these dramatic differences are present both in the United States and in Brazil indicate that the low variation in the number of workers per farm in Brazil is likely driven by agricultural technology rather than local labor market characteristics.

## C Details of the Model

### C.1 Deriving the farm-gate price

Here, I derive the price in equation 5 considering the use of temporary workers ( $n_{ki}^T(k)$ ). The production function for agriculture is

$$q_{ki}(v) = \begin{cases} z_{ki}(v)n_{ki}^T(v)^{\phi_k}l_{ki}(v)^{\gamma_k}C_{ui}(v)^{\alpha_k} & \text{if } n_{ki}(v) \geq f_k \\ 0 & \text{if } n_{ki}(v) < f_k \end{cases},$$

where  $q_{ki}(v)$  is the output, the inputs are  $l_{ki}(v)$  with cost  $r_i$ ,  $C_{ui}(v)$  with cost  $P_{ui}$ , temporary workers ( $n_{ki}^T(v)$ ) with cost  $w_i$ , permanent workers ( $n_{ki}(v)$ ) with cost  $w_i$ .  $z_{ki}(v)$  is an efficiency term and all parameters are positive  $\gamma_k, \alpha_k, f_k$ , where  $\phi_k + \alpha_k + \gamma_k \in (0, 1)$ .

To solve for the maximization problem of a farm. Assume that  $n_{ki}(v) = f_k$ . The cost minimization problem is given by

$$\min_{n_{ki}^T(v), l_{ki}(v), C_{ui}(v), \lambda} r_i l_{ki}(v) + P_{ui} C_{ui}(v) + w_i n_{ki}^T(v) + \lambda (q_{ki}(v) - z_{ki}(v)n_{ki}^T(v)^{\phi_k}l_{ki}(v)^{\gamma_k}C_{ui}(v)^{\alpha_k}).$$

The first order conditions are given by

$$\begin{aligned} r_i &= \lambda \gamma_k z_{ki}(v) n_{ki}^T(v)^{\phi_k} l_{ki}(v)^{\gamma_k-1} C_{ui}(v)^{\alpha_k}, \\ w_i &= \lambda \phi_k z_{ki}(v) n_{ki}^T(v)^{\phi_k-1} l_{ki}(v)^{\gamma_k} C_{ui}(v)^{\alpha_k}, \\ P_{ui} &= \lambda \alpha_k z_{ki}(v) n_{ki}^T(v)^{\phi_k} l_{ki}(v)^{\gamma_k} C_{ui}(v)^{\alpha_k-1}, \\ q_{ki}(v) &= z_{ki}(v) n_{ki}^T(v)^{\phi_k} l_{ki}(v)^{\gamma_k} C_{ui}(v)^{\alpha_k}, \end{aligned}$$

and using the first and second we get

$$\begin{aligned}
l_{ki}(v) &= \lambda \frac{\gamma_k q_{ki}(v)}{r_i}, \\
n_{ki}^T(v) &= \lambda \frac{\phi_k q_{ki}(v)}{w_i}, \\
C_{ui}(v) &= \lambda \frac{\alpha_k q_{ki}(v)}{P_{ui}},
\end{aligned}$$

Substituting back in the production function gives

$$q_{ki}(v) = z_{ki}(v) \left( \lambda \frac{\phi_k q_{ki}(v)}{w_i} \right)^{\phi_k} \left( \lambda \frac{\gamma_k q_{ki}(v)}{r_i} \right)^{\gamma_k} \left( \lambda \frac{\alpha_k q_{ki}(v)}{P_{ui}} \right)^{\alpha_k},$$

and the langrangian multiplier is given by

$$\lambda = \left[ \frac{q_{ki}(v)^{1-\gamma_k-\alpha_k-\phi_k}}{z_{ki}(v)} \left( \frac{w_i}{\phi_k} \right)^{\phi_k} \left( \frac{r_i}{\gamma_k} \right)^{\gamma_k} \left( \frac{P_{ui}}{\alpha_k} \right)^{\alpha_k} \right]^{\frac{1}{\phi_k+\alpha_k+\gamma_k}}.$$

Multiplying the first and second order conditions by  $l_{ki}(v)$ ,  $C_{ui}(v)$  and  $n_{ki}^T(v)$ , we get

$$r_i l_{ki}(v) + w_i n_{ki}^T(v) + P_{ui} C_{ui}(v) = \lambda (\phi_k + \alpha_k + \gamma_k) q_{ki}(v).$$

Inserting the langrangian multiplier gives

$$r_i l_{ki}(v) + w_i n_{ki}^T(v) + P_{ui} C_{ui}(v) = \left[ \frac{1}{z_{ki}(v)} \left( \frac{w_i}{\phi_k} \right)^{\phi_k} \left( \frac{r_i}{\gamma_k} \right)^{\gamma_k} \left( \frac{P_{ui}}{\alpha_k} \right)^{\alpha_k} \right]^{\frac{1}{\phi_k+\alpha_k+\gamma_k}} (\phi_k + \alpha_k + \gamma_k) q_{ki}(v)^{\frac{1}{\phi_k+\alpha_k+\gamma_k}}.$$

Now, if the output price is given by  $p_{ki}(v)$ , then variable profits are given by

$$\begin{aligned}
\Pi_{ki}^V &= p_{ki}(v) q_{ki}(v) - r_i l_{ki}(v) - w_i n_{ki}^T(v) - P_{ui} C_{ui}(v), \\
&= p_{ki}(v) q_{ki}(v) - \left[ \frac{1}{z_{ki}(v)} \left( \frac{w_i}{\phi_k} \right)^{\phi_k} \left( \frac{r_i}{\gamma_k} \right)^{\gamma_k} \left( \frac{P_{ui}}{\alpha_k} \right)^{\alpha_k} \right]^{\frac{1}{\phi_k+\alpha_k+\gamma_k}} (\phi_k + \alpha_k + \gamma_k) q_{ki}(v)^{\frac{1}{\phi_k+\alpha_k+\gamma_k}}.
\end{aligned}$$

The optimal  $q_{ki}(v)$  maximizes this expression and the first order condition is then given by

$$p_{ki}(v) = \left[ \frac{1}{z_{ki}(v)} \left( \frac{w_i}{\phi_k} \right)^{\phi_k} \left( \frac{r_i}{\gamma_k} \right)^{\gamma_k} \left( \frac{P_{ui}}{\alpha_k} \right)^{\alpha_k} \right]^{\frac{1}{\phi_k+\alpha_k+\gamma_k}} q_{ki}(v)^{\frac{1-\alpha_k-\gamma_k-\phi_k}{\phi_k+\alpha_k+\gamma_k}}.$$

The level of production is given by

$$q_{ki}(v) = p_{ki}(v)^{\frac{\phi_k+\alpha_k+\gamma_k}{1-\phi_k-\alpha_k-\gamma_k}} \left[ \frac{1}{z_{ki}(v)} \left( \frac{w_i}{\phi_k} \right)^{\phi_k} \left( \frac{r_i}{\gamma_k} \right)^{\gamma_k} \left( \frac{P_{ui}}{\alpha_k} \right)^{\alpha_k} \right]^{\frac{-1}{1-\phi_k-\alpha_k-\gamma_k}}.$$

Substitute this equation in the maximization problem to obtain

$$\Pi_{ki}^V = (1 - \phi_k - \alpha_k - \gamma_k) p_{ki}(v)^{\frac{1}{1-\phi_k-\alpha_k-\gamma_k}} \left[ \frac{1}{z_{ki}(v)} \left( \frac{w_i}{\phi_k} \right)^{\phi_k} \left( \frac{r_i}{\gamma_k} \right)^{\gamma_k} \left( \frac{P_{ui}}{\alpha_k} \right)^{\alpha_k} \right]^{\frac{-1}{1-\phi_k-\alpha_k-\gamma_k}}.$$

Free entry of farms drives profits to zero. Therefore, in equilibrium,  $p_{ki}(v)$  must be such that  $\Pi_{ki}^V = w_i f_k$ . This gives

$$p_{ki}(v) = \frac{f_k^{1-\phi_k-\alpha_k-\gamma_k} r_i^{\gamma_k} P_{ui}^{\alpha_k} w_i^{1-\alpha_k-\gamma_k}}{(1 - \phi_k - \alpha_k - \gamma_k)^{1-\phi_k-\alpha_k-\gamma_k} \phi_k^{\phi_k} \gamma_k^{\gamma_k} \alpha_k^{\alpha_k} z_{ki}(v)}.$$

To obtain the formula presented in the paper, define  $\kappa_k \equiv \frac{f_k^{1-\alpha_k-\gamma_k}}{(1-\alpha_k-\gamma_k)^{1-\alpha_k-\gamma_k} \gamma_k^{\gamma_k} \alpha_k^{\alpha_k}}$  and set  $\phi_k = 0$ .

Finally, note that, for the equilibrium price  $p_{ki}(v)$ ,  $n_{ki}(v) > f_k$  and  $n_{ki}(v) \in (0, f_k)$  leads to negative profits. Therefore,  $n_{ki}(v) = f_k$  in this equilibrium. Note that the price of variety  $v$  in county  $i$  is associated with a constant returns to scale production where the labor share is  $1 - \gamma_k - \alpha_k$  and does not depend on  $\phi_k$ .

## C.2 Equations to Calculate the Equilibrium

To calculate the equilibrium, I set the supply of workers in each county to be equal to the implied demand for workers. To obtain the equation I use to calculate the equilibrium, first solve the land market clearing condition, which would give

$$r_i = w_i s_i \frac{N_i}{L_i},$$

where  $s_i \equiv \sum_k \frac{\gamma_k}{1-\gamma_k-\alpha_k} \frac{N_{ki}}{N_i}$ . Now, we can write total income in each county compactly as  $I_i = w_i(1+s_i)N_i$ . Now, use the fact that total revenues in each sector must equal total sales

$$\begin{aligned} \frac{w_i N_{ki}}{(1-\gamma_k-\alpha_k)} &= \sum_{n=1}^C \pi_{kin} \left( \mu_k w_n (1+s_n) N_n + 1(k=u) \sum_{n=1}^C \pi_{kin} \sum_{k'=1}^K \alpha_{k'} \frac{w_n N_{k'n}}{(1-\gamma_{k'}-\alpha_{k'})} \right), \\ N_{ki} &= (1-\gamma_k-\alpha_k) \bar{T}_{ki} w_i^{-1} \tau_{kin}^{-\theta_k} \left[ \sum_n P_{kn}^{\theta_k} (\mu_k E_n + 1(k=u) E_{un}^A) \right], \end{aligned}$$

The equation above defines the implicit demand for workers in county  $i$  and sector  $k$ . With this equation, I can recover the distribution of workers up-to-scale. I use the final condition that the economy must satisfy the full employment condition to establish the level of  $N_{ki}$ . The following equations characterize the spatial equilibrium:

$$\begin{aligned} P_{ki} &\equiv \left( \sum_{n=1}^C \bar{T}_{kn} \tau_{kin}^{-\theta_k} \right)^{-\frac{1}{\theta_k}}, \quad P_{ui} \equiv \left( \sum_{n=1}^C T_{ui} (w_i \tau_{uin})^{-\theta_u} \right)^{-\frac{1}{\theta_u}}, \quad W = \frac{w_i}{\prod_k P_{kn}^{\mu_k}} a_i, \quad s_i = \sum_k \frac{\gamma_k}{1-\gamma_k-\alpha_k} \frac{N_{ki}}{N_i}, \\ N_{ki} &= (1-\gamma_k-\alpha_k) \bar{T}_{ki} w_i^{-1} \tau_{kin}^{-\theta_k} \left[ \sum_{n=1}^C \tau_{kin}^{-\theta_k} P_{kn}^{\theta_k} (\mu_k E_n + 1(k=u) E_{un}^A) \right], \\ E_i &= w_i(1+s_i)N_i, \quad E_{un}^A = \sum_{k=1}^K \alpha_k \frac{w_i N_{ki}}{(1-\gamma_k-\alpha_k)}, \quad \text{and} \quad \sum_{i=1}^C \sum_{k=1}^K N_{ki} = \bar{N}. \end{aligned}$$

The solution for this system of equations involves two fixed point problems: one for the allocation of workers into sectors and counties, and, for each different allocation of workers, a fixed point that defines the vector of wages. Therefore, the algorithm has an outer loop for the allocation of agricultural workers, and an inner loop for the vector of wages. I solve this system of equation using a Gauss-Seidel shooting algorithm, where I guess an initial vector for the allocation of workers, and I then solve the system of equations for the new vector of agricultural workers, and I update the guess using a weighted average of the initial and new values.

### C.3 Equations to Calculate the Equilibrium with no Mobility

With no mobility, welfare does not equalize across counties. In this case, I substitute the previous equations with the following conditions

$$W_i = \frac{w_i}{\prod_k P_{kn}^{\mu_k}} a_i, \text{ and } \sum_k N_{ki} = \bar{N}_i.$$

Solving this system of equations requires a fixed point on the distribution of wages that equalizes the implicit demand for workers in each county with the supply of workers in these counties. I start with a guess of wages. If the demand for workers in a county is larger than  $\bar{N}_i$ , then I increase wages in the county and update the guess. I keep with this algorithm until I find the set of wages that equalize the implicit demand for workers with the actual supply.

## D Details of the Structural Estimation

### D.1 Alternative Approach for Estimation of Trade Costs

An alternative approach for the estimation of trade cost is to, instead of looking at price differential between counties, to look at how much price decreases with distance to ports. Most ports are located in highly urbanized cities with little agricultural production. Therefore, the price at the port mainly reflects the price of transporting a product from an origin. Denoting  $P$  as the subscript for ports, we have

$$\log p_{ymki}(v) - \log p_{ymkP}(v) = \tau_{ymkiP}.$$

Assuming that  $\tau_{ymkiP} = \exp(\beta_{ym} + \beta_{yk} + \beta_{mk} + \delta_k MTD_{iP} + \varepsilon_{ymkiP})$ , then we have

$$\log p_{ymki}(v) = \beta_{ym} + \beta_{yk} + \beta_{mk} + \delta_k MTD_{iP} + \varepsilon_{ymkiP}.$$

Here, the fixed effect parameters absorb the price of variety  $v$  at the port ( $p_{ymkP}(v)$ ). Table A.6 shows results using this alternative approach. The costs of transporting non-perishables is slightly smaller than the baseline estimation and, for cereals, we get moderately larger trade costs. For perishables, however, trade costs are much smaller. This can be attributed to the fact that perishables are not internationally traded, therefore, distance to port has a smaller influence on the price.

### D.2 Statistics to Estimate the Production Function

To estimate the parameters from the production function, I explore four sets of statistics to apply the classical minimum distance estimation.

**Set of Statistics 1** I use the fact that there is an additive linear relationship between the total number of agricultural workers and the number of farmers in each county, and I assume that deviations from this relationship are given by a uncorrelated measurement error term

$$N_i^{ag} = \sum_k f_k N_{ki}^F,$$

where  $N_{ki}^F$  is the number of farms in each activity, which is observed in the data. This provides me 11



moments.

**Set of Statistics 2** The model predicts that total payments to wages increases linearly with revenues. In that case, I estimate

$$w_i N_i^{ag} = \sum_k (1 - \alpha_k - \gamma_k) R_{ki}.$$

This provides another 11 moments in the estimation.

**Set of Statistics 3** The model predicts a relationship between the ratio of revenues per land between activities given by

$$\frac{R_{ki}}{L_{ki}} = \frac{1/\gamma_k R_{k'i}}{1/\gamma_{k'} L_{k'i}}.$$

This expression can also be estimated by ordinary least squares. Given that cattle is present in nearly every county, I compare the revenues per land in each activity with the revenues per land in cattle production. This provides 10 moments.

**Set of Statistics 4** As additional moments, I target the average farm size. In this case, note that the farm size in each county is given by

$$\frac{L_{ki}}{N_{ki}^F} = \frac{(\gamma_k f_k / 1 - \gamma_k - \alpha_k)}{\sum_{k'} (\gamma_{k'} f_{k'} / 1 - \gamma_{k'} - \alpha_{k'}) N_{k'i}^F} L_i.$$

In this case, I take the average farm size in each county, I then estimate the average farm size predicted by the model using the formula above. Therefore, I have 43 moments and 11 constraints given by the fact that I force the estimation to perfectly match the aggregate share of wages relative to revenues in each sector.

### D.3 Expenditure shares ( $\mu_k$ )

Expenditure shares can be obtained directly from revenue information and the estimated technology parameters. To see this, use the condition that total sales for consumers must equal total expenditure

$$\begin{aligned} \sum_{i=1}^C \frac{w_i N_{ki}}{(1 - \gamma_k - \alpha_k)} &= \sum_{i=1}^C \sum_{n=1}^C \mu_k \pi_{kin} E_n, \\ R_k &= \mu_k I, \\ \mu_k &= \frac{R_k}{I}, \end{aligned}$$

where  $I$  defined the total income of agents in the economy ( $I$ ), the total revenues of the production sector ( $R$ ), and the total revenue of sector  $k$  ( $R_k$ ). Because expenditure shares must equal to 1, I estimate expenditure shares for agricultural commodities and I infer the demand for urban goods from  $1 - \sum_{k=1}^{K_{ag}} \mu_k$  where  $K_{ag}$  is the set of agricultural activities. To introduce the international sector, I add the total exports from Brazil to external markets in the formula ( $\bar{E}_k^F$ ). This gives

$$\mu_k = \frac{R_k - \bar{E}_k^F}{I - \bar{E}^F},$$

I subtract the exports from total income because this part of income is taxed.

#### D.4 Proof of Proposition 1

I use the proof strategy presented in [Allen and Arkolakis \(2014\)](#) to show that the natural advantages and amenities are identified given the spatial distribution of wages and agents in the economy.

**Proposition 2.** (*Recovering Natural Advantages and Amenities*) *Given an observed equilibrium distribution of economic activity given by a vector of wages and farmers and workers allocation  $\{\mathbf{N}, \mathbf{w}\}$ , a vector of parameters defining consumers' preferences  $\{\boldsymbol{\sigma}, \boldsymbol{\mu}\}$ , a vector of parameters defining the production technology  $\{\boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\alpha}, \mathbf{f}\}$ , a vector of parameters defining the total world consumption and domestic exports for each activity  $\{\mathbf{E}^F, \mathbf{X}^F\}$ , a matrix of bilateral trade costs  $\{\boldsymbol{\tau}\}$ , a vector of land supply  $\{\mathbf{A}\}$ , there is a unique (up-to-scale by activity) vector of natural advantages and amenities  $\{\mathbf{T}, \mathbf{a}\}$  as well as foreign markets competitiveness terms  $\{\mathbf{T}_F\}$  such that  $\{\mathbf{N}, \mathbf{w}\}$  defines the spatial equilibrium.*

*Proof.* The proof is based on the same strategy used to prove the uniqueness of a price vector defining a general equilibrium elaborated in [Mas-Colell et al. \(1995\)](#). This proof is based on the gross substitution properties of the excess demand functions. Here, I just show that the conditions for the uniqueness to hold are valid (see the full proof in [Mas-Colell et al. \(1995\)](#) or [Allen and Arkolakis \(2014\)](#)). Define the following type of excess demand function but for the productivities

$$f_{ki}(\mathbf{T}) = R_{ki} - \bar{T}_{ki} \left[ \sum_{n=1}^C \frac{\tau_{kin}^{-\theta_k}}{\sum_{n'=1}^C \tau_{kn'n}^{-\theta_k} \bar{T}_{kn'}} (\mu_k E_n + 1(k=u) E_{un}) + \frac{\tau_{kiF}^{-\theta_k}}{\sum_{n'=1}^C \tau_{kn'F}^{-\theta_k} \bar{T}_{kn'} + \bar{T}_{kF}} E_{kF} \right].$$

Where I use the previous formulas for  $E_n$  and  $E_{un}$ , and data on wages to obtain revenues  $R_{ki} = w_u f_k N_{ki} / (1 - \gamma_k - \alpha_k)$ . Here, we have the value of  $\bar{T}_{kF}$  that must be substituted with the following equation

$$\begin{aligned} X_{kF} = \sum_{i=1}^C X_{kiF} &= \sum_{i=1}^C \frac{\tau_{kiF}^{-\theta_k} \bar{T}_{ki}}{\sum_{n'=1}^C \tau_{kn'F}^{-\theta_k} \bar{T}_{kn'} + \bar{T}_{kF}} E_{kF}, \\ \bar{T}_{kF} &= \left( \frac{E_{kF}}{X_{kF}} - 1 \right) \sum_{i=1}^C \tau_{kiF}^{-\theta_k} \bar{T}_{ki}. \end{aligned} \quad (19)$$

Finally, we can work with the following excess demand function

$$f_{ki}(\mathbf{T}) = R_{ki} - \bar{T}_{ki} \left[ \sum_{n=1}^C \frac{\tau_{kin}^{-\theta_k}}{\sum_{n'=1}^C \tau_{kn'n}^{-\theta_k} \bar{T}_{kn'}} (\mu_k E_n + 1(k=u) E_{un}) + \frac{\tau_{kiF}^{-\theta_k}}{\sum_{n'=1}^C \tau_{kn'F}^{-\theta_k} \bar{T}_{kn'} + (s_{kF} - 1) \sum_{n'=1}^C \tau_{kn'F}^{-\theta_k} \bar{T}_{kn'}} E_{kF} \right],$$

where I defined  $s_{kF} \equiv E_{kF} / X_{kF}$ . Note that  $s_{kF}$  is observed. This excess demand function have the following properties: (1) it is continuous; (2) it is homogeneous of degree 0; (3)  $\sum_{i=1}^C f_{ki}(\mathbf{T}) = 0$ ; (4) it presents the gross substitution property, i.e.,  $\partial f_{ki}(\mathbf{T}) / \partial \bar{T}_{ki} < 0$  and  $\partial f_{ki}(\mathbf{T}) / \partial \bar{T}_{k'i} > 0$  for any  $i' \neq i$ . These four properties allow one to apply the proof strategy for the uniqueness of a vector of prices that defines a general equilibrium presented in [Mas-Colell et al. \(1995\)](#). As in the case of a vector of prices, the distribution of  $\bar{T}_{ki}$  is defined only up-to-scale, which can be seen from the fact that the demand excess function is homogeneous of degree 0. While in [Allen and Arkolakis \(2014\)](#) there is one sector in the economy, here, the distribution of the competitiveness parameters are identified

up-to-scale within each activity  $k$ . I still have to show that given a distribution of  $\bar{T}_{ki}$ , there is a unique distribution of  $T_{ki}$ . Note that the equation to recover  $T_{ki}$  is given by

$$T_{ki} = \left( \frac{N_i s_i}{L_i} \right)^{-\theta_k \gamma_k} \left( P_{ui}^{\alpha_k} w_i^{1-\alpha_k} \right)^{-\theta_k} / \bar{T}_{ki}.$$

Because  $P_{ui}$  is uniquely determined by the distribution of  $T_{ui}$  (and  $\alpha_k = 0$  for the urban sector). There is no need for a fixed point theorem to identify  $T_{ki}$  from  $\bar{T}_{ki}$ :  $T_{ki}$  is already uniquely determined from the equation above. Finally, to recover the distribution of amenities, once with  $T_{ki}$ , we can construct the price indexes and insert them in the following equation

$$a_i = W \frac{\prod_k P_{ki}^{\mu_k}}{w_i}.$$

Where I normalize welfare to 1. This final step completes the proof of the proposition.  $\square$

$\square$

## D.5 Estimation of Internal Rates of Return

The internal rate of return is the interest rate that makes the net present value of the investment equal to zero

$$NP(r) = \sum_{t=1975}^{2006} (B_t - C_t)(1+r)^{2006-t} + (B_{2006} - C_{2006})/r.$$

Here,  $B_t$  is the benefit in year  $t$ ,  $C_t$  is the cost of research in year  $t$ .

## E Additional Tables and Figures

Table A.1: Market Access and Agricultural Specialization with Additional Controls (2006)

	Log of Farm Size			Log of Land Price		
	(1)	(2)	(3)	(4)	(5)	(6)
Log of Market Access	-0.48 (0.11)	-0.43 (0.07)	-0.43 (0.05)	0.49 (0.03)	0.32 (0.04)	0.33 (0.03)
Obs	5523	5523	5523	5523	5523	5523
$R^2$	0.17	0.42	0.56	0.27	0.61	0.63
	Log of Vegetables Farms (%)			Log of Cattle Farms (%)		
	(1)	(2)	(3)	(4)	(5)	(6)
Log of Market Access	0.16 (0.05)	0.13 (0.03)	0.12 (0.02)	-0.14 (0.06)	-0.13 (0.04)	-0.12 (0.04)
Obs	5453	5453	5453	5507	5507	5507
$R^2$	0.06	0.34	0.40	0.03	0.22	0.41
<b>Controls</b>						
Ag Suitability	-	Yes	Yes	-	Yes	Yes
Land & Credit Markets	-	-	Yes	-	-	Yes
Farmers' Education	-	-	Yes	-	-	Yes

Notes: Standard errors clustered at the state level (27 units) in parenthesis. \* denotes significance at 1% level. Agricultural suitability includes 44 measures of suitability given by FAO. Land and credit markets includes variables related to the source of resources used to purchase a farm and the legal arrangement of the farm.

Table A.2: Market Access and Agricultural Activity (Variation Over Time)

	Log of Farm Size			Log of % of Pasture		
	OLS (1)	OLS (2)	IV (3)	OLS (4)	OLS (5)	IV (6)
Log of Market Access	-0.21 (0.05)	-0.11 (0.07)	-1.37 (0.19)	-0.08 (0.10)	-0.06 (0.12)	-0.38 (0.28)
$R^2$ or $F$	0.87	0.88	163	0.55	0.61	200
Obs	20995	20995	20457	16968	16968	16424
<b>Controls</b>						
Year and County FE	Yes	Yes	Yes	Yes	Yes	Yes
Dist. to State Capitals	-	Yes	Yes	-	Yes	Yes
>250 km from Federal Capital	-	Yes	Yes	-	Yes	Yes
Education	-	Yes	-	-	Yes	-
Agricultural Suitability	-	Yes	-	-	Yes	-

Notes: This table shows regressions of farm size on log of market access using previous data from the agricultural census. Share of pasture is the land used in the production of cattle relative to other activities in each county. All regressions control for year and county fixed effects, therefore, variation in market access comes uniquely from variation in population over time. I use as instrument the distance of a county to a set of roads constructed during the 1960's to connect the Federal capital to State capitals in different regions of Brazil.

Table A.3: Elasticity of County FE on Market-Access

	Log of County FE	
	(1)	(2)
Log of Land Prices	-0.254 (0.011)	-0.155 (0.016)
Log of Wages	0.373 (0.007)	0.280 (0.008)
Log of Market Access	-0.187 (0.010)	-0.102 (0.012)
Obs	5494	5494
$R^2$	0.33	0.45
<b>Controls</b>		
Agricultural Suitability (FAO)	-	Yes

Notes: Regressions for this table uses county fixed effects from regressions of farm size per activity on county and activity fixed effects. Agricultural suitability includes measures of suitability from FAO. Land prices is the price of land divided by total land use. See main text for measure of market access.

Table A.4: Elasticity of Workers w.r.t. Number of Farms and Farm Size in Brazil and the US

	Log of Ag Workers			
	BR (1)	BR (2)	US (3)	US (4)
Log of Number of Farms	0.936 (0.023)	1.167 (0.028)	1.064 (0.025)	1.094 (0.028)
Log of Avg Farm Size		0.419 (0.053)		-0.022 (0.068)
Log of Number of Farms $\times$ Log of Avg Farm Size		-0.055 (0.008)		0.028 (0.010)
Obs	5537	5537	2813	2813
$R^2$	0.881	0.886	0.519	0.593

Notes: The table shows the relationship between the number of agricultural workers, the number of farms and the size of farms. It shows that, once we have information on the number of farms in each county, additional information on the average farm size marginally increase our ability to predict the data.

Table A.5: Workers per Farm and Farm Size according to Farm Size Groups in Brazil and the US

Farm Size Decile (1)	Workers per Farm BR (2)	Farm Size (in <i>he</i> ) BR (3)	Workers per Farm US (4)	Farm Size (in <i>he</i> ) US (5)
1	2.9	9.1	0.5	0.4
2	2.9	15.9	0.5	1.1
3	3.1	21.5	0.7	2.0
4	3.1	28.1	0.8	3.4
5	3.2	37.7	0.8	5.7
6	3.4	50.1	1.1	10.4
7	3.6	67.6	1.3	22.7
8	3.6	97.9	1.4	52.1
9	4.3	163.2	1.8	139.8
10	5.9	508.6	1.5	488.4

Notes: The table shows the total number of workers per farm and the average farm size per decile of the distribution of farm size in Brazil and the United States. The unit of observation is a county in Brazil and in United States. Workers in Brazil capture all workers in a farm including the administrator. Workers in United States include hired labor.

Table A.6: Estimates of Trade Costs based on Distance to Port-Counties

Estimates of Trade Costs ( $\delta_k$ )	$\log p_{ymki}$ (1)
Non-Perishables $\times MTD_{ip}/1000$	-0.024 (0.002)
Cereals $\times MTD_{ip}/1000$	-0.217 (0.007)
Perishables $\times MTD_{ip}/1000$	-0.054 (0.018)
Obs	31150
$R^2$	0.98

Notes: Robust standard errors in parenthesis.  $MTD_{ip}$  is the minimum travel distance to the closest port. Cereals include varieties of rice, corn, soybeans and wheat. Perishables include varieties of fruits, vegetables and sugarcane. Non-perishables include all the remaining varieties. See appendix for a complete description of the varieties included in each group. Every regression controls for year-variety, month-variety, and month-year fixed effects.

Table A.7: Summary Statistics - Price Data

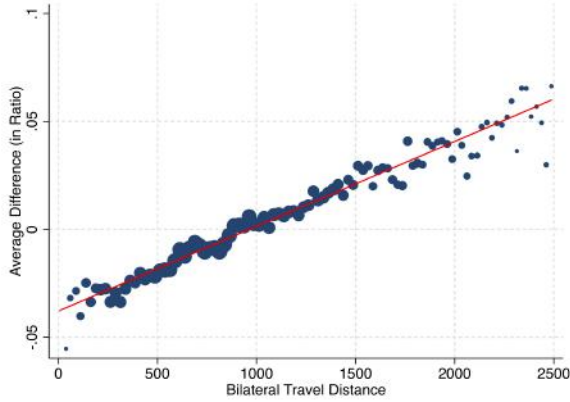
Commodity (1)	Name of Variety (2)	Category (3)	Avg. Price (4)	Std. (5)	Obs (6)
banana	nanica	perishables	0.7905	0.3725	266
banana	pacovan	perishables	0.8524	0.2470	152
banana	prata	perishables	140.47	80.357	232
beef	boi gordo	non perishables	850.85	180.73	11429
black beans	black	non perishables	1020.8	260.30	84
black beans	carioca	non perishables	1160.2	530.98	1025
black beans	generic	non perishables	700.11	410.65	578
black beans	mulato	non perishables	620.07	360.73	168
cocoa		perishables	500.85	270.20	204
coffee	cereja	non perishables	3830.1	1040.2	494
coffee	duro	non perishables	2660.5	1110.5	1224
coffee	pulpe	non perishables	2630.3	1050.8	226
coffee	rio	non perishables	1760.0	560.24	273
corn	seco	cereals	190.03	60.699	2371
cotton	caroco	non perishables	130.78	30.458	462
cotton	pluma	non perishables	500.33	170.65	2122
grapes	italy	perishables	140.17	80.208	408
guava		perishables	30.391	10.166	153
guava	paluma	perishables	120.87	60.121	153
mango		perishables	20.159	20.492	256
melon		perishables	0.6576	0.3806	401
onions		perishables	140.60	100.38	931
orange	for industry	perishables	1440.6	630.73	369
orange	pera	perishables	100.87	40.399	876
passionfruit		perishables	120.42	50.009	296
rice	casca	cereals	350.46	70.637	1280
soy		cereals	360.87	150.22	1764
sugarcane		perishables	270.84	20.456	256
tomatoes	for table	perishables	0.0613	0.0419	1631
watermelon		perishables	0.2825	0.1823	408
wheat	seeds	cereals	300.65	90.130	658
<b>TOTAL</b>					<b>31150</b>

Notes: This table shows summary statistics of the price data used to calculate transportation costs. Data comes from five different States of Brazil: Mato Grosso (MT), São Paulo (SP), Rio Grande do Sul (RS), Goiás (GO) and Bahia (BA). Information on the variety is provided whenever available. Information is available for 212 different counties.

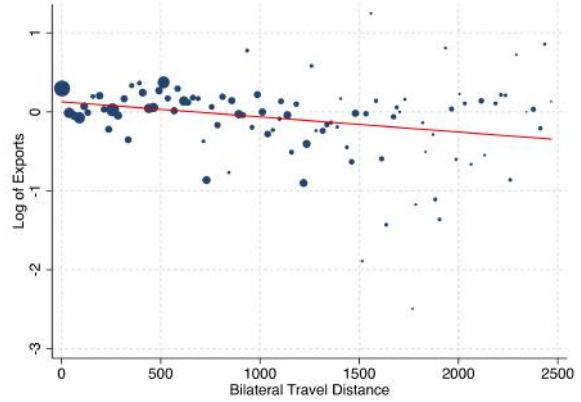


Figure A.1: Price Wedges and Bilateral Distance between Counties

A. Log of Price Wedge on Distance



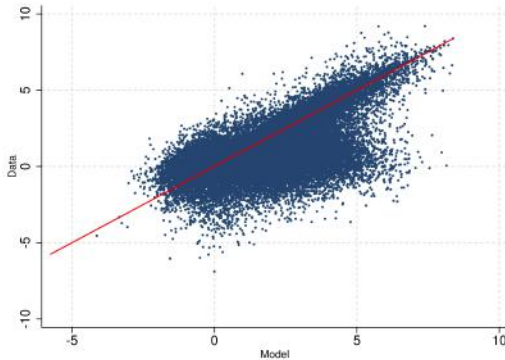
B. Log of Exports on Distance



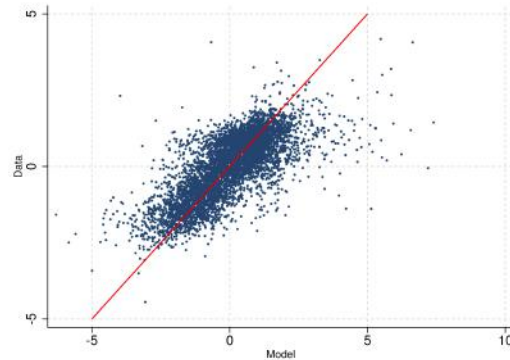
Notes: To construct this figure, I first absorbed a set of year-month and product fixed effects from the absolute wedge in the price dispersion  $|\log p_{ki} - \log p_{kn}|$  for the log of price wedge. For the log of exports I absorb a year-origin-product and a year-destination-product fixed effects. Then, I divided the bilateral travel distance in 100 equal bins of 25 km. Each dot shows the average of the absorbed price in each bin. The black line shows the best prediction of the dependent variable on bilateral trade after aggregating the data into bins and weighting according to the number of observations in each one of them. The size of the dots represent the number of observations.

Figure A.2: Fit of the Model (Non-Targeted)

A. Farm Size per Activity



B. Land Price



Notes: The figure investigates the correlation between two non-targeted statistics in the estimation. The red line shows the 45 degree line.