Migration, Specialization and Trade: Evidence from the Brazilian March to the West [Preliminary Draft: Please Do Not Cite or Circulate]

Heitor S. Pellegrina
NYU Abu Dhabi

Sebastian Sotelo
University of Michigan*

February 2018

Abstract

We study how the knowledge that migrants carry over space shapes specialization and trade. Using Brazilian census data, we first document that, upon migration, farmers originating in regions specialized in a crop are more likely to grow that same crop and earn higher incomes than other farmers doing so. Second, we show that the composition of workers in terms of their region of origin correlates with regional exports, after controlling for total sectoral employment. Informed by these facts, we develop and estimate a quantitative dynamic model of trade and migration in which a region's specialization is determined, in part, by the knowledge that migrants in that region carried with them. Applying our model to the large migration of agricultural workers to the west of Brazil since the 1980s, we find that the knowledge carried by migrants contributed substantially to Brazil's recent specialization in exporting commodities, such as soybean.

*We thank Jonathan Eaton, Andrei Levchenko, and John Ham for helpful comments. We also thank participants in the seminar in Purdue Trade Brown Bag. Natalia Martinez Camelo provided excellent research assistance. E-mail: ssotelo@umich.edu and heitor.pellegrina@nyu.edu. All errors are our own.

1 Introduction

Trade and migration are key to understanding welfare and the spatial organization of economic activity. In recent quantitative research, comparative advantage has emerged as a central force driving trade and the spatial patterns of specialization, while spatial frictions of many sorts deter the flows of goods and migrants. Migrants thus shape comparative advantage by changing the abundance of labor across regions, affecting factor proportions and fueling agglomeration effects on productivity. Yet, this line of research has not contemplated that migrants may influence the spatial patterns of specialization by carrying knowledge about how to produce specific goods.[1]

Meanwhile, a large literature has documented historical episodes where migrants carried good-specific knowledge with them as they moved. This was the case, for example, with the introduction of new varieties of wheat in the northeast of the US in the 19th century (Olmstead, Rhode, et al., 2008), the diffusion of Islam and the introduction of wheat in North Africa (Watson, 1983), and the recent formation of several industries in Latin America (Sabel, Fernandez-Arias, Hausmann, Rodriguez-Clare, and Stein, 2012). According to this evidence, migrants deploy knowledge from their region of origin in their destination region, shifting local supply of knowledge and influencing spatial patterns of specialization.[2] As such, migrants shape comparative advantages by changing not only the abundance of labor in a region, but also the stock of good-specific knowledge embodied in its pool of workers.

This paper quantifies the influence of good-specific knowledge acquired by migrants in their region of origin on their production choices in their destination region, and embeds this mechanism into a spatial economy model to examine its aggregate importance to spatial patterns of specialization. We apply our framework to a recent episode of large internal migration of agricultural workers in Brazil. In our setting, workers are endowed with crop-specific knowledge, which they acquire through exposure to economic activity in their region of origin. They 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 region's comparative advantage in a crop reflects a combination of natural advantage, such as land quality, the abundance of labor and the stock of good-specific knowledge in its pool of workers.

[1] Examples of this literature are discussed in Eaton and Kortum (2012), Allen and Arkolakis (2014a) and Allen and Donaldson (2018). See Redding and Rossi-Hansberg (2017) for a review of current models of economic geography allowing for labor migration.

[2] Historical evidence of the diffusion of knowledge via migration also includes the production of flowers by Dutch refugees in England in the late 16th century (Scoville, 1951). More recently, Bazzi, Gaduh, Rothenberg, and Wong (2016) document in Indonesia that migrants knowledge is crop-specific and that skills are not perfectly transferable across locations.

Brazil provides a unique empirical setting to investigate how migrants carry knowledge across space. Besides being an agriculturally diverse country, it has been the stage of a massive migration of workers in its recent history. 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 4.5 million families 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.[3] This period coincided with Brazil's emergence as an important global exporter of crops such as soy and livestock, crops in which the West has a comparative advantage.

To capture the empirical relationship between migration, specialization, and trade, we first assemble a detailed data set, composed of several waves of the Brazilian demographic and agricultural census. We use these data to establish three facts. 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 revealed comparative advantages in crops such as soy and livestock, reflecting in large part the intensity with which the West specializes in these crops.

The next two facts document the influence of a migrant's origin on her crop-choices and earnings. First, we find that a 1 percent increase in the number of farmers growing a crop in an origin region is associated with a 0.14 percent increase in the number of emigrants from that region producing the same crop, compared to all farmers growing that crop in the same destination. Second, a 1 percent increase in the number of workers growing a crop in an origin region is associated with a 0.03 percent increase in the earnings of that region's emigrants, compared again to all other farmers growing that crop in a given destination. Additionally, we document that, as our theory would suggest, regional crop revenues and exports increase when the regional mix of workers favors migrants from regions specialized in those crops (after controlling for the total number of farmers). These observations are consistent with the idea that a farmers' good-specific knowledge is shaped by her sending region's specialization, and with the farmer carrying her knowledge as she migrates.

Motivated by these facts, we embed the diffusion of knowledge via migration in a dynamic quantitative model of migration and trade. 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

[3]In terms of total magnitude, the Braziliam march to the west is comparable to the US Great Migration during the 20th century, a period in which roughly 6 million americans moved away 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 .

to work and which crop to grow. In equilibrium, the comparative advantage of each region is determined by migration 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.

We next bring our model to the data. We show that the two key parameters of the model, namely, (i) the elasticity of a farmer's productivity to crop intensity in her region of origin and (ii) the elasticity of occupational choice to income, can be cleanly identified from reduced form elasticities. These two parameters govern migrants' responsiveness to new opportunities and the extent to which they are capable of shaping aggregate outcomes. We follow standard procedures in the spatial economy literature to estimate migration costs and calibrate the rest of parameters as to match exactly cropping patterns, and regional exports in the year 2010. To understand the impact of migration on specialization and trade, we then compare our model to a counterfactual in which we forbid East-West migration. Relative to the data, the counterfactual largely reduces exports in the West, especially in soy, one of the crops that led the agricultural export boom in Brazil in previous decades. The model suggests that approximately one fifth of this drop is due to our knowledge mechanism.

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: Buera and Oberfield (2016), Levchenko and Zhang (2014), Hanson, Lind, and Muendler (2015) (see Lind and Ramondo (2018) for a summary). More broadly, we relate to the quantitative literature at the intersection of international trade and economic geography, which includes Allen and Arkolakis (2014a), Redding (2016), Caliendo, Dvorkin, and Parro (2015), and Allen and Donaldson (2018). To the best of our knowledge, this literature has not examined migration as a channel of knowledge diffusion, nor its implications for comparative advantage.

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), Pellegrina (2018), and Sotelo (2018). Most of this literature treats comparative advantage as fixed – either arising from quality of land, factor proportions, or both – but does not study the origins of comparative advantage, which is the focus of this paper.

Finally, we relate to a literature that studies the transferability of worker's skills over space and their implications for individual productivity, e.g., Bazzi, Gaduh, Rothenberg, and Wong (2016). In the agricultural context, Olmstead and Rhode (2011) have documented the role of geography and migration in the expansion of wheat 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 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 documents the facts that form the empirical basis of the paper. Section 4 introduces our model of trade and knowledge diffusion through migration. Section 5 uses the model to quantify the strength of our mechanism. Section 6 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.[4] 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

[4] 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.

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 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). Importantly, migrants came from all parts of the East. Figure 7 shows that 10 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.

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 migration of farmers to the East of Brazil. The following passage from Sabel C., et al. (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 the East to the West. 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 the next section, in fact, 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 advantages, 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-intensity 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 557 micro-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 present systematically 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, the information on agricultural production comes from the agricultural census of 1970, 1980, 1996 and 2006. These census provide information on land use, revenues and number of establishments per crop. To obtain production in non-agricultural sectors, we collected data from the Brazilian Statistical Bureau (IBGE)[5] on value added of services and manufacturing.

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 Diffusion

Fact 1 (Comparative Advantages)*: 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 share of crop k in Brazil's exports is larger than that crop's share in world's exports. A number above one suggests Brazil has a comparative advantage in crop k.

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 advantages 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.

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

Fact 2* (Influence of the Origin)*: Upon migration, farmers originating in regions where a crop is prevalent are more likely to grow that same crop and earn higher incomes than other farmers doing so.

To investigate the role of the region of origin of a migrant farmer on her crop-choices and earnings, we estimate

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

where i indexes the origin region, j the destination region, k the agricultural activity and t the year. 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. The term $\iota_{i,t-1}$ is an origin fixed effect that captures any factor that is common for farmers from region i in the previous year $t-1$ such as previous investments in technology. The parameter β 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 2 using two dependent variables: the total number of agricultural workers and the average income of agricultural workers. We use as $L_{i,kt-1}$ the number of farmers lagged by thirty years and experiment with alternatives in the appendix. To avoid including the same farmers in the left and right hand side of the equation, we restrict our sample to migrant farmers. Finally, we present poisson estimates of equation 2 due to the presence of zeros.

Panel A in Table 2 shows OLS estimates of equation (2) on the total number of agricultural workers. Column 1 shows 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 0.15%. Column 2 shows that controlling for socio-economic characteristics of each group of farmers have a small effect on our point estimates, indicating the differences in overall education is unlikely to be driving our results. Column 3 confirms that the elasticities exist for the sample of farmers producing in the west. Finally, in column 4, we restrict our analysis only to migrants who are more than 500 km away from their previous region. Even in this case, where we would expect substantial differences in natural advantages to minimize the effect of the origin, we still find a strong relationship. Panel B shows that the elasticity of earnings with respect to the number of farmers in the origin is between 0.03%-0.035% across specifications.

Fact 3* (Farmers Composition)*: Controlling for the abundance of farmers in a region and crop, agricultural revenues is strongly associated with the composition of farmers in terms of their region of origin.

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

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

where $\lambda_{ij,kt}$ is the share of workers from i in j producing crop k in year 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 that a 1 percent increase in the abundance of farmers in a region is associated with a 0.65% increase in revenues. Perhaps more surprising is the fact that a 1 percent increase in the average number of farmers in the origin is associated with a 0.25% 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. This 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.

Taken together, the last two micro-economic facts 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. Our results alsosuggest that this good-specific knowledge tends to be associated with higher labor productivity in in agriculture. These two micro-economic facts indicate that good-specific knowledge brought about by migrants to the West of Brazil might have played an important role in recent changes in Brazil's RCA. However, the facts from this section do not allow us to gauge the aggregate importance of migrants' knowledge. In the next section, we address this question by introducing the diffusion of good-specific knowledge via migration into a general equilibrium model that allows us to answer questions about economic aggregates.

4 Model

In this Section, we introduce into a dynamic spatial economy model the diffusion of good-specific knowledge via migration. Our goal in the formulation of the model is twofold. First, to give a structural interpretation to the reduced form regressions in the previous section. Second, to use the model to examine the aggregate importance of the diffusion of knowledge via migration. Below, we present a simplified version of the model to save on notation. Later, we bring additional features to take the model to the data.

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 of good 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 vector of productivities, and endowment of land, a vector of bilateral trade costs and a vector of bilateral migration costs $\{A_{j,kt}, H_{j,t}, \tau_{ij,kt}, \mu_{ij,kt}\}$. We drop 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

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

where γ_k measures the land intensity of crop k.

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 children utility. L_{jt} is adult population at time t in j. 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.

Children born in i in year $t-1$ are characterized by a vector of location and occupation preference draws, $\varepsilon = (\varepsilon_{ik})$. Reflecting our empirical findings, a worker's skill depends on the crop and on the region he comes from. Thus, a worker from region i has skill s_{ik} in crop k.

Given the wage per efficiency unit of labor, $w_{j,kt}$, workers choose where to live and work in t by maximizing welfare:

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

where preference shocks are drawn iid from $G(\varepsilon) = \exp(-\varepsilon^{-\kappa})$ and W_{ijkt} is the systematic component of welfare. This systematic component is equal to earnings $w_{j,kt}s_{i,kt}/\mu_{ij,kt}$ divided by the CES price index of consumption P_{jt}. The CES price indexes are given by $P_{j,t} = (\sum_k a_{k,t}^{\frac{1}{\sigma}} P_{j,kt}^{\frac{\sigma-1}{\sigma}})^{\frac{\sigma}{\sigma-1}}$ and $P_{j,kt} = (\sum_k a_{j,kt}^{\frac{1}{\sigma}} P_{j,kt}^{\frac{\eta-1}{\eta}})^{\frac{\eta}{\eta-1}}$, where $a_{k,t}$ and $a_{j,kt}$ are preference shifters.

Knowledge Endowment

Workers knowledge to grow particular crops depends on their region of origin: $s_{i,kt} \equiv s(L_{i,kt-1})$. 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:

$$s_{i,kt} = \bar{s}_{i,kt-1}L_{i,kt-1}^{\beta}. \tag{4}$$

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 and output are linear 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 optimal output is

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

where $\kappa_q = (\gamma_k)^{\gamma_k/(1-\gamma_k)}$.

The earnings of the farmer, which we will match to the data, correspond to the profits in the model, which are also linear in the farmer's knowledge:

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

To define the equilibrium, it will be useful to write output as a function of output prices. Defining efficiency wages as $w_{j,k} \equiv \pi_{j,k}(s)/s$, i.e. as earnings per unit of knowledge, equilibrium prices are proportional to the marginal cost of producing with one unit of knowledge:

$$p_{j,k} = \frac{w_{j,k}^{1-\gamma_k} r_j^{\gamma_k}}{\kappa_\pi A_{j,k}}.$$

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

$$\pi_{ij,kt} = \frac{\left(w_{j,kt}^{1-\gamma_k} r_{j,t}^{\gamma_k}/A_{j,kt}\right)^{(1-\eta)}}{\sum_{i'} \left(w_{i',kt}^{1-\gamma_k} r_{i',t}^{\gamma_k}/A_{i',kt}\right)^{(1-\eta)}}.$$

Next, optimal worker sorting gives the share of workers from i choosing to work in region j and sector k, $\lambda_{ij,kt}$:

$$\lambda_{ij,kt} = \frac{\left[w_{j,kt} s_{i,kt}/\left(\mu_{ij,kt} P_{j,t}\right)\right]^\kappa}{\Xi_{i,t}^\kappa} \tag{5}$$

where and $\Xi_{i,t}^\kappa \equiv \left(\sum_j \sum_k \left[w_{j,kt} s_{ikt}/\left(\mu_{ij,kt} P_{j,t}\right)\right]^\kappa\right)$. 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

$$E_{ij,kt} \equiv s_{i,kt} \lambda_{ij,kt} L_{i,t-1}. \tag{6}$$

Finally, to simplify our exposition, we assume that every region has a landowner who spends her income locally. Later, we collect land rents throughout the country in a single national asset that is distributed proportionally to labor income, thus avoiding assigning land ownership to new migrants into 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.[6]

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

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.

Goods market equilibrium in period t.

Given parameters, 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:

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

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

2. Land markets clear in region j:

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

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

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

   where the function $s_{i,kt}$ is defined in equation (4).

In the definition above, we note explicitly that the 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 parameters 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:$\{L_{ij,kt}$, $w_{i,kt}(L_{ij,kt})$, $r_{i,t}(L_{ij,kt})$, $E_{i,kt}(L_{ij,kt})\}$ 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:

$$L_{ij,kt} = \lambda_{ij,kt}L_{i,t-1}, \tag{7}$$

exposition.

where $\lambda_{ij,kt}$ is given by equation 5.

Finally we define a the equilibrium for this dynamic model.

Competitive equilibrium.

Given a sequence parameters 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, labor allocations, and prices, $\{L_{ijkt}, w_{i,kt}, r_{i,t}, E_{i,kt}\}_{t=1}^{\infty}$ that satisfy the goods market and migration equilibria in each period t.

Steady State Equilibrium.

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

Intuitively, we have a number of dispersion forces in the model. First, the idiosyncratic draws are a force towards populating all region-crop cells. The strength of this force is governed by the inverse of κ: as κ decreases, individuals have stronger idiosyncratic tastes. 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 σ: as σ 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 go there.

The opposing, agglomeration force, is given by the diffusion 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.

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.

5 Taking the Model to the Data

This section describes how we quantify the exogenous parameters of the model $\{\kappa\ ,\ \beta,\ \theta,\ \gamma_k,\ \tau_{ijk,t},\ \mu_{ij,kt},\ A_{j,kt},\ H_{j,t}\}$. In doing so, we map the model's goods 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. First, we connect 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. Taking these elasticities as given, we then 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 start by using the model to uncover two key parameters, one of which is new to our theory. First, using equation (4), and letting $\mu_{ij,kt} = \mu_{j,kt}\mu_{ij,t}\epsilon^{\mu}_{ij,kt}$, and $\bar{s}_{i,kt-1} = \bar{s}_{i,t-1}\epsilon^{s}_{i,kt}$, our model suggests that the income of migrants from i into j growing crop k ($I_{j,kt}$) relates to the scale of that crop at the migrants' origin:

$$\begin{aligned}\log I_{j,kt} &= \log\left(w_{j,kt}s_{i,kt}\right)\\ &= \iota_{j,kt} + \iota_{i,t-1} + \beta\log L_{i,kt-1} + u^{earnings}_{i,kt},\end{aligned} \tag{8}$$

where we defined $\iota_{i,t-1} \equiv \log\bar{s}_{i,t-1}$, $\iota_{j,kt} \equiv \log w_{j,kt}$ and $u^{earnings}_{ij,kt} = \log\epsilon^{s}_{i,kt}$. 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 migration costs and knowledge in $\tilde{s}_{i,kt}$ and $\tilde{\mu}_{ij,kt}$ are 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 (4) and (5) into (7), we obtain our econometric specification:

$$\begin{aligned}\log L_{ij,kt} &= \log(\lambda_{ij,kt}L_{i,t-1})\\ &= \iota_{j,kt} + \iota_{ij,t} + \kappa\beta\log L_{i,kt-1} + u^{migration}_{ij,kt},\end{aligned} \tag{9}$$

where $\iota_{ij,t} = \log L_{i,t-1} + \kappa(\log\bar{s}_{i,t-1} - \log\mu_{ij,t} - \log\Xi_{it})$, $\iota_{j,kt} = \kappa(\log w_{j,kt} + \log\bar{s}_{k,t-1} - \log\mu_{j,kt})$ and $u^{migration}_{ijkt} = \kappa\log\epsilon^{\mu}_{ij,kt} - \kappa\log\epsilon^{s}_{i,kt}$. Knowledge of β allows us to disentangle κ from the coefficient in the previous regression.[7]

As in common in the empirical trade literature, our data contain a number of zero migration flows[8]. Therefore, we estimate equations (8) and (9) by Poisson Pseudo ML (Santos Silva and Tenreyro, 2006), which allows us to include observations where the dependent variable is zero.[9] Since our data set is comprised by samples of the Brazilian census, we interpret the errors in the equations above as sampling errors, thus keeping the estimation consistent with our theory.

Table 4 shows results from PPML estimates of equations (8) and (9). As in the OLS estimates, we focus on the sample of migrants to avoid including the same farmers in the right and left hand side of the equation. We estimate these regressions using only data from 2010. Due to computational constraints, we include only origin fixed effects in each specification, but we control for bilateral distance between regions to partially control for unobserved shocks that are origin and destination specific. We find point estimates that are substantially larger than our OLS results in table 2. 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. For our estimates, we pick a value inbetween our OLS and PPML estimates $\beta = 0.05$. For the dispersion in preferences, our OLS and PPML results are similar and we pick $\kappa = 2$.

[7]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

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

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

[8]We observe positive flows for about 10% of all the potential flows of migrants from a given origin i and any destination j and crop k.

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

5.2 Calibration

We proceed in four steps. First, we calibrate regional transfers and consumption parameters so that our model is consistent with data on trade flows and revenues. Second, using our estimates of β, we construct the stock of efficiency labor in each region and sector and calibrate land productivity to match data on revenues. With productivities in hand, we compute real income per region and sector implied by the model and calibrate the migration costs to match migration flows in the data.

Data for calibration.

We calibrate the model to match the economic activity of 2010 and the previous location of workers of 1980. We observe revenues, trade flows, foreign expenditure, and total land use $(R_{j,k}, X_{r,k}, E_k^F, H_j)$ in the year 2010. We adopt the following procedure to make migration flows consistent with our modeling choices. We use data from 2010 to compute the number of workers in each origin in 1980, we then multiple each of these aggregates by the share of workers in each activity and region given by the census of 1980 to obtain the total employment of workers in 1980 $(L_{j,kt-1})$.

Transfers.

We introduce transfers between regions in the model to account for the trade imbalances in the data and to remove the direct influence of land rents on migration decisions.[10] Each region j belongs to a state $r = 1, ..., R$ and I_r is the set of regions included in state R. There is a national portfolio that aggregates transfers from different sources and redistribute them to workers in the form of three transfers that are proportional to wages. First, land rents are transfered to a national portfolio and then redistributed equally to workers so that $t = \sum_j r_j H_j / \sum_j \sum_k w_{j,k} L_{j,k}$. Second, we treat the aggregate trade imbalance of Brazil as a transfers from foreign markets to the national portfolio that is equally redistributed to workers so that $t^F = \sum_r \sum_k (X_{r,k} - M_{r,k}) / \sum_{j \in I_r} \sum_k w_{j,k} L_{j,k}$, where $X_{r,kt}$ are exports and $M_{r.kt}$ are imports from state r in sector k. Third, the remaining trade imbalance between the foreign market and Brazilian states becomes additional foreign transfers that are unequally distributed to workers according to their state of residence so that $t_r^F = \sum_k (X_{r,k} - M_{r,k}) / \sum_{i \in I_r} \sum_k w_{i,k} L_{i,k}$. Wages after transfers are $w_j(1+t)(1+t^F)(1+t_r^F)$. This structure gives flexibility in our calibration to perfectly match our trade and revenue data.

[10] This approach avoids migration based on the payments to factors of production other than labor.

Production parameters, trade elasticities and trade costs between regions

Finally, we must also pick parameters for the transmission of knowledge (β), the migration elasticity (κ), the factor shares (γ_k), the trade elasticity (η) and the trade cost (δ). Following our reduced form estimates, we set the transmission of knowledge (β) to 0.035 and the migration elasticity (κ) to 2, which is the ratio of the reduced form elasticities of employment and wages with respect to the number of workers in the origin producing the same crop. Note that, while we use a different source of variation for the estimation of the migration elasticity, relative to the literature (Monte, Redding, and Rossi-Hansberg, 2015; Suárez Serrato and Zidar, 2016; Bryan and Morten, 2015), our result is in the ballpark of recent estimates.[11] We set the trade elasticity (η) to 4, following Simonovska and Waugh (2014). Finally, we set $\delta = 0.25$ and distance $dist_{ij}$ in thousands of kilometers which is the average trade cost across crops obtained in Pellegrina (2018) using price data from Brazil.

Land productivity and international trade costs.

For the calibration, we assume that $\sigma \to 1$ so that preferences collapse to Cobb-Douglas between sectors with α_k representing the share of expenditure and we normalize $\bar{s}_{i,kt-1} = 1$. Following Allen and Arkolakis (2014b), we derive the following result for the calibration of our model in the appendix: given parameters $(\gamma_k, \beta, \kappa, \alpha_k, \eta)$, bilateral trade costs and transfers from the national asset and the foreign sector $(\tau_{ij,t}, t_t, t^F_t, t^F_{r,t})$ and data on revenues, trade flows, foreign expenditure, bilateral migration, land endowments, and previous employment $(R_{j,kt}, X_{r,kt}, X^F_{kt}, \lambda_{ij,kt}, H_{j,t}, L_{j,kt-1})$, there exists unique values of migration costs (up-to-scale within each origin) and sectoral productivities (up-to-scale within each sector) $(\mu_{ij,kt}, A_{j,kt})$ that are consistent with the data being an equilibrium of the model. Using this result, we calibrate the model in three steps. First, we calibrate expenditure shares and transfers. Second, we parametrize trade costs and calibrate sectoral productivities. Third, we calibrate migration costs given sectoral productivities.

To calibrate expenditure shares (α_k) and transfers (t, t^F, t^F_r), we start by computing total foreign transfers (t^F) as the ratio of the trade deficit to total revenues in Brazil $(t^F = \sum_{k \in \mathcal{K}} \sum_{s \in \mathcal{S}} (X_k - M_k) / \sum_k \sum_j R_{j,k})$ and use it to calculate national income as total revenues after foreign transfer $(I = \sum_j \sum_k R_{j,k}(1 + t^F))$.[12] We then compute the appar-

[11] 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.

[12] Since our model does not contain an input-output structure, total revenues in our model correspond to total value added.

ent national consumption ($E_k = \sum_j \sum_k R_{j,k} - \sum_r \sum_k (X_{rk} - M_{rk})$ and take the ratio of the apparent consumption to total income as expenditure shares (α_k). We calibrate the transfer from the national asset as the ratio of total revenues to total payments to labor ($t = \sum_j \sum_k R_{j,k} / \sum_i \sum_k (1 - \gamma_k) R_{j,k}$). Once with these two transfers, we compute the total apparent consumption of each state ($E_r = \sum_j \sum_k R_{j,k} - \sum_k (X_{r,k} - M_{r,k})$) and calibrate the federal transfer to each state as the ratio of apparent consumption to total state income before federal transfers ($t_r^F = E_r / \sum_{j \in I_r} \sum_k R_{j,k}(1 + t^F)$).

With expenditure shares and transfers in hand, we move to the calibration of sectoral productivities ($A_{j,k}$). We start by parameterizing the bilateral trade cost within Brazil according to $\tau_{ij} = (1 + dist_{ij})^{\delta}$, where $dist_{ij}$ is the travel distance between region i and j. To trade with the foreign sector, we assume that every region must first 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 $\tau_{Fi,k} = \tau_{Fr,k}\tau_{iF}$.[13] With this parametrization, sectoral productivities ($A_{j,k}$) and international trade costs ($\tau_{rF,k}$ and $\tau_{Fr,k}$) are calibrated to match data on revenues and aggregate trade flows between states and the foreign region through a numerical algorithm. In practice, before recovering $A_{j,k}$, we first recover $\overline{T}_{j,k} = A_{j,k} c_{j,k}^{-\theta}$ using the numerical algorithm, we then back-out $A_{j,k}$ using unit costs of production $c_{j,k}$ that are model consistent. We assume that shocks to the Brazilian economy do not affect the costs of production in the rest of the world, so that $c_{Fk} = \bar{c}_{Fk}$.

Migration costs

We calibrate migration costs so that it perfectly matches the data on migration flows. To do so, we compute wages in each sector and region implied by the model as well as the price index. We then use our parametrization of the good-specific knowledge to measure the real income for workers from every origin region working in every destination and crop. Finally, given real income, we calibrate the migration costs that would rationalize the mass of workers from this region migration to each destination and crop.

6 The Aggregate Effects of Knowledge Diffusion

In this section, we quantify the effects of technological diffusion via migration to agricultural activity in Brazil. 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 abundance of raw

[13] We assume that the distance of the foreign sector to itself is zero so that $\tau_{FFk} = 1$.

labor from the contribution of knowledge diffusion.

6.1 Quantifying the mechanism

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 migration flows between 1980 and 2010, and the 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, as we explain next, 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_{jkt}/L_{jkt}. 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.

Table 5 shows the effects of banning East-West migration on the export share of each crop (in percentage changes). Under the assumption that the world's response to this shocks is negligible, this table presents changes in Brazil's comparative advantage. The first column shows the total change in going from the baseline to the no-migration scenario. A main message of the table is that migration contributed to the soybean export boom, helping soybean's export share grow. In the next two columns, we break down the contribution of each mechanism, as we discussed in the previous paragraph. The column labeled "Labor Abundance" presents the change in exports that comes from removing knowledge effects, expressed as a fraction of the total change induced by banning migration. Across crops, changes in labor abundance account for about 80 percent of the total. The column labeled "Knowledge" measures the strength of the mechanism we propose, as it abstracts from raw labor movements. As the table shows, migrants' knowledge accounts for a varying fraction of total changes, and about 30 percent on average –the same as for soybeans. The sum of the two columns is close to one, suggesting there are relatively small interactions between these two effects.

To highlight the role of regional specialization in these aggregate changes, Table 6 reproduces this calculation focusing now on the West's export shares. The table shows that the aggregate changes are driven in large part by the changes experienced by regions in the West. The West's comparative advantage in soybeans drops substantially, as this crop was disproportionately affected by the lack of migrants' knowledge.[14] Part of the reason we observe this drop for soybeans and not other crops is that there was limited production of soybeans in the East previous to 1980. As a consequence, there are few workers from the East with knowledge to employ in the production of soybeans in the East itself.

7 Conclusions

We have embedded a new mechanism into an otherwise standard, dynamic model of trade and migration, to study how the knowledge carried by migrants shapes aggregate specialization and trade. We find that, during the March to the West – which entailed a large transformation of Brazil's spatial organization and international trade profile – this mechanism is quantitatively important. Had there been no migration to the West, Brazil's comparative advantage in soybean would be substantially weaker, and at least 20 percent of that reduction would be due to migrants not bringing their knowledge along.

[14] The share of other exports increases noticeably, reflecting that their export shares are quite low at the baseline in the year 2000. In the Appendix we show that export shares in the East are much more stable, reflecting the fact that knowledge flowed from previously populated regions to the agricultural frontier.

References

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

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

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

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

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.

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.

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.

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

Ireland, C. T., and S. Kullback (1968): "Contingency tables with given marginals," *Biometrika*, 55(1), 179–188.

Levchenko, A. A., and J. Zhang (2014): "Ricardian productivity differences and the gains from trade," *European Economic Review*, 65(C), 45–65.

Lind, N., and N. Ramondo (2018): "Innovation, Knowledge Diffusion, and Globalization," Working Paper 25071, 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*.

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

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

Redding, S. J., and E. Rossi-Hansberg (2017): "Quantitative spatial economics," *Annual Review of Economics*, 9, 21–58.

Rogers, A. (1968): *Matrix analysis of interregional population growth and distribution*. Univ of California Press.

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.

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

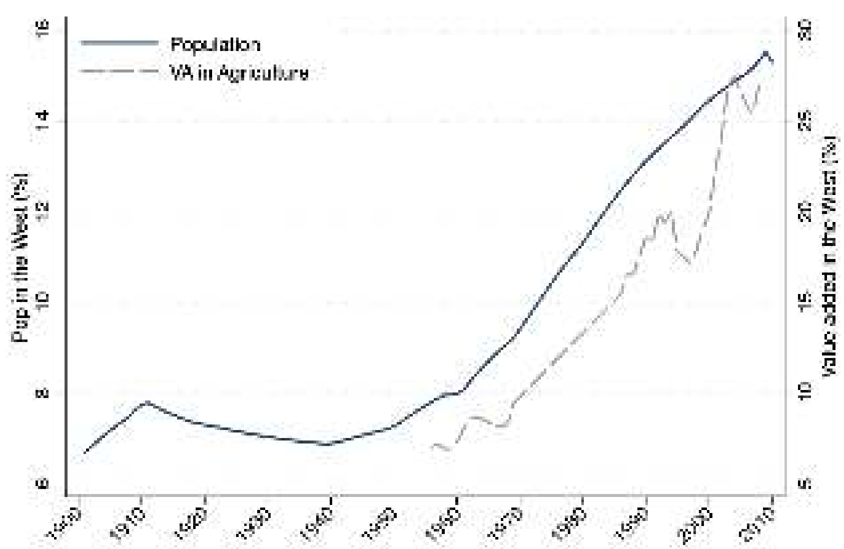
Suárez Serrato, J. C., and O. Zidar (2016): "Who benefits from state corporate tax cuts? A local labor markets approach with heterogeneous firms," *American Economic Review*, 106(9), 2582–2624.

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

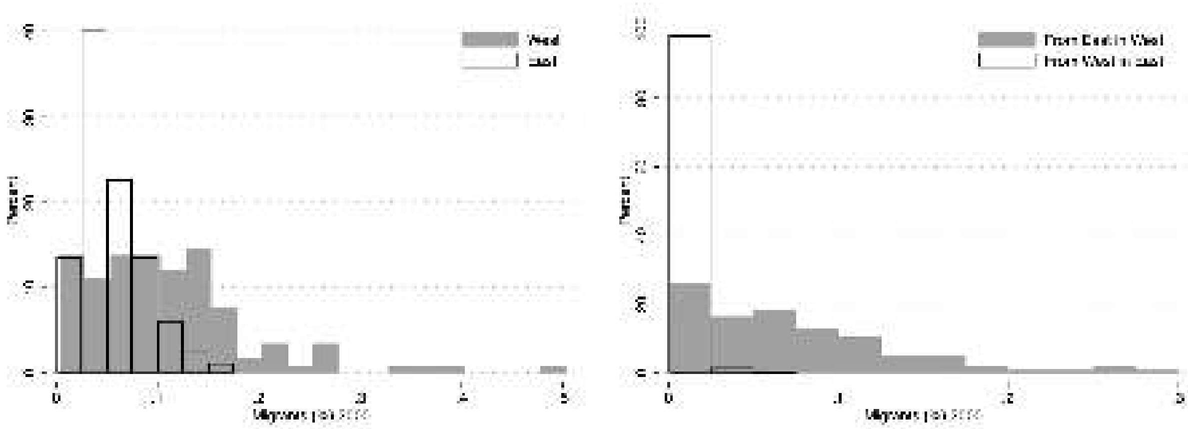


Notes: This figure highlights the increase in the share of the Brazilian population living in the west. This figure shows 3 years moving average.

Figure 3: Distribution of Migrants among Agricultural Workers across Meso Regions in 2000

(a) Migrants in the East and in the West

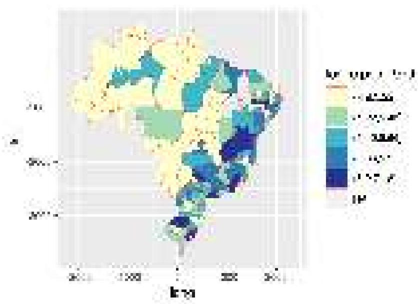
(b) Migrants from the East in the West



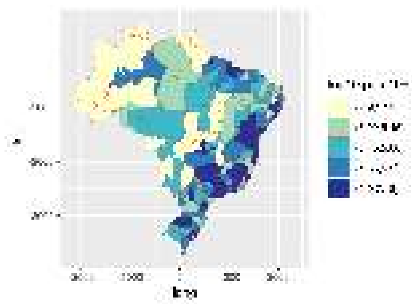
Notes: Panel a shows the distribution of the share of farmers that are a state migrant across regions in the West and in the East of Brazil. Panel b shows the share of farmers in the West who migrated from the East and the share of farmers in the East that migrated from the West. The average share of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

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

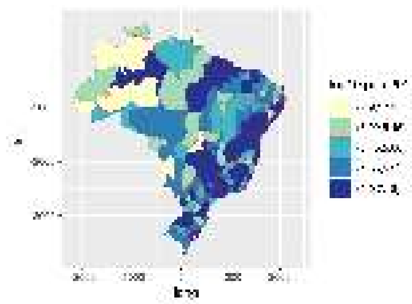
(a) Pop 1970



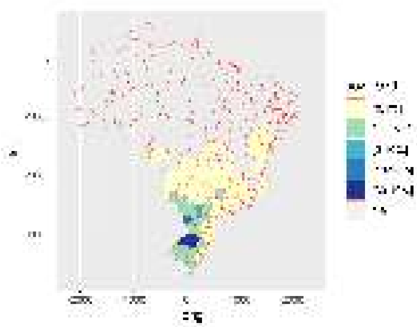
(b) Pop 1990



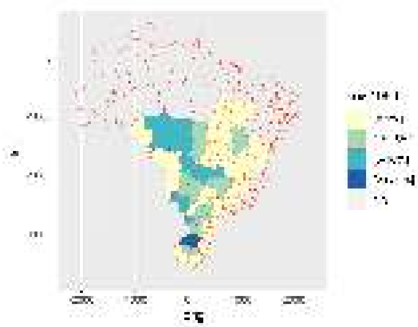
(c) Pop 2010



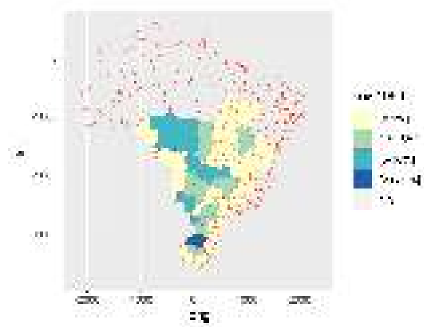
(d) Soy 1970



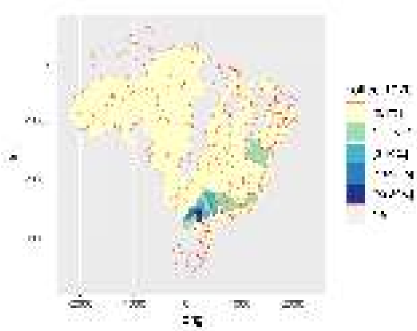
(e) Soy 1990



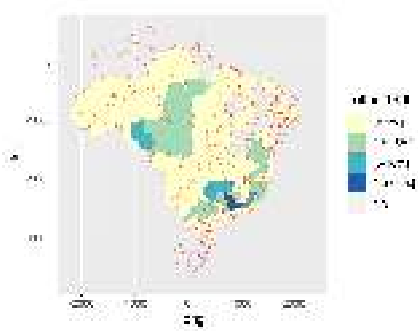
(f) Soy 2010



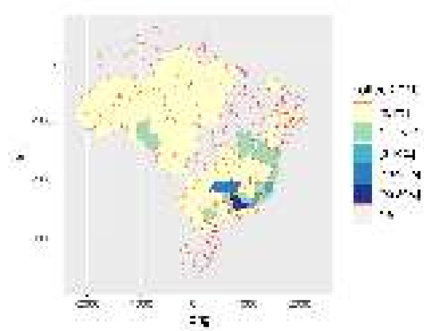
(g) Coffee 1970



(h) Coffee 1990



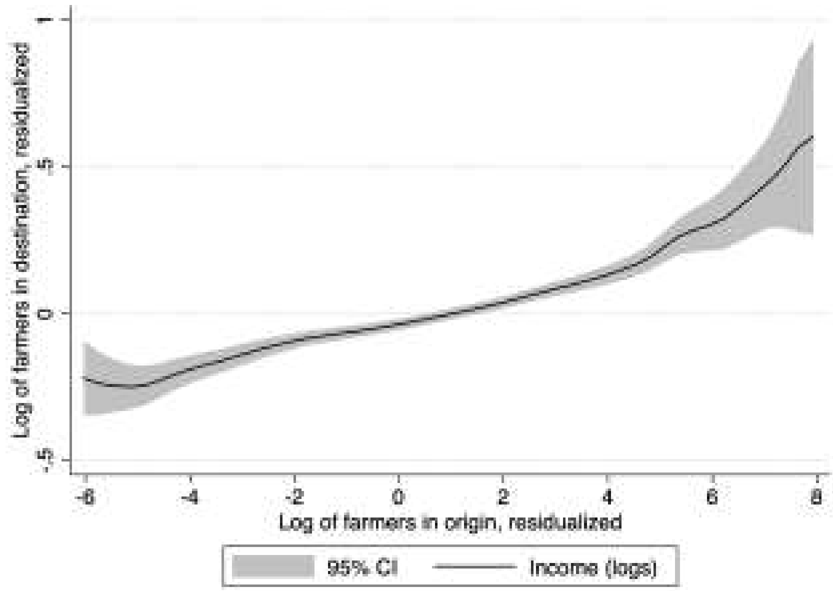
(i) Coffee 2010



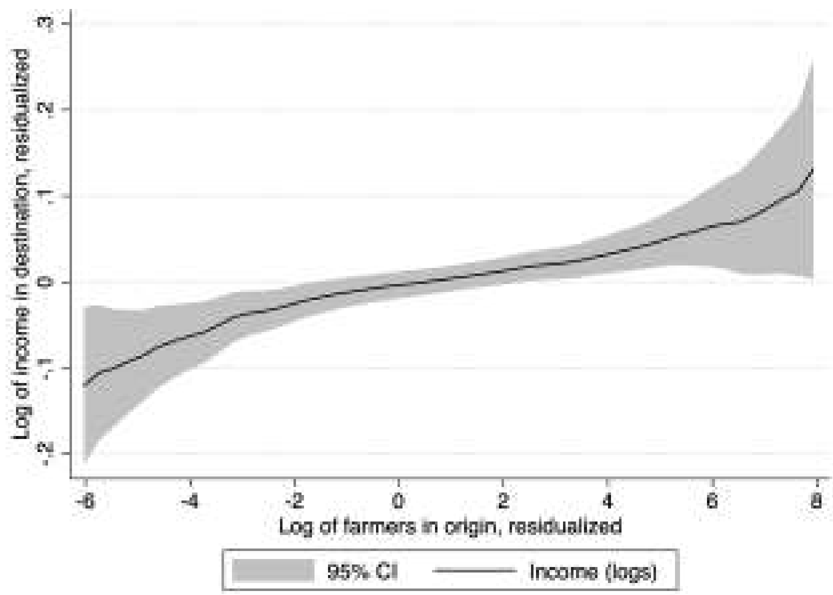
Notes: This figure highlights changes in the spatial allocation of the population and the production of two major crops in Brazil, soybeans and coffee. It shows that the diffusion of crops was not directly associated with the expansion of the population to the region.

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.

Table 1: The Revealed Comparative Advantages of Brazil

	(1) Banana	(2) Cassava	(3) Chicken	(4) Cocoa	(5) Coffee	(6) Corn
RCA Brazil 1980	1.928	0.107	5.468	16.25	18.24	0.717
RCA Brazil 2010	0.453	0.321	21.08	1.573	10.06	5.769
Change in RCA (%)	-76.4	200.4	285.4	-90.3	-44.8	704.2
East 2010	0.499	0.324	23.13	1.734	11.08	5.583
West 2010	0	0.296	1.006	0	0.100	7.591
	(7) Cotton	(8) Cattle	(9) Rice	(10) Soy	(11) Sugarcane	(12) Tobacco
RCA Brazil 1980	0.390	1.364	0.770	7.619	5.732	4.757
RCA Brazil 2010	4.221	9.726	0.830	26.40	28.37	9.330
Change in RCA (%)	980.4	612.8	7.748	246.5	395.0	96.12
East 2010	4.365	8.760	0.881	19.44	31.02	10.28
West 2010	2.810	19.15	0.325	94.31	2.566	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 Region of Destination

	Sample			
	All	All	West	Long Dist
	(1)	(2)	(3)	(4)
Panel A: Dep var is log of farmers				
Log(Farmers in origin)	0.143***	0.126***	0.141***	0.097***
	(0.002)	(0.002)	(0.004)	(0.003)
	10782	10029	2854	5686
Panel B: Dep var is log of income				
Log(Farmers in origin)	0.027***	0.024***	0.036***	0.015***
	(0.001)	(0.001)	(0.001)	(0.001)
	9460	8771	2522	4895
Destination-Crop-Year FE	✓	✓	✓	✓
Origin-Year FE	✓	✓	✓	✓
Controls		✓		

Notes: * / ** / *** denotes significance at the 10 / 5 / 1 percent level. Estimates from Poisson regressions. Standard errors clustered at meso-region level in parenthesis. This table shows estimates of equations (2). The unit of observation is a given at the destination-activity-origin-year. Explanatory variable is the log of agricultural workers in the same activity in the previous period (thirty years before) ($L_{ij,k(t-1)}$). We include the census of 2000 and 2010 in our regressions. We exclude non-migrants from sample.

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 4: PPML Estimates of the on Earnings and Employment of Agricultural Workers in their Region of Destination

	Dep variable	
	Log(Income)	Log(Farmers)
	(1)	(2)
Log(Farmers in origin)	0.09	0.17
	-	-
Log(Distance)	-1.52	-2.21
	-	-
Crop-Year FE	✓	✓
Origin FE	✓	✓
Year	2010	2010

Notes: This table shows results from PPML estimates of equations 8 and 9. We used the condensed likelihood function to partial out the effects of the fixed effects before estimating the coefficients.

Table 5: Effects of Banning East-West Migration on Aggregate Export Shares

Crop	Δ Exp. Share	Abundance	Knowledge
banana	6.02	0.84	0.22
cassava	6.76	0.83	0.23
chicken	7.46	0.92	0.12
cocoa	3.49	0.61	0.53
coffee	5.80	0.84	0.22
corn	1.64	1.07	-0.24
cotton	5.75	0.75	0.30
fish	8.66	0.90	0.14
hortifruti	6.41	0.85	0.21
livestock	-0.15	7.91	-9.80
rice	9.90	0.86	0.18
soy	-9.72	0.79	0.30
sugarcane	4.55	0.84	0.24
tobacco	6.70	0.87	0.18

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: Effects of Banning East-West Migration on the West's Export Shares

Crop	Δ Exp. Share	Abundance	Knowledge
banana			
cassava	33.64	0.82	0.17
chicken	28.55	0.86	0.14
cocoa	13.90	0.50	0.56
coffee	26.47	0.75	0.25
corn	22.40	0.84	0.14
cotton			
fish	21.99	0.76	0.25
hortifruti	29.20	0.81	0.19
livestock	21.17	0.74	0.28
rice	28.98	0.71	0.29
soy	-14.61	0.75	0.26
sugarcane	17.70	0.80	0.22
tobacco			

Notes: This table shows the results only for the west region of Brazil. See explanation in 5.

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 and their municipality of residence in 1995, but not their previous municipality of residence. In this case, we found that in 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 their previous municipality. Since less than 0.1% of the population 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 census, 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 be found in the agricultural census: banana, cassava, chicken, cocoa, coffee, corn, cotton, fish, hortifruti, 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 revenues of a worker instead of their information on wages. As a first test to check if the classifications of activities in the census follow the actual spatial distribution production in Brazil, figure **??** shows two examples of the distribution of agricultural workers across meso-regions of Brazil. We can see that the share of workers employed in the production of fish is concentrated along the Atlantic coast and in the Amazon basin where there is direct access to water. The regions with higher share of workers in the production of fish in the interior of Brazil correspond to areas near important rivers. In Panel B in the same figure we show the distribution in the employment of coffee producers. It shows that most coffee producers are located in the southeast where the production is historically dominant. The figure also captures the more recent expansion of the production of coffee in the state of Rondonia in the Amazon.

Agricultural Production. We bring data on agricultural revenues, land use, number of farms and output by crop from the agricultural census of 1960, 1970, 1996 and 2006. The information from the agricultural census comes at the municipality level. We organized the agricultural activities defined in the agricultural census 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. Figure 6 shows the correlation between our measures of agricultural activity in the demographic and agricultural census.

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 so that it matches 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 the 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 numbers from FAO.

FAO-GAEZ. The data from FAO-GAEZ is publicly available.

B Complementary Facts

Complementary Fact 1: The influence of migrants on the specialization of the West depends on the specialization of their region of origin.

To study the specialization in factors of production, we use the following measure of specialization

$$\mathcal{S}_{j,k} = \frac{\sum_i L_{ij,k} \Big/ \sum_i \sum_k L_{ij,k}}{\sum_i \sum_j L_{ij,k} \Big/ \sum_i \sum_j \sum_k L_{ij,k}}, \tag{10}$$

where $L_{ij,k}$ is the number of workers from origin i in j producing k. To compute the contribution of migrants from the East to the patterns of specialization of the West, we proceed as follows. For each crop, we select the micro-regions in the East in the top decile of the distribution of $\mathcal{S}$ and compute two measures of specialization for the West: one that includes the farmers from the East in the top decile of $\mathcal{S}$ and another that excludes them. Figure 10 report the ratio of these measures of specialization.

Figure 10 shows that the inclusion of migrants from the top decile of soybean producing areas in the East makes the specialization of farmers in soybeans in the West 9 percent higher. Farmers from these same regions have no effect on our measure of specialization for other crops. Similarly, the inclusion of farmers from sugarcane producing regions rises the specialization of the West in sugarcane by 3.5 percent. Figure 11 shows the disaggregated effect for the main crops in Figure 10. Migrants from soybean regions do not affect the specialization of sugarcane, and migrants from sugarcane regions do not affect the specialization of soybeans. Panel B in Figure 10 shows that if we carry the same exercise using the bottom decile in the distribution of specialization in the East there is no effect of migrants in any crop.

C 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. First, we recover expenditure shares (α_k) and transfers (t^A, t^A_s, t^F). Second, we calibrate the sectoral productivities (A_{ik}). Third, we calibrate the bilateral migration costs (μ_{jik}).

C.1 Recovering expenditure shares (α_k) and transfers (t^A, t^A_s, t^F)

We have $K + S + 2$ parameters to calibrate in this section. We use data on revenues per region and sector, R_{ik} $\forall i \in \mathcal{I}_s$, $\forall s \in \mathcal{S}$ and $\forall k \in \mathcal{K}$, where $\mathcal{I}_s$ is the set of regions in state s,

$\mathcal{S}$ the set of states in Brazil and $\mathcal{K}$ the set of sectors. In addition, we use data on imports and exports between states and foreign countries, M_{sk} and X_{sk} for $\forall s \in \mathcal{S}$ and $\forall k \in \mathcal{K}$.

First, we note that total income in region i is

$$I_i = W_i(1+t^A)(1+t_s)(1+t^F),$$

where W_i is the total payments to wages in region i ($W_i = \sum_{k\in\mathcal{K}}(1-\gamma_k)R_{ik}$). The expenditure shares are then calculated according to

$$\alpha_k = \frac{\sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{I}_s} R_{ik} + \sum_{s\in\mathcal{S}}(M_{ks} - X_{ks})}{\sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{I}_s} I_i}, \quad \forall k \in \mathcal{K}. \tag{11}$$

The nominator is the apparent consumption of the Brazilian economy by sector and the denominator is the total income. To calculate I_i, we assume that the interstate taxes sum to zero

$$0 = \sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{I}_s}\sum_{k\in\mathcal{K}}(1-\gamma_k)R_{ik}t_s,$$

that the foreign transfer is

$$1+t^F = \frac{\sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{I}_s}\sum_{k\in\mathcal{K}} R_{ik} + \sum_{s\in\mathcal{S}}\sum_{k\in\mathcal{K}}(M_{ks} - X_{ks})}{\sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{I}_s}\sum_{k\in\mathcal{K}} R_{ik}}, \tag{12}$$

and that the transfer from the national asset is given by

$$1+t^A = \frac{\sum_{i\in\mathcal{I}_s}\sum_{s\in\mathcal{S}}\sum_{k\in\mathcal{K}} R_{ik}}{\sum_{i\in\mathcal{I}_s}\sum_{s\in\mathcal{S}}\sum_{k\in\mathcal{K}}(1-\gamma_k)R_{ik}}. \tag{13}$$

The foreign transfer can be calibrated directly using the data on revenues and trade. Using the equation for interstate transfer, we can obtain the following equation for income in Brazil

$$I = \sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{I}_s}\sum_{k\in\mathcal{K}}(1-\gamma_k)R_{ik}(1+t^A)(1+t^F).$$

Now, substituting the formula for national asset and foreign transfer, we obtain

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

which is the aggregate apparent consumption of Brazil. Using this formula for total income, we compute the expressions for expenditure shares (α_k). Finally, with these expression, we can compute the interstate transfer as

$$(1+t_s) = \frac{\sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{I}_s} R_{ik} + \sum_{k\in\mathcal{K}}(M_{ks} - X_{ks})}{\sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{I}_s}(1-\gamma_k)R_{ik}(1+t^A)(1+t_s)(1+t^F)}. \tag{14}$$

12 and 13 provide 2 equations, 11 provides K equations and 14 provides S equations. Therefore, we have the same number of equations and parameters are we are just identified.

C.2 Recovering sectoral productivities (A_{ik})

To recover sectoral productivities, we first have to parametrize our iceberg trade costs. As explained in the main body of the paper, we set $\tau_{ij} = (1 + dist_{ij})^{\delta}$, 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}) and $2SK$ international trade costs ($\tau_{sF,k}$ and $\tau_{Fs,k}$) 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 $\forall k \in \mathcal{K}$, and model implied expenditure per region $E_{nk} = \alpha_k I_n$.

Our numerical algorithm to calibrate the structural productivities is as follows. We start with an initial guess of composite sectoral productivities defined by ($\overline{T}_{ik}{=}A_{ik}(c_{ik})^{-\theta}$). 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 keep iterating until expenditures implied by the model match revenues in the data. Once these composite sectoral produtictivities converge, we then back-out the values for the sectoral productivity A_{ik} using a model consistent measure of unit cost of production.

We explain this procedure in detail now. Starting with a guess $\bar{T}_{ik}$, $\forall i, k$. To match the model to state-level exports, aggregate region i's exports to F in good k in the model

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

to obtain state exports:

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

This term can be written as

$$X_{skF} = (\tau_{sFk})^{-\theta} \frac{\sum_{i\in\mathcal{I}_s} \overline{T}_{ik} (\tau_{iF})^{-\theta}}{\sum_{s\in\mathcal{S}} \sum_{i'\in\mathcal{I}_s} \overline{T}_{i'k} (\tau_{sFk}\tau_{i'F})^{-\theta} + \overline{T}_{Fk}} E_{kF}.$$

Using this expression, we can recover the relative international trade costs between states of Brazil and the rest of the world, using the ratio of state level exports and our guess of composite sectoral productivities ($\overline{T}_{ik} = A_{ik}(c_{ik})^{-\theta}$). To see this, take the ratio of exports between states

$$\frac{X_{s'kF}}{X_{skF}} = \frac{(\tau_{s'Fk})^{-\theta}}{(\tau_{sFk})^{-\theta}} \frac{\sum_{i\in\mathcal{I}_s} \overline{T}_{ik} (\tau_{iF})^{-\theta}}{\sum_{i\in\mathcal{I}_{s'}} \overline{T}_{ik} (\tau_{iF})^{-\theta}}.$$

Isolating the international trade costs gives

$$\frac{\tau_{s'Fk}}{\tau_{sFk}} = \left(\frac{X_{s'kF}}{X_{skF}} \Big/ \frac{\sum_{i\in\mathcal{I}_s} \overline{T}_{ik} (\tau_{iF})^{-\theta}}{\sum_{i\in\mathcal{I}_{s'}} \overline{T}_{ik} (\tau_{iF})^{-\theta}} \right)^{-\frac{1}{\theta}}. \tag{15}$$

This equation provides the international trade costs between states relative to a baseline state. Note that we are not able to distinguish the level of international trade costs, τ_{sFk}, fromthe composite productivity of the foreign sector $\overline{T}_{Fk}$. Therefore, we normalize one of the international trade costs to one. With this normalization and equations 15, we have $S \times K$ equations and $S \times K$ international trade costs parameters.

Once with have τ_{sFk}, we also obtain the composite sectoral productivity of the external sector ($\overline{T}_{Fk}$) within each iteration using the following expression for total exports

$$\sum_{s\in\mathcal{S}} X_{skF} = \sum_{s\in\mathcal{S}} \sum_{i\in\mathcal{I}_s} \frac{\overline{T}_{ki}(\tau_{sFk}\tau_{iF})^{-\theta}}{\overline{T}_{Fk} + \sum_{s'\in S} \sum_{i'\in N_{s'}} \overline{T}_{ki'}(\tau_{s'Fk}\tau_{i'F})^{-\theta}} E_{kF}.$$

Isolate $\overline{T}_{Fk}$ to obtain

$$\overline{T}_{kF} = \left(\frac{E_{kF} - X_{kF}}{X_{kF}} \right) \sum_{s'\in S} \sum_{i\in N_s} \overline{T}_{ki}(\tau_{s'Fk}\tau_{i'F})^{-\theta}. \tag{16}$$

We have K equations for 16 and K composite sectoral productivities. Therefore, these equations identify the composite sectoral productivity of the external sector.

We are now left with the composite sectoral productivities of each region ($\overline{T}_{ik}$) and the international trade costs from the foreign market to each state (τ_{Fsk}). The equation for imports is given by

$$M_{ikF} = \frac{\overline{T}_{Fk}(\tau_{Fsk}\tau_{iF})^{-\theta}}{\sum_{s'\in S}\sum_{i'\in\mathcal{I}_{s'}}\overline{T}_{i'k}(\tau_{i'F})^{-\theta} + \overline{T}_{kF}(\tau_{Fsk}\tau_{iF})^{-\theta}} E_{ik},$$

adding up to the state level gives

$$M_{sk} = \sum_{i\in\mathcal{I}_s}\frac{\overline{T}_{kF}(\tau_{Fsk}\tau_{iF})^{-\theta}}{\sum_{s'\in S}\sum_{i'\in\mathcal{I}_{s'}}\overline{T}_{i'k}(\tau_{i'F})^{-\theta} + \overline{T}_{kF}(\tau_{Fsk}\tau_{iF})^{-\theta}} E_{nk},$$

which can be written as

$$\tau_{Fsk} = \left(M_{sk}\Big/\frac{\sum_{i\in\mathcal{I}_s}\overline{T}_{kF}(\tau_{iF})^{-\theta}}{\sum_{s'\in S}\sum_{i'\in\mathcal{I}_{s'}}\overline{T}_{ki'}(\tau_{i'F})^{-\theta}. + \overline{T}_{kF}(\tau_{Fsk}\tau_{iF})^{-\theta}} E_{nk}.\right)^{-\frac{1}{\theta}}. \tag{17}$$

Within every iteration, we can obtain the trade costs from the foreign sector to each state that rationalizes the model implied expenditure per state.

Finally, with τ_{Fsk}, τ_{sFk} and $\overline{T}_{kF}$ in hand, we can obtain the composite sectoral productivities. To do so, we use

$$R_{ik} = \left[\sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{I}_s}\frac{\overline{T}_{ik}(\tau_{ij})^{-\theta}}{\overline{T}_{kF} + \sum_{s'\in S}\sum_{i'\in\mathcal{I}_{s'}}\overline{T}_{ki'}(\tau_{i'j})^{-\theta}} E_{jk} + \frac{\overline{T}_{ik}\tau_{iFk}}{\sum_{i'\in\mathcal{I}_s}\overline{T}_{i'k}\tau_{i'Fk}} X_{sFk}\right].$$

Isolating $\overline{T}_{ik}$, we get

$$\overline{T}_{ik} = R_{ik}\Big/\left[\sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{I}_s}\frac{\overline{T}_{ik}(\tau_{ij})^{-\theta}}{\overline{T}_{kF} + \sum_{s'\in S}\sum_{i'\in\mathcal{I}_{s'}}\overline{T}_{ki'}(\tau_{i'j})^{-\theta}} E_{jk} + \frac{\overline{T}_{ik}\tau_{iFk}}{\sum_{i'\in\mathcal{I}_s}\overline{T}_{i'k}\tau_{i'Fk}} X_{sFk}\right]. \tag{18}$$

Note that the level of the composite sectoral productivity is not identified. Therefore, we normalize one of them to one. With this normalization, we have $I \times K$ of this expression since the level of the composite sectoral productivities is not identified. In summary, 15 and a normalization of international trade costs give $S \times K$ equations, 17 gives SK equations, 16 gives additional K equations and 18 and a normalization of composite sectoral productivities give $I \times K$ equations.

Once with composite sectoral productivities ($\overline{T}_{ik}$), we can use the unit cost of production c_{ik} to recover A_{ik}. To compute the unit cost of production that is model consistent we proceed as follows. First, we use the land market clearing condition

$$r_i H_i = \sum_{k\in\mathcal{K}} \gamma_k R_{ik}.$$

Using the fact that $w_{ik}E_{ik} = (1-\gamma_k)R_{ik}$, we write land rents as

$$r_i H_i = \sum_{k\in\mathcal{K}} \frac{\gamma_k}{1-\gamma_k} w_{ik}E_{ik},$$

which become

$$r_i = \sum_{k\in\mathcal{K}} \frac{\gamma_k}{1-\gamma_k} \frac{w_{ik}E_{ik}}{H_i}.$$

The unit cost of production ($c_{ik} = (w_{ik})^{1-\gamma_k}(r_i)^{\gamma_k}$) can then be written as

$$c_{ik} = (w_{ik})^{1-\gamma_k} \left(\sum_{k\in\mathcal{K}} \frac{\gamma_k}{1-\gamma_k} \frac{w_{ik}E_{ik}}{H_i} \right)^{\gamma_k}. \tag{19}$$

The second term captures the congestion in micro-region i. For the wage rate, we construct wages that are model consistent. In particular, we use the total efficiency of agricultural workers of in i producing k ($\tilde{E}_{ik}$)

$$\tilde{E}_{ik} = \sum_{j\in\mathcal{I}} L_j \lambda_{ijk} L_{jk}^{\beta}.$$

Our wage rate per unit of efficiency labor is then given by

$$w_{ik} = \frac{(1-\gamma_k)R_{ik}}{\sum_{j\in\mathcal{I}} L_j \lambda_{ijk} L_{jk}^{\beta}}. \tag{20}$$

Finally, substituting 20 into 19, we obtain the unit cost of production as a function of the data and previously estimated parameters, which can then be used to recover A_{ik} from the composite sectoral productivity $\overline{T}_{ik}$.

C.3 Recovering bilateral migration costs (μ_{ijk})

To recover the bilateral migration cost, we use the following expression

$$\lambda_{ijk} = \frac{\mu_{ijk}[I_{ijk}/P_i]^{\kappa}}{\sum_{j'\in\mathcal{I}} \sum_{k\in\mathcal{K}} \mu_{ij'k}[I_{ij'k}/P_{j'}]^{\kappa}},$$

where λ_{ijk} is the share of workers from j migrating to i in activity k. I_{ijk} is the nominal income that the worker obtains in activity i and P_i is the price index. Using the previous calibration of A_{ik}, we construct price indexes that are consistent with the model. The income is given by $I_{ijk} = w_{ik}(L_{jk})^{\beta}(1+t^A)(1+t^F)(1+t_s^A)$. The bilateral migration flows are up-to-scale within an origin region. In other words, we are not able to identify the level of bilateral migration costs. Therefore, we can normalize $\sum_{j' \in \mathcal{I}} \sum_{k \in \mathcal{K}} \mu_{ij'k}[I_{ij'k}/P_{j'}]^{\kappa} = 1$. The expression of the bilateral migrations costs is then given by

$$\mu_{ijk} = (I_{ijk}/P_i)^{\kappa}/\lambda_{ijk}.$$

C.4 Adjusting Migration Flows

This section describes the adjustment of migration flows that we use to match the allocation of workers between different waves of the census. We make this adjustment because we do not observe the full history of migration of a worker in the data. As a consequence, when workers migrate more than once between census, the estimates of migration flow based on the current and previous location of workers do not necessarily match the observed location of workers between two census, even if we correct for the birth and death rate of workers. To illustrate the problem, consider the following example. There are three regions A, B and C and two waves of the census t and $t-1$. In $t-1$, there are 1000 workers in each region and, in t, we have 3000 workers in A and 0 elsewhere. However, if workers from C migrated to B before reaching their final destination, the reported information in the census in t will understate the flow of migrants from C to A between census.

We adjust our measures of migration flows based on widely used techniques employed in the construction of datasets where flows must match aggregate information. A typical example is the construction of input-output datasets where the sales between sectors must match total sales. Here, we adjust migration flows in order to match aggregate number of workers. Our method is based on Ireland and Kullback (1968) and Rogers (1968).

First, write the total number of workers in destination j in year t as

$$L_{j,kt} = \sum_i p_{ij,kt}(1 + b_i - d_i)L_{i,t-1},$$

where $p_{ij,k}$ is the share of workers who migrate from i to j to produce k, d_i is the death rate of workers from i and b_i is the birth rate. We assume that the death rate and birth rate are constant across regions. In that case, we have

$$L_{j,kt} = (1 + b - d) \sum_i p_{ij,kt} L_{i,t-1}$$

we can measure these terms as

$$\begin{aligned} L_t &= (1+b-d)\sum_j \sum_k \sum_i p_{ij,kt} L_{i,t-1} \\ &= (1+b-d) L_{t-1} \end{aligned}$$

Therefore, we can estimate $1+b-d$ by looking at the ratios of the overall population. We call $\tilde{L}_{i,t-1} = (1+b-d)L_{i,t-1}$. Now, we have the following problem

$$L_{j,kt} = \sum_i p_{ij,kt} \tilde{L}_{i,t-1}.$$

Write the equation above as

$$p_{j,kt} = \sum_i p_{ij,kt} p_{i,t-1},$$

where $p_{j,kt} = L_{j,kt}/L_t$ and $p_{i,t-1} = \tilde{L}_{i,t-1}/L_t$.

We now turn to the problem of estimating $p_{ij,kt}$. In what follows, we drop the time index. Our goal is to recover the probability $p_{ij,k}$ given values for $p_{j,k}$ and p_i and an imperfect proxy for $p_{ij,k}$ denoted by $\pi_{ij,k}$. Following Ireland and Kullback (1968), the difference between $p_{ij,k}$ and $\pi_{ij,k}$ is decomposed into two adjustment terms a_i and b_{jk}

$$p_{ij,k} = a_i b_{jk} \pi_{ij,k}.$$

Therefore, we have

$$p_{j,k} = b_{j,k} \sum_i a_i \pi_{ij,k},$$

and

$$p_i = a_i \sum_j \sum_k b_{j,k} \pi_{ij,k}.$$

We choose a_i and b_{jk} to minimize

$$\sum_j \sum_i \sum_k p_{ij,k} \log\left(\frac{p_{ij,k}}{\pi_{ij,k}}\right),$$

Ireland and Kullback (1968) shows that the choice of a_i and b_{jk} that minimizes the criterion above is such that the sum of $p_{ij,k}$ matches exactly the marginals $p_{j,k}$ and p_i. The intuition is as follows. When the flow of migrants from all i to destination j,k in the data is low and not enough to match the actual number of workers observed in j,k, we increase $b_{j,k}$. An analogous intuition goes for the choice of a_i. The limitation of this method is that it gives equal weights for migrants coming from all regions when it does this correction. In practice, Ireland and Kullback (1968) shows that the choice of $p_{ij,k}$ that is consistent with the choice of a_i and $b_{j,k}$ that minimizes the criterion can be obtained using the following iterative procedure

$$
\begin{aligned}
p^1_{ij,k} &= \frac{p_i}{\pi_i}\pi_{ij,k} \\
p^2_{ij,k} &= \frac{p_{j,k}}{p^1_{j,k}}p^1_{ij,k} \\
p^3_{ij,k} &= \frac{p_i}{p^2_i}p^2_{ij,k} \\
p^4_{ij,k} &= \frac{p_{j,k}}{p^3_{j,k}}p^3_{ij,k} \\
&\vdots \\
p^{2n-1}_{ij,k} &= \frac{p_i}{p^{2n-2}_i}p^{2n-2}_{ij,k} \\
p^{2n}_{ij,k} &= \frac{p_{j,k}}{p^{2n-1}_{j,k}}p^{2n-1}_{ij,k}.
\end{aligned}
$$

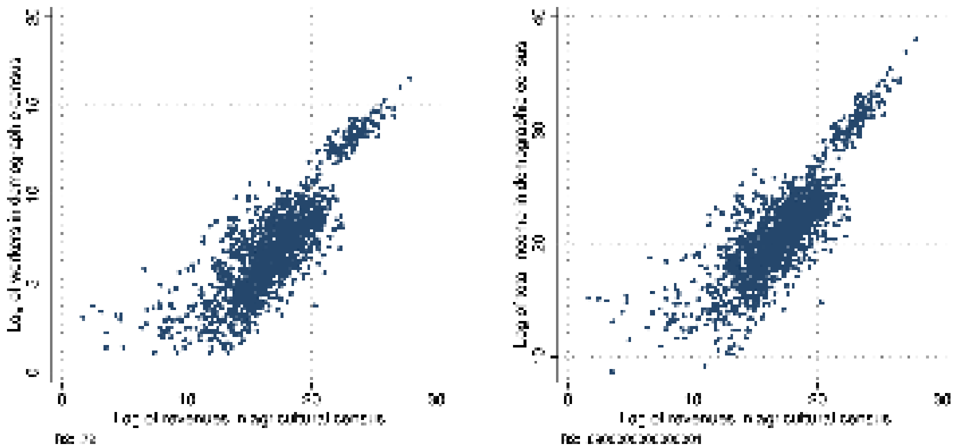
We apply the method using as $\pi_{ij,k}$ the information collected by the census on the current and previous micro-region reported by workers. The marginals p_i and $p_{j,k}$ come directly from the share of workers in each micro-region in the origin and in the destination. We estimate $p_{ij,k}$ so that our data perfectly matches the marginals p_i and $p_{j,k}$.

To check the validity of our method, we compare the aggregate migration flow between states predicted by the flow of migrants $p_{ij,k}$ at the micro-region level with the data on migration of workers between their current state and their state of birth. Unfortunately, since we do not have the micro-region of birth, we can only check our method based on the aggregate flow. Figure 9 shows that when we look at migration between the current and previous location without any adjustment, the data understates substantially the magnitude of the migration between the current state and the state of birth of workers. This occurs because many workers move within their current state after leaving their state of birth. The slope of a regression of the log of the share of workers in a state coming from another given state on the log of the share of workers in a state who were born in another given state is only 0.75. The second figure shows that with our adjustment we match the magnitude of migration between states. Here, the analogous regression gives a slope of 1.03.

D Additional Figures and Tables

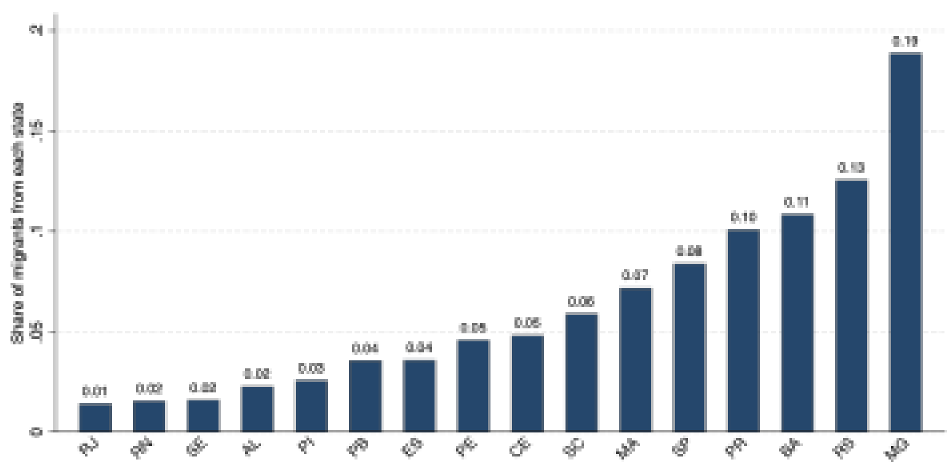
Figure 6: Correlation between Measures of Agricultural Production in the Demographic and the Agricultural Census

a. Workers on Revenues b. Income on Revenues



Notes: These figures show the correlation between our measure of revenues and number of agricultural workers in each crop calculated according to the demographic census and our measure of revenues computed from the agricultural census. They indicate that our classification of agricultural activities are consistent across the two datasets.

Figure 7: State of Origin of Migrants in the West coming from the East

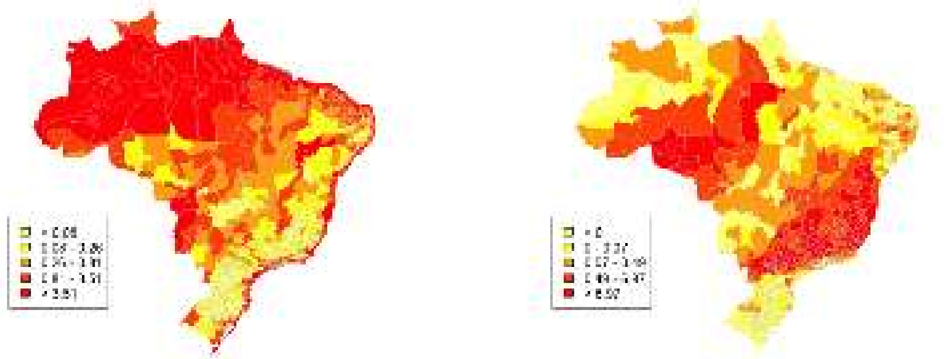


Notes: This figure shows the proportion of migrants in the East of Brazil coming from each state of origin in the West. It indicates that migrants come from a large variety of places. For example, 13% of migrants in the west come from Rio Grande do Sul, which is located to the extreme south of Brazil,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

Figure 8: Examples of the Distribution of Agricultural Workers By Crop in 2000

a. Fish

b. Coffee

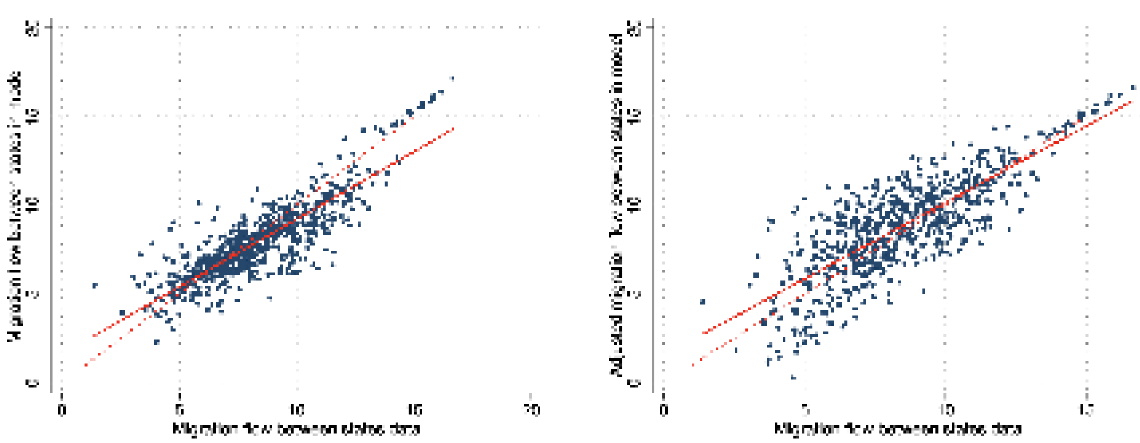


Notes: Warmer colors in the map indicate higher share of workers employed in the agricultural activity. Panel A shows that there is a much larger share of workers producing fish in micro-regions next to the Atlantic coast and in the Amazon basin to the northeast where there is large access to water. The larger share of agricultural workers in the middle of the country correspond to regions with major rivers in Brazil. Panel B shows the distribution of workers in the production of coffee. Most of them are located in the southeast of the country where the production was historically located. The figure also picks up the recent expansion in regions in the Amazon.

Figure 9: Adjustment of Migration Flows

a. Non-adjusted

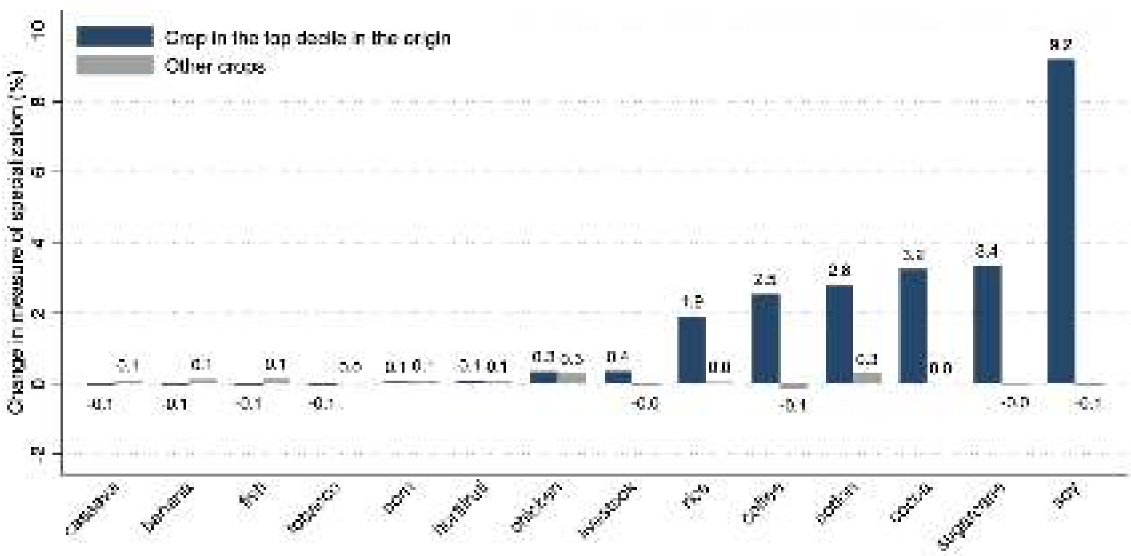
b. Adjusted



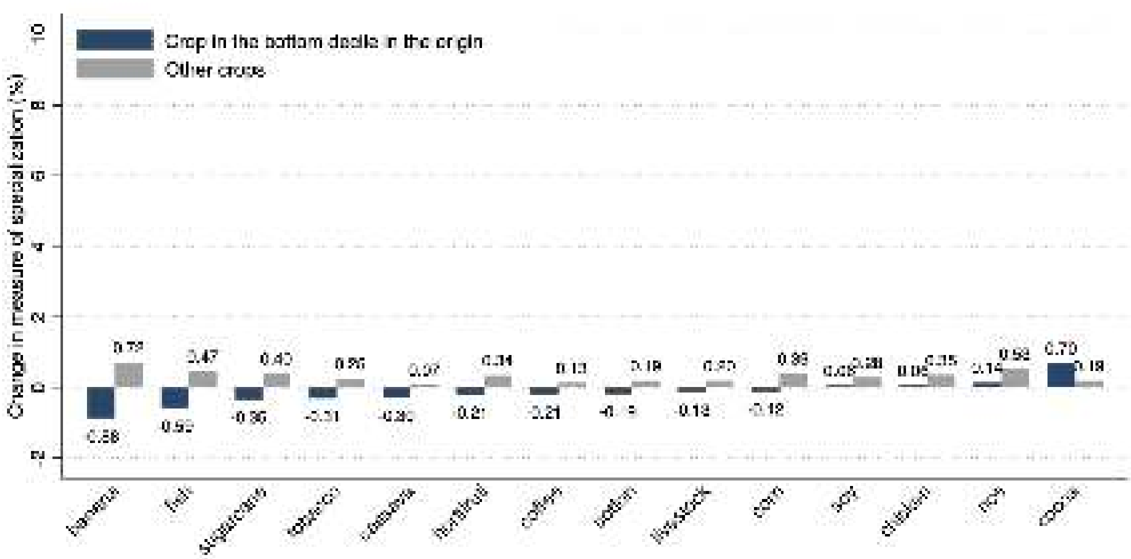
Notes: This figure shows the flow of migrants between states. On the x-axis, we show this measure of migration based on the state of birth of migrants. On the y-axis in Panel A, we show the state migration according to actual data on the previous location of workers. On the y-axis in Panel B, we show the state migration after the adjustment procedure. After the adjustment, the level of interstate migration becomes closer to the one in the data.

Figure 10: The Specialization of Migrants from the East living in the West according to the Crop-Intensity in the Origin

(a) Migrants from regions in the top decile of the specialization of the crop

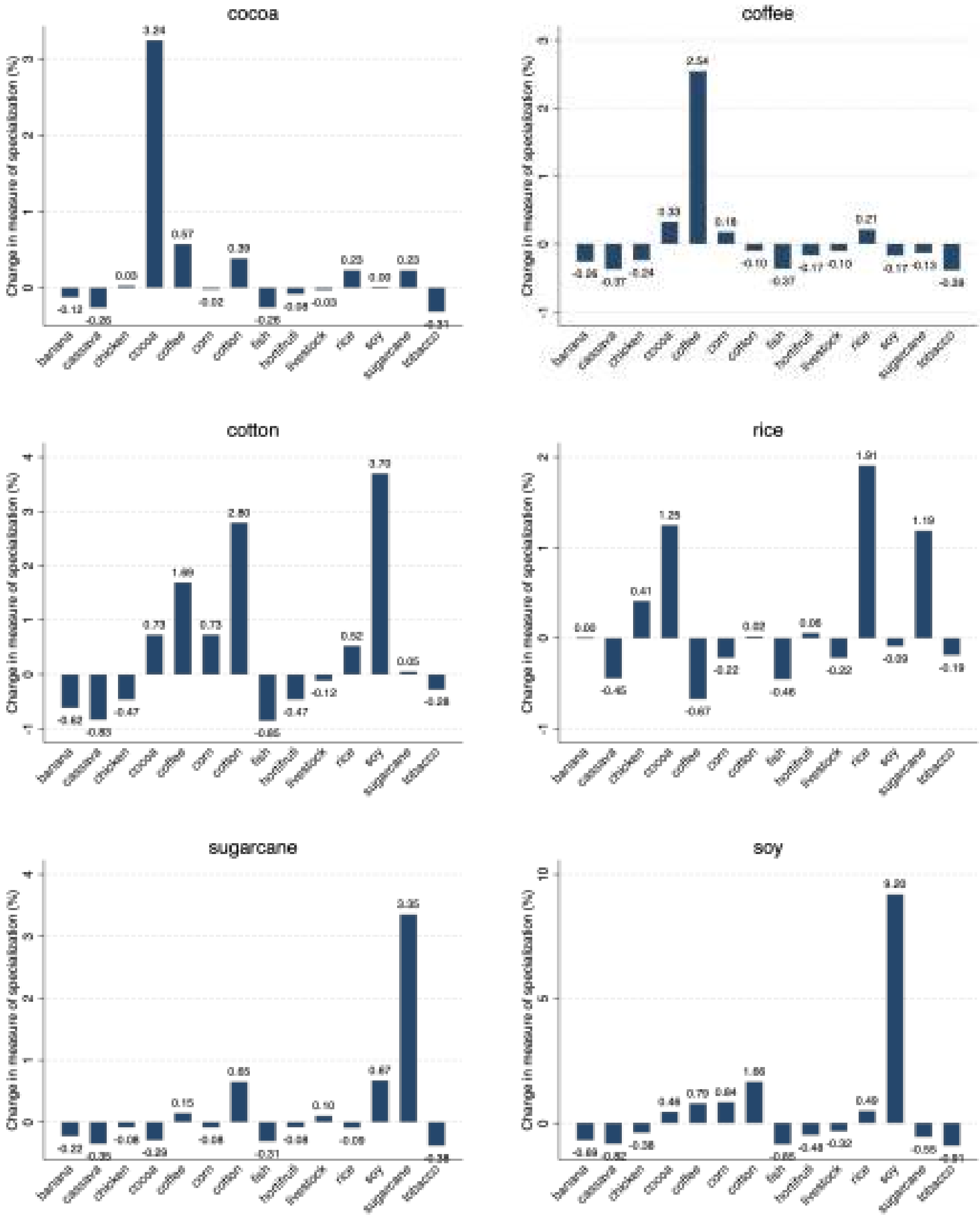


(b) Migrants from regions in the bottom decile of the specialization of the crop



Notes: To construct Panel A we proceed as follows. We first select the micro-regions in the East that are the in the top decile in terms of the specialization of factors of production for a given crop (as given by equation 10). We then compute two measures of specialization in the West: one that includes the migrants from the top decile and another that excludes them. This figure reports the ratio of these measures minus one and multiplied by 100. Panel B shows the same procedure but using migrants from the bottom decile of the distribution of each crop.

Figure 11: Contribution of Migrants by Top Deciles in the Origin in the Crops with Largest Contribution



Notes: This table shows the results from Figure 10 broken down by each crop in which the migrants from a specific origin are contributing to.

Table 7: Effects of Banning East-West Migration on the West's Export Shares

Crop	Δ Exp. Share	Abundance	Knowledge
banana	-0.27	1.71	-0.91
cassava	0.44	0.12	1.02
chicken	1.16	1.10	-0.13
cocoa	-2.63	1.23	-0.28
coffee	-0.47	1.36	-0.43
corn	-1.30	0.51	0.97
cotton	-0.52	2.27	-1.19
fish	2.73	0.97	0.05
hortifruti	0.24	0.16	1.04
livestock	-0.45	1.51	-0.60
rice	3.58	0.84	0.19
soy	2.37	1.16	-0.19
sugarcane	-1.03	1.04	-0.06
tobacco	0.38	0.65	0.42