Wp

WORKING PAPERS

THE CAUSES AND CONSEQUENCES OF THE SPATIAL ORGANIZATION OF AGRICULTURE IN BRAZIL
10/2017
N° 2017/21

THE CAUSES AND CONSEQUENCES OF THE SPATIAL ORGANIZATION OF AGRICULTURE IN BRAZIL

Pellegrina, H.

CAF DEVELOPMENT BANK OF LATIN AMERICA

THE CAUSES AND CONSEQUENCES OF THE SPATIAL ORGANIZATION OF AGRICULTURE IN BRAZIL

Pellegrina, H.

CAF – Working paper N° 2017/21
10/2017

ABSTRACT

Why are there vast differences in agricultural activity across space? How do these differences shape the aggregate impact of agricultural shocks? To address these questions, I build a quantitative general equilibrium model that accounts for rich spatial differences in agriculture and use comprehensive county-level data from Brazil to estimate the model. I find that differences in natural advantages and factor intensities are key causes of the spatial patterns of agricultural specialization but that differences in trade costs across crops play a minor role. In addition, I study two major shocks in Brazilian agriculture: the adaptation of soybeans to tropical regions and the rise in the Chinese demand for commodities. The results show that general equilibrium effects substantially shaped the returns to agricultural research, the impact of tropical soybeans on urbanization, and the gains from trade with China.

CAUSAS Y CONSECUENCIAS DE LA ORGANIZACIÓN ESPACIAL DE LA AGRICULTURA EN BRASIL

Pellegrina, H.

CAF - Documento de trabajo N° 2017/21
10/2017

RESUMEN

¿Por qué hay grandes diferencias en la actividad agrícola a través del espacio? ¿Cómo influyen estas diferencias en el impacto agregado de los shocks agrícolas? Para abordar estas preguntas, construyo un modelo de equilibrio general cuantitativo que explica las diferencias espaciales en la actividad agrícola y estimo el modelo con datos de Brasil a nivel de municipio. Encuentro que las diferencias en las ventajas naturales y en la intensidad de los factores son causas clave de los patrones espaciales de la especialización agrícola, pero que las diferencias en los costos comerciales entre cultivos juegan un papel menor. Además, estudio dos grandes shocks en la agricultura brasileña: la adaptación de la soja a las regiones tropicales y el aumento en la demanda china de commodities. Los resultados muestran que los efectos de equilibrio general afectaron sustancialmente los retornos a la investigación agrícola, el impacto de la soja tropical sobre la urbanización y los beneficios del comercio con China.

The Causes and Consequences of the Spatial Organization of Agriculture in Brazil

Heitor S. Pellegrina*
New York University Abu Dhabi

October, 2017

Abstract

Why are there vast differences in agricultural activity across space? How do these differences shape the aggregate impact of agricultural shocks? To address these questions, I build a quantitative general equilibrium model that accounts for rich spatial differences in agriculture and use comprehensive county-level data from Brazil to estimate the model. I find that differences in natural advantages and factor intensities are key causes of the spatial patterns of agricultural specialization but that differences in trade costs across crops play a minor role. In addition, I study two major shocks in Brazilian agriculture: the adaptation of soybeans to tropical regions and the rise in the Chinese demand for commodities. The results show that general equilibrium effects substantially shaped the returns to agricultural research, the impact of tropical soybeans on urbanization, and the gains from trade with China.

Keywords: Agriculture, Trade, Spatial Economics, Nature, Agricultural Research
JEL classification: N5, N7, O1, Q1

---

*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 and to Sebastian Sotelo and Trevor Tombe for detailed discussion. I appreciate comments from Alexei Abrahams, George Allessandria, Guillermo Alves, Juliano Assunção, Joaquin Blaum, Arthur Bragança, Lorenzo Caliendo, Farid Farrokhi, David Kohn, Samuel Kortum, Kim Ruhl, Lucas Scottini, Dejanir Silva, Meredith Startz, Daniel Tanis, and participants in presentations held at Brown, PSU, Conoco-Phillip Seminar, NEUDC, FREIT-EIIT, RIDGE, CAF, UCSC, UCSD-GPS, NYUAD, FGV/EESP, INSPER, FGV/EPGE and New Faces of International Trade. Part of this paper was written while I was visiting the Economics department at PSU. 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. This paper circulated previously as "The Causes and Consequences of Agricultural Specialization in Brazil". All errors are my own. Email: heitor.pellegrina@nyu.edu.

1 Introduction

Even within small regions, there are enormous differences in agricultural activity across space. As one drives away from large urban centers, the agricultural landscape typically changes quickly from small farms producing goods for local markets to large and export-oriented establishments. While natural advantages may explain part of these patterns, the theoretical literature in spatial economics has long emphasized the role of proximity to urban centers because of the perishability of agricultural goods and the price of land (Fujita et al., 2001; Samuelson, 1983).[1] This paper empirically evaluates causes of the spatial differences in agricultural activity and how these differences shape the aggregate impact of agricultural shocks.

To illustrate how spatial differences in agriculture matter in the aggregate impact of agricultural shocks, consider the impact of a technology that expands the production a crop in a specific region of an economy on the reallocation of workers from rural to urban activities. Assume that this crop is export oriented and labor saving. The expansion of this crop may reallocate labor from agriculture to urban activities in this region but may also reduce the local production of domestic-oriented crops, which incentivizes a reallocation of workers from urban activities to the production of these crops in other parts of the economy. The size of this countervailing effect depends on how different regions produce and trade agricultural goods. As such, evaluating the aggregate impact of this technology on urbanization demands a framework that quantifies these differential effects across the economy.

In this paper, I develop a quantitative general equilibrium model that captures these economic forces the model is estimated with comprehensive county-level data for Brazil. I use the model to quantify causes of the spatial organization of agriculture and general equilibrium effects of two major shocks in Brazilian agriculture: the adaptation of soybeans to tropical regions and the rise in the Chinese demand for commodities. I show how general equilibrium effects shaped the returns to agricultural research, the impact of tropical soybeans on urbanization, and the gains from exports to China in these two episodes.

The basic structure of the model is as follows. The model includes multiple counties, multiple agricultural sectors and one urban sector. Counties trade in an Eaton and Kortum (2002) fashion and have different natural advantages for each sector. Sectors have different trade costs and dispersion of efficiencies that controls trade elasticities, as well as differences in factor intensities in terms of land, labor and intermediate inputs. Farms have decreasing returns to scale and require the employment of a manager so that the farm size is well defined.[2] Workers have heterogeneous

---

[1] This literature often refers to natural advantages as the "first nature" causes of specialization and to market conditions related to the location of the population as the "second nature" causes.

[2] The approach used in this paper is equivalent to assuming a span-of-control production function as in Lucas (1978) where managers are homogeneous in terms of their skills and have an opportunity cost equal to wages earned

preferences for counties, which in turn governs labor supply elasticities. There is a representative external market that has exogenous expenditures and absolute advantages for each sector, but external prices are still endogenous to shocks in the domestic economy. This structure leads to four main sources of comparative advantages driving the spatial organization of agriculture: (i) differences in natural advantages, (ii) differences in factor intensity, (iii) differences in trade costs across crops, and (iv) differences in trade elasticities.

To quantify the model, I use a combination of estimation and calibration. Similar to Donaldson (2012), I infer trade costs from the price differentials between counties and trade elasticities from gravity equations. I estimate the production function using agricultural census data on production. I calibrate the labor supply elasticity according to the literature. I set natural advantages, amenities and foreign absolute advantages so that the model perfectly matches the data in terms of workers, wages, aggregate exports and imports from Brazil for 2006.

Once with the model quantified, I shut down each source of heterogeneity driving comparative advantages and ask how much of the correlation between the model and the data remains. In doing so, I find that differences in natural advantages and factor intensities are key for the model to rationalize the data but that differences in transportation costs across crops play a minor role. An analysis of the costs of production shows that this result occurs because there is huge spatial variation in land rents and as a consequence even small differences in land-use intensity across crops generate large variations in the costs of production.[3] In line with this analysis, a decomposition exercise shows that 59% of the spatial variation in the costs of serving agricultural goods to state capitals comes from natural advantages, 30% from differences in factor prices, and only 1% from differences in trade costs across crops.

I then use the model to simulate the effects of the development of new seeds that allowed the production of soybeans in tropical regions in the Brazilian savanna, *Cerrado*.[4] To put this technology in perspective, before the 1970s when it was developed, there was virtually no production of soybeans in *Cerrado*, but nowadays, this region accounts for 15% of the global production of this crop. I find that the fact that soybeans are land-intensive and export-oriented is crucial for the general equilibrium effects. First, the international price of soybeans falls by 10% with its adaptation to the tropics, which generates economic losses in sub-tropical regions of Brazil that were already able to produce soybeans. Second, the production of soybeans requires large amounts of land in

---

by workers.

[3] For example, on average, counties that are 200 *km* from their state capital have land rents that are one-tenth of the land rent in counties that are 50 *km* away.

[4] Soybeans originate from temperate regions in China, and traditional soybean cultivars tend to flower prematurely in tropical regions because of the short length of days during the summer. The economic and biological aspects of the adaptation of soybeans to tropical regions have been largely studied in a series of articles (Assuncao and Braganca, 2015; Freire de Sousa and Busch, 1998; Gasparri et al., 2016; Pardey et al., 2003,0; Spehar, 1995). Some of these aspects are discussed in section 5.

Cerrado, generating substantial reductions in the local production of alternative crops.

These general equilibrium effects have key implications for how we measure the returns to the adaptation of soybeans to the tropics. Constructing a series of costs of the research on soybeans using data from the Brazilian agricultural research department, Embrapa, I find that the measured return to research using the gains generated within the soybean sector is 64%, whereas the aggregate return is 42%. Furthermore, these general equilibrium effects are key for the impact of tropical soybeans on urbanization: the share of workers employed in urban activities increases in *Cerrado*, but it falls in the aggregate. In particular, using agricultural data from the 1970s, I show that a difference in difference estimation of the local impact of soybeans using either actual or simulated data yields a similar positive effect on urbanization. In addition to serving as an out-of-sample test of the model, this exercise shows that the local impact on urbanization captured by a reduced form technique is quantitatively compatible with a negative effect in the aggregate.

I close the analysis by investigating the rise in the Chinese demand for agricultural commodities. Because China's agricultural imports are largely concentrated in soybeans and cotton, the bulk of the increase in domestic exports comes from regions where these crops can be grown. However, there are substantial indirect gains for regions of the economy that do not have natural advantages for the production of these crops. In particular, some regions gain by increasing their specialization in the production of domestically oriented goods that meet the new demand generated in the parts of the country where exports rise.

This paper contributes to recent research on agricultural trade in different ways (Allen, 2014; Costinot and Donaldson, 2014; Costinot et al., 2016; Donaldson, 2012; Donaldson and Hornbeck, 2016; Sotelo, 2016; Tombe, 2015). First, this work confronts the empirical relevance of different theoretical sources of comparative advantages in agriculture. Second, while previous models integrating internal geography with external markets assume exogenous international prices (Fajgelbaum and Redding, 2014; Sotelo, 2016), this paper provides an alternative approach in which the elasticity of international prices with respect to shocks to the domestic economy are controlled by trade elasticities. The model has the convenient property that the domestic economy can be large for some crops but not others, which is likely the case in many settings. Third, this paper is the first to introduce farm size within an agricultural trade model. In contrast with from previous papers that explain farm size distribution based on misallocation and natural advantages (Adamopoulos and Restuccia, 2014; Eastwood et al., 2010; Foster and Rosenzweig, 2011), the farm size distribution in this paper is driven by variations in factor prices across space.

By developing a general equilibrium framework to evaluate the effects of agricultural shocks, this paper provides a complementary approach to that found in a wide literature that evaluates the local impact of agricultural technologies with reduced form techniques (Bustos et al., 2016; Fiszbein, 2015; Foster and Rosenzweig, 2004; Hornbeck and Keskin, 2014; Hornbeck and Naidu, 2013). Closest to this paper, Bustos et al. (2016) investigate the introduction of Argentinian varieties of soybeans in the late 1990s in Brazil. The positive local impact of soybeans on urbanization found in their study are consistent with the effects that I estimate for the introduction of soybeans in *Cerrado* during the 1970s, but here, the general equilibrium framework indicates that the countervailling effects across the economy substantially minimize the local effect of this crop.

This paper also relates to an extensive literature on the returns to agricultural research (Alston et al., 2000; Evenson, 2001; Griliches, 1958,6; Pardey et al., 2016).[5] This is the first paper that uses a quantitative general equilibrium trade model to calculate the returns to agricultural research. The results suggest that accounting for general equilibrium effects can explain part of the seemingly implausible high returns to agricultural research typically found in the literature. For example, in the case of tropical soybeans, when I do not account for output reductions in other crops, the implied return to research is about 50% larger than the return implied by the aggregate gain.

Finally, I draw directly from growing research formulating quantitative general equilibrium models based on the work of Eaton and Kortum (2002) and their extensions to multiple sectors (Allen and Arkolakis, 2014; Baum-Snow et al., 2016; Caliendo et al., 2014; Desmet et al., 2015; Fajgelbaum and Redding, 2014; Monte et al., 2015; Morten and Bryan, 2015; Ramondo et al., 2016; Tombe and Zhu, 2016). Relative to this literature, this paper focuses on the interactions between agricultural sectors and introduces differences in land and labor intensity. I found this dimension of heterogeneity to be crucial in the agricultural context since there are huge differences in land and labor intensity across crops and because this heterogeneity can generate out-migration from regions and sectors receiving positive productivity shocks.[6]

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

2 Data

The data contain 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*). There are 5564 counties and the median county has an area of approximately 400 km^2. Panel (a) in Figure 1 shows the territorial division of counties.

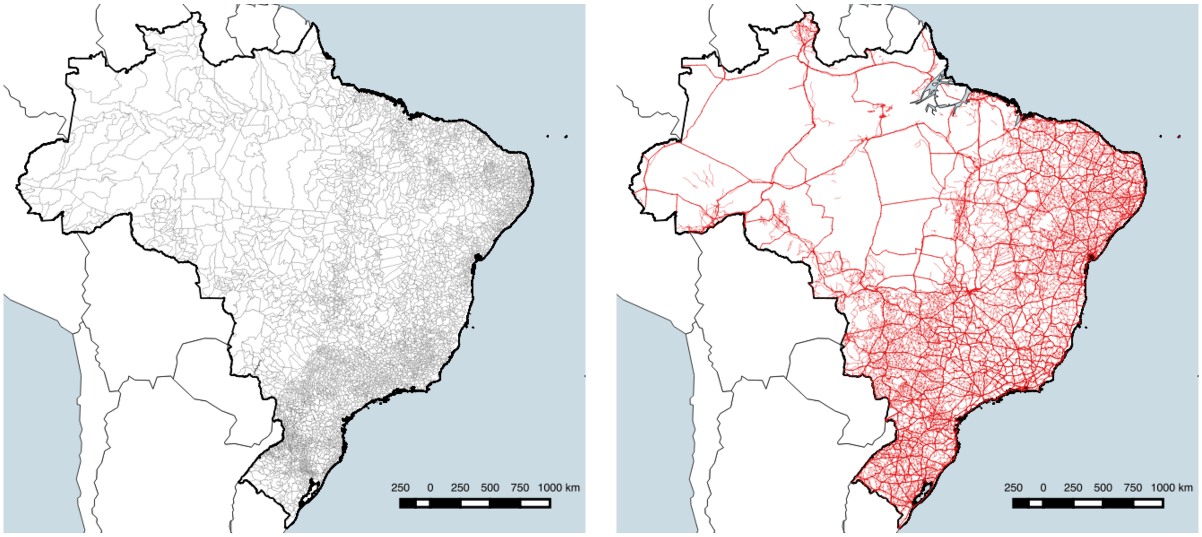
First, my primary data source is the Brazilian agricultural census of 2006, which provides

---

[5] For example, Alston et al. (2000) assemble 292 studies measuring a total of 1886 returns to agricultural research. They found a median return to research of 48% per year.

[6] Without this heterogeneity, when a region is affected by a positive productivity shock in a sector, congestion limits the influx of workers, but the model does not generate out-migration from it.

Figure 1: Spatial Components of the Data



a. Territorial Division of Counties

b. Road Network

Notes: Panel (a) highlights the territorial division of counties used in this article. Panel (b) shows the complete road network used to calculate the minimum travel distance between counties.

information on revenues, land use, agricultural workers and the number of farms at the crop and county level. I defined 11 agricultural sectors that can be linked with the trade data. In addition, I use the demographic census of 2010 to obtain the total number of workers per county and to define employment in non-agricultural sectors.

Second, I combine data on the Brazilian road network with the official location of administrative center of counties. With this data, I construct a matrix of nodes and arcs based on the road network that allows me to calculate the travel distance between counties and ports.[7] To connect counties to external markets, I identify 14 major ports in Brazil and assume that counties trade with countries overseas using the closest port and with neighboring countries such as Argentina and Paraguay using the internal road system.

Third, I collected annual information on exports and imports (F.O.B.) that were disaggregated by product from every county to different external markets from 2000 to 2015, which is provided by the Brazilian Ministry of Development, Industry and External Commerce.

Fourth, I organized county-level data on the wholesale prices of 31 agricultural products from 1998 to 2015. Information is provided at a monthly frequency in the official statistics of the state Secretaries of Agriculture.

---

[7] An alternative procedure is to use travel distances in *km* or time according to Google maps. However, Google maps is not able to calculate the bilateral travel distance for a number of counties that are closer to the Amazon region of Brazil. In the appendix, for the subset of counties that I use to estimate trade costs and elasticities, I show that variation in the distances estimated with my matrix of nodes and arcs can explain 98% of the variation in distance obtained from Google. Also, I show that estimates of trade costs and trade elasticities are largely unaffected by whether I use distance in hours, kilometers or that derived from Google maps.

3 Facts

This section presents three facts about Brazilian agriculture that motivate the building blocks of the model that I present next. First, the observed differences in natural advantages alone cannot explain strong correlations between market access and the spatial patterns of agricultural specialization, which suggests the presence of alternative causes of specialization related to proximity to markets. Second, there is large heterogeneity in farm size across crops, which indicates technological differences in the use of land. Third, there are large differences in the importance of external markets across crops.

Fact 1: The observed differences in natural advantages are insufficient to explain strong correlations between market-access and the spatial patterns of agricultural activity.

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). Market access in county i (MA_i) is the sum of income in every other county n (I_n)[8] weighted according to a exponential decay of the minimum travel distance between county i and n (MTD_{in})

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

where C is the set of counties in the data. I set $\lambda = 0.01$, which is consistent with the magnitude of trade costs in Brazil.[9]

Figure 2 shows the south of Brazil where land regulation and natural advantages are more homogeneous relative to the rest of the country. Counties closer to the Atlantic coast have better market access compared to regions to the west. Panel (b) indicates that counties with better market access have a higher ratio of workers producing vegetables relative to cattle. Panel (c) shows that lower market access is associated with larger farms, and Panel (d) shows that there is a positive correlation between market access and the average price of land in a county. Table 1 shows that these correlations are statistically significant, large and robust to the inclusion of a rich set of controls for natural advantages based on measures of agricultural suitability from the FAO.[10]

While the model that I formulate in the next section captures the influence of natural advantages on market access and agricultural specialization, the model also contains mechanisms that explain

---

[8] I augment the total income in port cities according to total exported value in the port.

[9] An alternative approach would be to estimate λ to maximize the fit of the regressions with key variables in the data. For example, setting the decay parameter to 0.016 would minimize the residuals of a regression of the log of farm size against the log of market access.

[10] I constructed measures of agricultural suitability from pixel level maps from the FAO-GAEZ. I organize data for 22 crops, and for each crop, two types of technologies are available: one is related to the high use of intermediate inputs and another is for low use. See Costinot et al. (2016) for a detailed description of the FAO-GAEZ measures of agricultural suitability.

Table 1: Relationships between Market Access and Agricultural Activity

	Dependent Variable (in Logs)		
	Land Price (1)	Farm Size (*ha*) (2)	$\frac{\text{Workers in Vegetables}}{\text{Workers in Cattle}}$ (3)
(a) No Controls			
$\log MA_i$	0.473***	-0.212**	0.466***
	(0.044)	(0.083)	(0.097)
R^2	0.47	0.08	0.15
(b) Controls for Agricultural Suitability			
$\log MA_i$	0.305***	-0.260***	0.533***
	(0.023)	(0.047)	(0.083)
R^2	0.73	0.42	0.29

Notes: *** denotes significance at the 1% level and ** at 5%. Sample size is 5200. Panel (a) shows estimates of the raw elasticity of variables associated with agricultural agricultural and market access. Panel (b) shows these correlations once we control for 44 measures of agricultural suitability from the FAO.

why these links hold when we control for observed differences in natural advantages.[11]

Fact 2: There are large differences in farm size across crops.

Using county-level information on farm size disaggregated by agricultural activity, Figure 3 shows that the average farm size in each activity is drastically different: soybean uses on average 212 *ha,* whereas coffee uses 11 *ha*. Table 2 reports the regressions of farm size against different sets of fixed effects to study the sources of variation in farm size. County fixed effects explain 25% 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 approximately 47% of the variation, which indicates that a large part of the variation is explained by factors that are common to each crop. Together, both sets of fixed effects explain 67% of the variation. Therefore, a large part of the variation that is explained by county-fixed effects is not already absorbed by activity fixed effects since the explanatory power of the regression increases by 20 percentage points. These patterns motivate the introduction of differences in factor intensity across crops in the model.

Fact 3: There are large differences in the importance of external markets across crops.

Figure 4 shows the distribution of exports and revenues in different crops. There are large dif-

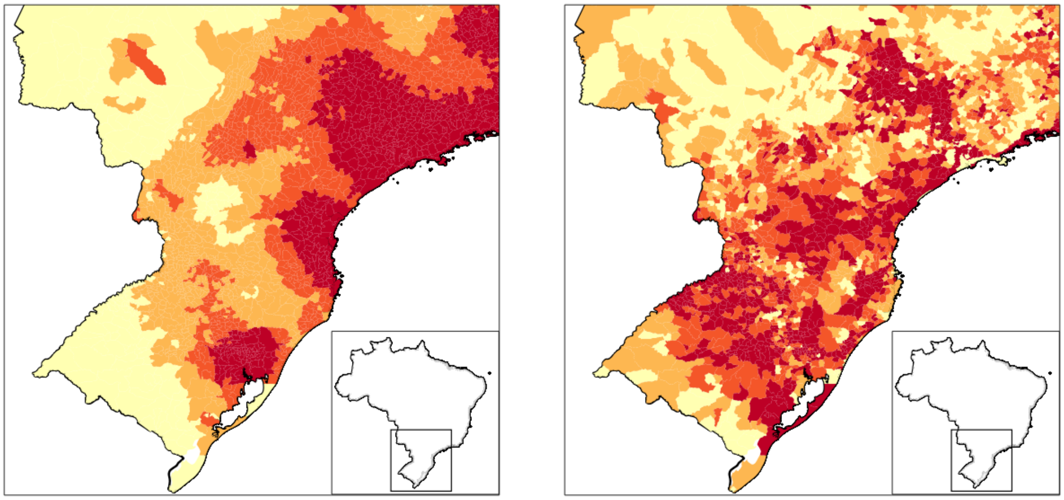
---

[11] In the appendix, I investigate the role of market access in the specialization in land intensive crops following the approach used in Nunn (2007), Costinot (2009) and Levchenko (2007). I find that in a regression of the log of total revenues from crop k in county i with crop fixed effects and county fixed effects, a coefficient associated with the interaction between the log of market access and the log of land-use intensity as measured in the structural section of the paper is large, negative and statistically significant.

Figure 2: Market Access and Agricultural Activity across Space (2006)

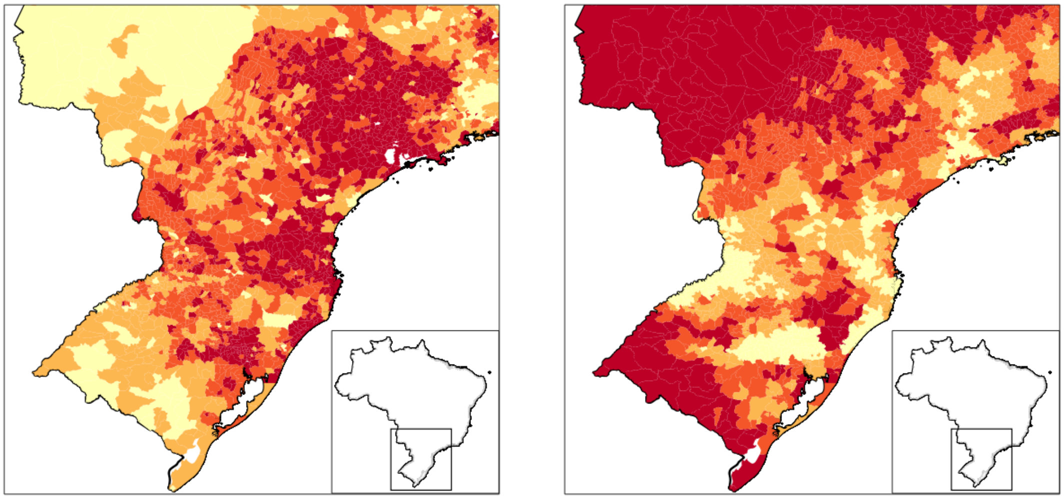
(a) Market Access

(b) $\frac{\text{Workers in Vegetables}}{\text{Workers in Cattle}}$



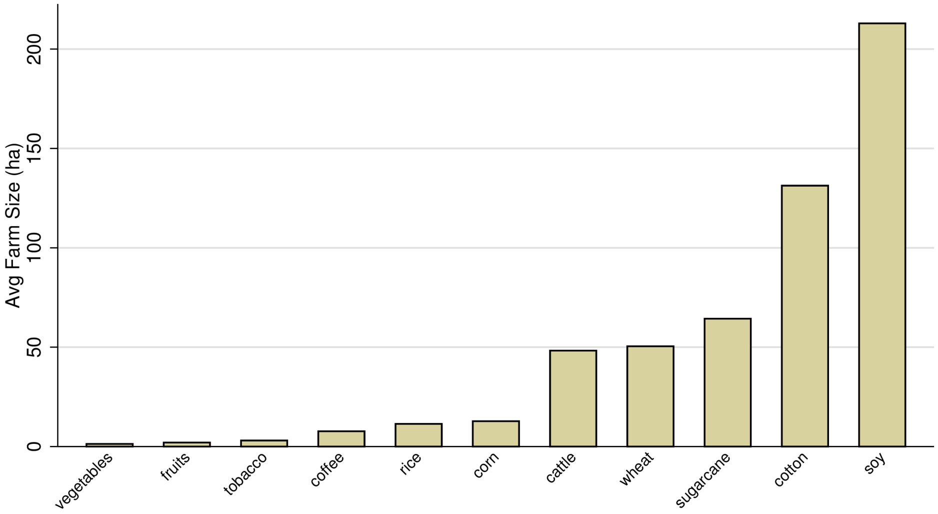
(c) Avg Land Price

(d) Avg Farm Size



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 income elsewhere $MA_i \equiv \sum_{n \in C} exp(-0.01 \times MTD_{in}) \times I_n$. Average farm size is the total land used in agriculture divided by the total number of farms in each county. Land price is the total value of farms divided by the total amount of land in each county.

Figure 3: Average Farm Size (*ha*) per Agricultural Activity



Notes: To construct average farm size per agricultural activity, for the counties with a positive production, I take the average of the total land use divided by the total number of farms in each county and crop.

Table 2: Sources of Variation in Farm Size (*ha*) across Activities and Counties

	Regression of Farm size (*ha*) on		
	County FE (1)	Activity FE (2)	County & Activity FE (3)
R^2	0.25	0.47	0.67
Obs	30333	30333	30333

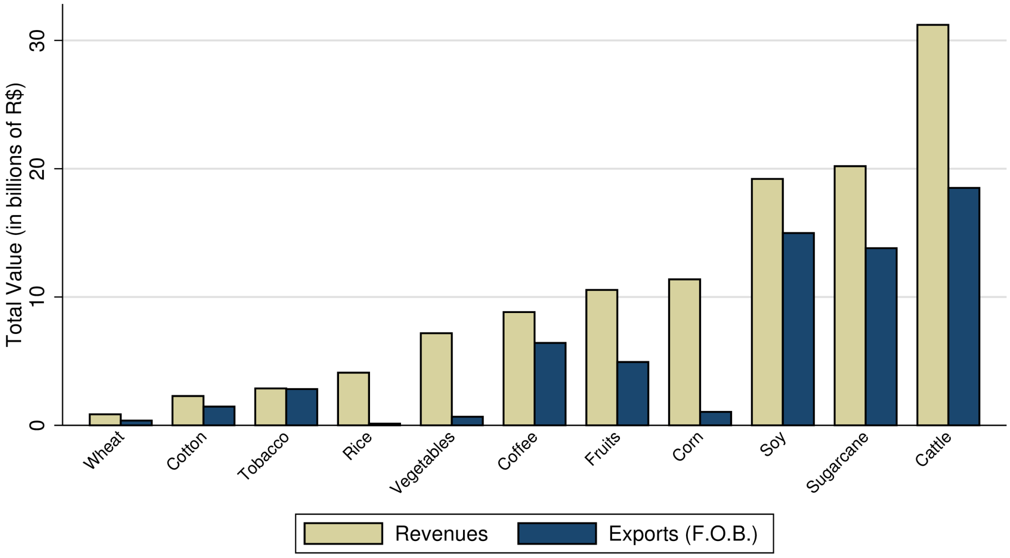
Notes: The regressions for this table use data on average farm size per activity and county. Farm size is constructed by dividing total land use by the total number of farms in each activity and each county.

ferences in the share of revenues attributed to external markets: 60% of Brazilian soybeans are exported compared to 30% for fruits. Additionally, Brazil is a large global producer of many commodities. For example, Brazil accounts for 16% of the global production of beef, 30% of that of soybeans, 35% of that of coffee, and 25% of that of sugarcane. The model in the next section allows shocks to the domestic economy to affect the international price of commodities.

4 The Model

My goal is to develop a quantitative general equilibrium model that accounts for spatial differences in terms of what crops are produced, how they are produced, and how they interact with external markets. To do so, I build a spatial model with multiple locations, multiple agricultural sectors and one urban sector where locations trade in an Eaton and Kortum (2002) fashion. I allow agricultural

Figure 4: Revenues and Total Exports per Agricultural Activity (2006)



Notes: Each bar shows the aggregate revenues and exports per agricultural activity in the data. Exports in dollars are multiplied by 2.175, the conversion rate for the Brazillian currency (Reais) in 2006.

sectors to be different in terms of their factor intensities and geographic trade costs. This structure implies tractable gravity equations defining trade activities that I use to highlight the sources of comparative advantages captured by the model. I introduce international trade by assuming a representative county for the external market. To save on notation, I first present the model without international trade.

4.1 Setup

The domestic economy has C counties denoted by i or n. There are K_A agricultural sectors and one urban sector u, both of which are tradeable. I denote sectors by $k \in \{1, ..., K_A, u\}$ and let K be the set of all sectors. Within each sector, there is a *continuum* of varieties $v \in [0,1]$. For example, fruit is a broad sector producing different varieties of fruits, and coffee is a sector that accounts for different varieties of coffee beans. There are two types of agents: workers who earn wages and can move between counties and sectors, and landowners who earn land rents and are immobile.[12][13] The economy has two factors of production, labor and land. There is a total mass of labor N, and each county has a land endowment of L_i. In equilibrium, counties may produce goods in both

---

[12] General statistics on migration indicate large spatial mobility in Brazil. According to the Brazilian annual household survey of 2006, PNAD, approximately 55% of the heads of households lived in a county where they were not born and 40% lived in a state where they were not born.

[13] Another common approach is to assume that land rents become part of a national asset that is equally redistributed across regions (see Caliendo et al. (2014)). Using this alternative procedure leads to similar aggregate results, but it shuts down distributional impacts between landowners and workers.

urban and agricultural sectors and differ in terms of their degree of urbanization, i.e., the share of labor employed in urban activities.

Preferences Workers have Cobb-Douglas preferences over final goods from each agricultural sector k

$$U_{i\omega} \equiv a_{i\omega} \prod_{k=1}^{K} (C_{ki\omega})^{\mu_k}, \tag{2}$$

where ω denotes a worker, $U_{i\omega}$ is the utility, C_{ki} is the consumption of final goods, μ_k ($\sum_k \mu_k = 1$) is the share of consumption on goods from each sector k and $a_{i\omega}$ is an idiosyncratic amenity shock. Idiosyncratic amenities are independently drawn from a Fréchet distribution

$$G_i(a) = exp(-A_i a^{-\kappa}),$$

where A_i governs the average of amenities a and κ governs their dispersion. Workers in i earn the same wage and face the same prices. Hence, in what follows, I suppress the subscript ω except where important. The consumption of final goods in each sector k is

$$C_{ki} \equiv \left(\int c_{ki}(v)^{\frac{\sigma_k - 1}{\sigma_k}} dv \right)^{\frac{\sigma_k}{\sigma_k - 1}},$$

where $c_{ki}(v)$ is the consumption of variety v and σ_k is the elasticity of substitution between varieties. Landowners have the same preferences as workers.

Urban Sector Firms produce varieties of urban goods with the following technology

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

where $z_{ui}(v)$ is the efficiency of technology in each county, $n_{ui}(v)$ is the employment of labor 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 $w_i/z_{ui}(v)$.[14]

Agricultural Sector Farms produce varieties of agricultural goods with the following technology

$$q_{ki}(v) = z_{ki}(v) n_{ki}(v)^{\phi_k} l_{ki}(v)^{\gamma_k} C_{ui}(v)^{\alpha_k}, \tag{4}$$

---

[14] In the appendix, I show the results using an alternative specification in which urban actitivies also use intermediate inputs as follows

$$q_{ui}(v) = z_{ui}(v) n_{ui}(v)^{1-\alpha_k} C_{ui}(v)^{\alpha_k},$$

where $C_{ui}(v)$ is a composite of goods produced by the urban sector and α_k is the share of intermediate inputs. I obtain similar conclusions with this alternative specification.

where $q_{ki}(v)$ is the output, $l_{ki}(v)$ is land use, $C_{ui}(v)$ is a composite of the final goods from the urban sector, $n_{ki}(v)$ is the employment of variable labor, $z_{ki}(v)$ is an efficiency term, and γ_k and α_k are the factor shares ($\phi_k + \gamma_k + \alpha_k < 1$).[15]. $C_{ui}(v)$ captures the use of intermediate inputs such as tractors, technical assistance and fertilizers. There is a fixed cost of f workers to set up a farm, which I assume to be the cost of a manager and equal to 1. Since technology is decreasing returns to scale, farm size is well defined. Assuming that farms make zero profits, in the appendix I demonstrate that the farm gate price of variety v in sector k ($p_{ki}(v)$) is

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

where κ_k is a constant.[16] Note that the aggregate production function at the county level implied by this farm-gate price exhibits constant returns to scale with a labor share (variable and fixed) of $1 - \alpha_k - \gamma_k$ and does not depend on the variable labor intensity ϕ_k. This formulation reconciles micro data on farm size with the fact that estimates of agricultural production functions using aggregate data typically do not reject the hypothesis of constant returns to scale (Mundlak et al., 1999).[17]

Trade Following Eaton and Kortum (2002), the efficiency $z_{ki}(v)$ 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)

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

where T_{ki} is a scale parameter that increases the probability of receiving good efficiency draws for z ($T_{ki} \geq 0$), and θ_k drives the dispersion of the draws. T_{ki} is specific to county i and relates to the absolute advantage in the production of varieties for sector k. θ_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

---

[15] With this formulation, 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.

[16] Here, $\kappa_k \equiv \frac{f^{1-\alpha_k-\gamma_k-\phi_k}}{(1-\alpha_k-\gamma_k-\phi_k)^{1-\alpha_k-\gamma_k-\phi_k} \phi_k^{\phi_k} \gamma_k^{\gamma_k} \alpha_k^{\alpha_k}}$.

[17] This production function is consistent with the following micro-level correlations presented in Foster and Rosenzweig (2011) for India: labor costs per acre, output per acre and labor per acre decrease with farm size. In addition, this production function is also consistent with additional micro-level correlations presented in Adamopoulos and Restuccia (2014): the capital-land ratio decreases with farm size. In this formulation, farms producing the same crops and in the same county have the same size, which is the finest level of disaggregation available in the data.

variety k to county n is $p_{kin}(v) = \tau_{kin} p_{ki}(v)$.

Eaton and Kortum (2002) demonstrate that under this formulation, the total sales X_{kin} in sector k from county i to county n is given by

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

where E_n is the total expenditure from consumers in n and π_{kin} is the probability that county i is the lowest-cost provider of a variety to destination n. In addition, the price index becomes[18]

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

where Γ_k is a constant.[19] This price index can be used to simplify the total sales in (7)

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

This gravity equation defines the value exported from county i to n as a function of a set of the origin county's characteristics ($T_{ki}(P_{ui}^{\alpha_k} r_i^{\gamma_k} w_i^{1-\gamma_k-\alpha_k} \tau_{kin})^{-\theta_k}$), the total expenditure of the destination county on sector k ($\mu_k E_n$), bilateral trade costs ($\tau_{kin}^{-\theta_k}$), and the price index of the destination ($P_{kn}^{\theta_k}$), which is a multilateral resistance term capturing competition between counties. Here, the dispersion parameter (θ_k) governs the elasticity of trade with respect to trade costs (τ_{kin}). Note that the gravity equation for the urban sector is given by setting $\alpha_k = \gamma_k = 0$ and adjusting the total expenditure to include the expenditure on urban goods coming from agricultural sectors.

For the next section, it is useful to define an overall absolute comparative advantage of county i in sector k which is a combination of natural advantage and factor prices

$$\overline{T}_{ki} \equiv \underbrace{T_{ki}}_{\textit{Natural Advantages}} \underbrace{\left(P_{ui}^{\alpha_k} r_i^{\gamma_k} {w_i}^{1-\alpha_k-\gamma_k}\right)^{-\theta_k}}_{\textit{Factor Prices}}. \tag{10}$$

---

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

[19] Here, $\Gamma_k \equiv \kappa_k \Gamma\left(\frac{\theta_k+1-\sigma_k}{\theta_k}\right)^{1/(1-\sigma_k)}$ and Γ is the gamma function. To have a well-defined price index, I assume that $\theta_k > \sigma_k - 1$. In practice, since Γ_k and T_{kn} are not separately identified and because σ_k only enters in the equations defining the equilibrium through Γ_k, I do not need to separately estimate σ_k.

4.2 International Trade

To introduce international trade, I assume the existence of a representative county for the external market (F) that has an exogenous expenditure in each sector (E_{kF}),[20]as well as an exogenous absolute advantage for exporting to itself ($\overline{T}_{kX}$) and to the domestic economy ($\overline{T}_{kM}$). With these assumptions, the exports to external markets are given by

$$X_{kiF} \equiv \pi_{kiF} E_{kF} = \frac{\overline{T}_{ki}\tau_{kiF}^{-\theta_k}}{\overline{T}_{kX} + \sum_{i'=1}^{C} \overline{T}_{ki'}\tau_{ki'F}^{-\theta_k}} E_{kF}, \tag{11}$$

and imports from external market by

$$X_{kFi} \equiv \pi_{kFi} E_{kF} = \frac{\overline{T}_{kM}\tau_{kFi}^{-\theta_k}}{\overline{T}_{kM}\tau_{kFi}^{-\theta_k} + \sum_{i'=1}^{C} \overline{T}_{ki'}\tau_{ki'i}^{-\theta_k}} E_{ki}, \tag{12}$$

where τ_{kiF} is the distance of i to the closest port and E_{ki} is the total expenditure of county i in sector k.[21] An increase in the absolute advantage of the domestic economy increases the total sales to the external market relative to the external competitors. In particular, the elasticity of total exports with respect to shocks in the cost of production of county i is controlled by θ_k. Note that, despite the fixed expenditure from the external market, the domestic economy still affects the international price index, which is a function of the denominator in equation 11. I allow for an aggregate trade imbalance via an *exogenous* international transfer (Υ) to the external market of a lump-sum tax on income as in Deakle et al. (2008).[22]

4.3 Spatial Equilibrium

In spatial equilibrium, market clearing for each county i and sector k is given by

$$X_{ik} = \sum_{n} \pi_{kin} E_{kn} + \pi_{kiF} E_{kiF}. \tag{13}$$

where E_{kn} is the total expenditure of county n in sector k. The aggregate exports from the domestic economy to the external market must equal aggregate imports and the aggregate trade imbalance

---

[20] This is equivalent to assuming that the external market has Cobb-Douglas preferences for each sector and that shocks to the Brazilian economy do not affect its total global income.

[21] Here, $E_{kn} = \mu_k E_n$ for the agricultural sector and $E_{un} = \mu_u E_n + E^A_{ui}$ for the urban sector, where E^A_{ui} is the sum of the expenditure of the agricultural sector on urban goods ($E^A_{ui} \equiv \sum_{k=1}^{K} \alpha_k \frac{w_i N_{ki}}{(1-\gamma_k-\alpha_k)}$).

[22] Note that, since some counties trade more with the external market than others, the fact that I am assuming the same lump-sum tax across counties implicitly imposes trade imbalances between counties in the domestic economy, because counties that trade less with the external market implicitly transfer to counties that trade more.

(Υ)

$$\sum_k \sum_i \pi_{kiF} X_{kF} = \sum_k \sum_i \pi_{kFi} X_{kFi} + \Upsilon. \tag{14}$$

Workers choose to live where their utility is maximized. A worker in county i with a realization of amenity $a_{ki\omega}$ chooses to live in i if and only if $a_{ki'\omega}(w_{i'}/P_{i'}) \leq a_{ki\omega}(w_i/P_i)$ for all $i' \in C$, where $P_i \equiv \prod_k (P_{ki}/\mu_k)^{\mu_k}$ is the price index of county i. In this case, Ramondo et al. (2016) shows that the share of workers living in county i, π_i^M, is

$$\pi_i^M = \frac{A_i \left(w_i/P_i\right)^{\kappa}}{\sum_n A_n \left(w_n/P_n\right)^{\kappa}}. \tag{15}$$

Here, κ captures the role of the non-monetary costs of migration since this is the elasticity of workers' supply with respect to changes in real wages. When $\kappa \to \infty$, the labor supply becomes perfectly elastic. The welfare of workers W is given by $\gamma(\sum_n A_n (w_n/P_n)^{\kappa})^{1/\kappa}$, where $\gamma \equiv \Gamma(1 - 1/\kappa)$, and the welfare W_i of landowners in i is given by $\gamma A_i^{1/\kappa}(r_i/P_i)$.

Land market clearing satisfies $L_i = \sum_{k=1}^K L_{ki}$, where L_{ki} is the demand for land in each sector k, and labor market clearing requires $N = \sum_k \sum_i N_{ki}$, where N_{ki} is the mass of workers in sector k in county i.

Definition 1. (*Spatial Equilibrium*) Given a set of counties C, parameters defining the preferences $\{\sigma_k, \mu_k, \kappa\}$, production technologies $\{\theta_k, \gamma_k, \alpha_k, f\}$, natural advantages and amenities $\{T_{ki}, A_i\}$, absolute advantages of the external market $\{T_{kX}, T_{kM}\}$, expenditures of external markets $\{E_{kF}\}$, bilateral trade costs $\{\tau_{kin}\}$, total population N, and land supplies $\{L_i\}$, the spatial equilibrium is defined by endogenous vectors of worker allocation, wages, land rents, and welfare $\{N_{ki}, w_i, r_i, W\}$ such that: (1) land markets clear in each county; (2) farms and firms make zero profits; (3) labor is fully employed; (4) workers maximize utility; (5) goods market clear; (6) trade after transfers with the external market is balanced; (7) the total payments to factors of production equal the total sales.

4.4 Absolute and Comparative Advantage

Here, I discuss the sources of comparative advantages between sectors in the model. To do so, first take two counties i and i' and two sectors k and k', we can then combine (9) and (10) to obtain the following equation governing the patterns of specialization

$$\frac{X_{kin}}{X_{k'in}} \Bigg/ \frac{X_{ki'n}}{X_{k'i'n}} = \frac{\overline{T}_{ki} \tau_{kin}^{-\theta_k}}{\overline{T}_{k'i} \tau_{k'in}^{-\theta_{k'}}} \Bigg/ \frac{\overline{T}_{ki'} \tau_{ki'n}^{-\theta_k}}{\overline{T}_{k'i'} \tau_{k'i'n}^{-\theta_{k'}}}. \tag{16}$$

This equation shows that when the absolute advantage $(\overline{T}_{ki})$ of a sector k increases relative to the absolute advantage in another sector k' $(\overline{T}_{k'i})$, 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 (τ_{kin}) and productivity dispersion (θ_k) can also induce specialization. While the dispersion of efficiencies (θ_k) govern the comparative advantage *within* sectors in a single-sector version of the model, the comparative advantage *between* sectors is driven by differences in $\overline{T}_{ki}$ and transportation costs.

Equations (10) and (16) reveal four major sources of heterogeneity driving specialization.

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

Second, differences in factor intensity generate forces of specialization related to Hecksher-Ohlin models. Differences in land intensity induce specialization because of variations in land rents across counties. Note that even though an increase in land rents (r_i) in a county affects all crops, due to technological differences in land-use intensity (γ_k), this has a disproportionate effect on the absolute advantage ($\overline{T}_{ki}$) of land-intensive activities.[23] Similarly, access to cheap intermediate input-intensive (P_{ui}) induces specialization in crops that are intermediate inputs intensive (higher α_k) and lower wages (w_i) make counties competitive in labor-intensive crops (lower $\alpha_k + \gamma_k$).[24]

Third, the proximity to consumers is beneficial for perishable activities, which is captured by the fact that a sector k may have a larger τ_{kin} for the same pair of i and n.

Fourth, counties with lower costs of a bundle of inputs or delivery have a competitive advantage in crops with high trade elasticity. To understand this mechanism, note that trade elasticities (θ_k) govern the relative importance of natural advantages with respect to the costs of a bundle of inputs or delivery. For example, when θ_k is large, variations in factor prices (10) translate into larger variations in absolute advantage, reducing the relative importance of variation in natural advantages (T_{ki}). Intuitively, efficiency draws ($z_{ki}(v)$) are more homogeneous across counties for crops with high θ_k, which means that counties have less incentive to trade based on differences in their efficiency draws.

---

[23] 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 Figure 2.

[24] 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).

4.5 Example: The Economy on a Line

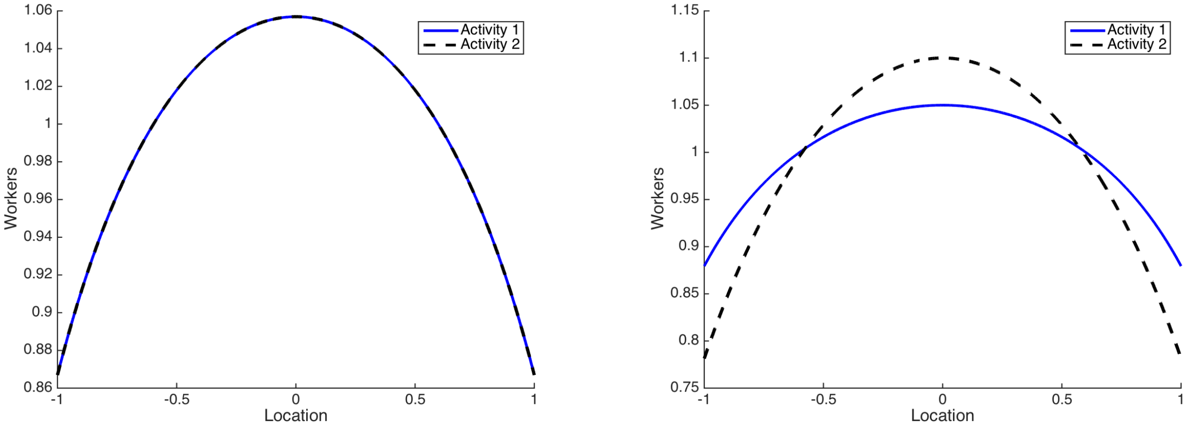
In this section, I illustrate the causes of agricultural specialization in the model by simulating an economy with the simple spatial structure of a line, two agricultural sectors, no international trade and no heterogeneity in natural advantages, amenities or land endowment. In this economy, consumers spend half of their income on each sector.

Panel (a) shows the allocation of workers when both agricultural sectors have the same land-use 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 because the Fréchet distribution ensures that every location is sufficiently competitive in a positive measure of varieties, employing 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 (10) and (16). 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 5, 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 parameters that can explain the observed patterns of specialization, such as those discussed in Fact 1.

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 the line due to transportation costs. In the data, heterogeneity in the productivity of urban activities (T_{ui}) is a major factor driving differences in population density. Note that, because some agricultural sectors demand more intermediate inputs from urban activities than others, the patterns of agricultural specialization in a region have an influence on urban activities. Therefore, the model captures circular causations in which agricultural activity fosters urban activity and vice versa.

Figure 5: Economic Activity on a Line


(a) No Technological Differences

(b) Technological Differences

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 Taking the Model to the Data

This section takes the model to the data. To do so, I first estimate the trade costs and trade elasticities (τ_{kin} and θ_k). Second, I recover the production function parameters (γ_k, α_k and ϕ_k). Third, I calibrate the labor supply elasticity (κ) and the expenditure shares (μ_k). Finally, I set the natural advantages (T_{ki}), external market advantages (T_{kX} and T_{kM}) and amenities (A_i) so that the model exactly matches the data in terms of wages, allocation of workers, aggregate exports and aggregate imports for 2006. I close this section with tests of the fit of the model.

5.1 Estimation

Trade Costs (τ_{kin}) and Trade Elasticity (θ_k) 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, here I do not observe the origin or the destination of a variety. In this case, a potential concern discussed in Eaton and Kortum (2002) and Anderson and van Wincoop (2004) is that the trade cost between two locations places only an upper bound on price differentials and, as such, 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 concept, I follow Allen (2014) and use the maximum realization of the monthly price differential within a year, variety v, and bilateral pair (i and n), which I define by $p^{Max}_{ymkin}(v) \equiv \max\{p_{y1ki}(v)/p_{y1kn}(v)\ ,\ldots, p_{y12ki}(v)/p_{y12kn}(v)\}$, where m stands for month and y for year. Hereafter, I drop the index for variety v to simplify notation. With these assumptions and a parametrization of trade costs (τ_{kin}),[25] I estimate the geographic trade cost (δ_k) with

$$\log p^{Max}_{ymkin} = \beta_{yk} + \beta_{mk} + \delta_k \log MTD_{in} + \varepsilon_{ymkin}, \tag{17}$$

where β_{yk} and β_{mk} capture annual and seasonal fluctuations in trade costs and ε_{ymkin} captures unobserved determinants of trade costs. In the appendix, I test a number of alternative approaches to the estimation of trade costs and show that these methods lead to similar or smaller estimates of trade costs.[26]

To derive the gravity equation that I use to estimate θ_k, log-linearize equation (9), parametrize trade costs and add the subscripts for year to obtain

$$\ln X_{ykin} = \beta_{yki} + \beta_{ykn} - \theta_k \hat{\delta}_k \log MTD_{in} + \varepsilon_{ykin}, \tag{18}$$

where β_{yki} captures the absolute advantage ($\overline{T}_{ki}$) and β_{ykn} captures the terms related to the destination ($P^{\theta_k}_{kn} E_n$). Using estimates of geographic trade cost ($\hat{\delta}_k MTD_{in}$), the gravity equation identifies the dispersion parameter (θ_k). To estimate equation (18), I assume that all counties face the same trade cost to reach markets overseas after reaching the port. Therefore, the variation that identifies θ_k in equation 18 comes from the fact that the same origin county i may export to both neighboring countries and to countries overseas.[27] Finally, statistical power allows me to estimate the transportation costs and trade elasticities for three major groups: perishable agricultural sectors (vegetables and fruits), cereals (rice, soybeans, corn and wheat) and non-perishable goods (remaining crops). For the urban sector, since I do not have price information, I use the estimates of the trade costs of non-perishable goods.

Table 3 shows that a 1% increase in bilateral distance is associated with a 0.05% increase in the price gap for non-perishable goods, a 0.08% increase for cereals, and a 0.10% increase for

---

[25] 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.

[26] I consider the following alternative specifications for the estimation of trade costs: (1) the use of the second highest price differential between counties within a year; (2) the use of all price differentials between counties; (3) the use of travel distance in terms of time as the explanatory variable with adjustments for speed of travel according to the quality of the road; (4) the use of the euclidean distance between counties as an instrument for travel distance between counties to minimize potential attenuation bias given by measurement error; and (5) the use of the travel distance in *km* and in time according to Google maps. Furthermore, to deal with the fact that I do not observe the origin or the destination of the products, I estimate the trade cost of two products that are largely exported (cattle and soybeans) using the travel distance to the closest port instead of travel distance between counties.

[27] About one third of the observations for sales comes from neighboring countries such as Argentina and Paraguay.

perishable products.[28] Table 3 reports results for trade elasticities. For perishables (vegetables and fruits), the elasticity is much smaller than cereals and non-perishables, which is consistent with the fact that there is likely more heterogeneity in varieties of fruits than in varieties of soybeans and wheat.

Table 3: Estimates of Trade Costs and Elasticities

		By Sector			
	All		Agriculture		
	Sectors	Urban	Non-Perishable	Cereals	Perishables
	(1)	(2)	(3)	(4)	(5)
(a) Dependent Variable: Price Differentials					
Geographic Trade Cost ($\hat{\delta}_k$)	0.066	-	0.051	0.082	0.104
	(0.002)	-	(0.002)	(0.009)	(0.009)
Obs	8081	-	5182	1723	1176
R^2	0.64		0.50	0.40	0.62
(b) Dependent Variable: Exports					
Trade Elasticity ($\hat{\theta}_k$)	4.26	5.27	4.60	5.60	2.45
	(0.548)	(0.840)	(1.598)	(1.715)	(1.265)
Obs	11691	5377	2557	1841	1916
R^2	0.92	0.86	0.93	0.96	0.93

Notes: Robust standard errors in parenthesis. Reported values are the negative of the estimated elasticity using the gravity equation. Estimates of geographic trade costs include year-product and month-product fixed effects. Estimates of trade elasticities include year-product-origin and year-product-destination fixed effects.

Production Function (γ_k, α_k and ϕ_k) I begin the estimation of the production function by setting the labor share in each crop ($\phi_k^L \equiv 1 - \alpha_k - \gamma_k$) equal to the aggregate ratio of payments to workers (variable and fixed) relative to revenues. I then calculate the aggregate labor and intermediate input share in agriculture, which equal 0.19 and 0.52, respectively, and imply an aggregate cost share of land of 0.29. To estimate the land-intensity parameters γ_k, I construct a set of moments derived from the first-order conditions of farmers and apply the generalized method of moments. These moments come from regressions of the log of revenues per hectare, the log of farm size and the log of workers per hectare against a set of crop fixed effects. In each regression, crop fixed effects identify land-use intensity relative to a base crop. The intuition for this identification is that crops with systematically lower revenues per hectare, farm size or workers per hectare are more land-intensive. Details about the construction of these moments are included in the appendix. With

---

[28] Cereals is the group with the highest weight per value in the data, which may explain why the geographic transportation costs are larger than the costs for non-perishable products. See Duranton et al. (2013) for an analysis of the relationship between weight per value and trade costs.

estimates of the relative parameters for land intensity, I adjust their levels to match the aggregate cost share of land of 0.29. Finally, with the land intensities (γ_k) and labor intensities (ϕ_k^L) for each crop, I recover the intermediate input share per crop (α_k).[29]

Table 4 shows large heterogeneity in factor shares across crops. For example, while the share of land for soybeans is 0.55, the share for fruits it is 0.05. Cattle production has a small share of intermediate inputs and a large share of land. Therefore, better market access provides low benefits for cattle ranchers and high land rents have a large negative effect on their costs of production. The combination of these two technological properties explain why cattle farms are often located in remote regions.

Table 4: Estimated and Calibrated Parameters

	Factor Intensity				Expenditure
	Labor $1-\hat{\gamma}_k-\hat{\alpha}_k$ (1)	Variable Labor $\hat{\phi}_k$ (2)	Land $\hat{\gamma}_k$ (3)	Intermediates $\hat{\alpha}_k$ (4)	Share $\hat{\mu}_k$ (5)
Cattle	0.272	0.207	0.557	0.170	0.045
Coffee	0.279	0.227	0.110	0.610	0.008
Corn	0.196	0.082	0.329	0.474	0.037
Cotton	0.036	0.033	0.043	0.920	0.004
Fruits	0.131	0.069	0.046	0.821	0.022
Rice	0.107	0.054	0.108	0.784	0.015
Soy	0.086	0.065	0.615	0.298	0.015
Sugarcane	0.169	0.154	0.078	0.751	0.023
Tobacco	0.117	0.089	0.106	0.776	0.000
Vegetables	0.420	0.209	0.068	0.510	0.025
Wheat	0.078	0.029	0.641	0.279	0.009
Urban	1	0	0	0	0.791

Notes: Labor share and variable labor are calibrated using information on payments to labor and revenues in each crop. Land intensity is estimated using the generalized method of moments. Intermediate input shares are calibrated according to the parameters estiamted for labor and land intensity. Expenditure share is calibrated to match the observed domestic absorption in each sector of the economy.

Labor Supply Elasticity (κ) I calibrate the labor supply elasticity κ implied by workers' preferences according to the literature. Oliveira and Morten (2014) estimate an elasticity of 3.4 using data for Brazil between 1940 and 1980. Diamond (2016), Suarez Serrato and Zidar (2016) and

---

[29] The aggregate production function implied by my estimates are consistent with previous estimates of aggregate agricultural production functions. Sotelo (2016) uses results from Avila and Evenson (2010), who calculate factor shares of 55% for labor and 22% for land in Peru between 1981 and 2001 for cropland. Mundlak et al. (1999) reports estimates of factor shares from several studies and finds a median factor share of 40% for labor, 40% for capital, and a factor share of 24% for labor. See Mundlak et al. (1999) for a review of this literature.

Monte et al. (2015) find elasticities between 1.3 and 4 using recent data for the US economy. I thus set κ equal to 3, which is between the values obtained in the literature, and experiment with lower and higher values in the appendix.

Expenditure Shares (μ_k) I set the expenditure shares μ_k equal to the aggregate absorption of the Brazilian economy in each sector, adjusting for aggregate exports and imports from external markets.

Natural Advantages (T_{ki})**, External Market Advantage** (T_{kX}**,** T_{kM}) **and Amenities** (A_i) Once with the parameters for preferences and technologies in hand, I can recover the distribution of natural advantages (T_{ki}), amenities (u_i) and absolute advantages of the external market ($\overline{T}_{kX}$ and $\overline{T}_{kM}$) that rationalize the data in terms of wages (w_i), workers (N_{ki}), aggregate exports (X_k) and aggregate imports (M_k) for a given cross-section.[30] The technique that I use to recover these values follows Allen and Arkolakis (2014) who prove that for a spatial economy with a single sector and no international trade there is a unique distribution of amenities and productivities that rationalizes the data in terms of workers and wages. In the appendix, I extend their proof to a spatial economy with multiple sectors and exogenous absolute advantages for the external market.

Proposition 1. *(Competitive Advantages and Amenities Inversion) Given the observed data on wages, land area, worker allocation, aggregate exports and imports from the domestic economy and the total expenditure of external markets* $\{w_i,\ L_i,\ N_{ki},\ X_k,\ M_k,\ E_{kF}\}$, *as well as parameters for preferences, production technology, trade elasticities and trade costs* $\{\kappa, \mu_k, \gamma_k, \alpha_k, f_k, \theta_k, \tau_{kin}\}$, *there is a unique (up-to-scale by sector) vector of natural advantages, amenities and absolute advantages of external markets* $\{T_{ki}, T_{kX}, T_{kM}, A_i\}$ *that is consistent with the data being an equilibrium of the model.*

5.2 Goodness of Fit

Before considering how the model behaves under counterfactuals, I discuss how the model fits within sample. I focus on four dimensions that are the focus of the model: the patterns of agricultural specialization and production, the price of land, and the foreign exports.

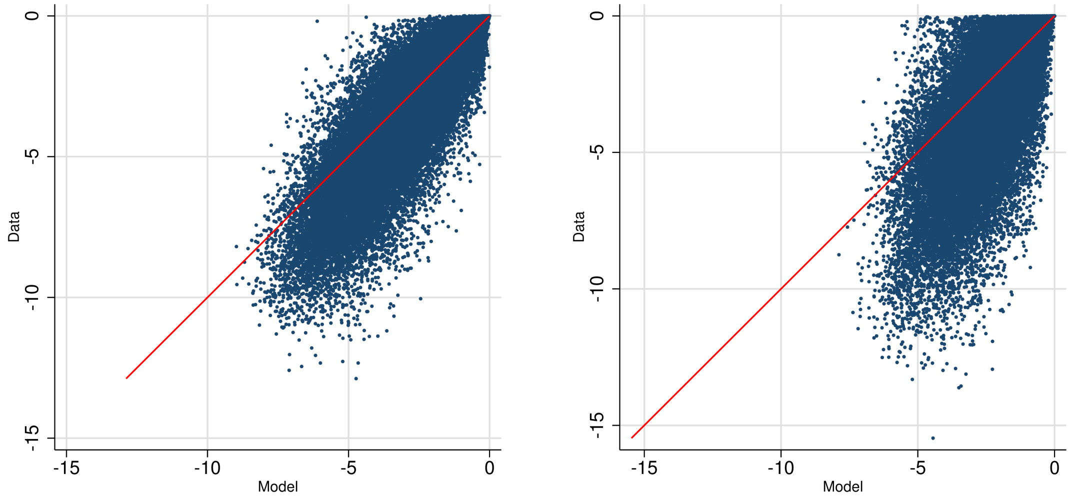
Agricultural Specialization By construction, the model correlates perfectly with the share of workers in different crops and counties. Figure 6 shows the fit of the model with two measures of specialization: the share of the land employed by each crop and the share of revenues of each crop

---

[30] In the structural model, I assume the existence of only one external market. Note, however, that for the estimation of trade elasticities θ_k, I used information for different foreign destinations.

within a county. To construct these measures, I only need the distribution of workers and wages and the estimates of the production function parameter. Since the model perfectly matches the distribution of workers and wages, divergences between the model and the data can be attributed to misspecification of the production function and to measurement error.[31] In Panel (a), a regression of the log of the share of area in each crop in the data against the log of the share of area in the model yields a coefficient of 1.15 with an R^2 of 0.72. For Panel (b), an analogous regression gives a coefficient of 1.26 with an R^2 of 0.46. The fit of the model is good even though these moments are not directly targeted in the estimation, but the model captures fewer small shares of land use and revenues than are in the data.

Figure 6: Non-Targeted Measures of Crop Specialization


(a) Share of Land Use

(b) Share of Revenues

Notes: Panel (a) shows the share of land employed in each crop within counties and Panel (b) the share of revenues of each crop within counties. The red line shows the 45 degree line.

Farm Size (*ha*) Panel (a) in Figure 7 shows that the model correlates well with the farm size distribution in each county and crop: the predictions from the model are concentrated in the 45°-degree line, and a regression of the log of farm size in model on the data provides a coefficient of 0.88 with an R^2 of 0.53. In addition, Figure 8 shows that the model predicts well the average farm size per activity as well, despite the fact that the average farm size moment is not directly used in the estimation.[32]

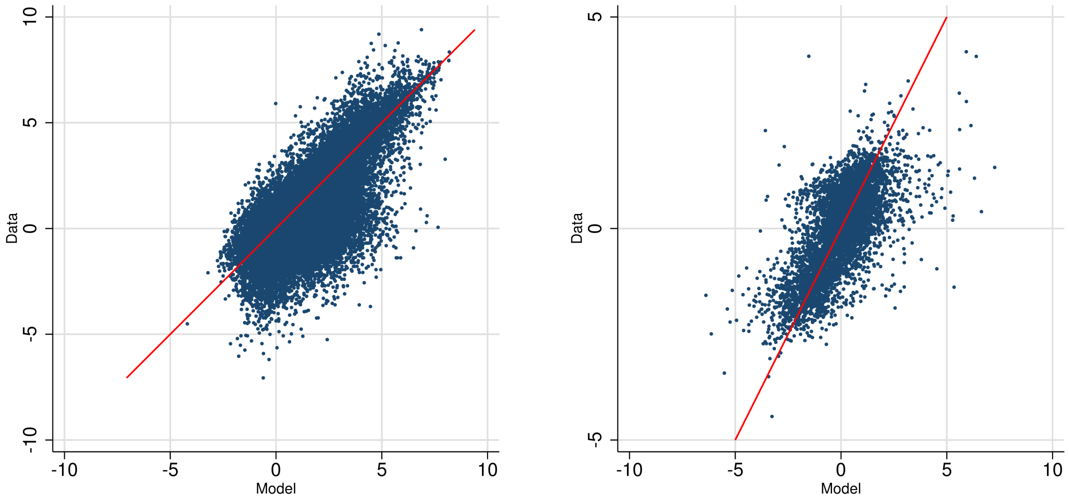
---

[31] Implicit in the estimation procedure is the assumption that the distributions of workers and wages are observed without measurement error. Alternative approaches rely on similar assumptions. For example, Costinot et al. (2016) assume that data on productivity from the FAO is measured without error and Sotelo (2016) assume that output is measured without error.

[32] If the fixed cost f is different from one and is crop specific, the model can perfectly match the average farm size per crop. However, results are unaffected if I include this level of heterogeneity.

Figure 7: Additional Non-Targeted Moments

a. Farm Size (Activity & County) b. Land Rents

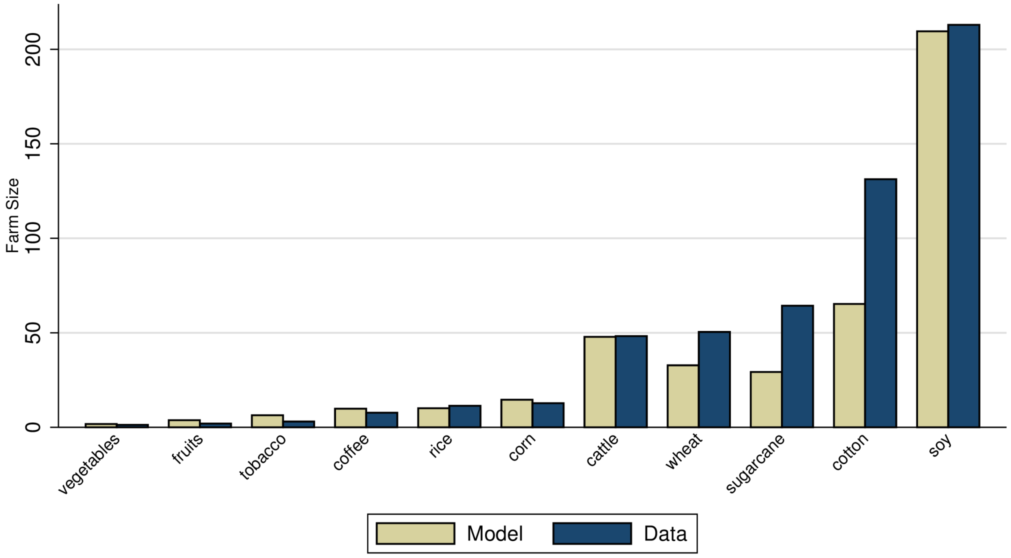


Notes: Panel (a) shows the fit of the model with the actual average farm size per activity in each county. Panel (b) shows the fit of land rents in the model with the average price of land in each county in the data. The red line shows the 45 degree line.

Land Rents Panel (b) in Figure 7 shows the correlation between the average price of land in the data and land rents in the model. In the appendix, I show that the price of land predicted by the model in a county is equal to wages multiplied by its factor proportion (total land and workers) and an adjustment term that averages the land-use intensity of different crops according to the share of workers. Despite this simple formulation of land rents, it is reassuring that a regression of the log of land prices in the data on the log of land rents in the model has a slope of 0.52 with a R^2 of 0.46, even though the price of land observed in the data is an imperfect *proxy* of land rents.

Output and Exports Since one of the goals of this paper is to study the interaction between the external and domestic market, I also check the fit of the model with respect to data on trade. Figure 9 shows that the model predicts well the total revenues in different states of Brazil, as well as the total exports. In particular, a regression of the log of exports in the model and that in the data gives a coefficient of 0.96 with a R^2 of 0.41, whereas an analogous regression for output gives a coefficient of 1.15 and an R^2 of 0.80. Figure 9 reveals that the discrepancy between the model and the data is particularly large for states and crops with small revenues, with the model generally predicting more revenues than the data.

Figure 8: Average Farm Size in the Data and the Model



Notes: The figure shows the fit of the model with the average farm size across crops in the data.

6 Causes of Agricultural Specialization

Having estimated the model, I investigate the causes of the spatial differences in agricultural activity. To do so, I shut down different sources of heterogeneity in the model, simulate the economy without these sources and ask how much of the correlation between the model and the data remains. This exercise indicates the importance of each source of heterogeneity for the model to rationalize the data. Column 1 in Table 5 presents the results from this exercise. The model correlates perfectly with the observed share of farms in each crop and it has a correlation of 0.49 with the farm size distribution across crops and counties. There is a large reduction in the correlation when I assume that all counties with a positive production of a crop have the same average natural advantage in T_{ki}. Shutting down the differences in factor intensity has a similar effect on this correlation, but when I assume that crops have the same trade costs, there is a negligible effect on this correlation, indicating that differences in trade cost are not crucial for the model to rationalize the data.[33]

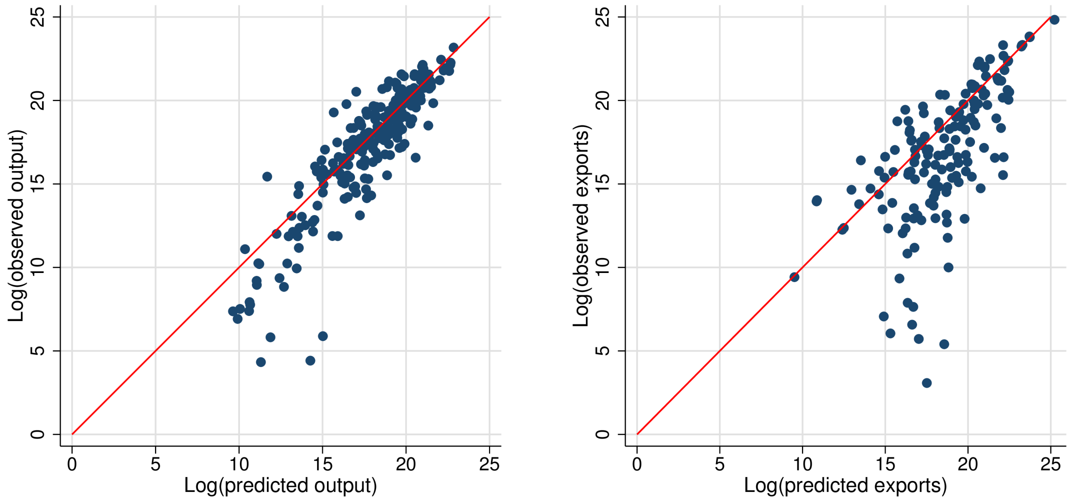
To understand the variations in the data driving these results, Table 6 examines the costs that counties face in serving agricultural goods to their respective state capital.[34] The table compares counties that are 50 *km* away from state capitals with those that are 200 *km* away. There are enormous differences in agricultural activity between these two groups: the average farm size in

---

[33] A regression of the log of natural advantages obtained by the model on the measures of agricultural suitability from the FAO fully interacted with dummies for each crop explains one third of the variation.

[34] State capitals in Brazil are always the largest urban centers in each state.

Figure 9: Exports and Revenues by State


a. Revenues

b. Exports

Notes: The figure shows the fit of the model with total revenues and total exports. The red line shows the 45 degree line.

counties that are 50 *km* away is half of those that are 200 *km*, and the ratio of vegetable farms relative to cattle farms is four times larger. The mobile inputs, labor and intermediate goods, have similar prices between these two groups of counties, but the price of land, which is immobile, is about 10 times higher in the 50 *km* counties. This difference generates large variation in the cost of producing land-intensive crops: for example, differences in land rents make the price of cattle products 3 times more expensive in the 50 *km* group. Travel distances generate much smaller differences in cost advantages: the counties that are 50 *km* away from state capitals have a cost advantage of only 7% over counties that are 200 *km* away.

Finally, denoting τ_{kiC} as the trade costs of i with its state capital, I decompose the variation in

Table 5: Evaluating the Causes of Agricultural Specialization

		Heterogeneity in			
	Baseline (1)	Natural Advantages (2)	Factor Intensity (3)	Trade Costs (4)	Trade Elasticity (5)
Dropping each Source of Heterogeneity					
- % of farms in crop	1	0.438	0.487	0.989	0.828
- Farm size per crop	0.548	0.161	0.289	0.548	0.544

Notes: The table shows the correlation between the baseline model and simulations when I shut down different sources of technological heterogeneity. Column 2 shows the correlation with a simulated economy where all counties with a positive production have the same average natural advantage. Columns 3 to 5 present results when I assume that all crops have the same technology in terms of factor intensity, trade costs or trade elasticity.

Table 6: Effects of Different Components on the Price of the Final Output Delivered at State Capitals

	Factor Prices			Distance to
	Land (1)	Intermediates (2)	Wages (3)	State Capital (4)
∼50 *km* to state capitals	1	1	1	1
∼200 *km* to state capitals	0.10	1.03	0.50	3.94
Effect of each component on the price delivered at State Capitals				
- Cattle	0.28	0.96	0.82	1.07
- Coffee	0.77	0.88	0.82	1.07
- Corn	0.47	0.93	0.87	1.12
- Cotton	0.90	0.97	0.97	1.07
- Fruits	0.90	0.92	0.91	1.15
- Rice	0.78	0.94	0.92	1.12
- Soy	0.24	0.98	0.94	1.12
- Sugarcane	0.83	0.91	0.88	1.15
- Tobacco	0.78	0.93	0.92	1.07
- Vegetables	0.85	0.86	0.74	1.15
- Wheat	0.23	0.98	0.94	1.12

Notes: The table compares the cost of producing agricultural goods and delivering them at State capitals for two groups: counties that are in approximately 50 *km* away from state capitals (40 to 60 *km*) and counties that are 200 \$km\$ away (190 to 210 *km*). The table multiplies the ratio of factor prices according to estimated factor shares and the difference in distance according to estimated trade costs.

the cost advantage of counties in serving their State Capitals using the following term

$$\Delta_{ikk'} = \underbrace{\log T_{ki} - \log T_{k'i}}_{\text{Natural Advantages}} + \underbrace{\theta_k \log c_{ki} - \theta_{k'} \log c_{k'i}}_{\text{Factor Prices}} + \underbrace{\theta_k \log \tau_{kiC} - \theta_{k'} \log \tau_{k'iC}}_{\text{Transportation Cost}}, \tag{19}$$

where $\Delta_{ikk'}$ captures the advantage of county i in the production of k relative to k' and $c_{ki} \equiv P_{ui}^{\alpha_k} r_i^{\gamma_k} w_i^{1-\alpha_k-\gamma_k}$. The variation in $\Delta_{ikk'}$ drives the comparative advantage of counties in the delivery of agricultural goods to state capitals. A Shapley decomposition of equation (19) shows that 59% of the variation in $\Delta_{ikk'}$ is driven by natural advantages, 29% is driven by variation in factor intensity term, and only 1% is driven by variation in trade costs.[35]

---

[35] The Shapley decomposition determines the marginal contribution of each component and can be used to decompose the contribution of each group of variables to the R^2 of a regression. I run regressions of $\Delta_{ikk'}$ on all the permutations of the components on the right hand side of equation 19 and I compute the contribution of each component to the R^2 across all these permutations. See Allen and Arkolakis (2014) for an application of the Shapley decomposition to studying the role of geography on economic activity.

7 Consequences of Agricultural Shocks

In this section, I study the general equilibrium effects of two major shocks in Brazilian agriculture: the adaptation of soybeans to the Brazilian savanna (*Cerrado)* at the end of the 1970s and the rise in the Chinese demand for commodities in the mid-1990s. I investigate the importance of general equilibrium effects in the evaluation of the returns to agricultural research, the impact of tropical soybeans on urbanization, and the gains from exports to China.

7.1 The Adaptation of Soybeans to the Brazilian Savanna, *Cerrado*

Context Soybeans originated in parts of Asia at a latitude of 40°N. Soy is a photoperiod-sensitive crop, which means that variations in the length of daylight exposure over the year govern the flowering period of the plant. A key problem for the application of traditional soybean cultivars in lower latitude regions is that due to the shorter daylight exposure during the summer in these regions, the plant flowers prematurely at a low height, resulting in low yields. In addition, acid soils with high aluminium concentrations and low calcium concentrations often prevail in tropical regions, which also limits the growth of the plant. Historically, these factors have kept soybeans from expanding in tropical regions. Nowadays, however, about 15% of the global production of soybeans comes from acid soils in the Brazilian savanna (*Cerrado*) at a tropical latitude of 15°S.

The breakthrough technology for the production of soybeans in tropical regions came in the 1970s with the development of soybean cultivars with late maturity, Al tolerance and low-Ca efficiency (Assuncao and Braganca, 2015; Freire de Sousa and Busch, 1998; Spehar, 1995). The introduction of cultivars carrying these treats in *Cerrado* was in large part the result of public investments in breeding programs from Embrapa, the Brazilian agricultural research department.[36] Figures 10 and 11 show a fast adoption of this technology: although in 1975, the production of soybeans in *Cerrado* was negligible, by 1985, more than 50% of the cropland in the region was already used in the production of soybeans. This expansion occurred despite a large reduction in the price of soybeans, indicating the existence of a positive supply shock in the production of this crop.[37] Also, in regions in the south of Brazil, there is a slight reduction in the share of land used

---

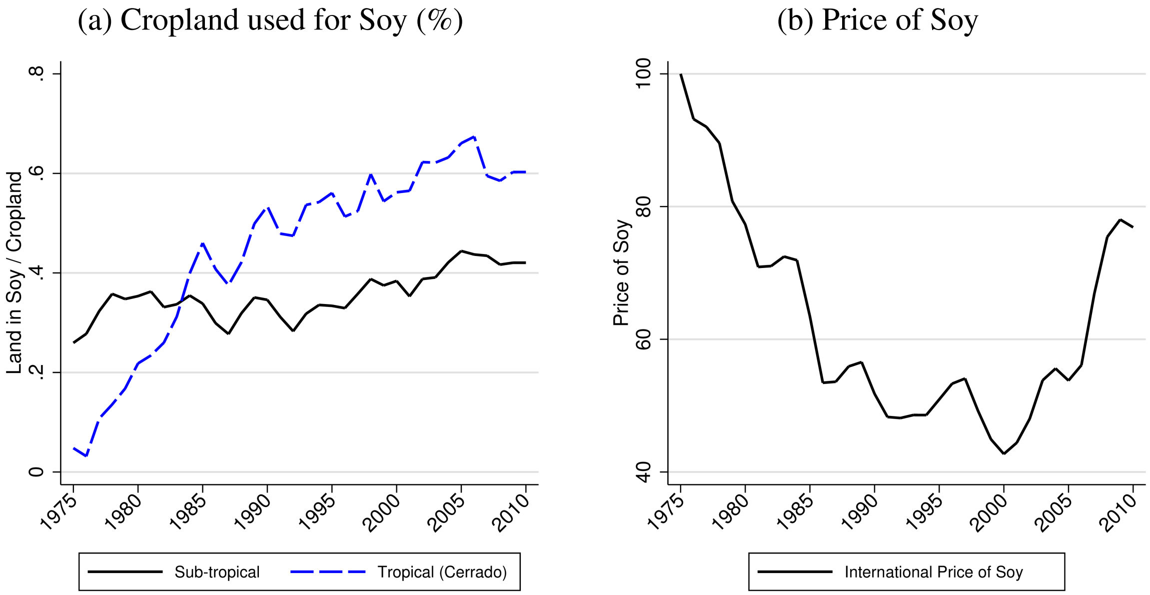
[36] As discussed in Pardey et al. (2003) and Pardey et al. (2006), there were collaborations between Embrapa and international agricultural research institutions. In the beginning of the 1950s, researchers from USDA developed less day-length sensitive soybeans in Illinois. During the 1960s and the 1970s Embrapa tested these varieties in collaboration with U.S. foreign programs. Later, when I measure the internal rate of return to investments in tropical soybeans, I focus on the domestic returns to soybeans. I do not account for the agricultural research costs for the USDA that preceded those incurred by Embrapa.

[37] During this period, the Brazilian government promoted a minimum price policy that generated a wedge between the domestic and the international prices of commodities. As discussed in Freire de Sousa and Busch (1998), however, this policy was typically not biding for soybean producers. Non-reported results show the same qualitative patterns for domestic prices constructed from information on revenues and quantities.

for soybeans between 1978 and 1985, which is consistent with the supply shock being specific to *Cerrado* and generating new competition for soybean producers in the south.[38]

To evaluate the introduction of tropical soybeans, I simulate a counterfactual economy in which I set the production of soybeans in counties above the Tropic of Capricorn to zero by assuming no natural advantage in the production of soybeans (that is $T_{ik} = 0$ if $k = soy$ and $i \in C_{tropics}$ where $C_{tropics}$ is the set of counties located above the Tropic of Capricorn), and I compare the outcomes of this counterfactual economy with the model calibrated for 2006. I present the results as changes from the counterfactual to the factual economy.

Figure 10: Expansion of Cropland used in the Production of Soy and the International Price of Soy


(a) Cropland used for Soy (%)

(b) Price of Soy

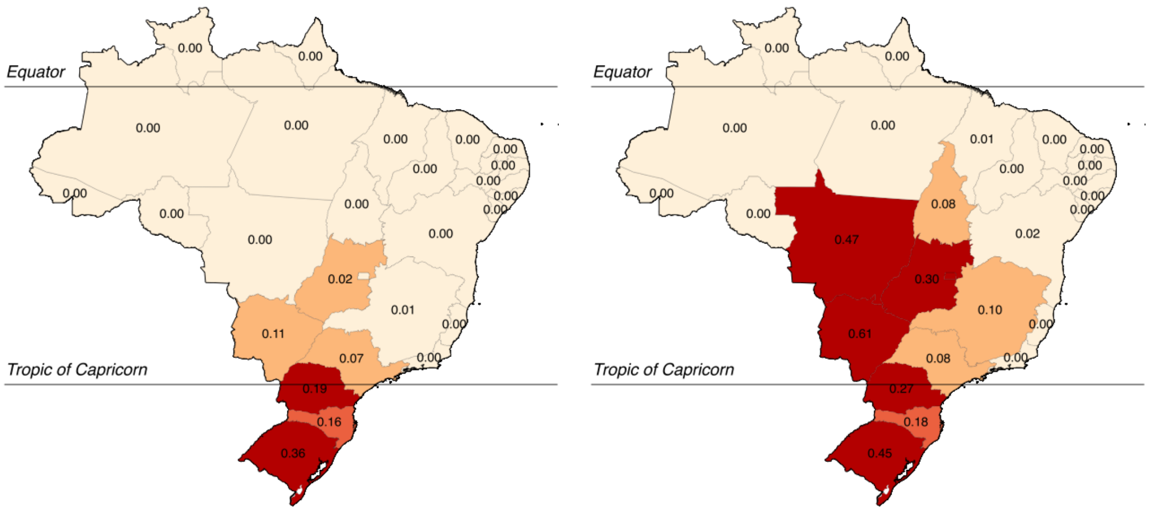
Notes: Panel (a) shows the share of cropland used in the production of soybeans in the Cerrado (latitude of ∼ 15°S) and in the south of Brazil (∼ 35°S). Panel (b) shows data from the World Bank on the real price of soybeans.

Effects per Activity Panel (a) in Figure 12 shows counties with a relative decrease in the production of soybeans in red and presents those with a relative increase in blue. The figure indicates a reallocation in the production of soybeans from the south to the north of Brazil. Table 7 provides details about sectoral changes. First, note that while tropical regions account for roughly 50% of the GDP of soybeans, the overall GDP of this sector increases by 40%. The aggregate effect for the economy is smaller because the average price of soybeans falls by 11%, inducing a reallocation of land and workers away from soybeans in southern counties. Even though total land used

---

[38] In the appendix, I show that productivity as captured by yields was stable in the south of Brazil between 1975 and 1985. Also, I show that there was no substantial increase in the use of land between these years.

Figure 11: Expansion of Soybean Production to the *Cerrado* between 1975 and 1985



(a) Share of Cropland for Soy 1975

(b) Share of Cropland for Soy 1985

Notes: Panel (a) shows the share of cropland in soybeans before the adaptation to the tropical area, and panel (b) shows that after the introduction of soybean cultivars that are adapted to the conditions of the Brazilian savanna (*Cerrado*).

in the production of soybeans increases by 90%, there is a reduction in the amount of workers employed in this sector. Since the relative price of land is low in *Cerrado* and production is less land-intensive, the reduction in the number of workers in soybeans in the south of Brazil is not sufficiently compensated by an increase in the number of workers in the north. For similar reasons, there is a 5-fold increase in the average size of farms producing soybeans.

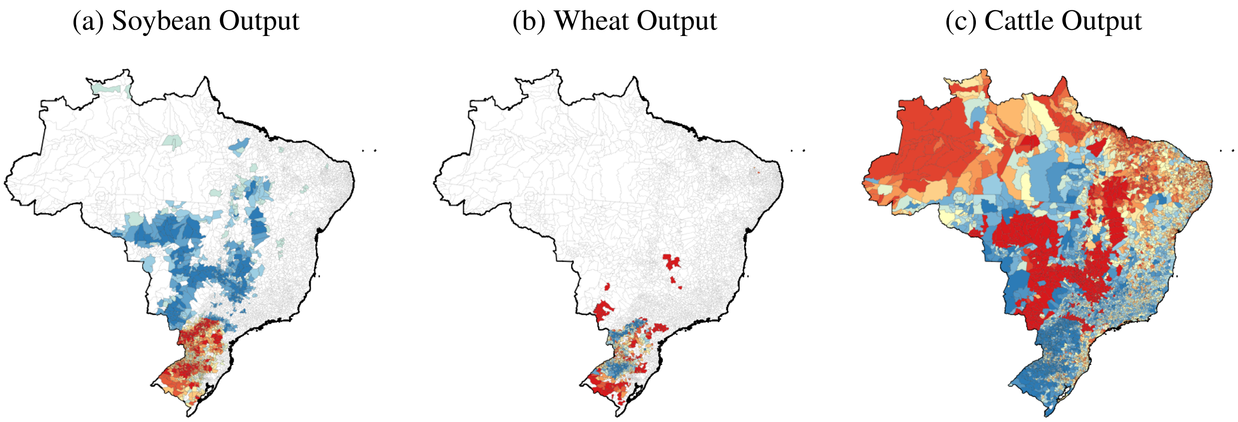
The relocation of soybeans to tropical regions releases the pressure on land markets in the south. The production of wheat, which originates in temperate regions and has not been as successfully adapted to tropical regions as soybeans, increases by 5.5%. However, in *Cerrado,* the pressure on land increases, generating a negative effect on the production of cattle and cotton, two important activities in tropical areas. These results show that the propagation of shocks to other crops is more pronounced in land-intensive ones, in particular those that compete geographically with soybeans for land.[39]

Effects by Region Table 8 presents effects according to three types of regions: (1) the tropical area to which soybeans expanded, which occurs mostly in *Cerrado*, (2) the sub-tropical region below the Tropic of Capricorn where soybeans were produced before the adaptation, and (3) the rest of

---

[39] The effect of soybeans in the *Cerrado* on other activities also depends on how productive the land is in the production of alternative crops. Non-reported results show that the natural advantage (T_{ki}) of the *Cerrado* in the production of other commodities is low relative to other parts of the economy, which is consistent with the fact that the region has an acid soil.

Figure 12: Effects of no Adaptation of Soybeans to *Cerrado* on Selected Activities


(a) Soybean Output (b) Wheat Output (c) Cattle Output

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 soybeans expand in the tropical regions.

the country where there is no production of soybeans.

In *Cerrado*, the production of soybeans increases the local GDP by 2.2%. The land employed in soybeans crowds out the production of labor-intensive crops, generating a reduction in the mass of agricultural workers and an out-migration from this region. In the sub-tropical region the effects go in the opposite direction: the production of soybeans and the regional GDP fall, but the mass of agricultural workers increases as well as due to an in-migration of workers from other parts of the country. In the tropical region not producing soybeans, there is also a small outflow of workers, but GDP increases.

The expansion of soybeans in *Cerrado* reduces the pressure on the local labor market, reducing wages in this region and increasing its comparative advantage in urban activities. Note that I still find an *aggregate negative* effect on urban workers despite the *local positive* effect on the employment of workers in the urban sector. The reallocation of land to the use of soybeans in *Cerrado* reduces the supply of domestic-oriented crops in Brazil, which is compensated by an increase in the production of these goods in both the domestic and external markets. Quantitatively, imports are small and most of this compensation effect comes from an increase in the domestic production.

Welfare Effects The aggregate welfare increases by 0.02% with the adaptation of soybeans to tropical regions. This is a substantial effect compared to other trade and productivity shocks found in the literature. For example, Caliendo et al. (2014) estimate a welfare gain of 0.2% per year from the productivity gain in computers and electronics in California between 2002 and 2007 and Caliendo and Parro (2014) found a welfare gain of 0.1% as a consequence of NAFTA's tariff

Table 7: Effect of Adaptation of Soybeans by Sector (in percentage change)

	GDP (1)	Export (2)	Imports (3)	Workers (4)	Land (5)	Farm Size (6)	Prices (7)
Agriculture	-4.06	1.51	0.09	1.32	0	0	-0.62
Urban	-0.08	-0.32	0.70	-0.23	0	0	0.01
Agricultural Sectors							
- Cattle	-7.90	-8.04	15.8	1.55	-10.9	-8.30	3.22
- Coffee	-1.19	-1.34	1.74	-0.90	-3.67	-6.37	0.32
- Corn	-2.67	-7.52	11.9	6.44	-12.3	-10.0	2.12
- Cotton	-2.17	-2.47	2.10	-0.18	-21.1	-23.7	0.41
- Fruits	-0.10	-0.36	0.65	-0.14	-5.11	-7.58	0.16
- Rice	-0.24	-2.36	2.42	0.62	-5.78	-9.48	0.48
- Soy	38.4	41.6	-45.2	-4.47	90	347	-10.6
- Sugarcane	-0.82	-0.68	1.40	-0.10	-6.22	-7.72	0.47
- Tobacco	-0.23	-0.21	0.42	-0.99	-0.64	0.14	0.02
- Vegetables	-0.08	-0.76	1.08	0.22	-6.26	-8.49	0.34
- Wheat	5.57	5.94	-0.26	11.1	-8.78	-2.79	-0.15

Notes: The table shows the effect of going from the counterfactual economy with no production of production of soybeans in the tropical region to the factual economy. Nominal GDP in column 1.

reductions. There are substantial distributional impacts across agents and regions. The aggregate welfare of landowners increases by 1.4%, but, in the sub-tropical region, landowners' welfare falls by 2.3%.[40]

Cost-Benefit Analysis Here, I evaluate the return to research on the introduction of soybeans in *Cerrado.* To do so, I close the books in 2006 and construct a cost and benefit series beginning in 1975. The benefit series is based on the results from the model. The Brazilian GDP in 2006 was R$2.4 trillion, and since preferences are homothetic, the model implies an annual monetary value of soybeans that is equivalent to the increase in real income. To account for the phase-in of benefits before 2006, I take a simple approach used in the literature: I assume a maturity period of 6 years for the benefits of research to start, and a linear phase-in of benefits thereafter.

To construct a cost series, I use information on the costs of Embrapa from 1975, when it was created, until 2006. As discussed in Pardey et al. (2006), it is important to account for the research costs of other domestic research institutions as well. Between 1975 and 1999, for example, 208 of 330 soybean varieties introduced in the market were not developed by Embrapa. I use the share

---

[40] The model accounts for the possibility of immiserizing growth (Bhagwati, 1958), where economic growth based on a heavily export-oriented sector leads to a decrease in the terms of trade of the country and makes the country worse off in the aggregate. Non-reported experiments with the parameters of the model show that this would be the case if Brazil were an even larger producer of soybeans in the international market.

Table 8: Effect of Adaptation of Soybeans to *Cerrado* by Region (in percentage changes)

	Real	Workers				
	GDP (1)	Ag (2)	Urban (3)	All (4)	Rents (5)	Wages (6)
Sub-Regions						
- Tropical *Cerrado*	2.19	-5.28	0.29	-0.36	30.5	-0.18
- Sub-tropical region	-0.74	2.25	-0.48	0.30	-2.90	0.26
- Tropical non-*Cerrado*	0.01	1.52	-0.25	-0.02	2.45	0.12
All Regions	0.30	1.32	-0.23	0	5.14	0.08

Notes: The table shows the results of the effect of the introduction of soybeans in different regions on selected variables. *Cerrado* is the region where soybeans expand, Sub-tropical is the region where soybeans can be produced without the new technology, and tropical non-*Cerrado* is the tropical region where soybeans are not produced in either the counterfactual or the factual economy. Column 1 is GDP divided by the price index in each county.

of Embrapa's attribution given by Pardey et al. (2006), who collect information from datasets and interviews to construct their cost series.[41]

Table 9: Estimates of the Internal Rate of Return of Agricultural Research on the Adaptation of Soybeans to Tropical Regions

	Different Measures			No Soybeans Exports	
	All Regions (1)	Only Soybeans (2)	Only *Cerrado* (3)	To China (4)	To RoW (5)
IRR	42.1	63.6	49.1	36.3	30.2
Benefit/Cost	7.84	20.8	11.0	5.78	4.07

Notes: The table shows the estimated returns to agricultural research on the adaptation of soybeans. Columns 1 to 3 use different measures of benefits based on a counterfactual with no production of soy in the tropics. Columns 4 shows the benefit for an economy with no exports of soybeans to China and column 5 with no exports of soybeans to the rest of the world.

Table 9 presents estimates of the annualized internal rate of return (IRR) to research on soybeans. This measure is the interest rate that would make the net present value of all the costs and benefits of soybeans between 1975 and 2006 equal to zero. Table 9 also reports estimates of the benefit cost ratio using an interest rate of 10% to calculate the present value benefits and costs.

---

[41] The data that I use to measure the costs of Embrapa come from CNPSo, the soybean research department. According to Pardey et al. (2006), when we account for the costs incurred by Embrapa headquarters and CENARGEN, Embrapa's biotechnology center, the budget cost of CNPQSo should be augmented by 35%. In addition, the research costs of Embrapa are then estimated to be 40% of the total costs of research once we include other domestic and foreign institutions. Therefore, I augment the observed costs of CNPSo by 35% and then by 150%.

Column 1 shows that the IRR is 42% if we measure the aggregate gains and losses. While high, this return is at the low end in the distribution of IRR found in the literature (Alston et al., 2000). Accounting only for the gains in the soybean sector, which is a more common approach, the return is 64%. If we only measure the local gains in *Cerrado*, the estimated effect would be 49%, because it does not account for the indirect output loss in other regions of Brazil.[42]

Finally, I simulate counterfactuals to investigate the complementarities between trade activities and the investments in agricultural research. Trade substantially enhances the returns from the development of new varieties of soybeans: the IRR is approximately 15% lower if Brazil does not to export soybeans to China and 30% lower if Brazil does not export any soybeans to external markets.[43]

Reduced Form Impact To close my analysis of the impact of tropical soybeans, I bring data from the agricultural census of 1975 to compare the reduced form impact of soybeans observed in the data with the impact obtained from the simulations.[44] The goal of this exercise is to verify whether the model generates correlations that are consistent with those observed in the data once we control for covariates that capture the influence of non-modeled shocks in the model. Since the data from 1975 were not used in the estimation, I interpret this exercise as an out-of-sample test of the model. To estimate the reduced form impact of soybeans, I use the following equation:

$$y_{it} = \beta_i + \beta_t + \beta^{Soy} L_{it}^{Soy} + \boldsymbol{X}_{it} \boldsymbol{\gamma} + \varepsilon_{it} \tag{20}$$

where y_{it} is different dependent variables in county i in year t; β_i and β_t are county and period fixed effects; L_{it}^{Soy} is the share of land in county i used in the production of soybeans; $\boldsymbol{X}_{it}$ is a set of covariates; and ε_{it} is the error term. I control for initial conditions related urban workers, population, and for total land used in agriculture before and after the adoption of soybeans since I do not model changes in the supply of land.[45]

Table 10 shows that the expansion of soybeans is negatively associated with the mass of agricultural workers per hectare, positively correlated with the share of workers in urban activities, and negatively correlated with the total population such that the expansion of soybeans generates out-migration. The coefficients are consistent in direction and magnitude with those found using the simulated data, which indicates that the model captures the economic forces changing the patterns of reallocation of workers and land during this period. Furthermore, this exercise shows that the

---

[42] Note that the effect for Brazil understates the aggregate benefits since other countries are benefited by the production of soybeans.

[43] To do so, I subtract from the total world expenditure the consumption from China ($E_{kF} - E_{kCHINA}$).

[44] I use a mapping of counties provided by *Instituto de Pesquisa Economica e Aplicada*a, IPEA, to construct county boundaries that are consistent over time.

[45] Since I have only two periods, I estimate equation 20 in first-difference, which provides the same result.

local positive effect of soybeans on the allocation of workers in the urban sector is quantitatively compatible with general equilibrium forces, resulting in a negative impact in the aggregate.

Table 10: Reduced Form Estimates of the Impact of Soybeans in the Data and in the Simulations

	Dependent Variables					
	Worker per Hectare		Urban Share		Population	
	Data (1)	Model (2)	Data (3)	Model (4)	Data (5)	Model (6)
Δ Soy Area Share	-0.360	-0.323	0.122	0.117	-0.090	-0.034
	(0.113)	(0.007)	(0.043)	(0.007)	(0.102)	(0.002)
N	3748	3748	3748	3748	3748	3748
R^2	0.51	0.92	0.73	0.34	0.10	0.56

Notes: All regressions control for initial conditions in 1975 in terms of log of total urban population, log of total population, and for log of total land use in 1975 and 2005. Dependent variables are in logs. Robust standard errors clustered at the county level in parenthesis. Information from IPEA is used to construct county boundaries that are consistent over time.

7.2 The Rise in the Chinese Demand for Commodities

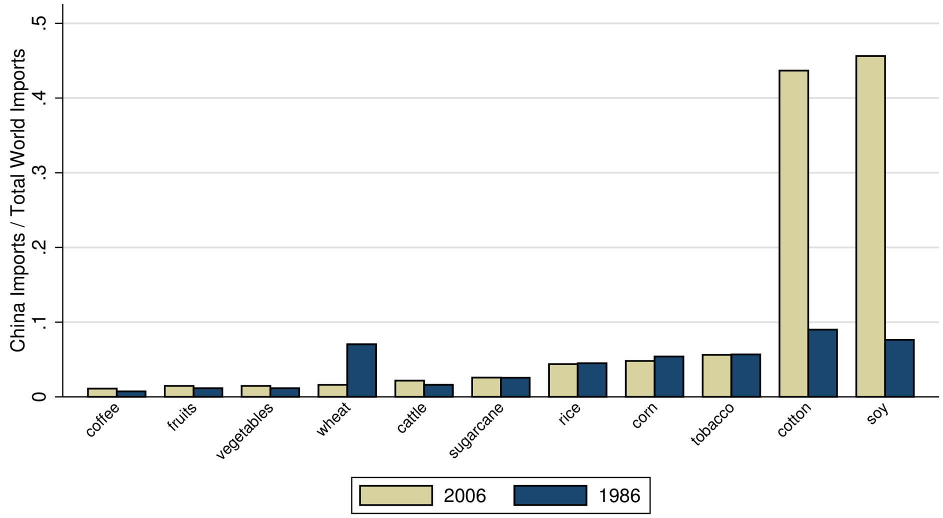
Context The rise of the Chinese economy after the 1990s had a dramatic effect on the international demand for commodities. This phenomenon was particularly important for the recent increase in commodity exports of Latin American and African countries (Gollin et al., 2016; Jenkins et al., 2008). The demand shock from China, however, was far from uniform across agricultural commodities: for example, the share of global imports of soybeans attributable to China went from less than 1% to 40% between 1990 and 2006, but for refined sugar this share remained between 1 and 2% during this period.[46] Figure 13 shows data from the FAO on the share of global imports in each sector that are attributable to China in 2006. China accounts for more than 40% of the global imports of soybeans and cotton, but a much smaller share of imports for all the other agricultural goods.

To study the propagation of this demand shock, I compute a counterfactual economy where I remove the demand from China from the global demand ($E_{kF} - E_{kCHINA}$). For simplicity, I abstract from general equilibrium effects that China might have generated on other countries. As in the analysis of the adaptation of soybeans to the tropics, I present results as changes from the counterfactual economy to the factual economy.

---

[46] More recently, there was an increase in the imports of particular commodities. For example, China became a larger importer of corn after 2010.

Figure 13: Share of Global Imports Attributed to China



Notes: The figure shows the share of total imports in each agricultural sector that are attributed to China according to trade data from FAO.

Effects per Sector Table 7 shows that agricultural exports increase by 4.5% and the urban sector by 1.5%. Note that, even though the Chinese demand shock is positive for every commodity, GDP and exports fall for many of them. The negative effect is large for land-intensive activities such as cattle and wheat. Columns 4 and 5 show that this effect is due to a large concentration of land and workers in the production of cotton and soybeans. Compensating for the negative effect that a reduction in the domestic production causes on the absorption of many agricultural goods, there is a substantial increase in imports from external markets in activities that are important for domestic consumption: for example, Brazilian imports of meat and corn both increase by more than 12% after the demand shock from China.

Effects by Region Table 12 presents the effects of China according to each quartile in the distribution of increase in exports per county. In the first quartile, exports increase by 8.5%, whereas in the bottom quartile, exports actually fall. Changes in the relative price o factors across space change the spatial patterns of comparative advantage, which induces some regions to specialize in the production of domestic oriented crops. Even though a restricted part of Brazil is directly benefited by trade opportunities with China, the benefits from exports propagate to other regions of the economy. In the bottom quartile, for example, exports fall but GDP still increases.

Welfare Effects The rise of China generates an aggregate welfare gain of 0.30%. Since the demand shock is largely concentrated in land-intensive activities, this effect is larger for landowners, who

Table 11: Effect of Exports to China per Sector (in percentage change)

	GDP (1)	Export (2)	Imports (3)	Workers (4)	Land (5)	Farm Size (6)	Prices (7)
Agriculture	-2.03	4.26	8.17	0.63	0	0	1.77
Urban	0.42	1.34	3.14	-0.11	0	0	0.26
Agricultural Sectors							
- Cattle	-5.20	-5.17	13.7	-2.74	-4.80	-3.05	2.60
- Coffee	-1.76	-1.93	4.61	-1.79	-3.57	-3.10	0.71
- Corn	-1.75	-5.07	12.9	2.23	-1.39	-3.40	2.09
- Cotton	35.9	72.2	3.23	38.9	28.7	-6.51	0.46
- Fruits	0.34	0.37	2.27	-0.08	-2.55	-2.92	0.44
- Rice	0.37	0.87	4.32	0.35	-2.35	-3.46	0.64
- Soy	23.8	46.7	49.1	35.6	21.6	-6.97	7.41
- Sugarcane	0.81	1.60	2.87	0.62	-2.08	-2.97	0.66
- Tobacco	1.46	2.14	4.16	1.16	-3.13	-4.59	0.68
- Vegetables	0.23	-0.32	2.90	0.11	-2.93	-3.04	0.71
- Wheat	-30.5	-32.2	9.58	-28.2	-33.7	-7.42	1.65

Notes: The table shows the effects of going from a counterfactual economy with no exports to China to the factual economy in different sector for selected variables.

obtain a welfare gain of 4.6%.

8 Conclusion

In this paper, I studied causes of the spatial differences in agricultural activity, and how these differences shape the aggregate impact of agricultural shocks. To do so, I formulated a quantitative general equilibrium model that accounts for rich spatial patterns of agricultural production and estimated the model using comprehensive county-level data from Brazil. The first finding of this paper is that differences in natural advantages across regions and differences in factor intensity across crops are key for the spatial organization of agriculture, but that differences in trade costs across crops play a minor role. The second finding is that general equilibrium effects can be crucial for our understanding of agricultural shocks. In particular, I show that general equilibrium effects substantially shaped the returns to agricultural research, the impact of agricultural shocks on urbanization, and the gains from trade.

Many agricultural policies and technologies, especially those that are transformative of an economy, generate effects that propagate across the economy due to trade activities and migration. The results indicate that measuring these effects is important. The framework developed in this paper is flexible and, in principle, could be adapted for the evaluation of general equilibrium

Table 12: Effect of Exports to China per Region (in percentage changes)

		Real	Workers			Land	
	Exports	GDP	Ag	Urban	All	Rents	Wages
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Sub-Regions in terms of Increase in Exports							
- 1th Quartile	8.73	2.25	0.49	-0.33	-0.02	9.37	0.85
- 2nd Quartile	1.32	0.21	0.82	-0.05	-0.02	2.59	0.82
- 3rd Quartile	0.64	0.46	0.93	-0.34	0.02	2.15	0.86
- 4th Quartile	-0.27	0.71	0.78	-0.65	57.1	2.08	0.90
All Regions	2.66	0.63	0.72	-0.12	0	5.93	0.85

Notes: The table shows results of exports to China on selected variables. Results are disaggregated according to four quartiles relative to changes in exports to investigate the heterogeneity of the impact across regions. Column 2 is GDP divided by the price index in each county.

effects in other contexts. For example, future work could extend the model in this paper to study the use of natural resources and the effect of migration policies on agricultural activity.

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.

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.

Anderson, J. E. and van Wincoop, E. (2004). Trade costs. *Journal of Economic Literature*, 42(3):691–751.

Assuncao, J. and Braganca, A. (2015). Technological change and deforestation: Evidence from the brazilian soybean revolution. *Working Paper*.

Avila, A. F. D. and Evenson, R. E. (2010). Total factor productivity growth in agriculture: the role of technological capital. *Handbook of agricultural economics*, 4:3769–3822.

Baum-Snow, N., Henderson, V., Turner, M. A., Zhang, Q., and Brandt, L. (2016). Highways, market access, and urban growth in china. *Working Paper*.

Bhagwati, J. (1958). Immiserizing growth: a geometrical note. *The Review of Economic Studies*, 25(3):201–205.

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.

Caliendo, L., Parro, F., Rossi-Hansberg, E., and Sarte, P.-D. (2014). The impact of regional and sectoral productivity changes on the us economy. Technical report, National Bureau of Economic Research.

Costinot, A. (2009). On the origins of comparative advantage. *Journal of International Economics*, 77(2):255–264.

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.

Deakle, R., Eaton, J., and Korum, S. (2008). Global rebalancing with gravity: Measuring the burden of adjustment. *IMF Staff Papers*.

Desmet, K., Nagy, K. D., and Rossi-Hansberg, E. (2015). The geography of development: Evaluating migration restrictions and coastal flooding. *Working Paper*.

Diamond, R. (2016). The determinants and welfare implications of us workers' diverging location choices by skill: 1980–2000. *The American Economic Review*, 106(3):479–524.

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.

Fiszbein, M. (2015). Agricultural diversity, structural change and long-run development: Evidence from us counties. *Working Paper*.

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. *Econojmic Growth Center, Yale University New Haven CT*.

Freire de Sousa, I. S. and Busch, L. (1998). Networks and agricultural development: The case of soybean production and consumption in brazil. *Rural Sociology*, 63(3):349.

Fujita, M., Krugman, P. R., and Venables, A. (2001). *The spatial economy: Cities, regions, and international trade*. MIT press.

Gasparri, N. I., Kuemmerle, T., Meyfroidt, P., Waroux, Y., and Kreft, H. (2016). The emerging soybean production frontier in southern africa: Conservation challenges and the role of south-south telecouplings. *Conservation Letters*, 9(1):21–31.

Gollin, D., Jedwab, R., and Vollrath, D. (2016). Urbanization with and without industrialization. *Journal of Economic Growth*, 21(1):35–70.

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.

Hornbeck, R. and Keskin, P. (2014). The historically evolving impact of the ogallala aquifer: Agricultural adaptation to groundwater and drought. *American Economic Journal: Applied Economics*, 6(1):190–219.

Hornbeck, R. and Naidu, S. (2013). When the levee breaks: Labor mobility and economic development in the american south. *The American Economic Review*.

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.

Levchenko, A. A. (2007). Institutional quality and international trade. *The Review of Economic Studies*, 74(3):791–819.

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

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.

Nunn, N. (2007). Relationship-specificity, incomplete contracts, and the pattern of trade. *The Quarterly Journal of Economics*, 122(2):569–600.

Oliveira, O. and Morten, M. (2014). Migration, roads and labor market integration: Evidence from a planned capital city. *Working Paper*.

Pardey, P. G., Alston, J. M., Chan-Kang, C., Magalhães, E. C., Vosti, S. A., et al. (2003). *Assessing and attributing the benefits from varietal improvement research: Evidence from Embrapa, Brazil*. Centre for International Economic Studies.

Pardey, P. G., Alston, J. M., Chan-Kang, C., MagalhÃ£es, E. C., and Vosti, S. A. (2006). International and institutional r&d spillovers: Attribution of benefits among sources for brazil's new crop varieties. *American Journal of Agricultural Economics*, 88(1):104–123.

Pardey, P. G., Chan-Kang, C., Dehmer, S. P., and Beddow, J. M. (2016). Agricultural r&d is on the move. *Nature*, 537:301–303.

Ramondo, N., Rodríguez-Clare, A., and Saborío-Rodríguez, M. (2016). Trade, domestic frictions, and scale effects. *The American Economic Review*, 106(10):3159–3184.

Samuelson, P. A. (1983). Thünen at two hundred. *Journal of Economic Literature*, pages 1468–1488.

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.

Suarez Serrato, J. C. and Zidar, O. (2016). Who benefits from state corporate tax cuts? a local labor markets approach with heterogeneous firms. *American Economic Review*, 106(9):2582–2624.

Tombe, T. (2015). The missing food problem: Trade, agriculture, and international productivity differences. *American Economic Journal: Macroeconomics*, 7(3):226–258.

Tombe, T. and Zhu, X. (2016). Trade, migration, and productivity: A quantitative analysis of china. *Working Paper*.