Migration, Specialization, and Trade: Evidence from the Brazilian March to the West

Heitor S. Pellegrina
NYU Abu Dhabi

Sebastian Sotelo
University of Michigan*

June 2019

Please click here for most recent version

Abstract

We study how internal migration shapes aggregate and regional comparative advantage. Using Brazilian census data, we document that farmers emigrating from regions with high employment in a given crop are more likely to grow that crop and earn higher incomes than other farmers doing so. We also show that regional production of a crop increases with immigration from regions employing many farmers in that crop, after controlling for total crop employment. We incorporate these facts into a quantitative, dynamic model of trade and migration in which a region's specialization is determined by natural advantage, labor availability, and the knowledge of the labor force. Applying our model to the large migration of agricultural workers to the west of Brazil in the second half of the 20th century, we find that the reallocation of workers reshaped Brazil's export profile, contributing to its rise as a leading commodity exporter, and that a substantial part of this change is due to the knowledge carried by migrants. Since the change in comparative advantage involved primarily the agricultural sector, aggregate gains from trade were little affected by migration.

Keywords: International Trade, Migration, Comparative Advantage

*We thank Jonathan Eaton and Andrei Levchenko for helpful comments. We also thank participants in seminars at Purdue Trade Brown Bag, University of Michigan, NYU Abu Dhabi, St. Louis Fed, IADB, TIGN-LACEA, Middlebury, RMET and UEA. Natalia Martinez Camelo provided excellent research assistance. E-mail: ssotelo@umich.edu and heitor.pellegrina@nyu.edu. All errors are our own.

1 Introduction

A central task in international trade, and in spatial economics more generally, is to understand the impact of trade on welfare and the patterns of specialization across locations. Understanding the impact of trade, moreover, goes beyond purely academic interest, as policymakers often attempt to influence their own countries' export and import activity, against the backdrop of increasing globalization. Seeking to quantify the impact of trade, a recent literature has incorporated comparative advantage—the notion that differences in relative costs across locations drive trade—into quantitative models and established it as a major determinant of trade flows.[1] But while a lot of progress has been made in measuring how comparative advantage shapes the impact of trade, there has been comparatively less progress in quantifying the sources of comparative advantage itself.[2]

In this paper, we start by noting that large migrations within countries are not uncommon, and that they are often accompanied by shifts in specialization and trade patterns. Consider, for example, the U.S. westward expansion and, more recently, the large migration of Chinese workers to export clusters. The question we ask is: Does the migration of workers within a country shape regional and aggregate comparative advantage? We consider two main mechanisms. First, migration determines the relative abundance of land and labor across regions and improves the match between regional productivity and labor allocation. Second, migrants have knowledge related to specific goods and, when new opportunities to deploy their knowledge emerge in a region, they migrate to take advantage of them.[3]

To quantify the impact of migration on comparative advantage, we extend a dynamic model of trade and migration to incorporate these two mechanisms through which migration shapes comparative advantage. In our model, workers are endowed with crop-specific knowledge, which they acquire through exposure to economic activity in their region of origin. Workers carry this knowledge along with them as they migrate, optimally choosing which crop to grow and where to produce, based on which region is better suited to their knowledge. In equilibrium, a regional and aggregate comparative advantage in a crop reflects a

[1] Examples of this literature are discussed in Eaton and Kortum (2012), Costinot and Rodriguez-Clare (2015).

[2] A recent line of work studies knowledge creation and diffusion in open economies, thereby exploring the origins of comparative advantage (see e.g., Lind and Ramondo, 2018).

[3] For example, Sabel, Fernandez-Arias, Hausmann, Rodriguez-Clare, and Stein (2012) describe how migrants have used their knowledge to form new export sectors in Latin America. Opala (1987) document that plantation owners in South Carolina and Georgia were willing to pay higher prices for slaves from this Sierra Leone and Liberia, denominated the "Rice Coast", due to their knowledge about the production of rice. Other historical examples include the diffusion of crops during the Columbian Exchange (Crosby, 1973), the introduction of new varieties of wheat in the northeast of the US in the 19th century (Olmstead, Rhode, et al., 2008), the introduction of wheat in North Africa during the diffusion of Islam (Watson, 1983), and the production of flowers by Dutch refugees in England in the late 16th century (Scoville, 1951).

combination of natural advantage (such as land quality), the abundance of labor, and the knowledge of the labor force.

We quantify the forces in our framework by applying it to Brazil, a country that provides us with a rich setting to examine our research question. During the XX century, Brazil experienced a large domestic migration, the "March to the West", that coincided with a transformation into an important global exporter of crops such as soy, corn, and livestock. Following a series of public initiatives to integrate the country's West to cities in the East during the 1950s, farmers from all parts of Brazil migrated to the West and, as we argue in this paper, brought along with them the agricultural knowledge from their region of origin. As a result, from 1950 to 2010, about 7 million people from the East migrated to the West of Brazil, raising the share of the Brazilian population living in the West from 5 to 15 percent.[4]

To capture the empirical relationship between migration, specialization, and trade, we assemble a detailed data set, composed of several waves of the Brazilian demographic and agricultural census. We use these data to establish a set of facts that guide our modeling approach. First, we document the transformation of Brazil's aggregate and regional agricultural landscape. We show that Brazil's exporting pattern shifted dramatically during this period: Since the 1980s, the country as a whole developed a comparative advantage in crops such as soy, corn and livestock, reflecting in large part the intensity with which the West specializes in these crops.

We also document the influence of a migrant's origin on her cropping choices and earnings. To do so, we compare farmers who, upon migrating, grow the same crop in the same destination, but who emigrate from different regions. First, we find that a 1 percent increase in the number of farmers growing a crop in the origin region is associated with a 0.16 percent increase in the number of emigrants from that region producing that crop. Second, a 1 percent increase in the number of workers growing a crop in the origin region is associated with a 0.12 percent increase in the earnings of that region's emigrants. Additionally, we document that, for a given crop, regional output and revenues increase when the regional mix of workers favors migrants from regions that employ more workers in that crop (after controlling for the total number of farmers). These observations are consistent with the idea that a farmer's good-specific knowledge is shaped by her origin region's specialization, and with the farmer carrying her knowledge as she migrates.

Motivated by these facts, we embed the allocation of workers and their knowledge as sources of comparative advantage in a dynamic quantitative model of migration and trade

[4]The magnitude of Brazilian march to the west is comparable to the US Great Migration during the 20th century, a period in which roughly 6 million Americans emmigrated from the south of the US. See Bazzi, Gaduh, Rothenberg, and Wong (2016) for an analysis of the Transmigration Program in Indonesia, which involved 2 million migrants.

(Caliendo, Dvorkin, and Parro, 2015; Allen and Donaldson, 2018). In our model, a worker born in a region has a vector of productivities for growing individual crops, which is a function of that region's employment in each crop. Given his crop-specific knowledge, the worker then decides where to work and which crop to grow. In equilibrium, migration determines national and regional comparative advantage in two complementary ways. First, relative productivity reflects a combination of natural advantage (tied to the characteristics of land) and the origins of farmers in that region. Second, migration allows labor-intensive crops to expand in areas that receive large migrant inflows and also allows for more productive regions to receive labor.

We next bring our model to the data. We show that the two key parameters of the model, (i) the elasticity of a farmer's productivity to crop employment in her region of origin and (ii) the elasticity of occupational choice to income can be transparently identified from reduced form elasticities. These two parameters govern migrants' responsiveness to new opportunities and also the extent to which they are capable of shaping aggregate outcomes. We calibrate the rest of parameters by combining the model with data on cropping patterns, employment and exports.

To understand the impact of migration on specialization and trade, we then conduct a policy experiment in which we ban East-West migration (but allow for migration within these regions) starting in 1980, when the March was already underway. Our first finding is that migration accounts for a sizable part of observed changes in comparative advantage. In particular, banning migration reduces agricultural RCA by approximately 20%, compared to the data. Moving on to specific commodities, we find that, relative to the data, banning East-West migration substantially weakens Brazil's comparative advantage in soy and livestock, new flagship commodities in Brazilian exports in the past decades. In the case of soy, the model suggests that approximately one third of this drop is due to the spatial reallocation of knowledge, while the rest is due to changes in raw labor allocations across regions. On the other hand, banning migration strengthens Brazil's RCA in traditional commodities, such as coffee.

Despite the effects of migration on international specialization within Brazil, its effects on the gains from trade are relatively small. In our baseline calibration, the gains from trade are about 8.1 percent, and the gains from agricultural trade are about half of that. These gains, however, are relatively insensitive to banning migration. The reason is that Brazil's gains from trade in this period are determined in large part by the exchange of agricultural goods for non-agricultural goods. While migration entailed large swings in specialization within the agricultural sector, on aggregate these reallocations allowed for a small increase in import shares of non-agricultural goods. The gains from allowing for internal migration, in turn, are about half as large as the gains from trade.

Our paper is mainly related to three strands of literature. The first is a recent literature that quantitatively studies the origins and evolution of Ricardian comparative advantage. (2016) and Hanson, Lind, and Muendler (2015) document substantial changes in comparative advantage over time and cross countries. Buera and Oberfield (2016) and Jie Cai and Maria (2019), among others, study the diffusion of ideas in an open economy and how it drives trade across countries (see Lind and Ramondo (2018) for a summary of this literature). Arkolakis, Lee, and Peters (2018), in particular, study the impact of migrants on the technological frontier in the United States in the 19th century. More broadly, we relate to a large quantitative literature at the intersection of international trade and economic geography, which includes Allen and Arkolakis (2014), Redding (2016), Bryan and Morten (2015), Tombe and Zhu (2019) and Morten and Oliveira (2016). This literature has applied quantitative frameworks to study the impact of infrastructure development and foreign trade shocks allowing for rich movements of people and goods. To the best of our knowledge, this literature has not examined how migration shapes aggregate and regional comparative advantage, which we do building on the recent dynamic approaches of Caliendo, Dvorkin, and Parro (2015) and Allen and Donaldson (2018).

Second, we also relate to a growing literature quantitatively examining the determinants and implications of trade in agriculture, including Costinot and Donaldson (2014), Costinot, Donaldson, and Smith (2016), Fajgelbaum and Redding (2014), Allen and Atkin (2016), Pellegrina (2018), Porteous (Forthcoming), Tombe (2015) and Sotelo (2018). Most of this literature treats comparative advantage as exogenous – either arising from quality of land, factor proportions, or both – but does not study the origins of comparative advantage or the factors shaping its evolution, which is the focus of our paper.

Finally, we relate to a literature that studies the transferability of worker's skills over space and their implications for individual productivity. Bazzi, Gaduh, Rothenberg, and Wong (2016) show that human capital depends on migrants' origins, and that it conditions how a migrant performs in a destination. In the agricultural context,Olmstead and Rhode (2011) have documented the role of geography and migration in the expansion of different crops in the US. Recently, Bahar and Rapoport (2016) provide evidence that international migration relates to comparative advantage across countries. We contribute to this literature by providing a new set of measurements of migration-led knowledge diffusion and, more importantly, embedding this mechanism into a quantitative general equilibrium model to evaluate its aggregate implications.

The rest of the paper is organized as follows. Section 2 describes the recent migration of workers to the West of Brazil. Section 3 describes our data and documents the facts that form the empirical basis of the paper. Section 4 introduces our model of trade and the spatial reallocation of knowledge due to migration. Section 5 and 6 uses the model to quantify the strength of our mechanism. Section 7 concludes the paper.

2 The March to the West

The West of Brazil is nowadays one of the World's major agricultural powerhouses. If the region were a country, it would be among the 15th largest agricultural exporters in the world.[5] This status, however, came rather recently. Despite the fact that the region accounts for 60% of Brazil's territory, in 1950, less than 5% of Brazil's value added in agriculture came from the West and 95% of the population lived in the East. This geographic concentration reflected the historical development of the Brazilian economy: with the exception of the gold extraction in the interior of Brazil during the 18th century and the exploitation of rubber in the Amazon forest in the late 19th century, the Brazilian economy was largely based on export-oriented crops such as sugarcane, coffee and cotton that required access to ports located along the Atlantic coast in the East of Brazil.

The rise of the West began in the 1950s, when urbanization and demographic transition took off in Brazil. Concerned with the population pressure in the urban centers of the southeast, the president at the time, Getulio Vargas, initiated a large-scale project to promote the migration of families to the Central-West. He named the project "March to the West" and, as stated by the government propaganda, the goal was to construct a nation that was free from the "vices of the coast". The project consisted of land grants for the formation of new agricultural colonies in the Central-West and investments in roads to connect them to the rest of the country. Economic growth allowed successive governments to expand the investments in this project. In 1960, the president Juscelino Kubitschek moved the capital of Brazil to a newly constructed city in the Central-West, Brasília. Between the 1960s and the 1980s, the military dictatorship expanded these highways to integrate the Amazon region in the North. The infrastructure investments in the West lost momentum in the mid-1980s, when Brazil entered a decade long period of economic depression and hyperinflation.

The "March to the West" deeply transformed the spatial organization of the Brazilian economy. As shown in Figure 2, the share of the Brazilian population in the West rose from 5 percent in 1950 to 15 percent in 2010 and the share of the agricultural value added coming from the West increased from 5 to almost 30 percent. The composition of migrants among

[5] The Brazilian States are officially divided in five broad regions based on socio-economic and geographic features: Central-West, North, Northeast, Southeast and South. Our analysis focuses on the occupation of the Central-West and the North, shown in Figure 1. For simplicity, we label "West" the broader region comprising the Central-West and the North and "East" the rest of Brazil.

agricultural workers in the West mirrors the region's increasing prominence. In 2010, the overall share of migrants in the west was 15 percent, with some regions having more than 40 percent of migrants (see Figure 3, Panel a). Importantly, migrants came from all parts of the East. Ten percent of migrants in the West came from Bahia, a state in the northeast that is a large producer of cocoa, whereas 13 percent of them came from Rio Grande do Sul in the farther south of Brazil in the frontier with Argentina, a region that specializes in the production of soybeans and cattle (See Appendix Figure 8.)

2.1 Crop diffusion during the "March to the West"

As in other historical episodes of crop diffusion, researchers have underscored the importance of migrants' knowledge in the expansion of the western agricultural frontier of Brazil. The following passage from Sabel, Fernandez-Arias, Hausmann, Rodriguez-Clare, and Stein (2012), p.181, for example, highlights the importance of migrants from the south of Brazil - called *gauchos*-, in the expansion of soybeans in the West:

> *The first movers had some experience with these crops in the southern part of Brazil, a region with a favorable climate and adequate conditions for soybean agriculture[...] Such experience and technical capabilities allowed them to experiment with soybean cultivation in other regions of the country at a time when international markets started to demand higher volumes of soybeans. The gaúchos also had experience with and knowledge about distribution channels for the product, since soybeans had already been sold in foreign markets using international trading companies, cooperatives, and national processors.*

Corroborating this passage, Figure 4 shows the pattern of expansion of soybean for each decade between 1970 and 2010. The figure shows that soybean production radiated from the southernmost part of Brazil – the historical origin of soybean production – towards the center of the country, first, and then to the northwest and even the northeast. In contrast, the same Figure shows that aggregate migration proceeded from many regions in the East. Figure 4 shows that these radiating patterns – which again, contrast with those of aggregate migration – also describe the expansion of other crops, such as coffee.

While productivity shocks could explain part of the patterns in Figure 4, such shocks, by themselves, are unlikely to explain why the expansion of specific crops is also associated with the origin of migrants. In fact, Figure 3, Panel (b) shows that migrants account for specialization patterns in the West with a simple accounting exercise. For each crop, it measures the decrease in the share of farmers producing that crop in the West, if one removes migrants originating in the top 10% of East regions in terms of employment. In the next section, moreover, we show that migrants coming regions where a crop is intensively grown are more likely to grow that crop upon migration. This link between the expansion of a crop and the presence of farmers from regions specialized in that same crop suggests a complementarity between the crop-specific knowledge of farmers and land quality. Unlike our theory, this complementarity would not be present in a theory that explained crop expansion patterns based solely on the exogenous evolution of productivities. Our theory, moreover, offers a parsimonious explanation for the expansion of crops away from high-production regions, based on the knowledge that farmers acquire there.

3 Data and Motivating Facts

This section describes the data and three facts about comparative advantage, specialization and migration in Brazil. The first fact describe changes in Brazil's revealed comparative advantage during the March to the West that we seek to explain. The following two facts study the effects of the crop employment in farmers' origin on their earnings and crop choice in the destination.

3.1 Data

We collected data on migration, employment, agricultural production and trade in Brazil between 1970 and 2010. Our final data contains 137 regions distributed across 27 States. Here, we provide an overview of the datasets employed in our analysis. See Appendix A for a detailed description of the datasets.

First, we source data on migration and employment from the Brazilian demographic censuses of 1980, 1991, 2000 and 2010. For each worker, we observe her current and previous municipality of residence. We also observe the activity in which an individual is employed and her earnings. There are roughly 170 activities in each year of the census. Within these activities, we identified 14 major agricultural crops that are systematically found across waves of the census: banana, cassava, chicken, cocoa, coffee, cotton, corn, fish, fruits and horticulture, livestock, rice, soy, sugarcane and tobacco. We call workers employed in agriculture as farmers.

Second, we bring data from PAM (*Produção Agrícola Municipal*), which provides annual information on agricultural revenues per crop at the municipality level since 1974 and state level since 1930s. To obtain production in non-agricultural sectors, we collected data from the Brazilian Statistical Bureau (IBGE)[6] on value added of services and manufacturing. We

[6] *Instituto Brasileiro de Geografia e Estatistica.*

adjust all the nominal values from Brazilian datasets according to dollar values for value added by sector from the UN National Accounts.

Third, the trade data comes from the Brazilian Ministry of Development, Industry and Foreign Trade (MDIC) and is available annually since 1995. It contains information on total exports and imports from every state of Brazil to external markets disaggregated according to the harmonized system at the 6 digit level. We classify these goods to match the 14 crops used in our analysis. We complement our dataset with information from FAOSTAT, which contains aggregate trade data for Brazil and the rest of the world, broken down by crops.

Finally, we combine our production and migration data with information on the highway network in Brazil to calculate the travel distance between micro-regions.

3.2 Three Facts about Migration and Knowledge

Fact 1: Comparative Advantage. Brazil gained comparative advantages in crops exported by the West.

In what follows, we use a common measure of revealed comparative advantage (hereafter, RCA)

$$RCA_k = \frac{X_k^{BR} \Big/ \sum_{k' \in \mathcal{K}} X_{k'}^{BR}}{X_k^{W} \Big/ \sum_{k' \in \mathcal{K}} X_{k'}^{W}}, \tag{1}$$

where k is an index for the goods from a sector, K the set of sectors and contains all crops and a non-agricultural activity, X_k^{BR} the exports of Brazil and X_k^{W} the global exports. The RCA measures the specialization of Brazilian exports in crop k, relative to the world's specialization in the same crop. A number above one suggests Brazil has a comparative advantage in crop k.[7]

Table 1 shows that, in 2010, Brazil exports 5-10 times more coffee, livestock and tobacco than the rest of the world, and 15-20 times more soybeans and sugarcane. Brazil has changed substantially its comparative advantage relative to the world. There was a large expansion in the RCA of cassava, chicken, livestock, soy, sugarcane and tobacco. For recent years, we can disaggregate the RCA by region. The West has substantially larger RCA in soy and cattle, which are key agricultural goods in the export basket of Brazil. These agricultural goods account for roughly 40% of Brazil's agricultural exports and 16% of its total exports.

[7] Since Balassa (1965), the revealed comparative advantages has been used extensively in the literature to examine the patterns of specialization. For example, Goldberg, Khandelwal, Pavcnik, and Topalova (2010) and Menezes-Filho and Muendler (2011).

Fact 2: Influence of Migrants' Origin Region. Upon migration, farmers originating in regions with high crop employment are more likely to grow that same crop and earn higher incomes than other farmers doing so.

To investigate the influence of a migrant farmer's origin region on her farming choices and her earnings, we estimate

$$\log\left(y_{ij,kt}\right) = \iota_{j,kt} + \iota_{ij,t} + \alpha \log(L_{i,kt-1}) + \epsilon_{ij,kt}, \tag{2}$$

where i indexes the origin region, j the destination region, k the agricultural activity and t the year.[8] The parameter α captures the elasticity of the outcome ($y_{ij,kt}$) with respect to the number of farmers in the origin i producing k in period $t-1$ ($L_{i,kt-1}$). We estimate equation (2) using two dependent variables: the total number of agricultural workers and the average income of agricultural workers. To gain precision, we stack data for the years $t = 2000$ and $t = 2010$. We use the number of farmers lagged by thirty years as our measure of $L_{i,kt-1}$.

Equation 2 is based on the idea that workers learn and are more productive in the production of a good when there are more workers producing this same good nearby, following a large literature in economic geography and urban economics studying external economies.[9] Our specification captures how much workers retain of the knowledge that they acquired in an origin as they migrate to new region.

Our identification of α is based on the assumption that, conditional on migration choices, the size of the workforce in a specific crop in the origin is not systematically correlated with local characteristics of the destination or to general characteristics of migrants from a specific origin in a specific destination that are not crop-specific. To implement this approach, our regression includes a rich set of fixed effects. The term $\iota_{j,kt}$ is a destination, crop and year fixed effect that captures the effect of any natural advantage or price shock that is common across agricultural workers in destination j producing crop k. The term $\iota_{ij,t}$ is a destination, origin and year fixed effect that captures any factor that is common for farmers from region i producing in j such as human capital and bilateral migration costs. We exlude non-migrants from our regressions to focus on how migrants shape their destination region and also to avoid including the same farmers in both sides of equation (2). Finally, we present Poisson estimates of equation (2) due to the presence of zeros.

Figure (5) shows the variation that identifies the coefficient α in both regressions, and suggests a log-linear relationship between both dependent variables and the crop employment

[8] Earnings is measured in the data by total income.

[9] We use a log-linear formulation, similar to that adopted in several papers in economic geography and urban economicswhen modeling external economics, including Ahlfeldt, Redding, Sturm, and Wolf (2015) and Bartelme, Costinot, Donaldson, and Rodriguez-Clare (2019).

in the origin. Panel A in Table 2 shows OLS estimates of equation (2) using the number of farmer migrants as the dependent variable. We find that an increase in the number of agricultural workers in the region of origin in a given activity of 1% increases the number of agricultural workers in the destination in this same activity by 0.06%. As we include origin and destination fixed effects and controls for socio-economic characteristics , we obtain similar point estimates. In column column 4, we narrow our estimation to the micro-region to better control for unobserved characteristics in the destination and find similar qualitative results. Panel B shows that the elasticity of migrants' earnings with respect to the number of farmers in the origin is between 0.02%-0.04% across the same specifications. We conclude that the employment structure in the region of origin shapes the decisions and earnings of migrants, consistent with crop-specific productivity increasing in the amount of employment in that crop at the origin.

Fact 3: Farmers Composition. Agricultural revenues increase with immigration from regions employing many farmers in that crop, after controlling for employment.

To study the relationship between the composition of farmers in terms of their region of origin and agricultural output in a region, we estimate

$$\log\left(y_{j,kt}\right) = \iota_{j,t} + \iota_{k,t} + \gamma_0 \underbrace{\log L_{j,kt}}_{Abundance} + \gamma_1 \underbrace{\log \sum_i \frac{L_{ij,kt}}{L_{j,kt}} L_{i,kt-1}}_{Composition} + \epsilon_{j,kt}, \tag{3}$$

where $L_{ij,kt}$ is the flow of workers from i to j producing crop k in period t. We examine the relationship with total revenues $y_{j,kt}$, and its decomposition into price and quantity effects. The first set of fixed effect on the right hand side captures any level effect such as the size of a region or the overall demand for agricultural goods and the second one captures any crop specific characteristic such as the land intensity. When the composition term is larger for a given destination and crop, then farmers come from origins that are more specialized in the production of this crop.

Table 3 shows, first, that a 1 percent increase in the abundance of farmers in a region is associated with a 0.9% increase in revenues. Second, it shows that a 1 percent increase in the average number of farmers in the origin is associated with a 0.15% increase in revenues, even controlling for the abundance of farmers, which indicates that the composition of farmers is also strongly associated with total size of the sector in a region. These results hold in magnitude when we look at the quantity produced. We find that the prices implied by revenues and quantity are negatively correlated with the level and the composition of farmers.[10]

Taken together, Facts 2 and 3 motivate the formulation of a model where farmers learn how to produce crops that are prevalent in the region where they live and that, upon migration, they carry and deploy their acquired knowledge in their destination region. In the next section, we introduce migration as a source of comparative advantage into a general equilibrium model and use it to assess the role that migration played in Brazil's transformation captured in Fact 1.

4 Model

In this section, we develop a dynamic spatial economy model in which comparative advantage is driven by land productivity, labor supplies, and the good-specific knowledge of migrants. We develop the model with two goals in mind. First, we give a structural interpretation to the reduced form regressions in the previous section. Second, we use the model to examine the mechanism through which migration shapes the evolution of comparative advantages. Below, we present a simplified version of the model to economize on notation.

4.1 Environment

Geography and Commodities.

We focus attention on a Home country, which we divide into $j = 1, \ldots, I$ regions, and a rest of the world composite, denoted by F. There are $k = 1, \ldots, K$ sectors and each region produces an unique variety in each sector. Time is discrete, and indexed by t. Iceberg trade and migration costs deter the flux of agents and goods across space. In each time, the geography of the economy is given by a set of natural advantages, total land endowments, a matrix of bilateral trade costs and a matrix of bilateral migration costs: $\left\{A_{j,kt}, \bar{H}_{j,t}, \tau_{ij,kt}, \mu_{ij,kt}\right\}$. We omit time indexes whenever unnecessary for our presentation.

Technology.

A continuum of farms produce in sector k, region i. A manager with ability s rents land and produces according to

$$q_{j,k}(s) = A_{j,k} s^{1-\gamma_k} l^{\gamma_k},$$

[10] A concern with Fact 2 above is that crop employment in the origin region does not increase productivity, but rather increases the prices migrants collect. This regression indicates that, while farmers from specific origins might have access to better prices in the market, this effect is not sufficiently large to overturn a local supply effect associated associated with higher productivity.

where γ_k measures the land intensity of crop k.[11]

Agents.

People live two periods, young and old. An adult at time t, upon observing her ability, decides where to live, what sector to work on and spawns a child. To simplify matters, only adults consume, and they ignore their child's utility. Adults have constant elasticity of substitution (CES) preferences between K sectors, with elasticity of substitution σ, and a CES preferences between varieties of each sector, with an elasticity of substitution η. Let L_{jt} denote the adult population at time t in j.

Adult workers maximize welfare by choosing where to live and in which sector to work at time t:

$$\max_{j,k} W_{ij,kt}\varepsilon_{i,kt},$$

where preference shocks are drawn i.i.d from $G(\varepsilon) = \exp(-\varepsilon^{-\kappa})$ and W_{ijkt} is the systematic component of welfare. This systematic component is given by

$$W_{ij,kt} = \frac{w_{j,kt}s_{i,kt}}{\mu_{ij,kt}P_{j,t}}, \tag{4}$$

where $w_{j,kt}$ is the wage per efficiency unit of labor (i.e., the return to a unit knowledge), $\mu_{ij,kt}$ represents iceberg migration costs that reduce utility directly, and $P_{j,t}$ is the CES price index of aggregate consumption in destination region j. Reflecting our empirical findings, $s_{i,kt}$ is a farmer's knowledge to produce in sector k, which depends on the region she comes from. The CES price indexes are given by $P_{j,t}^{\frac{\sigma-1}{\sigma}} = \sum_k a_{k,t}^{\frac{1}{\sigma}} P_{j,kt}^{\frac{\sigma-1}{\sigma}}$ and $P_{j,kt}^{\frac{\eta-1}{\eta}} = \sum_i p_{i,kt}^{\frac{\eta-1}{\eta}}$, where $a_{k,t}$ is a preference shifter.[12]

A child born in i at time $t-1$ is characterized by a vector of sector-specific productivities, $s_{i,kt}$, which depends on the employment structure in the region where he is born.

Knowledge Endowment

A worker's knowledge to grow each crop depends on her origin region: $s_{i,kt} \equiv s(L_{i,kt-1})$. In particular, motivated by our empirical findings, we assume that knowledge depends on crop-specific employment in the origin region through the following functional form:

$$s_{i,kt} = \bar{s}_k L_{i,kt-1}^{\beta}. \tag{5}$$

[11] For simplicity in presentation, we write down this technology as having unitary elasticity of substitution between land and labor. In the quantification, we work with a CES production function with elasticity ρ.

[12] Workers can migrate within the Home country, but not between Home and Foreign.

Land supply

We assume that every period there is an upward sloping supply of land for productive uses, given by

$$H_{j,t} = \frac{r_{j,t}^{\zeta}}{b_{j,t} + r_{j,t}^{\zeta}} \bar{H}_{j,t}. \tag{6}$$

In this equation, ζ is the partial elasticity of supply and $b_{j,t}$ stands for the opportunity use of land, which varies across regions and time.

4.2 Equilibrium

Taking the price of sector k, $p_{j,k}$, and the rental rate of land, r_j, as given, managers maximize profits. The optimal demand for land, $l_{j,k}(s)$,and output, $q_{j,k}(s)$, increase linearly in farmer's knowledge, s:

$$l_{j,k}(s) = s \left(\frac{\gamma_k}{r_j} p_{j,k} A_{j,k} \right)^{\frac{1}{1-\gamma_k}},$$

and

$$q_{j,k}(s) = \kappa_{q,k} A_{j,k} s \left(\frac{p_{j,k} A_{j,k}}{r_j} \right)^{\frac{\gamma_k}{1-\gamma_k}},$$

where $\kappa_{q,k} = (\gamma_k)^{\gamma_k/(1-\gamma_k)}$.

The farmer's earnings correspond to the profits of managing a farm, which are also linear in the farmer's knowledge:

$$\pi_{j,k}(s) = \kappa_{\pi,k} s \left(p_{j,k} A_{j,k} \right)^{\frac{1}{1-\gamma_k}} r_j^{\frac{\gamma_k}{1-\gamma_k}}.$$

To define the equilibrium, it will be useful to write the unit cost of producing a unit of the good as a function of rental rates of land and efficiency wages. Defining efficiency wages as $w_{j,k} \equiv \pi_{j,k}(s)/s$, i.e. as earnings per unit of knowledge, equilibrium unit costs are given by

$$\frac{c_{j,kt}}{A_{j,k}},$$

where we also define $c_{j,kt} \equiv \kappa_{\pi}^{-1} w_{j,k}^{1-\gamma_k} r_j^{\gamma_k}$.

As a result of utility maximization, the share of region j's expenditure in sector k goods produced in region i is given by:

$$\pi_{ij,kt} = \frac{\left(c_{j,kt}\tau_{ij,kt}/A_{i,kt}\right)^{1-\eta}}{\sum_{i'}\left(c_{i',kt}\tau_{i'j,kt}/A_{i',kt}\right)^{1-\eta}}.$$

Next, using the definition of the observable component of welfare (4), optimal worker sorting gives the share of workers from i choosing to work in region j and sector k, $\lambda_{ij,kt}$

$$\lambda_{ij,kt} = \frac{W_{ij,kt}^{\kappa}}{\Xi_{i,t}^{\kappa}} \tag{7}$$

where and $\Xi_{i,t}^{\kappa} \equiv \sum_j \sum_k \left[w_{j,kt}s_{ikt}/\left(\mu_{ij,kt}P_{j,t}\right)\right]^{\kappa}$. It follows that the flow of workers from i to region j, sector k is $L_{ij,kt} = \lambda_{ij,kt}L_{i,t-1}$. We define the effective units of labor migrating from i to region j, sector k as

$$E_{ij,kt} \equiv s_{i,kt}\lambda_{ij,kt}L_{i,t-1}. \tag{8}$$

Finally, reflecting the land grants program in Brazil, we assume that workers become landowners when they move to a region.

To close the model, we note that total expenditure in region j reflects payments to factors there

$$X_{j,t} = \sum_k w_{j,kt}E_{j,kt} + r_{j,t}H_{j,t},$$

and sectoral expenditure, $X_{j,kt}$ reflects the preferences described above.[13]

We are now ready to define an equilibrium for this economy. We break down the equilibrium in two parts, as in Caliendo, Dvorkin, and Parro (2015): a goods market equilibrium, which takes migration flows as given, and then the migration equilibrium. A competitive equilibrium is a sequence of allocations that satisfies both the goods market and the migration equilibrium. Finally, we also provide a definition of a steady state equilibrium.

Goods market equilibrium in period t.

Given the geography at time t, migration flows $\{L_{ij,kt}\}_{ijk}$, and past labor allocations, $\{L_{i,kt-1}\}_{ik}$, at time t, a goods market equilibrium is a set of factor prices and allocations of efficiency units of labor $\{r_{j,t}(\{L_{ij,kt}\}), w_{j,kt}(\{L_{ij,kt}\}), E_{j,kt}(\{L_{ij,kt}\})\}_{jk}$ such that:

[13]In taking the model to the data, we allow for trade imbalances, which we omit here to simplify the exposition.

1. The market for efficiency units of labor clears in region j and sector k:

$$w_{j,kt}E_{j,kt} = (1-\gamma_k)\sum_j \pi_{ij,kt}X_{j,kt}.$$

2. Land markets clear in region j:

$$r_{j,t}H_{j,t} = \sum_k \gamma_k \sum_j \pi_{ij,kt}X_{j,kt},$$

3. Total immigration into region j, sector k determine the effective supply of labor there:

$$E_{j,kt} = \sum_i s_{i,kt}\left(L_{i,kt-1}\right)L_{ij,kt},$$

   where the function $s_{i,kt}$ is defined in equation (5).

In the definition above, prices and allocations in the goods market equilibrium at time t depend on migration flows and past allocations, which introduce dynamics into the system.

Migration equilibrium in period t.

Given geography at time t and labor allocations in period $t-1$, $\{L_{i,kt-1}\}$, a migration equilibrium at time t is a set of migration flows, labor allocations and prices:$\{L_{ij,kt},\ w_{j,kt}\left(L_{ij,kt}\right),\ r_{j,t}\left(L_{ij,kt}\right),\ E_{j,kt}\left(L_{ij,kt}\right)\}$ for regions i and j, and sector k, such that migration flows evolve according to optimal sorting of workers for each pair of regions i and j, and each sector k:

$$L_{ij,kt} = \lambda_{ij,kt}L_{i,t-1}, \tag{9}$$

where $\lambda_{ij,kt}$ is given by equation 7.

Finally we define a competitive equilibrium for this dynamic model.

Competitive equilibrium.

Given a geography for $t = 1, \ldots, \infty$, and initial labor allocations in period 0, $\{L_{i,k0}\}_{i,k}$, a competitive equilibrium is a sequence of migration flows, efficient labor allocations, and prices, $\{L_{ijkt}, E_{i,kt},\ w_{i,kt},\ r_{i,t},\ \}_{t=1}^{\infty}$ that satisfy the goods market and migration equilibria in each period t.

To study the long-run behavior of this economy, we define a steady state equilibrium next.

Steady State Equilibrium.

Given a constant geography for $t = 1, \ldots, \infty$, a steady state equilibrium is a competitive equilibrium in which migration flows, labor allocations, and prices, are unchanged: $L_{ijkt} = \bar{L}_{ijk}$, $w_{i,kt} = \bar{w}_{i,k}$, $r_{i,t} = \bar{r}_i$, and $E_{i,kt} = \bar{E}_{i,k}$, $\forall t = 1, \ldots, \infty$.

Discussion of the equilibrium.

The properties of the equilibrium are shaped by the interaction of agglomeration and dispersion forces. First, the idiosyncratic draws are a force towards populating all region-crop cells. The strength of this force is governed by the dispersion in preference shocks κ: as κ decreases, individuals have stronger idiosyncratic tastes for working in different regions and activities. Second, the external sector has a downward sloping demand for the goods in Brazil; this acts as a force against full agglomeration in a given crop, within regions. The strength of this force is governed by η: as η grows, terms of trade turn against Brazil faster as output in a given crop increases. Third, our assumptions on technology yield high marginal values of labor when $L_{i,kt} = 0$, which provides an incentive for workers to be employed in each region-crop combination.

The opposing, agglomeration force is given by the spatial allocation of knowledge: if there is a large number of workers populating a region-crop cell, workers want to locate there because their productivity is larger. The strength of the agglomeration force is governed by β. Note that this force only operates in the steady state, since in each period past allocations are taken as given. In other words, at any given time, conditional on past labor allocations, ours is a standard model of migration and trade in which there are no agglomeration forces. Relatedly, there is a dynamic externality in the way we model knowledge diffusion, since workers do not internalize their impact on the productivity of the next generations.

4.3 Migration and Comparative Advantage

In this environment, migration shapes comparative advantage through two channels. First, by increasing the effective labor force, E_{ikt}, migration affects the productivity of land and raw labor. This force is captured in the model by a decrease in $w_{i,kt}$. Second, relative factor abundance shapes comparative advantage, as in the classic Heckscher-Ohlin model. Migration alters relative factor proportions across regions and, everything else constant, large migration inflows favor the production of labor-intensive crops.

To see how knowledge shapes comparative advantage, consider a simplified version of the model in which labor is the only factor of production. In that model, we say that a country

i as having comparative advantage in good k (relative to country i' and good k') if

$$\frac{A_{i,k}e_{i,k}}{A_{i,k'}e_{i,k'}} > \frac{A_{i',k}e_{i',k}}{A_{i',k'}e_{i',k'}},$$

where we define $e_{i,k} \equiv E_{i,k}/L_{i,k}$ as the average efficiency per worker in country i, sector k (see French, 2017). This expression shows how comparative advantage results from the interaction of natural advantage and worker efficiency. Our paper provides a theory of the formation of worker efficiency through labor reallocation.

5 Taking the Model to the Data

This section describes how we quantify the exogenous parameters of the model $\{\kappa,\ \beta,\ \theta,\ \gamma_k,\ \tau_{ijk,t},\ \mu_{ij,kt},\ A_{j,kt},\ \bar{H}_{j,t},\ b_{j,t}\}$. We map the model's goods-market equilibrium to the year 2010, while taking region and sector labor allocations $L_{i,kt-1}$ as we observe them in the year 1980. Thus, we effectively set a time period to 30 years. We choose these periods to strike a balance between the quality and availability of data, and a time period early enough that we can observe the transformation of the Brazilian economy due to the March to the West.

In what follows, we begin by connecting our model explicitly to the reduced form evidence in Fact 2. This approach allows us to measure directly the key elasticities β and κ in the micro data. With these elasticities in hand, we discuss our calibration procedure, including the data we match and the parameters we choose.

5.1 Elasticities: Connecting the Model to Reduced Form Evidence

We use the model to uncover two key parameters: the dispersion parameter, κ and the origin elasticity β (which is new to our theory). First, using equation (5), and letting $\mu_{ij,kt} = \mu_{ij,t}\epsilon^{\mu}_{ij,kt}$, and $\bar{s}_k = \tilde{s}_k\epsilon^{s}_{i,kt}$, our model suggests that the earnings of migrants from i into j growing crop k relates to the employment in that crop at the migrants' origin:

$$\begin{aligned} \log \text{earnings}_{j,kt} &= \log\left(w_{j,kt}s_{i,kt}\right) \\ &= \iota_{j,kt} + \iota_{ij,t} + \beta \log L_{i,kt-1} + u^{earnings}_{i,kt}, \end{aligned} \tag{10}$$

where we define $\iota_{j,kt} \equiv \log w_{j,kt} + \log \tilde{s}_k$ and $u^{earnings}_{i,kt} \equiv \log \epsilon^{s}_{i,kt}$. This regression allows us to uncover the parameter β and provides direct evidence for our mechanism. Note that to identify β we assume that, conditional on fixed effects, the remaining component of knowledge in $\epsilon^{s}_{i,kt}$ is uncorrelated with $L_{i,kt-1}$.

Second, we examine how employment shares within destination-crop relate to migrant origin. In our model these employment shares are linked to migration flows, $L_{ij,kt}$. Substituting equation (5) and (7) into (9), we obtain our econometric specification:

$$\begin{aligned}
\log L_{ij,kt} &= \log(\lambda_{ij,kt} L_{i,t-1}) \\
&= \log(\lambda_{ij,kt}) + \log(L_{i,t-1}) && (11) \\
&= \log w_{j,kt} + \log s_{i,kt} + \log P_{j,t} + \log \mu_{ij,kt} - \log \Xi_{i,t} + \log(L_{i,t-1}) && (12) \\
&= \iota_{j,kt} + \iota_{ij,t} + \kappa\beta \log L_{i,kt-1} + u_{ij,kt}^{migration}, && (13)
\end{aligned}$$

where $\iota_{ij,t} = \log L_{i,t-1} - \kappa\left(\log \mu_{ij,t} - \log \Xi_{it}\right)$, $\iota_{j,kt} = \kappa\left(\log w_{j,kt} + \log \bar{s}_k\right) - \log P_{j,t}$, and $u_{ijkt}^{migration} = \kappa \log \epsilon_{ij,kt}^{\mu} - \kappa \log \epsilon_{i,kt}^{s}$. Knowledge of β allows us to disentangle κ from the coefficient in the previous regression.[14]

As in common in the empirical trade literature, our data contain a number of zero migration flows. Therefore, we also estimate equations (10) and (13) nonlinearly via Poisson PML (Santos Silva and Tenreyro, 2006), which allows us to include observations where the dependent variable is zero.[15] Since our data set is comprised by samples of the Brazilian census, an interpretation of the errors in the equations above is sampling errors, which keeps the estimation consistent with our theory.

Table 2 shows results from PPML estimates of equations (10) and (13) in column 4. We find substantially larger point estimates than in our OLS results in columns 1 through 3. This occurs because the PPML gives a higher weight to large values and, as we can see in Figure 5, the influence of the number of farmers in the origin becomes steeper for higher values. Moreover, PPML also incorporates information from zeros in the dependent

[14] Our model also rationalizes our regressions for revenues $R_{j,kt}$. In the model, regional revenues in crop k are given by

$$R_{j,kt} = \kappa_k \left(\frac{p_{j,k} A_{j,k}}{r_j^{\gamma_k}}\right)^{\frac{1}{1-\gamma_k}} L_{j,kt} \sum_i \frac{E_{ij,kt}}{L_{j,kt}},$$

where κ_k is a constant. Taking the logs gives

$$\log\left(R_{j,kt}\right) = \underbrace{\log\left(L_{j,kt}\right)}_{Abundance} + \underbrace{\log\left\{\sum_i \frac{L_{ij,kt}}{L_{j,kt}} \times \bar{s}_{i,kt-1} L_{i,kt-1}^{\beta}\right\}}_{Composition} + u_{j,kt},$$

where we defined $u_{j,kt} \equiv \kappa_k \left(\frac{1-\gamma_k}{r_{j,t}} p_{j,kt} A_{j,kt}\right)^{1/\gamma_k}$.

[15] Specifically, since we have a large number of fixed effects, we estimated these equations using a concentrated likelihood function. This technique partials out the fixed effects from the likelihood function before the estimation of β and κ.

variable. For our calibration, we pick a value in between our OLS and PPML estimates $\beta = 0.1$. For the dispersion in preferences, our OLS and PPML results are similar and we pick $\kappa = 1.25$. Our estimate of κ is close to that obtained in recent papers (Monte, Redding, and Rossi-Hansberg, 2015; Suárez Serrato and Zidar, 2016; Bryan and Morten, 2015).[16]

5.2 Quantification

We quantify the model in two steps. First, with a parametrization of trade costs, we adjust composite productivities - which includes both natural advantages and factor prices - in each region to match state-level trade flows and regional revenues.[17] Second, we calibrate migration costs using estimates of composite productivities. We then disentangle natural advantages from composite productivities using factor prices implied by the model. We provide a short summary next, and relegate details to Appendix B. Table 4 lists the parameter values we use in our quantification.

Step 1. Composite productivities, preferences and trade costs

We calibrate composite productivities $\overline{T}_{jk} \equiv c_{j,k}/A_{i,k}$ and preferences shifters a_k in the model to match revenues by region and crop and state-level trade flows with the foreign region. To do so, we need a parametrization of trade costs and values for trade elasticities (η_k). We set η_k in the non-agricultural sector to 5.5 and the one in the agricultural sector to 9.5 following Caliendo and Parro (2015). We parametrize the bilateral trade cost between regions in Brazil using $\tau_{ij} = (\text{distance}_{ij})^{\delta_\tau}$, where distance_{ij} is the travel distance between region i and j and pick $\delta_T = 0.05$ following Pellegrina (2018). We parametrize the trade cost between domestic regions and the foreign sector as folows. We assume that every region must incur a trade cost of τ_{iF} to reach their state capital and, after then, they incur an additional international component that is state, crop and direction specific, $\tau_{iFk} = \tau_{rF,k}\tau_{iF}$ and $\tau_{Fi,k} = \tau_{Fr,k}\tau_{iF}$. All trade costs with the region itself is normalized to 1. We calibrate a_k to match aggregate absorption, $\overline{T}_{j,k}$ to match revenues and $\tau_{rF,k}$ to match trade state-level trade flows. Using trade costs, composite productivities and preference shifters, we can construct price indexes for each region, which we use next to calibrate migration costs.

[16] A common approach is to estimate the migration elasticity according to a gravity equation of the log of migration flows on the bilateral distance between the point of origin and destination of workers. Monte, Redding, and Rossi-Hansberg (2015) find a heterogeneity in commuting preferences of 3.3 using data for commuters in the US. Morten and Oliveira (2016) find a migration elasticity to wages of 1.9 using the Brazilian data and Bartik instruments.

[17] As a matter of accounting, before proceeding we calibrate transfers and consumption parameters to match trade imbalances and the aggregate absorption of the economy. See Appendix

Step 2. Migration costs, knowledge, and natural advantage

We assume that migration costs are not sector specific. We break it down into two parts: a migration cost between state r and s, $\mu_{r(i)s(j)}$, and a cost associated with the distance between region i and j which we parametrize using $(dist_{ij})^{\delta_M}$. We thus write migration costs as $\mu_{ij} = \mu_{r(i)s(j)} (dist_{ij})^{\delta_M}$.

We parametrize knowledge using $s_{i,kt} = \bar{s}_k L^{\beta}_{i,kt-1}$, where $\bar{s}_k$ represents the aggregate state of knowledge in the economy about the production of crop k and $L^{\beta}_{i,kt-1}$ is the origin-specific component acquired by farmers.

Based on the parametrization of knowledge and migration costs described above, we search for values of $\mu_{r(i)s(j)}$ that make migration choices in the model match the observed mass of workers flowing between states, values of δ_M that makes the model match the overall number of farmers living in their region of birth, and values of $\bar{s}_k$ that make the model match the number of farmers in each activity k. In matching these statistics, we compute model implied wages based on values for the knowledge parameter (β), the migration elasticity (κ) and the factor shares (γ_k). We pick β and κ from our reduced form procedure and set γ_k to match the value added share of land. Usingthe price index constructed in the previous step and guesses of $\mu_{r(i)s(j)}$, δ_M and $\bar{s}_k$, we thus get model implied wages from

$$R_{jk} = \frac{w_{jk} \sum_i \lambda_{ij,k} L_i s_{i,k}}{1 - \gamma_k},$$

where $R_{j,k}$ is the observed revenue in a region. Since $\lambda_{ij,k}$ is itself a function of wages, we use a numerical procedure to find model implied wages. From the equation above, we obtain predictions for migration, which we use to search for values of $\mu_{r(i)s(j)}$, δ_M and $\bar{s}_k$. Note that this procedure does not impose a perfect match between migration flows in the model and in the data, but it is still consistent with the perfect matching of the model in terms of trade end revenues that we imposed to recover $\overline{T}_{i,kt}$.

Finally, with model implied wages and land rents, we recover natural advantages $(A_{j,k})$ using $A_{j,k} \equiv c_{j,k} (w_{j,k}, r_j) / \overline{T}_{jk}$.

6 The Aggregate and Regional Effects of Migration

In this section, we evaluate the quantitative importance of migration as a driver of comparative advantage. To gauge the aggregate importance of this mechanism, we ask what regional and aggregate exports would have been in 2010, had there been no East-West migration since 1980. We then separate the contribution of changes in the allocation of raw labor over space from the contribution of workers' knowledge. Our choice of policy counterfactual is loosely guided by the policies instituted in Brazil around the mid twentieth century. We interpret it as asking: What would have happened if Brazil's central government had not provided incentives for workers to populate the West?[18]

6.1 Decomposing the Effects of Banning Migration to the West

To calibrate the model, we set $t-1 = 1980$ and $t = 2010$. We then consider two scenarios. In the first, we calibrate the model as to match perfectly the data on production and trade data in 2010. In the second, we set $\mu'_{ij,kt} \to \infty$ if i is in the East and j is in the West (or the converse). We allow, however, for migration within the East and within the West. We call these "baseline" and "no migration" simulations.

We use these two simulations to separate the effect of changes in land abundance from the effect of changes in the composition of workers. First, we compute a counterfactual *goods market* equilibrium (see 4.2) taking the stock of labor, $L_{ij,kt}$, from the baseline simulation, combined with the efficiency per worker of the "no migration" simulation, $e_{j,kt}$. That is, we compute an equilibrium using the stocks of labor we see in the data, assuming each worker's productivity is the counterfactual one, thus removing the effect of migrants' knowledge. Second, we compute another counterfactual *goods market* equilibrium, taking the stock of labor from the "no migration" counterfactual, but using the efficiency per worker of the baseline simulation. This counterfactual removes the effect of labor abundance.

Our first result concerns Brazil's agricultural sector as a whole. As discussed before, Brazil's agricultural RCA increased from 3.74 to 3.38, as discussed before. Absent migration, the RCA would rise counterfactually to 4.33, which we interpret as saying that migration accounts for 20 percent of the actual increase. Of this increase, about 20 percent comes from knowledge reallocation.

Moving on to individual goods, Table 5 shows the effects of banning East-West migration on the RCA of each crop, which speaks directly to the evolution of Brazil's external sector summarized in Fact 1. The first column shows the export share of each crop in the year 2010, highlighting the importance of soy in total agricultural exports. The second and third columns show Brazil's observed RCA in 1980 and 2010, while the fourth column shows the counterfactual RCA in 2010 after banning East-West migration. A main message of the table is that migration contributed to Brazil's export boom, precisely in the crops in which it experienced a gain in RCA, namely, soy, livestock, and corn. For example, migration accounts for about 36 percent of the RCA increase in soy, 11 percent of the increase in livestock, and

[18] We are currently working on recovering empirically the migration costs in previous periods as to obtain a more accurate sense of the role of these policies.

26 percent of the RCA increase in corn. Likewise, it accounts for 39 percent of the decrease in coffee. The columns labeled "Labor" and "Knowledge"present the contribution to these RCA changes that comes from removing the effect of raw labor movement and knowledge reallocation, expressed as a percent of the total change induced by banning migration. Across crops, changes in labor abundance account for at least 70 percent of the changes in RCA. As the table shows, migrants' knowledge accounts for a varying fraction of total changes, and about 20 percent for soy, the flagship crop of Brazil's external transformation. For livestock, however, the effect of knowledge reallocation is negative, since by 1980 a critical mass of ranchers was already present in the west. Across crops, the sum of the two last columns is close to one, suggesting there are relatively small interactions between these two effects in our decompositions.

The previous results show that Brazil's comparative advantage was substantially affected by migration. Our next set of results link these changes in specialization to changes in welfare (Table 6). We define aggregate welfare as the weighted average of expected utility across regions within Brazil,

$$\text{welfare} = \sum_i \frac{L_{i,1980}}{\sum_{i'} L_{i',1980}} \Xi_{i,1980},$$

where $\Xi_{i,1980}$, defined in Section (4), can be interpreted as the expected utility of workers born in region i. The first column shows the Gains from Trade (GFT) in our baseline calibration, defined as the welfare loss associated with going to autarky. Moving to a no-trade scenario leads to 8.1 percent losses, while moving to a scenario of no agricultural trade leads to 3.4 percent losses, which confirms the importance of the agricultural sector in Brazil. The second column repeats these calculations in the no East-West migration scenario, showing that the gains from trade decrease marginally relative to the baseline (about 3 percent lower). The reason is that Brazil's GFT in this period are determined in large part by the exchange of agricultural goods for non-agricultural goods. While migration entailed large swings in specialization within the agricultural sector, on aggregate these reallocations allowed for small increase in import shares of non-agricultural goods.

Columns three and four of Table (4) focus on the gains from internal migration, defined as the loss from banning migration, relative to the baseline. Note first that these gains are comparable to the gains from agricultural trade, which means that migration had a positive welfare effect in the aggregate by allowing workers to move to low density areas in the West. Second, the gains from migration are about 9 percent lower in the absence of trade, suggesting that international trade is key to how the West absorbs migrant workers.

6.2 Steady-State Effects on Employment and Exports

To understand the long run implications of knowledge reallocation on aggregates such as sectoral exports and employment, we calculate the steady state of two versions of our model. The first one is our baseline calibration, in which $\beta = 0.1$. The second shuts down the reallocation of knowledge due to migration, but allows for labor to allocate freely to the most productive regions, by setting $\beta = 0$. Both calibrations, by construction, fit exports exactly in $t = 2010$. Figure 7 shows large steady-state differences across states, in these two calibrations. We find permanent increases in aggregate exports of up to 50 percent for individual crops. In line with our previous results, exports of soy are permanently larger by more than 30 percent, when knowledge reallocation is present. Knowledge diffusion also permanently reduces the exports of other traditional crops, such as coffee.

7 Conclusions

We study how the allocation of labor within a country shapes comparative advantage at the regional and aggregate level. We disaggregate these effects into two channels: the allocation of raw labor and the spatial reallocation of workers' knowledge. In the case of Brazil, where large migration coincided with a large change in comparative advantage, we show, first, that migration transformed Brazil's external sector. Had there been no migration since 1980, Brazil's comparative advantage in new commodities, such as soy, livestock, and corn, would have been substantially weakened. We also show that the spatial reallocation of knowledge is important in the aggregate and for individual crops, such as soy, accounting for about 20 percent of the aggregate changes in exports due to migration.

Our results also contain policy lessons on the formation of new exporting sectors. We find that migrants can substantially boost the expansion of new exporting activities, specially if they bring knowledge about these activities. While our results have direct implications for developing countries that showed interest in replicating the Brazilian experience with the expansion of new agricultural sectors, we believe that the underlying mechanisms considered in this paper should also be relevant for other sectors of the economy.

References

Ahlfeldt, G. M., S. Redding, D. Sturm, and N. Wolf (2015): "The economics of density: evidence from the berlin wall," *Econometrica*.

Allen, T., and C. Arkolakis (2014): "Trade and the Topography of the Spatial Economy," *The Quarterly Journal of Economics*, 129(3), 1085–1140.

Allen, T., and D. Atkin (2016): "Volatility and the Gains from Trade," Working Paper 22276, National Bureau of Economic Research.

Allen, T., and D. Donaldson (2018): "The Geography of Path Dependence," .

Arkolakis, C., S. K. Lee, and M. Peters (2018): "European Immigrants and the United States' Rise to theTechnological Frontier in the 19th Century," .

Bahar, D., and H. Rapoport (2016): "Migration, knowledge diffusion and the comparative advantage of nations," *The Economic Journal*.

Balassa, B. (1965): "Trade liberalisation and revealed comparative advantage 1," *The manchester school*, 33(2), 99–123.

Bartelme, D., A. Costinot, D. Donaldson, and A. Rodriguez-Clare (2019): "External economies of scale and industrial policy: a view from trade," *Working Paper*.

Bazzi, S., A. Gaduh, A. D. Rothenberg, and M. Wong (2016): "Skill Transferability, Migration, and Development: Evidence from Population Resettlement in Indonesia," *American Economic Review*, 106(9), 2658–2698.

Bryan, G., and M. Morten (2015): "Economic development and the spatial allocation of labor: Evidence from Indonesia," *Manuscript, London School of Economics and Stanford University*, pp. 1671–1748.

Buera, F. J., and E. Oberfield (2016): "The Global Diffusion of Ideas," Working Paper 21844, National Bureau of Economic Research.

Caliendo, L., M. Dvorkin, and F. Parro (2015): "The impact of trade on labor market dynamics," Discussion paper, National Bureau of Economic Research.

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

Costinot, A., and D. Donaldson (2014): "How Large are the Gains from Economic Integration? Theory and Evidence from U.S. Agriculture, 1880-1997," .

Costinot, A., D. Donaldson, and C. Smith (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, 205–248.

Costinot, A., and A. Rodriguez-Clare (2015): "Chapter 4 - Trade Theory with Numbers: Quantifying the Consequences of Globalization," in *Handbook of International Economics*, ed. by K. R. Elhanan Helpman, and G. Gopinath, vol. 4 of *Handbook of International Economics*, pp. 197 – 261. Elsevier.

Crosby, A. W. (1973): *The Columbian Exchange. Biological and Cultural Consequences of 1492*, no. 2 in Contributions in American Studies. Greenwood.

Eaton, J., and S. Kortum (2012): "Putting Ricardo to Work," *Journal of Economic Perspectives*, 26(2), 65–90.

Fajgelbaum, P., and S. Redding (2014): "External Integration and Internal Development: Evidence from Argentina, 1870-1914," NBER Working Papers 20217, National Bureau of Economic Research, Inc.

French, S. (2017): "Revealed comparative advantage: What is it good for?," *Journal of International Economics*, 106(C), 83–103.

Goldberg, P. K., A. K. Khandelwal, N. Pavcnik, and P. Topalova (2010): "Imported intermediate inputs and domestic product growth: Evidence from India," *The Quarterly journal of economics*, 125(4), 1727–1767.

Hanson, G. H., N. Lind, and M.-A. Muendler (2015): "The Dynamics of Comparative Advantage," Working Paper 21753, National Bureau of Economic Research.

Jie Cai, N. L., and A. Maria (2019): "Knowledge Diffusion, Trade, and Innovation across Countries and Sectors," .

Levchenko, A. A., and J. Zhang (2016): "The evolution of comparative advantage: Measurement and welfare implications," *Journal of Monetary Economics*, 78(C), 96–111.

Lind, N., and N. Ramondo (2018): "Innovation, Knowledge Diffusion, and Globalization," Working Paper 25071, National Bureau of Economic Research.

Menezes-Filho, N. A., and M.-A. Muendler (2011): "Labor reallocation in response to trade reform," Discussion paper, National Bureau of Economic Research.

Monte, F., S. J. Redding, and E. Rossi-Hansberg (2015): "Commuting, migration and local employment elasticities," Discussion paper, National Bureau of Economic Research.

Morten, M., and J. Oliveira (2016): "The Effects of Roads on Trade and Migration: Evidence from a Planned Capital City," *NBER Working Paper*, 22158.

Olmstead, A. L., and P. W. Rhode (2011): "Adapting North American wheat production to climatic challenges, 1839–2009," *Proceedings of the National Academy of Sciences*, 108(2), 480–485.

Olmstead, A. L., P. W. Rhode, et al. (2008): "Creating Abundance," *Cambridge Books*.

Opala, J. A. (1987): *The Gullah: rice, slavery and the Sierra Leone-American connection.* USIS.

Pellegrina, H. S. (2018): "The Causes and Consequences of the Spatial Organization of Agriculture," *Working paper*.

Porteous, O. (Forthcoming): ""High Trade Costs and Their Consequences: An Estimated Dynamic Model of African Agricultural Storage and Trade"," *American Economic Journal: Applied Economics*.

Redding, S. J. (2016): "Goods trade, factor mobility and welfare," *Journal of International Economics*, 101(C), 148–167.

Restuccia, D., D. T. Yang, and X. Zhu (2008): "Agriculture and aggregate productivity: A quantitative cross-country analysis," *Journal of Monetary Economics*, 55(2), 234–250.

Sabel, C., E. Fernandez-Arias, R. Hausmann, A. Rodriguez-Clare, and E. Stein (2012): "Export Pioneers in Latin America," .

Santos Silva, J. M. C., and S. Tenreyro (2006): "The Log of Gravity," *The Review of Economics and Statistics*, 88(4), 641–658.

Scoville, W. C. (1951): "Minority migrations and the diffusion of technology," *The Journal of Economic History*, 11(4), 347–360.

Simonovska, I., and M. E. Waugh (2014): "The elasticity of trade: Estimates and evidence," *Journal of International Economics*, 92(1), 34–50.

Sotelo, S. (2018): "Domestic Trade Frictions and Agriculture," .

Suárez Serrato, J. C., and O. Zidar (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," *American Economic Journal: Macroeconomics*, 7(3), 1–33.

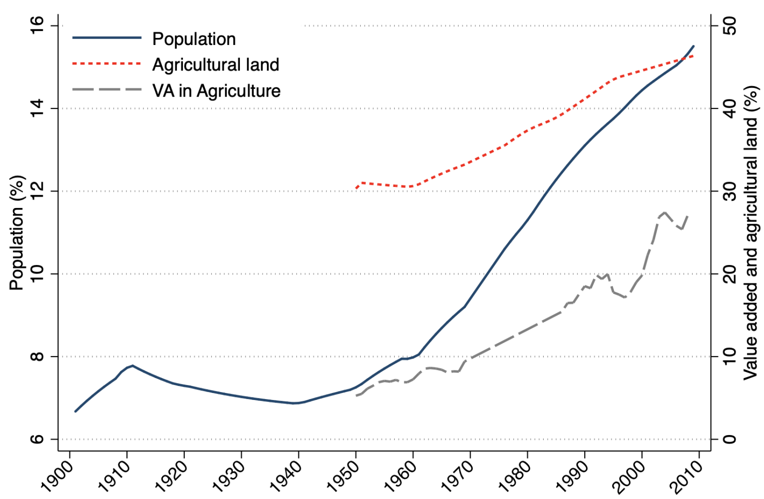
Tombe, T., and X. Zhu (2019): "Trade, migration, and productivity: A quantitative analysis of china," *American Economic Review*, 109(5), 1843–72.

Watson, A. M. (1983): "Agricultural innovation in the early Islamic world; the diffusion of crops and farming techniques, 700-1100," .

8 Tables and Figures

Figure 1: Meso regions and the West of Brazil

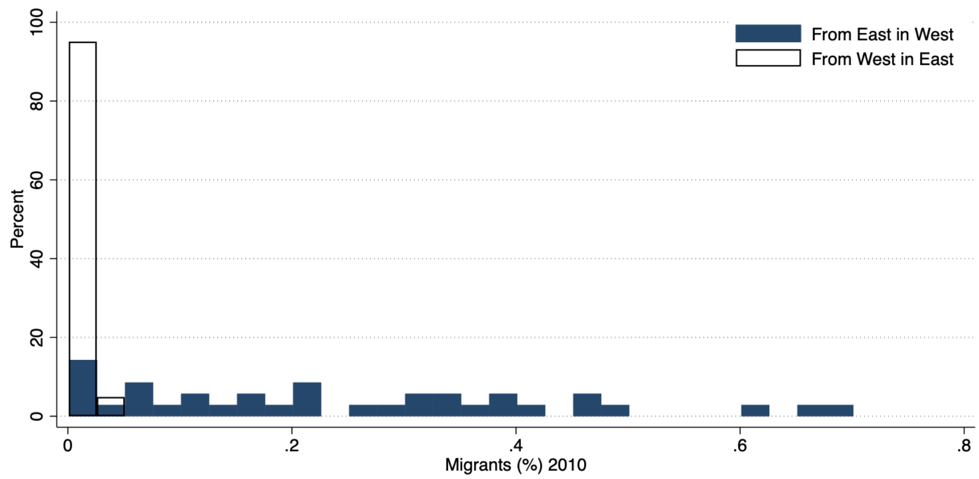
Figure 2: Migration and Agricultural Land Use in the West of Brazil



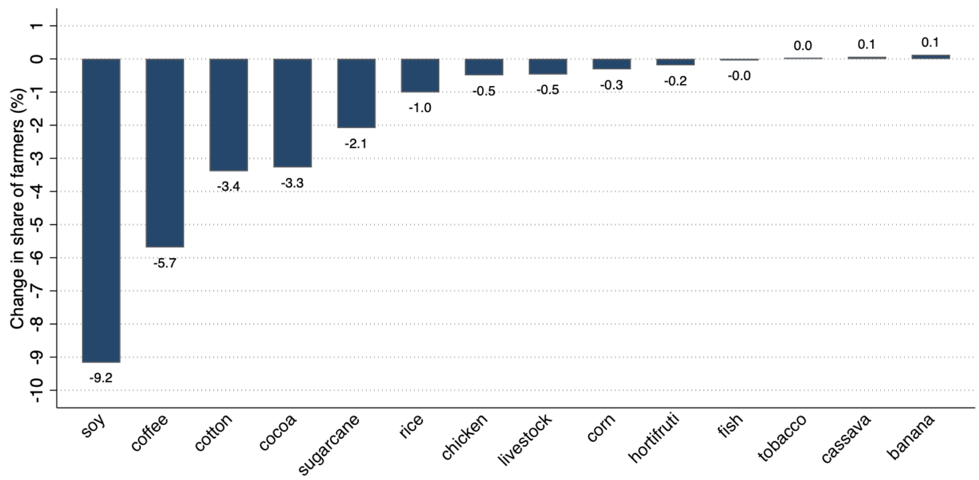
Notes: This figure highlights the increase in the share of the Brazilian population living in the west. This figure shows 3 years moving average.

Figure 3: Regional Migration and Crop-Choices in 2010

(a) Distribution of regional migrants between meso-regions

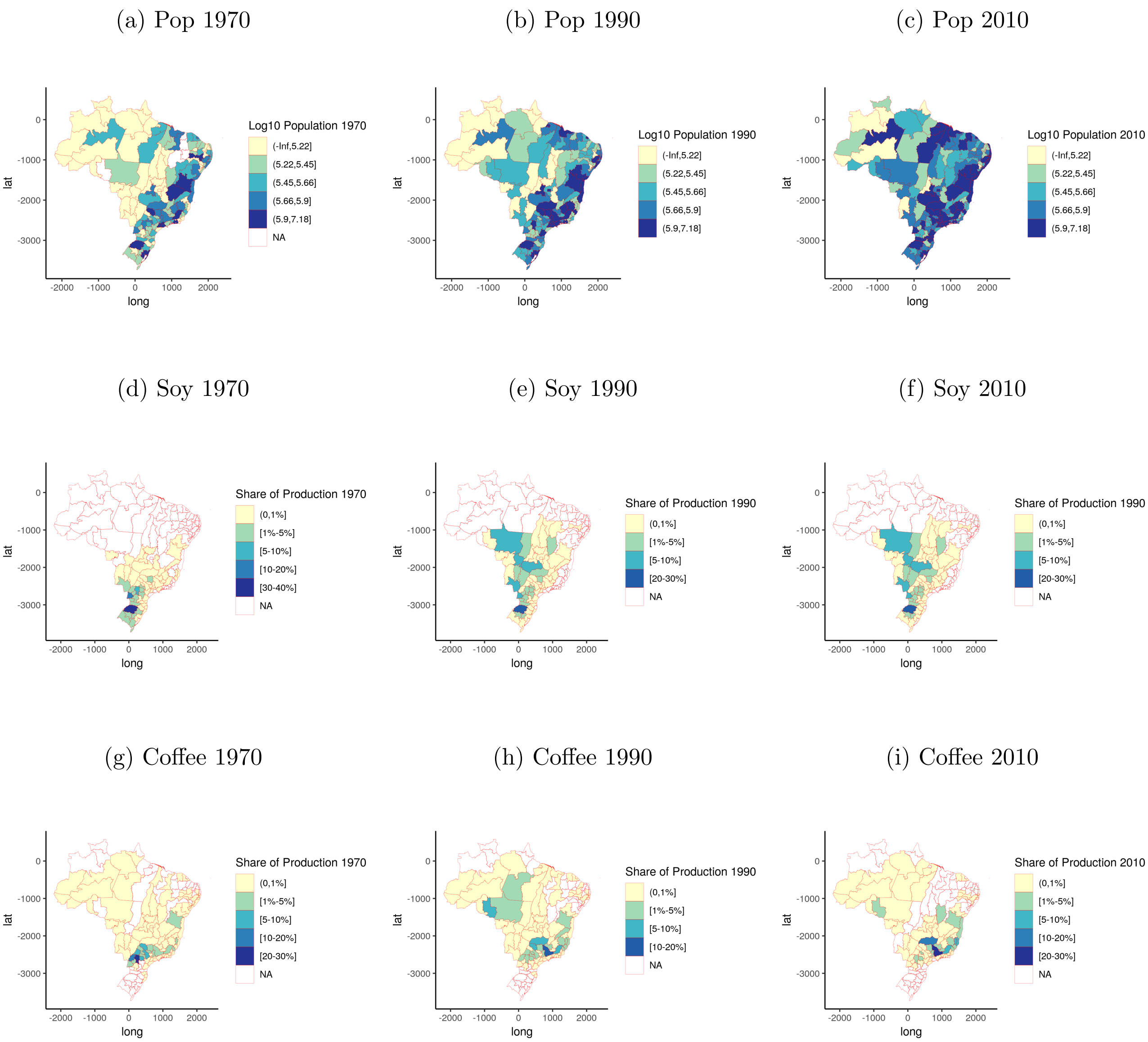


(b) Contribution of migrants from specialized regions in the East to share of farmers in the West



Notes: Panel a shows the distribution of the share of farmers that are a state migrant across regions in the West and in the East of Brazil. Panel b shows the share of farmers in the West who migrated from the East and the share of farmers in the East that migrated from the West. The average share of state migrants among agricultural workers in Brazil is 15%.

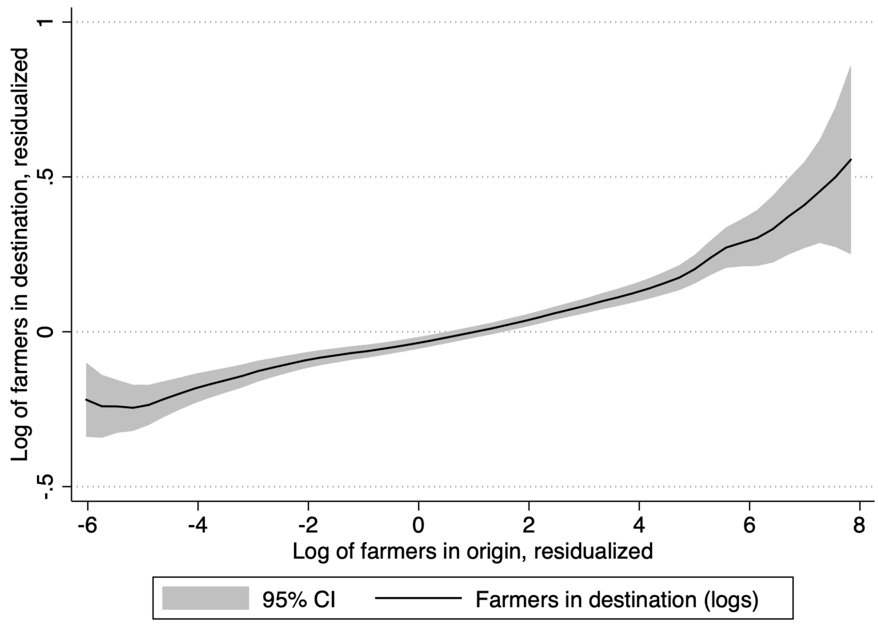
Figure 4: Population and Crop Expansion During the March to the West, 1970-2000



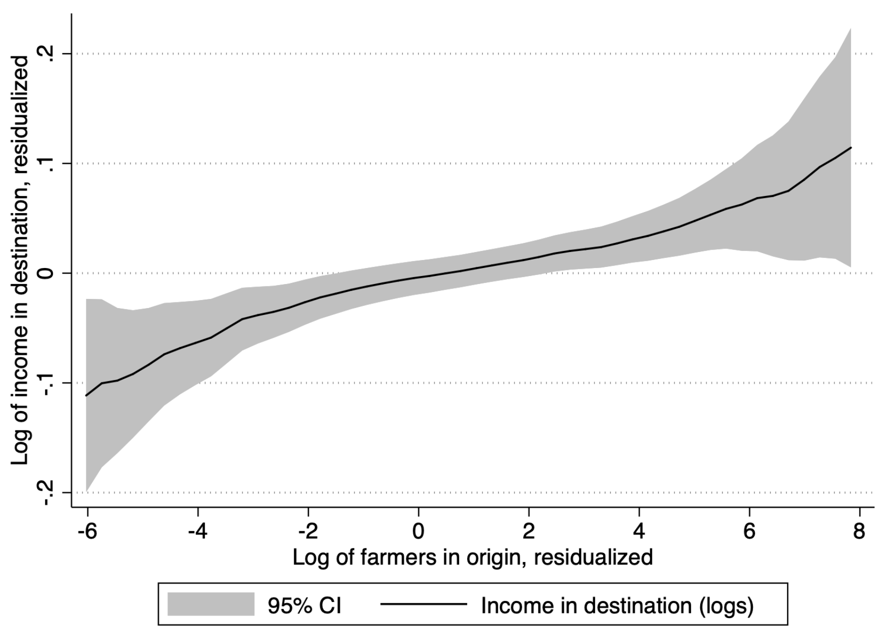
Notes: This figure highlights changes in the spatial allocation of the population and the production of two major crops in Brazil, soybeans and coffee. It shows that the diffusion of crops was not directly associated with the expansion of the population to the region.

Figure 5: Local Polynomial Regressions of the Influence of the Region of Origin on Crop Choice and Earnings of Farmers in their Destination Region

(a) Farmers in destination, residualized



(b) Earnings in destination, residualized



Notes: To compute these local polynomial regressions, we first absorb destination-crop-year fixed effects from each variable.

Figure 6: The Contribution of Knowledge Reallocation to Counterfactual Changes in Soy Exports

(a) Total Change in Exports (b) Contribution of Knowledge Diffusion

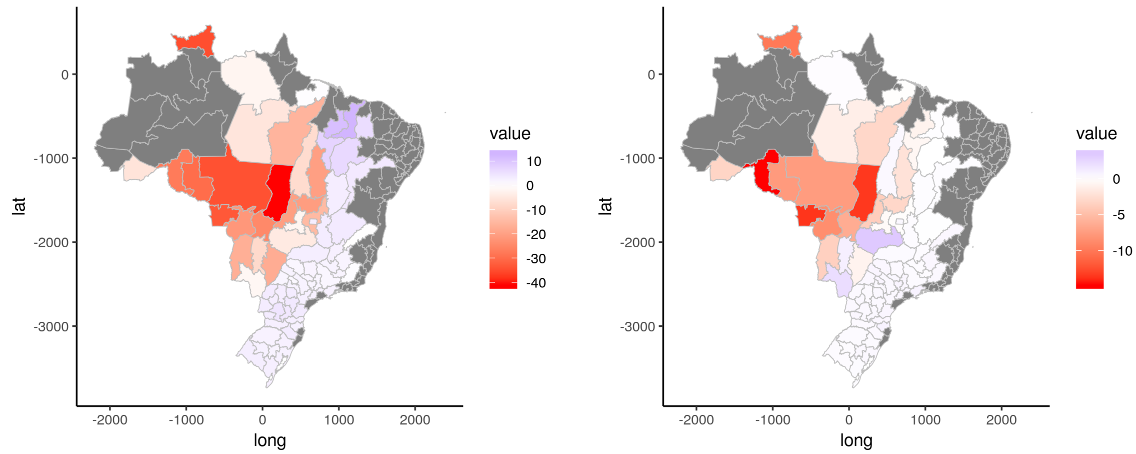
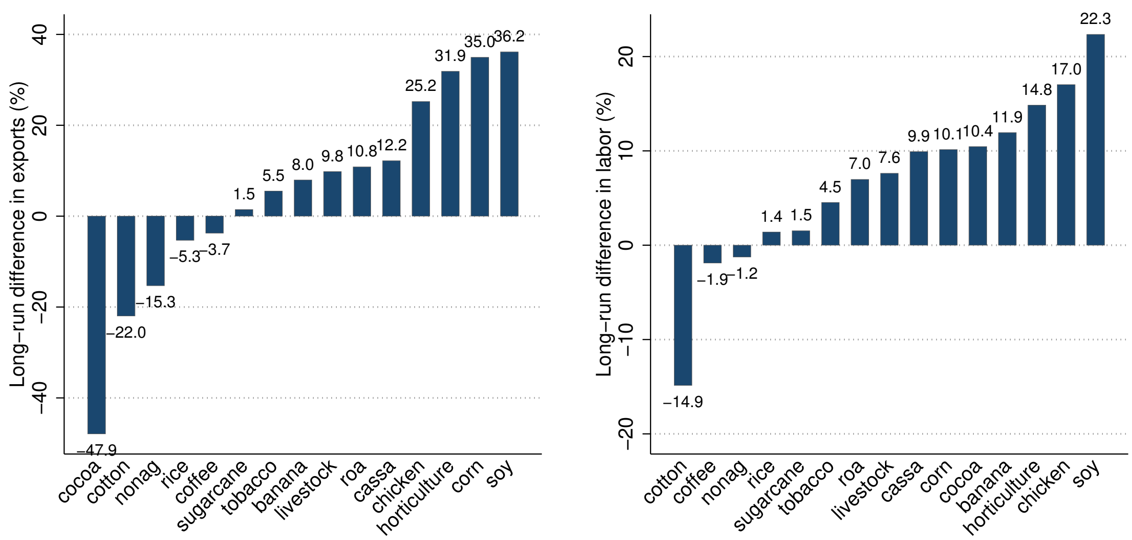


Figure 7: Effects of Knowledge in the Steady State

(a) Exports (b) Farmers



Notes: Results show the difference in the steady state of the model with the knowledge mechanism relative to the model without the mechanism.

Table 1: The Revealed Comparative Advantages of Brazil

Activity	Agriculture	Rest of Ag	Banana	Cassava	Chicken	Cocoa	Coffee	Corn
Brazil RCA 1980	3.74	1.22	1.25	0.2	6.5	15.59	17.28	0.23
Brazil RCA 2010	4.48	0.78	0.4	0.44	16.89	1.39	15.9	5.23
East RCA 2010	2.48	0.79	0.47	0.52	19.86	1.65	18.79	1.72
West RCA 2010	6.14	0.52	0	0	1.03	0	0.08	6.61
Activity		Cotton	Horticulture	Livestock	Rice	Soy	Sugarcane	Tobacco
Brazil RCA 1980		0.28	1.98	2.92	0.13	16.11	6.85	4.94
Brazil RCA 2010		4.95	1.19	10.2	0.98	22.61	29.67	9.5
East RCA 2010		3.36	1.39	6.29	1.14	15	33.57	11.26
West RCA 2010		2.8	0.11	14.54	0.06	43.37	2.07	0

Notes: This table computes the revealed comparative advantages (RCA) of Brazil. RCA 1980 uses data from FAO to compute the RCA in 1980 and 2010. We use the average of 1980-1985 and 2010-2015 to construct our measure of RCA.

Table 2: The Influence of the Region of Origin on Earnings and Employment of Agricultural Workers in their Destination Region

	OLS	OLS	OLS	PPML
	(1)	(2)	(3)	(4)
A. Farmers in destination (logs)				
Farmers in origin (logs)	0.089***	0.074***	0.064***	0.160***
	(0.007)	(0.007)	(0.006)	(0.007)
R^2	0.696	0.730	0.737	-
Obs	8308	8308	10989	46021
B. Earnings (logs)				
Farmers in origin (logs)	0.022***	0.023***	0.022***	0.124***
	(0.007)	(0.007)	(0.007)	(0.012)
R^2	0.716	0.740	0.784	-
Obs	6860	6860	8283	40089
β	0.022	0.023	0.022	0.124
κ	4.045	3.217	2.909	1.290
Destination-Crop-Year FE	✓	✓	✓	✓
Destination-Origin-Year FE	✓	✓	✓	✓
Controls		✓		
Micro-region			✓	

Notes: * / ** / *** denotes significance at the 10 / 5 / 1 percent level. Standard errors clustered at the destination-crop-year level in parenthesis. The unit of observation is a given at the destination-activity-origin-year. Columns 1, 2 and 4 use data at the meso-region. Columns 3 and 5 use data at the micro-region. Controls include: age, age squared, time in municipality, years of education, share of black and share of white. Explanatory variable is the log of agricultural workers in the same activity in the region of origin lagged by thirty years ($L_{ij,k(t-1)}$). We include the census of 2000 and 2010 in our regressions. We exclude non-migrants from sample.

Table 4: Summary of Calibrated Parameters

Parameter	Moment or Source	Parameter Value
β	Fact 2, Income regression	0.1
κ	Fact 2, Both regressions	1.25
γ_k	Value added share of land	crop-specific
ρ	Restuccia, Yang, and Zhu, 2008	0.5
η_k, Agric.	Caliendo and Parro, 2015	9.5
$\eta_{\text{Manufacturing}}$	Simonovska and Waugh, 2014	5.5
σ	Sotelo, 2018	2.5
ζ	Costinot, Donaldson, and Smith, 2016	2.8
τ_{ij}	Pellegrina, 2018	distance$^{0.05}$
τ_{Fs}, τ_{sF}	State-to-Foreign Trade	state-specific
μ_{ij}	Aggregate share of migrants	region-specific
μ_{rs}	State-to-state migration	region-specific

Table 3: Relationship between the Composition of Farmers and Agricultural Output

	(1)	(2)	(3)
	Revenues	Quantity	Prices
Dep var in logs:	(1)	(2)	(3)
Log(Farmers)	0.914***	0.981***	-0.067***
	(0.037)	(0.039)	(0.016)
Log(Composition of Farmers)	0.145***	0.176***	-0.030*
	(0.041)	(0.041)	(0.015)
R^2	0.861	0.834	0.889
Obs	1509	1509	1509
Crop-Year FE	✓	✓	✓
Destination-Year FE	✓	✓	✓

Notes: * / ** / *** denotes significance at the 10 / 5 / 1 percent level. Standard errors clustered at the mesoregion level in parenthesis. The unit of analysis is a mesoregion and crop. The explanatory variable is the log of the weighted average of $L_{i,kt-1}$ according to the share of migrants from this origin in the destination. Regressions include the years of 1990, 2000 and 2010.

Table 5: Effects of Banning East-West Migration on Brazil's Aggregate RCA

Activity	Shr 2010	RCA 1980	RCA 2010	RCA w/o mig	Labor	Knowledge
Non-Agri	68.4	0.63	0.74	0.75	0.88	0.18
Rest of Ag	2.84	1.22	0.78	0.77	0.86	0.13
Banana	0.02	1.25	0.4	0.41	0.9	0.13
Cassava	0	0.2	0.44	0.44	-1.36	3.26
Chicken	3.54	6.5	16.89	17.39	0.94	0.09
Cocoa	0.14	15.59	1.39	1.44	0.91	0.12
Coffee	2.38	17.28	15.9	16.45	0.88	0.14
Corn	0.91	0.23	5.23	3.9	0.87	0.2
Cotton	0.41	0.28	4.95	4.48	0.84	0.22
Horticulture	1.63	1.98	1.19	1.22	0.93	0.11
Livestock	2.37	2.92	10.2	9.38	1.07	-0.09
Rice	0.13	0.13	0.98	1	0.73	0.36
Soy	9.29	16.11	22.61	20.26	0.87	0.2
Sugarcane	4.82	6.85	29.67	30.26	0.92	0.1
Tobacco	3.09	4.94	9.5	9.75	0.92	0.12

Notes: This table shows the decomposition of our counterfactual analysis. The first column shows the change in exports when we ban migration between the east and the west. The second and third column computes the relative importance of abundance of farmers and the knowledge of farmers for the change in export shares.

Table 6: The Gains from Trade and the Gains from Migration

	Gains from Trade		Gains from E-W Migration	
	Base	No Migration	Base	No Trade
All sectors	8.09	7.84	3.18	2.91
Agriculture	3.44	3.35	3.18	3.09

Notes: This table shows the decomposition of our counterfactual analysis. The first column shows the change in exports when we ban migration between the east and the west. The second and third column computes the relative importance of abundance of farmers and the knowledge of farmers for the change in export shares.

A Data

This section describes in detail the collection of datasets and the construction of the variables used in our analysis

A.1 Data

Employment and Migration

Our data on migration and employment comes from the decadal demographic and economic census organized by the Brazilian statistical institute IBGE (*Instituto Brasileiro de Geografia e Estatistica*). We use information from the editions of 1980, 1991, 2000 and 2010. The information from the census is divided in two questionnaires, an universal one with basic questions about education and the family structure, and a sampled one with detailed information on migration and employment. In 1970 and 1980, 25% of the population was sampled for the detailed questionnaire. For 1990, 2000 and 2010, about 25% of the population was sampled in smaller municipalities and 10% for the larger ones. The municipality thresholds defining the sample size depend on the year of the census. To illustrate the final sample size, in 2000, the census included about 12 million individuals in the sampled questionnaire.

Since the census of 1980, we have information about the current and the previous municipality of residence of each invidual in the case of migration. The exception is the census of 2000 which asks individuals their previous state of residence, their municipality of residence in 1995, but not their previous municipality of residence. In this case, we found that for 80% of the observations the state of the municipality in 1995 was the same as the previous state. Therefore, we used the information on the municipality of residence in 1995 as a *proxy* for previous municipality. Since less than 0.1% of the population was born abroad, we ignore international migration.

The census contains a specific module of labor employment with questions about the sector of employment of the worker. In each edition, there are more than 150 sectors. About 25 of them can be classified as agricultural activities. We identified 14 agricultural activities that have definitions that are consistent over time and that can also be found in other datasets used in the paper: banana, cassava, chicken, cocoa, coffee, corn, cotton, fish, horticulture and fruits, livestock, rice, soy, sugarcane and tobacco. In addition, the census ask questions about total revenues of a worker. Since many agricultural workers do not receive their earning in the form of wages, which is the case of managers who are about 30% of our sample, we use data on total income instead of their information on wages. Importantly,this measurement is theoretically consistent with our model.

Agricultural Production

We used data from PAM (*Produção Agrícola Municipal*), which is organizede by the Brazilian census bureau. It contains municipality level data since 1974 for more than 20 crops and state level data since 1930s for a subset of these crops. We combine this data with information from the agricultural census on agricultural revenues from the agricultural census of 1960, 1970, 1996 and 2006, to obtain revenues in livestock and chicken. We organized the agricultural activities defined in the agricultural census and PAM so that they matched the agricultural activities that we identified in the demographic census. We found that the definitions were highly consistent across the datasets: the correlation between the log of total revenues by activity according to the agricultural census on the log of workers employed by agricultural activity according to the demographic census is equal to 0.70.

Trade Flows

The data on trade flows come from the Comexstat, a website organized by the Ministry of Development, Industry and Foreign Trade (MDIC). For each State, we observe how much was exported and imported from abroad. According to MDIC, the trade data at the state level is registred according to the location of production. The data is disaggregated by good according to the harmonized system at the 6 digit level. We classified the trade flows according to our 14 agricultural activities. We focused on the unprocessed versions of each good. For example, for tobacco, we excluded manufactured cigars and, for wheat, we excluded pastry related goods. In our structural estimation, to obtain the global imports of each good, we combined our trade flow data from Comexstat with trade data from FAO-STAT. Since the FAO-STAT does not contain a category for fish, we assumed that 10% of the global trade in agricultural commodities come from trade in fish and fishery products, which is consistent with reports on the fishery industry from FAO.

B Details about the Calibration of the Model

In this section, we describe in detail the equations used in the calibration of the model. As described in the main body of the paper, the calibration follows in three steps.

Recovering transfers (t, t_r)

We have $S+1$ tranfers to calibrate in this section. We use data on revenues per region and sector, R_{ik} and data on imports and exports between states and foreign countries, M_{sk} and X_{sk}.

We use the expression for total income in region i

$$I_i = R_i(1+t)(1+t_r),$$

where R_i is the total revenue in region i ($R_i = \sum_k R_{ik}$). To calculate I_i, we assume that interstate transfers sum to zero

$$0 = \sum_{s\in\mathcal{S}} \sum_{i\in\mathcal{I}_s} \sum_{k\in\mathcal{K}} R_{ik} t_s,$$

and adjust the values of t to match the following aggregate identity

$$\sum_{s\in\mathcal{S}} \sum_{i\in\mathcal{I}_s} \sum_{k\in\mathcal{K}} R_{ik}\,(1+t) = \sum_{s\in\mathcal{S}} \sum_{i\in\mathcal{I}_s} \sum_{k\in\mathcal{K}} R_{ik} + \sum_{s\in\mathcal{S}} \sum_{k\in\mathcal{K}} (M_{ks} - X_{ks}). \tag{14}$$

We can compute t using data on revenues and trade. With values for t, we compute the interstate transfer as

$$\sum_{k\in\mathcal{K}} \sum_{i\in\mathcal{I}_s} R_{ik}(1+t)(1+t_s) = \sum_{k\in\mathcal{K}} \sum_{i\in\mathcal{I}_s} R_{ik} + \sum_{k\in\mathcal{K}} (M_{ks} - X_{ks}). \tag{15}$$

Expression 14 provides 1 equation and 15 provides S equations. Therefore, we are exactly identified on transfers.

Recovering composite productivities ($\overline{T}_{jk}$), natural advantage ($A_{j,k}$) and preferences shifter (a_k)

To recover sectoral productivities, we first have to parametrize our trade costs. As explained in the main paper, we set $\tau_{ij} = (dist_{ij})^{\delta_\tau}$, where $dist_{ij}$ is the travel distance between region i and j. To trade with the foreign sector, we assume that $\tau_{iF,k} = \tau_{sF,k}\tau_{iF}$, $\tau_{Fi,k} = \tau_{Fs,k}\tau_{Fi}$, and that $\tau_{FF,k} = 1$, where τ_{iF} is the trade cost of region i to its respective State capital and $\tau_{sF,k}$ and $\tau_{Fs,k}$ are international trade costs. We have $(I+1)K$ sectoral productivities (A_{ik}), $2SK$ international trade costs ($\tau_{sF,k}$ and $\tau_{Fs,k}$) and K preference shifters to estimate. We use data on revenues per region and sector, $R_{i,k}$ $\forall i \in \mathcal{I}_s$, $\forall s \in \mathcal{S}$ and $\forall k \in \mathcal{K}$, data on imports and exports between states and foreign countries, $M_{s,k}$ and $X_{s,k}$ for $\forall s \in \mathcal{S}$ and $\forall k \in \mathcal{K}$.

Our numerical algorithm calibrates compositive productivities, trade costs and preference shifters as follows. We start with an initial guess of composite sectoral productivities ($\overline{T}_{jk} \equiv (w_{i,k}^{1-\gamma_k} r_i^{\gamma_k}/A_{i,k})^{1-\eta}$) and preference shifter a_k. For each guess, we back out the international trade costs so that the model matches exports and imports in the data. We then compute the implied expenditure of the economy on sector k from region i. Finally, we update the composite sectoral productivities and preference shifter, and keep iterating until expenditures implied by the model match revenues in the data.

We explain the identification of the parameters in detail below.

Starting with a guess of $\overline{T}_{i,k}$, $\tau_{Fj,k}$, $\tau_{jF,k}$, $\overline{T}_{F,k}$ and a_k, $\forall i,k$. We construct price indexes for each region using

$$P_{j,k} = \left(\sum_i \tau_{ij}^{1-\eta}\overline{T}_{i,k} + \tau_{Fjk}^{1-\eta}\overline{T}_{F,k}\right)^{\frac{1}{1-\eta}}.$$

We then construct expenditure shares

$$E_{j,k} = \frac{a_k P_{j,k}^{1-\sigma}}{\sum_k a_k P_{j,k}^{1-\sigma}} X_j,$$

Within each iteration of $\overline{T}_{i,k}$, we adjust a_k so that the model matches the apparent consumption

$$\sum_j E_{j,k} = \sum_{i\in\mathcal{I}_s} R_{ik} + \sum_{s\in\mathcal{S}} (M_{ks} - X_{ks}).$$

Note that the level of a_k is not identified. Therefore, we just normalize one of the values to one.

We calibrate τ_{sFk} to match state level exports as follows. Write region i's exports to the foreign sector $(E_{F,k})$ F in good k

$$X_{ik,F} = \frac{\left(\overline{T}_{ik}\tau_{sFk}\tau_{iF}\right)^{1-\eta}}{(\overline{T}_{F,k})^{1-\eta} + \sum_{s\in\mathcal{S}}\sum_{i'\in\mathcal{I}_s}\left(\overline{T}_{i'k}\tau_{sFk}\tau_{i'F}\right)^{1-\eta}} E_{F,k}.$$

The state level exports are given by

$$X_{sFk} = \sum_{i\in\mathcal{I}_s} \frac{\left(\overline{T}_{ik}\tau_{sFk}\tau_{iF}\right)^{1-\eta}}{(\overline{T}_{F,k})^{1-\eta} + \sum_{s\in\mathcal{S}}\sum_{i'\in\mathcal{I}_s}\left(\overline{T}_{i'k}\tau_{sFk}\tau_{i'F}\right)^{1-\eta}} E_{F,k}. \tag{16}$$

We have $S\times K$ equations for (16), which we use to identify $S\times K$ trade cost parameters τ_{sFk}. Similarly, we calibrate τ_{Fsk} to match state level imports. For imports, we have

$$X_{Fsk} = \sum_{j\in\mathcal{I}_s} \frac{(\overline{T}_{Fk}\tau_{Fsk}\tau_{Fj})^{1-\eta}}{\left(\tau_{Fsk}\tau_{Fj}\overline{T}_{Fk}\right)^{1-\eta} + \sum_{s'\in S}\sum_{i'\in N_{s'}}(\overline{T}_{i',k}\tau_{i'j})^{1-\eta}} E_{j,k}. \tag{17}$$

The equation above gives another $S\times K$ equations, which identify $S\times K$ parameters τ_{Fsk}.

In addition, we have the following set of equations defining revenues per region

$$R_{j,k} = \sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{I}_s} \frac{(\overline{T}_{j,k}\tau_{ji})^{1-\eta}}{\left(\tau_{Fsk}\tau_{iF}\overline{T}_{Fk}\right)^{1-\eta} + \sum_{s'\in S}\sum_{i'\in\mathcal{I}_{s'}}(\overline{T}_{ki'}\tau_{i'j})^{1-\eta}} E_{jk} + \frac{\left(\overline{T}_{ik}\tau_{sFk}\tau_{iF}\right)^{1-\eta}}{(\overline{T}_{F,k})^{1-\eta} + \sum_{s\in\mathcal{S}}\sum_{i'\in\mathcal{I}_s}\left(\overline{T}_{i'k}\tau_{sFk}\tau_{i'F}\right)^{1-\eta}} E_{F,k}$$

This provides $I\times K$ equations, which we use to identify $I\times K$ compositive productivities $\overline{T}_{j,k}$. We also have K equations for revenues in the external sector, which identify the K equations for the external sector composite productivity. Note that the level of the composite sectoral productivity is not identified. Therefore, we normalize all values to according to the average. Finally, to recover $A_{j,k}$, we construct model implied wages and land rents as described in the paper. With these factor prices, we can disentangle $\overline{T}_{j,k}$ from $A_{j,k}$. For the external sector, we assume that changes in the Brazilian economy do not affect factor prices in the world, so that $\overline{T}_{j,k}$ is fixed in our counter-factual scenarios.

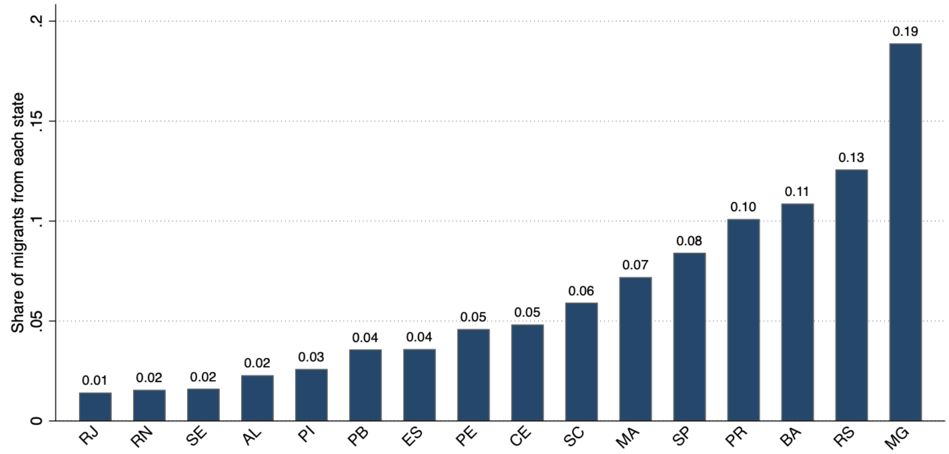
Table 7: The Gains from Trade and the Gains from Migration

	Gains from Trade		Gains from Sorting	
	Base	Random Mig.	Base	No Trade
All sectors	8.09	8.49	3.18	6.50
Agriculture	3.44	3.42	3.18	6.50

Notes: This table shows the decomposition of our counterfactual analysis. The first column shows the change in exports when we ban migration between the east and the west. The second and third column computes the relative importance of abundance of farmers and the knowledge of farmers for the change in export shares.

C Additional Figures and Tables

Figure 8: State of Origin of Migrants in the West coming from the East



Notes: This figure shows the proportion of migrants in the East of Brazil coming from each state of origin in the West. It indicates that migrants come from a large variety of places. For example, 13% of migrants in the west come from Rio Grande do Sul, which is located to the extreme south of Brazil, whereas 11% come from Bahia, which is located in the northeast of the country.

D Migration and Comparative Advantage

In this section, we explain how we derive the conditions for comparative advantage used in Section 4. Following French (2017), we focus on conditions that characterize the relative autarky costs of production between two regions and two crops.

In our trade environment, region j's marginal cost (which equals its price) of producing crop k is given by:

$$\begin{aligned} P^{AUT}_{j,k} &= p_{jj,k} \\ &= c_{j,k}/A_{j,k}. \end{aligned}$$

Therefore, a region i has CA in crop k (relative to region j and crop k') if its autarky relative cost is lower:

$$\frac{P^{AUT}_{i,k}}{P^{AUT}_{i,k'}} < \frac{P^{AUT}_{j,k}}{P^{AUT}_{j,k'}} \Leftrightarrow \frac{c_{i,k}/A_{i,k}}{c_{i,k'}/A_{i,k'}} < \frac{c_{j,k}/A_{j,k}}{c_{j,k'}/A_{j,k'}}. \tag{18}$$

To fix ideas, we specialize to a case in which labor is the only factor of production, meaning:

$$c_{i,k} = w_{i,k}.$$

Since wages are not equalized across sectors, we also need to solve for the equilibrium wages. For simpliciy, assume preferences are Cobb-Douglas, with expenditure shares α_k. Then equilibrium wages are determined by

$$\begin{aligned} \frac{w_{i,k}s_{i,k} \times \lambda_{i,k}}{w_{i,k'}s_{i,k'} \times \lambda_{i,k'}} &= \frac{\alpha_k}{\alpha_{k'}} \\ \left(\frac{w_{i,k}s_{i,k}}{w_{i,k'}s_{i,k'}}\right)^{\kappa+1} &= \frac{\alpha_k}{\alpha_{k'}} \Rightarrow \frac{w_{i,k}s_{i,k}}{w_{i,k'}s_{i,k'}} = \frac{s_{i,k'}}{s_{i,k}}\left(\frac{\alpha_k}{\alpha_{k'}}\right)^{\frac{1}{1+\kappa}}. \end{aligned}$$

Going back to the Comparative Advantage condition (18), we rewrite it as

$$\begin{aligned} \frac{w_{i,k}/A_{i,k}}{w_{i,k'}/A_{i,k'}} &< \frac{w_{j,k}/A_{j,k}}{w_{j,k'}/A_{j,k'}} \Leftrightarrow \\ \frac{A_{i,k}e_{i,k}}{A_{i,k'}e_{i,k'}} &> \frac{A_{j,k}e_{j,k}}{A_{j,k}e_{j,k}}, \end{aligned} \tag{19}$$

where $e_{i,k}$ is the efficiency per worker, which in a single-region model coincides with $s_{i,k}$. Expression (19) shows that CA depends on natural advantage and on efficiency per worker, $e_{i,k}$. Our model is a theory of how migration shapes $e_{i,k}$.