Migration, Specialization, and Trade: Evidence from Brazil's March to the West

Heitor S. Pellegrina — NYU Abu Dhabi

Sebastian Sotelo — University of Michigan[‡]

September 2021

Please click here for the most recent version

Abstract

We study how migration shapes aggregate and regional comparative advantage, exploiting a large migration of farmers to the West of Brazil between 1950 and 2010. Besides reallocating workers to regions where productivity was rising and land was abundant, we show that migration allows workers to sort according to their own comparative advantage. Incorporating these mechanisms into a quantitative model, we find that migration cost reductions reshaped Brazil's comparative advantage and contributed to its rise as a leading commodity exporter—accounting for 30 percent of observed changes in specialization. Migration opportunities, moreover, account for a substantial share of the gains from trade.

Keywords: International Trade, Migration, Comparative Advantage

[*]We thank Dominick Bartelme, Jonathan Eaton, Andrei Levchenko, Steve Redding, Esteban Rossi-Hansberg, and Matthew Turner for helpful comments, as well as Walter Steingress, João Paulo Pessoa, and Ferdinando Monte for useful discussions. We also thank participants in seminars at Purdue, Michigan, NYU Abu Dhabi, St. Louis Fed, IADB, TIGN-LACEA, Middlebury, RMET, UEA, FGV/EESP, Banco Central de Colombia, Insper, UFABC, Banco Central de Chile, UCLA, Berkeley ARE, Aarhus, the World Bank, NBER Agricultural Trade Policy Meeting, ETOS-FREIT, Princeton, CAF, SBE 2020 Meeting, the Econometric Society 2021 Meetings, Dartmouth, and Georgetown. Andrés Fajardo, Natalia Martinez Camelo, Brian Cevallos Fujiy, and Elliot Kang provided excellent research assistance. Financial support from the Michigan Institute for Teaching and Research in Economics (MITRE) is gratefully acknowledged. Sotelo thanks the IES at Princeton for its hospitality and funding during part of this research. E-mail: heitor.pellegrina@nyu.edu and ssotelo@umich.edu. All errors are our own.

[†]A Supplementary Appendix containing proofs and complementary quantitative results can be found here.

1 Introduction

A central task in international trade, and in spatial economics more generally, is to understand the impact of trade on welfare and on the patterns of specialization across locations. Seeking to quantify the consequences of trade, a recent literature has incorporated comparative advantage into quantitative models and established it as a major determinant of trade flows and welfare (Eaton and Kortum, 2012; Costinot and Rodríguez-Clare, 2015). But while recent work has also documented that comparative advantage itself evolves over time, there has been comparatively less progress in quantifying the mechanisms that drive that evolution.

The starting point of this paper is the observation that episodes of large internal migration are common, and that they are often associated with dramatic changes in the sectoral and spatial composition of economic activity of the countries experiencing them. Consider, for example, the U.S. westward expansion and, more recently, the large migration of Chinese workers to export-oriented clusters. Across the world, moreover, rural-urban migration is a hallmark of development.

Based on this observation, we ask: How does migration within a country shape regional and aggregate comparative advantage? We consider three channels. First, migration determines the allocation of labor across regions and activities that differ in their natural advantage. Second, migration alters the relative abundance of land and labor across regions. Third, workers whose knowledge is heterogeneous across activities sort according to their own comparative advantage, which determines the composition of the labor supply across economic activities and regions. To measure the strength of these mechanisms, we incorporate them into a quantitative spatial model, which we take to data from Brazil in the second half of the 20th century.

We focus on an episode often called the "March to the West" (Villas Bôas and Villas Bôas, 1994; Nehring, 2016). Starting in the 1950s and following a series of public initiatives to integrate the country's West to the urban centers in the East, approximately 8 million Brazilian workers migrated to low-density, high productivity-growth areas in the Cerrado and the Amazon. As a result, the share of Brazil's population living in the West grew from 7 to 15 percent between 1950 and 2010, while that region's participation in total agricultural production and total land use also rose sharply.

We document four empirical facts that guide our modeling approach, and to that end, we construct a data set that combines several waves of Brazil's demographic census with production and trade data since the 1950s. Over the course of the March, Brazil's agricultural sector was transformed. Using an index of comparative advantage, we first show that Brazil developed a new comparative advantage in crops such as soy, beef, and corn. Likewise, these commodities' share in total exports grew from nearly zero to about 13 percent, while the share of commodities that had been traditional mainstays of Brazil's external sector, such as coffee, cacao, and bananas, declined sharply. The repercussions of this transformation were felt worldwide: By 2010, Brazil was one of the top three world exporters of soy, beef, and corn (FAOSTAT). Second, we show that Eastern migrants comprise about 30 percent of the West's labor force in 2010, but this share varies greatly across agricultural activities and is particularly large for soy. Third, we document a strong link between a migrant's origin and her income and employment activity. Comparing farmers who emigrate from different origins but end up in the same destination and working in the same activity, we find that farmers coming from a region with 1 percent larger employment in that activity earn 0.02 to 0.08 percent larger incomes and that their number is 0.07 to 0.16 percent larger than those of other regions. Our conclusions regarding worker heterogeneity are robust to a number of specification checks, alternative hypotheses, and controls for selection. Fourth, we show that aggregate output is higher in regions and activities in which the labor force is composed by a larger proportion of farmers emigrating from origins with high-employment in that same activity.

Guided by these facts, we develop an overlapping generations model of trade with many activities and heterogeneous workers. Based on our empirical facts about worker productivity, workers in our model are endowed with good-specific knowledge, which they acquire through exposure to economic activity in their origin region, and they choose their location and activity according to their own comparative advantage. In equilibrium, regional and aggregate comparative advantage reflect a combination of natural advantage, relative abundance of labor, and knowledge of the labor force.

Using a pared-down version of our model, we derive two new analytical results relating migration to comparative advantage. Our first result shows that a reduction in migration costs may weaken or amplify the comparative advantage—defined as autarky opportunity costs—arising from exogenous differences in productivity. By spreading knowledge across regions, migration undoes the ex-ante spatial allocation of worker knowledge, and doing so may amplify or reduce opportunity cost differentials relative to the rest of the world. Our second result shows that the non-migrant employment share in a region-activity pair is a sufficient statistic for the impact of migrants on comparative advantage (along the lines of Arkolakis et al., 2012).

In bringing our model to the data, we identify key parameters that control worker heterogeneity and mobility, based on our reduced-form elasticities relating migrants' activities and income to their regions of origin. We also calibrate the rest of the parameters to match production by region and activity between 1950 and 2010. We then conduct an experiment to gauge the consequences of reductions in migration costs during this period. Specifically, we shut down the evolution of migration costs since the 1950s, while allowing all other exogenous factors to evolve over time, and we compare the resulting counterfactual economy to our baseline.

We find that the evolution of migration costs was critical in settling the West, accounting for half of the growth of that region's population share. Turning to trade, had migration costs not declined, trade specialization in agricultural goods in 2010 would have been pervasively lower across Western regions. In line with our theoretical results, the observed exposure of each region and activity to workers migrating from the East is strongly associated with these changes in specialization, because this exposure captures the direct impact of limiting migration on relative marginal costs. Aggregating these regional counterfactual changes yields large swings in specialization in new commodities for the West as a whole: specialization drops by 58 percent for soy and by 35 percent for beef and corn.

While migration is a key driver of the West's transformation, in Eastern regions the impact of migration is small because these regions were already densely populated in 1950. Nonetheless, reductions in migration costs are critical to understand Brazil's country-wide international specialization in agriculture. Soy and beef specialization are 29 and 25 percent lower in 2010, absent migration cost reductions. The observed drop in migration costs, moreover, accounts for up to 30 percent of the *observed* evolution of Brazil's specialization between 1950 and 2010 (or 27 percent of the export share growth of these commodities).

We next assess how our margins of comparative advantage shape the impact of migration. Due to the essential role that heterogeneity in land productivity plays in agriculture, we keep this feature in our next two exercises. First, we recalibrate our model shutting down knowledge heterogeneity and find that this heterogeneity amplifies the impact of migration on aggregate trade. In the West, worker heterogeneity aided substantially in the expansion of new agricultural activities, accounting for up to one-seventh of our baseline results. In the aggregate, the role of workers sorting is particularly strong for soybean—accounting for about one-ninth of our baseline results. Worker heterogeneity and migration also strengthen the link between trade specialization and natural advantage in the long-run, doubling aggregate soy exports and revenue shares, relative to a model without knowledge. Second, we allow for heterogeneity in land-intensity within agriculture and find that it alters the incidence of the labor supply shock induced by migration. Because cattle ranching is the most land-intensive activity, the impact of migration on this activity declines by about 5 percent in aggregate (17 percent for the West); the response of the other activities is modest.

We close our analysis by studying welfare. We offer a decomposition that shows that the full gains from trade (i.e. welfare costs of not being able to trade with any other region) depend critically on the ability of workers to migrate, as migration opportunities account for up to 60% of the gains from trade. In addition, regional comparative advantage has a large quantitative impact on how migration shapes the gains from trade. Finally, we show that the East-West migration observed during the March interacted with regional comparative advantage, often increasing the gains from trade with the rest of the world.[1]

Our paper contributes to a literature that measures the sources and evolution of comparative advantage in international trade. Levchenko and Zhang (2016) and Hanson et al. (2015) document substantial changes in Ricardian comparative advantage over time and across countries. Morrow (2010) and Chor (2010) show that forces such as relative productivity and factor abundance differences are key drivers of comparative advantage. Other work has studied the role of alternative sources of comparative advantage, such as institutional differences (Levchenko, 2007, Nunn, 2007, and Manova, 2013), the unobservable dispersion of workers' abilities (Grossman and Maggi, 2000, Bombardini et al., 2012, Ohnsorge and Trefler, 2007) and, related to our empirical findings, international migration (Bahar and Rapoport, 2016).[2] We contribute to this literature by establishing conditions under which domestic migration determines the evolution of regional and national comparative advantage, through Ricardian and Heckscher-Ohlin forces and by reallocating activity-specific knowledge over space—a hitherto unexplored mechanism.

This paper also relates to a recent literature that studies the impact of agricultural trade on welfare (Costinot and Donaldson, 2014, Allen and Atkin, 2016, Porteous, 2019, Pellegrina, 2020, Sotelo, 2020), development and structural change (Tombe, 2015, Farrokhi and Pellegrina, 2020, Porteous, 2020), and the implications of climate change (Costinot et al., 2016, Gouel and Laborde, 2018). In most of this literature, agricultural comparative advantage is static and arises from exogenous differences in the quality of land, factor proportions, or both. We show that taking advantage of those exogenous factors requires availability of workers and their knowledge. Our paper also clarifies our understanding of Brazil's emergence as a major global player in commodity markets, a unique episode that transformed international agricultural markets.[3]

[1]We allow for endogenous expansions of the agricultural frontier, but we do not incorporate their welfare cost due to deforestation. Although it is an important issue, carefully measuring those costs is challenging, and we leave it for future research.

[2]Buera and Oberfield (2016) and Cai et al. (2019), among others, study how the diffusion of ideas in an open economy shapes international comparative advantage (see Lind and Ramondo, 2018).

[3]Our results complement those of Bustos et al. (2020), who show that the arrival of GMO soy to Brazil in the 2000s expanded local production of low R&D sectors by releasing unskilled labor from agriculture. We take productivity growth as given and focus instead on how spatial labor mobility shaped comparative advantage within agriculture since the 1950s.

Our quantitative analysis deploys tools developed in the spatial economics literature, including Cosar and Fajgelbaum (2016), Redding (2016), Bryan and Morten (2019), Tombe and Zhu (2019), Nagy (2020), Porcher (2020), and Fujiwara et al. (2020) (see Redding and Rossi-Hansberg, 2017 for a recent review). Fajgelbaum and Redding (2014), in particular, explore how the domestic allocation of workers interacted with international comparative advantage to contribute to structural change and urbanization in 19th century Argentina. Different from that paper, we focus on the role of migration costs and worker heterogeneity.[4] In a closed economy model, Morten and Oliveira (2016) evaluate the equilibrium impact of road construction on intranational trade and migration in Brazil. Relative to their work, we introduce a new, observable source of worker heterogeneity into an open economy model to study how migration shapes international trade. We also develop analytical expressions relating the gains from trade in economies with and without internal migration. Building on the recent dynamic approaches of Allen and Donaldson (2020), Artuç et al. (2010), and Caliendo et al. (2019), we also add to this literature a new source of complementarity between trade and migration, which relies on the differential propensity of workers—driven by ability or geography—to migrate to different regions and pursue different activities.[5]

Lastly, we contribute to a literature documenting how migrants' characteristics shape their impact on the economy.[6] Studying a population resettlement program in Indonesia, Bazzi et al. (2016) show that regions that received migrants from origins with similar agroclimatic conditions perform better than others—indicative of migrants transferring their skills. Olmstead and Rhode (2008) document the role of geography and migration in the expansion of different crops in the United States. Arkolakis et al. (2018) study the impact of migrants on the technological frontier in the U.S. in the 19th century.[7] Our contribution to this literature is twofold. First, relying on detailed individual-level surveys, we provide new evidence

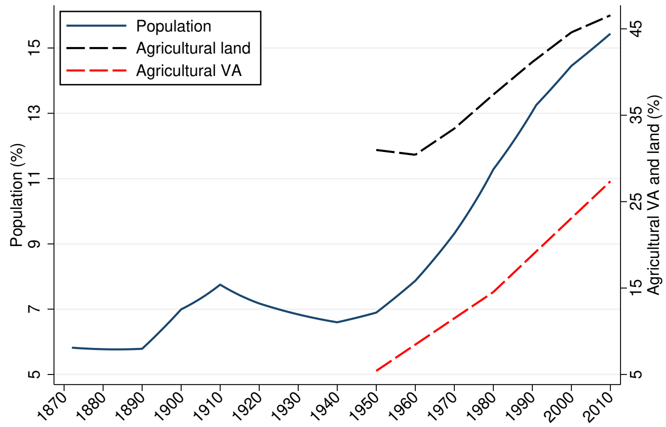
[4]In a Heckscher-Ohlin framework, Courant and Deardorff (1992) show that, if one factor of production is unevenly distributed across otherwise homogeneous regions within a country, aggregate comparative advantage can arise in the sector that uses that factor more intensively. In our framework, an uneven allocation of a single factor might affect aggregate comparative advantage depending on the distribution of natural advantages across regions.

[5]Burchardi et al. (2019) show that the nationality of international migrants is a determinant of FDI across US counties. Likewise, Cardoso and Ramanarayanan (2019) and Bonadio (2020) show that international migration increases international trade flows in Canada and in the US. Di Giovanni et al. (2015) and Klein and Ventura (2009) study the welfare effects of international migration, but abstract from sector-specific worker skills.

[6]Sabel et al. (2012) describe the role of migrants in the formation of new export sectors in Latin America. Other examples include the diffusion of crops during the Columbian Exchange (Crosby, 1973), the introduction of new varieties of wheat in the northeast of the US in the 19th century (Olmstead and Rhode, 2008), the introduction of wheat in North Africa during the diffusion of Islam (Watson, 1983), and the production of flowers by Dutch refugees in England in the late 16th century (Scoville, 1951).

[7]Recent work, including de la Roca and Puga (2017), has shown that a migrant's past environment shapes his learning and future productivity.

Figure 1: The Evolution of Economic Activity in Brazil's West



Notes: This figure shows the evolution of the the share of population, agricultural land, and agricultural value added in the West. The "March to the West" was first announced in 1937, but changes are only noticeable after 1950, following a series of economic and policy shocks.

relating the productivity of migrants to their origin. Second, different from these papers, we embed this heterogeneity into a quantitative general equilibrium framework to measure how migration shapes trade and sectoral specialization.

2 The March to the West

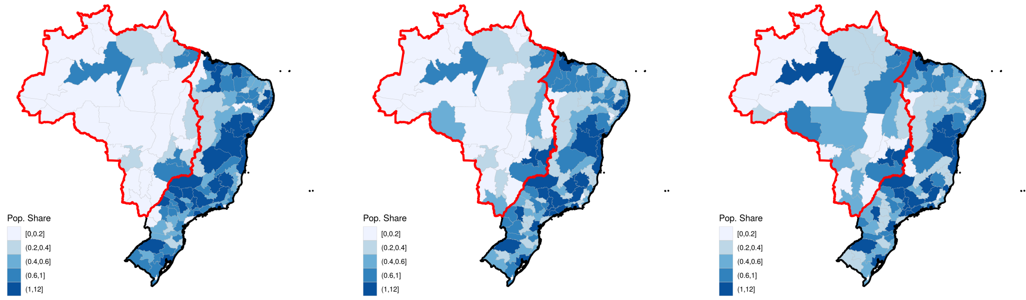
The West of Brazil is nowadays one of the World's major agricultural powerhouses, whose agricultural exports are comparable to those of large countries like Mexico and India. This status, however, came rather recently: As Figure 1 shows, the 1950s mark an inflection point in the evolution of Brazil's West. The share of Brazil's population living there remained around 7 percent before the 1950s, but it has doubled since then. Likewise, the share of agricultural land employed in the West increased by 12 percentage points, and the share of agricultural value added generated there increased by 20 percentage points. The sequence of maps in Figure 2 shows the area we term the West in a red contour and displays its uneven settlement, in which population grew fastest in regions close to the East.[8]

The Onset of the March. Until the 1950s, Brazil's population was concentrated in its eastern shores, mirroring historical economic development. The economy was organized mostly around export-oriented commodities such as sugarcane, coffee, and cotton, which benefited from easy access to the ports located along the Atlantic coast. Two noteworthy

[8] Brazilian States are officially divided in five broad regions based on socioeconomic and geographic features: Central-West, North, Northeast, Southeast, and South. Our analysis focuses on the occupation of the Central-West and the North, which we label the West for simplicity.

Figure 2: The Spatial Distribution of the Brazilian Population between 1950 and 2010

(a) 1950 (b) 1980 (c) 2010



Notes: The figure shows the percentage of total population which live in each meso-region, which is the geographic unit in our analysis. The red contour shows the meso-regions classified as the West. The West incorporates the North and the Central-West, two of the five official regions of Brazil.

exceptions were gold extraction in the state of Minas Gerais during the 18th century and rubber exploitation in the Amazon in the late 19th century. Beyond that, poor access to the interior severely limited economic activity there (Baer, 2001).

The March to the West began in the 1940s, when urbanization and demographic transition took off in Brazil. Concerned with food security and population pressure in the urban centers of the southeast, Getulio Vargas, the president at the time, initiated a large-scale project to occupy the Central-West region; he named the project the "March to the West". Part of the project aimed at changing Brazilians' perceptions about the West via public information campaigns, proclaiming for example that "the true sense of Brazilianness is the March to the West" (e.g., see Appendix Figure G.4 and Vargas, 1938). In addition, Vargas launched expeditions to the West, invested in expanding and improving railroads, and created institutions to establish new agricultural colonies (Villas Bôas and Villas Bôas, 1994).[9]

While Getulio Vargas' government gave the occupation of the West its initial thrust, it was not until the 1960s, when successive governments undertook larger investments in infrastructure to occupy the interior of Brazil, that the migration to the West became a large scale phenomenon.

The Progress of the March. The next major step of the March occurred in 1964, when the president Juscelino Kubitschek moved the Brazilian capital from the coastal city of Rio

[9]In 1941, Getulio Vargas launched the first expedition to the west called "Roncador-Xingu". The goal of this expedition was to chart unpopulated regions in the interior of Brazil that were amenable to the construction of new cities. To complement this expedition, in 1943 the government created the Fundação Brasil Central (FBC), an institution charged with creating new population nuclei in the Western and Central region of Brazil. FBC was also tasked with coordinating the construction of a railroad in Tocantins to cross the center of Brazil, near Brasilia.

de Janeiro to Brasília, a newly constructed city in the Central-West region. Complementing this political decision, the government built a series of new highways connecting the new capital to ports and cities along the coast.[10]

Between the 1960s and the 1980s, a military dictatorship in Brazil expanded Kubitschek's projects as to further integrate the North of Brazil. The government invested in new roads under a new national transportation plan (*Plano Nacional de Viação*), granted land to agricultural colonization companies, and created a free economic zone focused on assembly plants in Manaus, a city located along the Amazon river.[11] Alongside these investments in infrastructure, the Brazilian agricultural research institute EMBRAPA—founded in the 1960s—expanded its research on the adaptation of crops to regions closer to the tropics. An emblematic result of this effort was the adaptation of soybeans to tropical areas (Nehring, 2016; Sabel et al., 2012; Amann et al., 2018).[12]

Spurred by these public investments, the March progressed against the backdrop of rapid economic growth, structural transformation out of agriculture, and continued urbanization, especially in the Southeast. But as Brazil entered a decade-long period of economic depression and hyperinflation in the 1980s, the cycle of large-scale investments in infrastructure that started in the 1950s came to a halt.

Knowledge and Migration during the March. The knowledge of migrants has driven the expansion of new economic activities in several historical episodes (Olmstead and Rhode, 2008; Scoville, 1951; Watson, 1983). Sabel et al. (2012), p.181 document that this was also the case in Brazil during the 20th century, when *gauchos*, migrants from the South of Brazil, led 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." (Appendix B recounts similar anecdotes for coffee.)

The fact that Brazilian farms—even those having large acreage—are usually managed by small teams of workers is also suggestive of the important role of knowledge in the orga-

[10]See Morten and Oliveira (2016) and Bird and Straub (2020) for a detailed description of the roads that were constructed during this period to connect Brasília to peripheral regions.

[11]Many of the transportation investments and the land settlements during the military government were part of the federal government's regional development program for the Amazon, POLOAMAZONIA (Browder et al., 2008). POLONOROESTE, a regional development project that received financing and development assistance from the World Bank, focused on the forest area of Rondonia, near the Bolivian border.

[12]Pellegrina (2020) analyzes the general equilibrium impact of the expansion of soybeans to tropical areas during the 1970s. Bustos et al. (2016) show that the arrival of new varieties of soybeans in the 2000s released labor from agriculture and contributed to structural change in Brazil.

nization of production. In particular, the average number of workers per farm in the West is 3.5, with small variations across municipalities: In the 1st decile there are 2.5 average workers per farm; in the 10th decile, 4.5. Average land holdings, in contrast, vary substantially: In the 1st decile, farms have 40 hectares on average, whereas in the 10th decile they have an average of 650. Most farmers in the region, moreover, own and manage their farms. According to data from the agricultural census of 2006, approximately 90 percent of farm managers are land holders, as opposed to externally hired managers. Consistent with such numbers, the demographic census of 2010 indicates that 42 percent of agricultural workers classify themselves as a managers.

In summary, several factors promoted worker migration to the West: changes in migration and trade costs associated with the construction of roads and the federal government's propaganda, reductions in the price of land associated with land grants, productivity shocks associated with the adaptation of new crops, and the existing knowledge of workers to take advantage of them. The spatial economy model that we formulate will incorporate the influence of all these different shocks on the evolution of the economy.

3 Data and Motivating Facts

This section provides an overview of our data set (Appendix A contains the details.) It then presents four empirical facts that describe (i) Brazil's aggregate specialization patterns, (ii) the importance of migrants in the West, (iii) workers' heterogeneity, and (iv) the relation between a region's sectoral production and the geographic origins of its workforce.

3.1 Data sources

We collect data from various sources to construct a panel with information about employment, migration, gross output, and domestic and international trade for Brazil between 1950 and 2010. Our data contains 133 meso-regions—which can be aggregated into 26 States—,[13] two countries (Brazil and the rest of the world), and 13 economic activities (11 agricultural activities, manufacturing, and services).[14]

To measure migration flows and incomes, we use state-level migration and employment

[13]Meso-region is a political boundary defined by the Brazilian statistical bureau, IBGE, that combines municipalities according to similarities in their economic activity and labor markets. Previous research has also employed micro-regions, which are geographically more disaggregated (e.g. Adão, 2015). Assembling a panel of micro-regions over several decades brings several complications because new micro-regions were created, while existing ones changed boundaries. We therefore constructed our data set at the meso-region level, to strike a balance between consistency over time and geographic disaggregation.

[14]The 11 agricultural activities are: banana, cacao, coffee, cotton, corn, cattle, rice, soy, sugarcane, tobacco and a residual agricultural termed "rest of agriculture". Manufacturing also includes other tradables, such as mining.

variables coming from decadal Brazilian demographic censuses from 1950 until 2010. For 1970 onward, we also observe a sample of individual level micro-data. In particular, a detailed questionnaire is applied for approximately 25 percent of households. In addition to individual migration variables, such as origin and year of migration, this sample contains a detailed description of each individual's work: we observe total earnings and hundreds of activities (e.g. the crop a farmer grows).

We collect measures of gross output by meso-region and agricultural activity from *Produção Agrícola Municipal* (PAM) and Brazilian agricultural censuses. For non-agricultural activities, we use value added data from *Instituto de Pesquisa Economica e Aplicada* (IPEA) and generate gross output using value-added shares from WIOD.[15] We adjust sectoral value added in Brazil to match aggregate values reported in UN National Accounts. For the rest of the world, gross output and value added information come from FAO-STAT and UN National Accounts. International trade data comes from FAO, and domestic trade flows between Brazilian states from Brazilian statistical yearbooks and Vasconcelos (2001). Lastly, data on land use and total labor employment for Brazil are from IPEA and for the rest of the world from FAO.

3.2 Facts about Migration and Comparative Advantage

Fact 1: Since the 1950s, Brazil's aggregate exports have specialized in crops that the West exports more intensively than the East. Throughout the paper, we use relative bilateral exports (RBE) to describe the changes in a region's patterns of specialization. Given a common reference destination, which we take to be the rest of the world, we compute the specialization of region i in activity k (relative to crop k' and region i') as follows:

$$RBE_{ii',kk'} \equiv \frac{X_{iF,k}/X_{i'F,k}}{X_{iF,k'}/X_{i'F,k'}}, \tag{1}$$

where $X_{ij,k}$ are $i's$ sales to j in activity k. This index of relative bilateral exports is a useful indicator of comparative advantage for two main reasons. First, it is defined for pairs of regions and activities, as is the standard definition of comparative advantage based on autarky relative costs. Second, by fixing a destination market as reference, it focuses on supply-side sources of specialization.[16]

[15] Our paper focuses on the agricultural sector, which uses entirely non-imputed data from *Produção Agrícola Municipal* (PAM) and the agricultural census. In addition, our unit of analysis is the meso-region, which minimizes concerns about imputation at lower levels of disaggregation.

[16] French (2017) shows that this measure maps to Haberler's definition of comparative advantage, which is based on autarky opportunity costs, if trade costs are not sector-country specific. Supplementary Appendix Table S.1 shows that using instead Balassa's Revealed Comparative Advantage index provides similar conclusions. See French (2017) and Costinot et al. (2012) for a detailed discussion of different empirical measures

Table 1: The Evolution of Brazil's Trade Specialization (1950-2010)

	Brazil			East	West	Exports (%)	
	1950	1980	2010	2010	2010	1950	2010
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
corn	2.12	0.52	4.56	1.45	25.43	1.35	3.93
beef	1.12	1.77	5.69	3.40	21.05	2.17	6.89
soy	0.00	60.88	69.74	44.85	237.02	0.00	29.43
cotton	9.41	0.72	3.28	2.16	10.83	6.23	1.79
restofag	1.01	2.19	2.51	2.52	2.47	13.66	28.40
sugarcane	18.82	27.91	52.28	57.32	18.36	5.70	16.13
rice	0.19	0.07	0.43	0.48	0.06	0.44	0.56
coffee	655.55	339.56	166.11	190.24	3.95	59.22	8.01
cocoa	273.44	299.89	15.00	17.22	0.11	8.98	0.40
tobacco	6.70	19.92	64.27	73.84	0.00	1.76	4.39
banana	5.09	1.81	0.38	0.44	0.00	0.49	0.06
agriculture	4.98	5.08	6.14	5.35	11.50	79.77	31.96
mfg	1.00	1.00	1.00	1.00	1.00	20.23	68.04

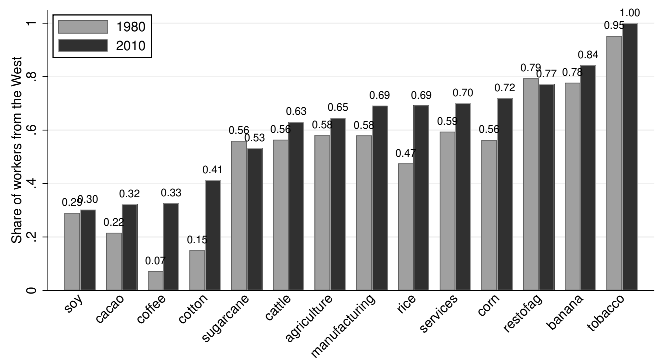
Notes: Columns (1) through (3) present our index of relative bilateral exports for each activity (relative to manufacturing and relative to the rest of the world) for 1950, 1980, and 2010. Columns (4) and (5) present the same index broken down by East and West of Brazil, for 2010. Columns (6) and (7) present the share of each activity in Brazil's agricultural exports in 1950 and 2010.

In Table 1, we apply this measure for each activity k in our data, fixing i = Brazil, i' = Rest of the World (F), and k' = Manufacturing. We sort crops by the intensity with which Brazil's West's specializes in them in the year 2010, for which we have exports data by Brazilian region. The table shows that by 2010, Brazil is intensively specialized in several commodities relative to ROW, among them coffee and soybean. Leading to 2010, however, Brazil's specialization changed substantially. For Brazil as a whole, RBE of commodities such as soy, beef, and corn — in which the West specializes — grew rapidly, while it plummeted for traditional ones such as coffee, cacao, and banana. Along with these changes in relative specialization came a large change in Brazil's export basket, in which these three crops now account for approximately 13 percent of total exports and 40 percent of agricultural exports.

Fact 2: Across regions and activities in the West, the participation of Eastern migrants was large and varied substantially. The dramatic rise of the West was driven, in no small part, by migration from the East: On average Eastern migrants account for 30 percent of the West's workforce in 2010. This share, however, varies substantially across locations and activities. Focusing on a statistic which in our theory will measure the impact of migration, Figure 3 shows that the share of non-migrants in employment is quite low for some activities, especially for soy, both in 1980 and 2010. Within activities, there is

comparative of advantage.

Figure 3: The Share of Workers in the West who were born there, by Economic Activity



Notes: The figure presents the fraction of the total employment in the West, in each year and economic activity, comprised by workers born in the West.

also large variation in the non-migrant share across regions, as shown in Appendix Figure G.5. These two facts suggest that migration had a differential impact across crop-region pairs and that workers followed clear sorting patterns.

Fact 3: Upon migration, farmers originating in regions with high employment in an agricultural activity are more likely to work in that activity and have higher income than other migrants doing so. We now document worker heterogeneity by comparing migrants working in the same activity and destination, but who come from origins with different patterns of agricultural specialization. Specifically, we estimate the following regressions:[17]

$$\log \text{income}_{ij,kt} = \iota^I_{j,kt} + \iota^I_{ij,t} + \alpha^I \log \text{workers}_{i,kt-1} + \epsilon^I_{ij,kt}, \tag{2}$$

$$\log \text{workers}_{ij,kt} = \iota^W_{j,kt} + \iota^W_{ij,t} + \alpha^W \log \text{workers}_{i,kt-1} + \epsilon^W_{ij,kt}, \tag{3}$$

where i indexes the origin region, j the destination region, k the agricultural activity and t the year. To gain precision, we stack data for the years $t = 2000$ and $t = 2010$. With an eye

[17]Throughout the paper, we focus on identifying the knowledge of migrants within the agricultural sector. We do so for several reasons, beyond the outsize importance of Brazil's agriculture in the world. Most importantly, the census provides quite granular sector definitions within agriculture, which allows us to compare workers within narrowly defined employment categories; for manufacturing, on the contrary, the sectoral classification is quite broad, especially if one is to create consistent categories across years. This is especially problematic in constructing gross output series, which we require for calibrating our model. Moreover, as discussed earlier, there is an abundance of anecdotal evidence on the diffusion of knowledge in agriculture, which we rigorously map to data here. That said, Appendix Table C.6 shows that when we focus on a sample of manufacturing workers and sectors our results broadly concur with those in this section.

toward our modeling strategy later, we use a thirty-year lag as our measure of $\text{workers}_{i,kt-1}$.[18] To avoid the inclusion of the same farmer in both sides of the equation, we exclude non-migrants from our sample.

Table 2: The Relation between Farmers' Income, Choices, and Region of Origin

	OLS	OLS	OLS	PPML	PPML	PPML
	(1)	(2)	(3)	(4)	(5)	(6)
a: Income (logs)						
Farmers in origin	0.016***	0.023**	0.047***	0.045***	0.083***	-
	(0.005)	(0.011)	(0.017)	(0.016)	(0.024)	-
R^2	0.342	0.702	0.729	-	-	-
Obs	6794	6794	5180	6794	5180	
b. Farmers in destination (logs)						
Farmers in origin	0.074***	0.075***	0.101***	0.120***	0.131***	0.165***
	(0.010)	(0.014)	(0.020)	(0.019)	(0.030)	(0.018)
R^2	0.183	0.751	0.774	-	-	-
Obs	7375	7375	5609	7375	5609	127950
c. Worker heterogeneity parameters						
κ	4.625	3.260	2.148	2.666	1.578	
β	0.016	0.023	0.047	0.045	0.083	
Dest-Act-Year FE	Y	Y	Y	Y	Y	Y
Dest-Orig-Year FE		Y	Y	Y	Y	Y
Above Q1			Y		Y	
Include zeros						Y

Notes: * / ** / *** denotes significance at the 10 / 5 / 1 percent level. Standard errors, multiway clustered at the destination-crop-year and origin-year level, are in parenthesis. The unit of observation is a destination-activity-origin-year cell. Columns 3 and 5 drop cells in the bottom quartile in the distribution of the dependent variable. "Farmers in origin" is the log of farmers in the same activity in the region of origin lagged by thirty years. Column 6 includes zeros in bilateral migration pairs for destinations with at least one producing farmer. We exclude return migrants and non-migrants.

In equation (2), α^I measures the association between a farmer's comparative advantage and the size of the workforce in his origin region. In equation (3), α^W captures how this comparative advantage translates into activity choice. By including destination-activity fixed effects ($\iota^W_{j,kt}$ and $\iota^I_{j,kt}$), we control for any factor that is common across workers in destination j and activity k, including natural advantages, local institutions, price shocks, access to inputs, and any local Marshallian externality, such as knowledge spillovers from other workers. By including origin-destination fixed effects ($\iota^W_{ij,t}$ and $\iota^I_{ij,t}$), we control for common factors that are origin and destination but not activity-specific, such as access to education and financial capital, bilateral migration costs, and weather similarity. To give an example, if demand for soy is relatively large in Mato Grosso, we identify worker heterogeneity from the fact that, facing the same local conditions in Mato Grosso, workers come disproportionately from Rio

[18]Data on a worker's meso-region of birth is unavailable. Therefore, to run the regressions at the meso-region level, we define migration based on a workers current and previous meso-region of residence. We discuss alternative measures of migration in the robustness section.

Grande do Sul where, 30 years earlier, soy employment was high relative to other regions and other crops.

Panel A in Table 2 shows estimates of α^I using our detailed census sample. In Column 1, an increase of 1 percent in the number of agricultural workers in the region of origin in a given activity increases the average income of workers in the destination in this same activity by 0.016 percent. This effect increases to 0.023 when we include destination-origin fixed effects (Column 2). Since origin-destination-crop cells containing zero workers are frequent in our sample, we explore three additional specifications. In Column 3 we drop observations below the bottom quartile in the distribution of $\text{workers}_{ij,kt-1}$. As one would expect if low values of our regressor make sampling zeros more likely, this specification increases our coefficients to 0.047. We repeat Columns 2 and 3 using a PPML estimator, to avoid the bias due to heteroskedasticity (Silva and Tenreyro, 2006). Doing so increases our estimate to 0.083 percent.

In Panel B, the elasticity of migrants' employment with respect to the number of farmers in the origin, α^W, rises from 0.074 to 0.131 across the same specifications. In this panel, we can also include all zeros in our sample; doing so increases the coefficient to 0.165 (Column 6). Taken together, the results in Panels A and B suggest that the comparative advantage of migrants is shaped by the employment structure of their origin region.[19] To capture the reduced-form elasticities presented in this section as parsimoniously as possible, our model will assume that workers acquire activity-specific knowledge in their origin region, through a learning externality.

We next demonstrate that our conclusions regarding worker heterogeneity are robust to alternative econometric specifications and interpretations of the evidence. We also provide supporting evidence coming from migration gravity regressions.

Fact 3: Robustness. Our results on worker heterogeneity are largely robust to a number of alternative econometric specifications and approaches to measurement. Appendix Tables C.3, C.5, and C.4 include the following checks. First, our results are not sensitive to defining regions at smaller (or larger) levels of regional aggregation, which suggests workers do not sort within regions into land of different quality. Second, our results remain if we define migrant origin based on the state of birth (rather than previous meso-region). Third, we estimate our equations dropping crops one at a time, showing that no single one drives our results. Fourth, we consider total revenues, total incomes, and total land use as alternatives

[19] Bazzi et al. (2016) exploit exogenous variation in migration to show that regions that receive migrants from regions with similar agroclimatic conditions tend to have higher rice yields. Our evidence comes instead from the fact that our very detailed data allow us to compare workers within narrowly defined regions and activities, but originating in regions with different employment structures.

to employment as a measure of economic activity in the origin. Fifth, to include earlier census years in our sample, we experiment with 20-year lags. Lastly, Figure C.1 in the Appendix shows non-parametrically the variation that identifies α^I and α^W. In both cases, a log-linear relationship provides a good approximation.

Fact 3: Alternative interpretations. We now discuss interpretations of the data that might be alternatives to our preferred one, namely, that workers' skills are heterogeneous and activity-specific. These results are collected in Appendix Table C.7. We start by emphasizing that our fixed effects strategy requires such alternatives to be origin-activity specific. One such alternative is that our results are driven by the composition of workers in each cell. Using our individual-level data, we therefore control for a host of socio-economic and demographic characteristics, and show that this is not the case (column 3). A second alternative is that early migrants come from particular regions and gain a first-player advantage in specific activities where they arrive. We therefore include a full set of fixed effects for the years of residence in the destination and show our results are not driven by the timing of migration (column 4).

Another alternative interpretation is that workers are ex-ante homogeneous, but those who migrated earlier guarantee better wages to newcomers from their own region. Column 5 reproduces our main results, controlling for the number of workers who previously migrated from the same origin to the same destination to produce the same activity (i.e., lagging $\text{workers}_{ij,kt-1}$, the dependent variable in equation (3)). Worker homogeneity, moreover, is inconsistent with the results we discuss below in Fact 4. A related alternative interpretation is that workers are homogeneous but differ in their physical capital endowment. Although in our theory one could relabel worker heterogeneity as "equipped workers" and our conclusions would remain, we think this is not a central mechanism for two reasons. First, as discussed before, no single crop drives our results, although they may have different capital requirements. Second, even if we drop the most capital-intensive activities, we still obtain similar results (column 1).[20]

We next address the possibility that migrants are selected in a way that the fixed effects in our main specification do not address. We do so in two different ways. Following Dahl's (2002) strategy for controlling for selection bias, we first estimate our earnings equation at the individual level, augmenting the equation to include a subset of the choice probabilities of individuals from origin i across options j, k. The identifying assumption is that the selectivity bias is a function of these probabilities, which we consistently estimate by computing

[20]We drop from our regressions the three activities with highest tractor-use per farmer (cotton, soy and tobacco). In Appendix Table C.6, we also show results for the manufacturing-sector and find similar results, despite the capital requirements in manufacturing being different agriculture.

population migration shares (i.e., $\lambda_{i,jk}$ in our model). These probabilities, which vary at the i, j, k level, capture forces both related and unrelated to wages that cause individuals to select into a location and activity (see Appendix C.1.2). Second, we include indicators for state of origin, activity, and year triplets to absorb any unobserved factor that is common among workers coming from the state of origin who end up producing the same activity. As reported in columns 6 and 7 of Table C.7, in both cases, our point estimates are quite stable. By limiting the scope for unobserved selection to drive our results, these two specifications make a causal interpretation more plausible.

Lastly, while we follow a common approach in introducing learning spillovers in our model, an alternative is that workers invest in activity-specific abilities when they are young. Although the method through which workers acquire knowledge differs, what is critical to our results is that in both cases knowledge be related to employment structure in the origin. In the Supplementary Appendix Section SJ, we discuss a theoretical framework in which workers acquire activity specific knowledge via a combination of these two economic mechanisms.[21]

Fact 3: Supporting evidence from migration and agricultural similarity. To provide supporting evidence that farmers' migration decisions relate to their skills, we show that crop-specific agricultural similarity is an important driver of migration. Motivated by Bazzi et al. (2016), we construct an agricultural similarity index using the GAEZ data set (IIASA/FAO, 2012) on agricultural suitability. First, we compute the similarity between region i and j in activity k, $\mathcal{A}_{ij,k}$:

$$\mathcal{A}_{ij,k} = -\left|y_{i,k}^{FAO} - y_{j,k}^{FAO}\right|$$

where $y_{i,k}^{FAO}$ is the average potential yield across cells contained in region i in crop k, coming from GAEZ (under high-input techniques). We normalize $\mathcal{A}_{ij,k}$ to be between zero and one for each crop, before computing the unweighted average of $\mathcal{A}_{ij,k}$ across activities k between regions i and j, $\mathcal{A}_{ij}$, which we use in our regressions.

Table 3, Column 1 shows that an increase in agricultural similarity is strongly associated with larger migration flows between regions. Column 2 shows that these results hold after controlling for distance between origin and destination. This specification also provides direct evidence that migration decreases with distance.

[21] The evidence that we show is inconsistent with a model where workers are homogeneous but have switching costs between sectors. Appendix C shows that a version of this model in which workers cannot switch jobs tend to generate α^W equal to 1 in equation (3) and α^I equal to 0 in equation (3).

Table 3: Migration of Agricultural Workers and Agricultural Similarity

	DV: Log of migration flows	
	(1)	(2)
$\log(\mathcal{A}_{ij})$	0.963***	0.398***
	(0.065)	(0.045)
$\log(dist_{ij})$		-1.097***
		(0.025)
R^2	0.185	0.462
Obs	16205	16205
Origin-Year and Destination-Year FE	Y	Y

Notes: * / ** / *** denotes significance at the 10 / 5 / 1 percent level. Robust standard errors clustered at the destination-year level. Agricultural similarity, $\mathcal{A}_{ij}$, averages the indexes $\mathcal{A}_{ij,k} = -\left|y_{i,k}^{FAO} - y_{j,k}^{FAO}\right|$ across crops, where $y_{i,k}^{FAO}$ averages potential yields in crop k for all cells contained in region i, using data from FAO-GAEZ.

Fact 4: Production is higher in regions and crops in which the labor-force composition favors farmers emigrating from origins with high-employment in that same crop. To study the impact of worker heterogeneity on aggregate measures of production, we document an aggregate counterpart to Fact 3. We estimate

$$\log\left(y_{j,kt}\right) = \iota_{j,t} + \iota_{k,t} + \alpha^A \underbrace{\log workers_{j,kt}}_{Abundance} + \alpha^C \underbrace{\log \sum_i \omega_{ij,kt} workers_{i,kt-1}}_{Composition} + \epsilon_{j,kt}, \tag{4}$$

where $workers_{j,kt}$ is the aggregate number of workers producing activity k in destination j and $\omega_{ij,kt}$ is the share of workers in destination j producing activity k who come from origin i. We estimate equation (4) using two dependent variables: quantity and revenues. The first set of fixed effects captures any level effect such as the size of a region or the overall demand for agricultural goods; the second set captures crop-specific characteristics, such as 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 4 shows, first, that a 1 percent increase in the abundance of farmers in a region is associated with a 0.9 percent increase in revenues. Second, it shows that a 1 percent increase in our measure of composition is associated with a 0.2 percent increase in revenues, which indicates that the composition of farmers is strongly associated with total size of the sector in a region (column 1). Our results are robust to controlling for workers' socioeconomic characteristics, the total share of migrants in a region, and a squared polynomial of the log of average farm size (column 2), and we obtain similar results when output quantity is our dependent variable (columns 4 and 5). Our results are also robust to an IV strategy that mitigates the impact of worker sorting in response to productivity shocks that are crop-destination specific—which would lead us to overestimate the impact of the composition of workers (columns 3 and 6).[22]

Table 4: Relationship between the Composition of Farmers and Agricultural Output

	Revenues			Quantity		
Explanatory Variable	(1)	(2)	(3)	(4)	(5)	(6)
Abundance	0.917***	0.691***	0.668***	0.983***	0.654***	0.643***
	(0.040)	(0.036)	(0.038)	(0.038)	(0.030)	(0.032)
Composition	0.191***	0.154***	0.201***	0.211***	0.203***	0.221***
	(0.046)	(0.037)	(0.043)	(0.044)	(0.029)	(0.034)
R^2	0.862	0.911	0.737	0.837	0.921	0.793
Obs	1381	1381	1375	1427	1427	1427
Region-Year FE	Y	Y	Y	Y	Y	Y
Activity-Year FE	Y	Y	Y	Y	Y	Y
Controls: Migrants		Y	Y		Y	Y
Controls: SES + Farm size		Y	Y		Y	Y
IV estimation			Y			Y

Notes: * / ** / *** denotes significance at the 10 / 5 / 1 percent level. Standard errors clustered at the meso-region level in parenthesis. The composition of farmers is the log of the number of farmers in the origin as weighted by the share of farmers in a destination. Regressions include the years of 2000 and 2010.

Lastly, in Appendix Table C.8, we show that our general conclusions remain, albeit with less statistical precision, when we use state-level variation. Using this level of variation, we can show that the composition of workers also shapes state exports specialization..

4 A Model of Migration and Comparative Advantage

Building on the dynamic approach by Allen and Donaldson (2020), we develop a model of trade and migration in which comparative advantage is driven by regional productivity, land abundance, and the activity-specific knowledge of workers. We use the model to obtain, first, an analytical result about the relation between migration and trade and, second, to derive a sufficient statistic that will guide our quantitative analysis. To simplify our exposition, we present a stripped down version of the model. (Supplementary Appendix SA-SE contains details and proofs.)

4.1 Environment

Geography and Commodities. Time is discrete and indexed by t. The world, $\mathcal{W}$, consists of many regions $j = 1, \ldots, I$ that comprise the Home country, and of a rest of the

[22]We instrument the composition term in (4) with $\sum_i \omega_{ij,t}^{IV} workers_{i,kt-1}$, where $\omega_{ij,kt}^{IV} = \left(dist_{ij}^{-1} / \sum_l dist_{lj}^{-1}\right) workers_{i,kt-1}$ and $dist_{ij}$ is the euclidean distance between meso-region i and j. This instrument captures variation in the composition of the labor force that is entirely driven by distance between origin and destination, and excludes variation associated to land productivity.

world composite, denoted by F. There are $k = 1, \ldots, K$ activities (or goods), and each region produces a unique variety of each good. At each time, the geography of the economy is given by a set of natural advantages, a matrix of iceberg trade costs, a matrix of iceberg bilateral migration costs, and a land endowment: $\{A_{j,kt}, \tau_{ij,kt}, \mu_{ij,kt}, H_{j,t}\}$. We omit time indexes whenever unnecessary for exposition.

Technology. In each region j, representative firms aggregate varieties of good k into aggregate good output with constant elasticity of substitution η_k.[23] Good k output is likewise aggregated into final output, with constant expenditure shares a_k. Only varieties are traded. Correspondingly, the price index for activity k in region j is given by $P_{j,kt}^{(1-\eta_k)} = \sum_{i \in \mathcal{W}} \left(\tau_{ij,kt} p_{i,kt}\right)^{(1-\eta_k)}$, while that of final output is $P_{j,t} = \prod_{k=1}^{K} P_{j,kt}^{a_k}$.

To produce q units of good k in region i, a worker with knowledge s combines land l and final output m, according to:

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

where α_k measures the share of value-added in production and γ_k measures the land intensity of good k. For non-agricultural activities, we set $\gamma_k = 0$.

Demography and Preferences. People live two periods, young and old. An adult at time t, upon observing her knowledge, chooses an activity-location pair and then spawns one child.[24] To avoid introducing forward-looking agents, we work with non-altruistic parents. Let $L_{j,t}$ denote the adult population at time t in region j.[25]

Adult workers born in $t-1$ maximize welfare by choosing where to live and in which activity to work at time t:

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

[23] For simplicity, we adopt an Armington formulation within activities k. Our focus is on intersectoral comparative advantage, which is still governed by differences in relative costs. We do not adopt the formulation in Sotelo (2020), which sets $\eta_k \to \infty$ for agricultural goods, to keep our simulations quantitatively tractable, since our model features dynamics and a large number of goods and regions.

[24] The difference in fertility rates between migrants and non-migrant families is small: according to the census of 2010, migrants families have on average 0.02 additional children. We therefore follow Allen and Donaldson (2020) and assume that each individual spawns one child. Note also that allowing workers to consume when young and allowing their parents to include the child's consumption, one can relax the extreme form of non-altruism we employ here.

[25] According to the census of 2010, migrants families have on average only 0.02 additional children compared to nonmigrant families. We therefore follow Allen and Donaldson (2020) and assume that each individual spawns one child. Note also that, allowing workers to consume when young and allowing the parent to care about the child's consumption when young, we can include a form of altruism that would retain the main structure our model.

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

$$W_{ij,kt} = \frac{w_{j,kt} s_{i,kt}}{\mu_{ij,kt} P_{j,t}}, \tag{5}$$

where $w_{j,kt}$ is the wage per efficiency unit of labor (i.e., the return to a unit knowledge), $\mu_{ij,kt}$ represents iceberg migration costs that reduce utility directly, and $P_{j,t}$ is the price index of final output consumption in destination j.[26] Reflecting our empirical findings, $s_{i,kt}$ is a farmer's knowledge to produce in activity k, which depends on the region she comes from. Workers can migrate within the Home country, but not between Home and Foreign.

Knowledge Endowment. A child born in i at time $t-1$ is characterized by a vector of activity-specific productivities, $s_{i,kt}$, which depends on the employment structure in the region where she is born, according to:

$$s_{i,kt} = \bar{s}_k L_{i,kt-1}^{\beta}. \tag{6}$$

That is, motivated by Fact 3 in Section 3, we assume that knowledge depends on good-specific employment.[27]

4.2 Equilibrium

We denote by $c_{j,kt}$ the cost of a unit of the input bundle, given by $c_{j,kt} = \bar{c}_k \left(w_{j,kt}^{1-\gamma_k} r_{j,t}^{\gamma_k}\right)^{\alpha_k} P_{j,t}^{1-\alpha_k}$, where $w_{j,kt}$ are efficiency wages, $r_{j,t}$ are land rents and $\bar{c}_k$ is a technology constant.

The share of region $j's$ sector k expenditure going to origin i is then given by $\pi_{ij,kt} = \left(c_{i,kt}\tau_{ij,kt}/A_{i,kt}\right)^{1-\eta} / P_{j,kt}^{1-\eta}$. Likewise, optimal worker sorting gives the share of workers from i choosing to work in region j and activity k:

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

where $W_{ij,kt}$ is given by equation (5) and $\Xi_{i,t}^{\kappa} \equiv \sum_j \sum_k W_{ij,kt}^{\kappa}$.[28] We define the effective units

[26] We introduce preference shocks, as opposed to productivity shocks, since they readily deliver the econometric formulation in equations (2) and (3). This formulation also recognizes that workers migrate for many reasons beyond productivity differentials.

[27] This formulation captures in a transparent way the relation between heterogeneity and migrant origin from Fact 3. Since our regressions include origin-destination fixed effects, we cannot determine if our parameterization of $s(L_{i,kt-1})$ should depend on the level or the share of agricultural employment in the origin region. Our formulation builds on a large literature emphasizing scale effects in productivity, both in growth and international economics (Ramondo et al., 2016). Supplementary Appendix Figure S.5 shows that our quantitative results remain almost unchanged when considering an alternative formulation in which knowledge depends on agricultural employment shares, $s_{i,k} = (L_{i,kt-1}/L_{i,t-1})^{\beta}$.

[28] Migration costs $\mu_{ij,kt}$ and knowledge $s_{i,kt}$ act together to rationalize migration data. Using our model,

of labor migrating from i to region j, activity k as

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

To close the model, we assume that land rents are paid to local landowners. Total expenditure in region j reflects final and intermediate expenditure:

$$X_{j,t} = \sum_k w_{j,kt} E_{j,kt} + r_{j,t} H_{j,t} + \sum_{k=1}^{K} (1 - \alpha_k) Y_{j,kt},$$

where $Y_{j,kt}$ denotes revenues in activity k, $Y_{j,kt} = \sum_i \pi_{ji,kt} a_k X_{i,t}$, and $E_{j,kt} = \sum_i E_{ij,kt}$ is the supply of efficiency units there.

Definition 1. (Equilibrium) Given the geography and initial labor allocations, $\{L_{i,k0}\}_{ik}$, a general equilibrium is a sequence of factor prices and labor allocations $\{r_{j,t}, w_{j,kt}, E_{j,kt}\}_{jkt}$, such that, for each region j and activity k and time t: (i) Workers choosing to migrate to j, k satisfy (7) given (6) and factor prices, (ii) the market for efficiency units of labor clears, (iii) land markets clear for region j, and (iv) trade is balanced.

The characterization of equilibria in spatial models with multiple activities is a current area of work, and, to the best of our knowledge, there are no general sufficient conditions that guarantee uniqueness. But we can build on current work to obtain a partial characterization. To begin, note that our model features no contemporaneous externalities, since $s_{i,kt}$ depends only on past allocations. Therefore, in the case of free trade, our trade model satisfies the conditions for uniqueness laid out by Kucheryavyy et al. (2016). Moreover, a single-activity version of this model satisfies the conditions for period-by-period uniqueness laid out by Allen and Donaldson (2020). In these cases it follows that, given the economy's initial conditions, the equilibrium path to the steady state is unique for all t.

In a steady state, however, our economy features contemporaneous externalities coming from the effect of equilibrium labor allocations on worker heterogeneity, and we cannot rule out equilibrium multiplicity. Because these externalities operate by increasing the effective supply of workers, they are distinct from Marshallian externalities or scale effects in models of imperfect competition, which instead raise the demand for labor (Bartelme et al., 2019; Kucheryavyy et al., 2016). Supplementary Appendix SI discusses two examples. First, a closed economy lacking internal geography will feature equilibria with corners of zero production, provided β is large enough relative to the elasticity of substitution between

in Section 5 we will be able to separate the contribution of skill specificity to what would otherwise look as migration costs.

activities. Second, in the limit in which Brazil is a small open economy, any region-activity pair with positive productivity will have positive employment, so we can rule out corners with zero employment.[29] The reason is that as long as at least one region produces a crop, next period there will be workers willing to go to any other region to do it.

4.3 How Migration Costs Shape Comparative Advantage

This section studies how migration shapes comparative advantage in our model, and connects our work to a long tradition studying how a country's cost structure relates to trade patterns. The general equilibrium interactions between marginal costs and migration in our model are, unfortunately, too complex to characterize them as explicit functions of exogenous parameters. We therefore derive analytical results using a pared down version of our model. We let labor be the only productive factor ($\gamma_k = 0, \forall k$) and impose equal value-added shares and trade elasticities across activities (i.e., $\alpha_k = \alpha$ and $\eta_k = \eta$, $\forall k$). Moreover—and in this section only—we assume that workers are born with an activity-specific type k and, hence, their only choice is their location, which is still subject to idiosyncratic shocks and migration costs μ_{ij}. While extreme, this formulation draws out clearly the implications of mobility in space.

4.3.1 Autarky opportunity costs

Haberler's classic definition of comparative advantage states that region i has a comparative advantage in activity k (relative to region j and activity k') if i's autarky opportunity cost of producing k in terms of k' is lower than that of j.[30] To study autarky opportunity costs, we fix $j = F$ and consider two different scenarios in which region i's trade costs with any other region are prohibitive. To avoid clutter, we drop time subscripts and denote all predetermined labor allocations with a "0" superscript.

First, in the case in which internal migration costs are prohibitively high, $\mu_{ij} \to \infty$, $\forall i \neq j$, stating that i has a comparative advantage in k (relative to k' and F) is equivalent to the following inequality:

$$\left(\frac{s_{i,k} L^0_{i,k}}{s_{i,k'} L^0_{i,k'}}\right)^{\alpha} \frac{A_{i,k}}{A_{i,k'}} > \left(\frac{s_{F,k} L^0_{F,k}}{s_{F,k'} L^0_{F,k'}}\right)^{\alpha} \frac{A_{F,k}}{A_{F,k'}}. \tag{9}$$

[29] In defining a small economy, we follow Alvarez and Lucas (2007) (see Supplementary Appendix SB). Intuitively, we study the limit of our economy as we let the labor endowments of each region in the Home economy vanish, while keeping constant for each region the ratios of total labor to sectoral productivity and total land.

[30] Deardorff (2005) discusses the application and generalization of this idea. French (2017) applies it to modern quantitative frameworks, but abstracts from internal geography.

Second, when internal migration is costless, $\mu_{ij} = 1, \forall i, j$, the same statement is equivalent to:

$$\left(\frac{\mathcal{S}_{H,k}}{\mathcal{S}_{H,k'}}\right)^{\alpha} \frac{A_{i,k}}{A_{i,k'}} > \left(\frac{\mathcal{S}_{F,k}}{\mathcal{S}_{F,k'}}\right)^{\alpha} \frac{A_{F,k}}{A_{F,k'}}, \tag{10}$$

where $\mathcal{S}_{H,k} \equiv \sum_i s_{i,k} L^0_{i,k}$.

Expressions (9) and (10) reveal that knowledge acts as a productivity shifter in our model and that migration determines its allocation. When migration costs are prohibitive, the knowledge available in a location comes solely from workers born there. With free mobility, in contrast, workers can bring their knowledge into a destination from any origin; this is reflected in the term $\mathcal{S}_{H,k}$, which measures the aggregate effective supply of workers at Home, for activity k.[31]

Migration therefore undoes any previous correlation between local worker productivity and local natural productivity, and its impact on regional comparative advantage is ambiguous: it will strengthen regional comparative advantage if high knowledge workers were born in low productivity locations, but it will weaken it the opposite is true. In our application, the West's rapid productivity growth in certain activities led migration to strengthen comparative advantage.

4.3.2 Regional and Aggregate Trade Specialization

How do autarky opportunity costs shape *realized* free-trade specialization, i.e., when $\tau_{iF,k} = \tau_{Fi,k} = 1$? We now turn to study the conditions under which $RBE_{iF,kk'} > 1$. It is straightforward to show that when migration costs are prohibitive, autarky opportunity costs fully determine trade specialization. This result is intuitive, since in this case a region really is simply a separate economy; the elasticity of trade flows to autarky costs is exactly $(\eta - 1)\left[1 + \alpha(\eta - 1)\right]^{-1}$ (see Supplementary Appendix Equation (52)).

Under free internal migration, however, this is no longer the case. $RBE_{iF,kk'} > 1$ is equivalent to the following relation between fundamentals:

$$\left[\left(\frac{\mathcal{S}_{H,k}\mathcal{A}_{i,k}}{\mathcal{S}_{H,k'}\mathcal{A}_{i,k'}}\right)^{\alpha}\left(\frac{A_{i,k}}{A_{i,k'}}\right)\right]^{\frac{\eta-1}{1+\alpha(\eta-1)}} > \left[\left(\frac{\mathcal{S}_{H,k}\mathcal{A}_{F,k}}{\mathcal{S}_{H,k'}\mathcal{A}_{F,k'}}\right)^{\alpha}\left(\frac{A_{F,k}}{A_{F,k'}}\right)\right]^{\frac{\eta-1}{1+\alpha(\eta-1)}}, \tag{11}$$

where $\mathcal{A}_{i,k} \equiv A_{i,k}^{\kappa\frac{\eta-1}{1+\kappa+\alpha(\eta-1)}} / \sum_h A_{h,k}^{\kappa\frac{\eta-1}{1+\kappa+\alpha(\eta-1)}}$.[32] This inequality provides three insights into

[31] The key assumption delivering this specific functional form is that workers have types. In that case, together with extreme-value idiosyncratic tastes, knowledge $s_{i,k}$ does not by itself generate differential sorting from region i into other regions. Allowing workers to also choose over activities, as we do in our main model, will tend to strengthen the correlation between natural advantage and comparative advantage. The reason is that sectoral productivity differences within a region will induce differences in wages, which in turn drive the equilibrium labor allocation.

[32] The exponent of $A_{i,k}$ in $\mathcal{A}_{i,k}$ increases in κ. As we let $\kappa \to \infty$ workers become homogeneous and therefore

specialization. First, just like with opportunity costs, the impact of migration on trade specialization in equation (11) is ambiguous and captured by $\mathcal{S}_{H,k}$. Second, direct comparison of expressions (10) and (11) reveals that autarky opportunity costs do not fully determine specialization, since the term $\mathcal{A}_{i,k}$ is absent from the costs inequality (10). The reason is that, with free migration, region i competes with other regions in its own country for the same workers, and the relative payments this region can offer depends on its sales. As opposed to autarky, having a high productivity relative to the rest of the country under free trade —i.e., a high $\mathcal{A}_{i,k}$— attracts a larger labor supply and reduces marginal costs.[33] Thus the relation between autarky opportunity costs and free trade specialization could even be overturned if a region is relatively productive in a crop when compared in isolation to F, but not relative to the rest of the country. Third, since aggregate knowledge $S_{H,k}$ affects all regions in Home equally, it does not shape comparative advantage nor trade flows across regions within the same country.

Turning to country-wide specialization, when internal migration costs are prohibitively high, autarky opportunity costs fully determine aggregate specialization, just as in the regional case (see Supplementary Appendix Equation (58)). Specifically, the Home economy specializes in exporting sector k if a weighted average of its autarky marginal costs is low enough.

But under free internal migration, $RBE_{HF,kk'} > 1$ is equivalent to

$$\left(\frac{S_{H,k}}{S_{H,k'}}\right)^{\frac{\alpha(\eta-1)}{1+\alpha(\eta-1)}} \frac{\sum_{i\in H} \mathcal{A}_{i,k}^{\frac{\alpha(\eta-1)}{1+\alpha(\eta-1)}} A_{i,k}^{\frac{\eta-1}{1+\alpha(\eta-1)}}}{\sum_{i\in H} \mathcal{A}_{i,k'}^{\frac{\alpha(\eta-1)}{1+\alpha(\eta-1)}} A_{i,k'}^{\frac{\eta-1}{1+\alpha(\eta-1)}}} > \frac{\mathcal{S}_{F,k}^{\frac{\alpha(\eta-1)}{1+\alpha(\eta-1)}} A_{F,k}^{\frac{\eta-1}{1+\alpha(\eta-1)}}}{\mathcal{S}_{F,k'}^{\frac{\alpha(\eta-1)}{1+\alpha(\eta-1)}} A_{F,k'}^{\frac{\eta-1}{1+\alpha(\eta-1)}}}. \tag{12}$$

Expression (12) shows that aggregate knowledge $\mathcal{S}_{H,k}$ also shapes aggregate specialization — an aggregation result that follows from the fact that every region within Home has equal access to this knowledge. Migration again has an ambiguous impact on comparative advantage, only now in the aggregate. If, on average, labor was initially allocated to regions where labor productivity is low, migration will undo that correlation, and amplify relative productivity differences.[34]

more elastic to wage differences across regions. In this limit, the competition for workers has its largest effect on marginal costs.

[33] The exponent of $A_{i,k}$ in $\mathcal{A}_{i,k}$ increases in κ. As we let $\kappa \to \infty$ workers become homogeneous and therefore more elastic to wage differences across regions. In this limit, the competition for workers has its largest effect on marginal costs.

[34] More broadly, as in Courant and Deardorff (1992) an uneven distribution of labor can generate comparative advantage; the difference is that in our framework it comes from the assignment of a single factor across heterogeneous regions (rather than two factors across homogeneous regions).

4.4 A Sufficient Statistic

We now introduce a statistic that summarizes the impact of migrants on specialization. We return to our framework where workers choose both location and activity, but focusing on a small open economy.

Proposition 2. *Assume that Home is a small open economy, and that $\eta_k = \eta$, $\gamma_k = \gamma$ $\forall k$. The change in specialization when migration to region i becomes prohibitively costly, $\mu'_{ij,k} \to \infty$, $\forall j \neq i$ is given by:*

$$\frac{\widehat{X_{iF,k}/X_{iF,l}}}{X_{FF,k}/X_{FF,l}} = \left(\mathcal{E}_{ii,k}/\mathcal{E}_{ii,l}\right)^{-\frac{\alpha(1-\gamma)(1-\eta)}{1+\kappa+\alpha(1-\gamma)(\eta-1)}}, \tag{13}$$

where $\mathcal{E}_{ij,k}$ is the share of workers from i in j's effective labor force in activity k, i.e., $\mathcal{E}_{ij,k} = E_{ij,k}/E_{j,k}$, and $\hat{z}$ denotes proportional changes in variable z.

The proposition shows that the relative share of domestic workers in total effective labor supply is a sufficient statistic for the impact of limiting immigration into i. This result is analogous to the well-known result by Arkolakis et al. (2012) on the gains from trade. In this statistic, the share of workers actually employed in the region contains information of how relative marginal costs responds to limiting migration.

There are two reasons why exposure $\mathcal{E}_{ij,k}$ varies across crops and regions in our setting. The first is worker heterogeneity, $s_{i,k}$, which makes workers from i sort differentially across activities k within destination j. Note that this result takes knowledge $s_{i,k}$ as given and therefore holds regardless of whether knowledge is acquired via an externality or individual investments. The second reason is migration costs that are origin-destination-activity specific. Without these two forces, exposure shares are equalized within region j, and migration has no impact on intersectoral specialization.[35]

The results of this section provide a sharp characterization of the role of migration costs in shaping comparative advantage and trade patterns. But there are two caveats in their application. First, observed changes in migration costs are finite. Second, sectors have different trade elasticities, while land and intermediate-input intensities are additional drivers of comparative advantage. Bearing such caveats in mind, we deploy these insights throughout the rest of the paper, which quantitatively evaluates the impact of observed changes in migration costs.

[35] An extension of this proposition shows that introducing capital that moves freely across activities only requires amending the exponent in equation (13) (See Supplementary Appendix SE). If instead workers are individually equipped with activity-specific capital in a way that correlates with their ability, Proposition 1 applies with α being the cost share of the bundle of capital and labor.

5 Taking the Model to Brazilian Data

We now take our model to Brazilian data, to simulate the impact of the migration to the West. We map the model to Brazil's economy in 1950, 1980, and 2010, thus setting a time period to 30 years. We thus start early enough to observe Brazil's transformation due to the March, but within the constraints imposed by data availability.

5.1 The Quantitative Model

Relative to the model presented in the last section, we add three elements: (i) Two-tiered CES preferences, in which agents choose first how much to consume of services, agriculture, and manufacturing ($s = S, A, M$), with an elasticity of substitution σ, and then how much to consume of each activity within agriculture, with an elasticity of substitution σ_A; (ii) CES technology in which farmers combine efficiency labor and land, with an elasticity of substitution ρ and land intensity γ_k, and (iii) Land endogenously supplied by local governments that use a decreasing returns technology that requires final output—with a productivity b_j and an elasticity of land supply to land rents is ζ. The profits from land development are rebated to farmers proportionally to their wages.[36]

5.2 Calibration

We provide now an overview of how we calibrate (i) worker heterogeneity, (ii) trade costs, (iii) technology and preferences, and (iv) migration costs. Supplementary Appendix SH specifies our calibration procedure.

Worker heterogeneity. The preference dispersion parameter, κ, and the worker productivity parameter, β, map to the reduced-form elasticities we estimated in Fact 3. Using equation (6), our model relates the income of migrants to the employment in the origin as follows:

$$\log\left(\text{income}_{ij,kt}\right) = \iota^I_{j,kt} + \iota^I_{ij,t} + \beta \log L_{i,kt-1} + u^I_{ij,kt}. \qquad (14)$$

To motivate the error $u^I_{ij,kt}$, we posit measurement error in income. The fixed effects correspond to $\iota^I_{j,kt} = \kappa \log w_{j,kt}$, $\iota^I_{ij,t} = \log L_{i,t-1}$—if knowledge depends on relative employment shares — but also absorb systematic components of measurement error in income.

Analogously, we substitute equation (14) into (7) to obtain our regression for activity

[36] We abstract from other forces common in the spatial economics literature, such as amenities or productivity spillovers, to provide a clear evaluation of the new mechanisms in this paper.

choice:

$$\log L_{ij,kt} = \iota^L_{j,kt} + \iota^L_{ij,t} + \kappa\beta \log L_{i,kt-1} + u^L_{ij,kt}, \tag{15}$$

where the fixed effects correspond to $\iota^L_{ij,t} = \log L_{i,t-1} - \kappa\left(\log \mu_{ij,t} + \log \Xi_{it}\right)$, $\iota^L_{j,kt} = \kappa(\log w_{j,kt} - \log P_{j,t})$, and $u^L_{ij,kt}$ captures a unmeasured component of migration flows or of migration costs orthogonal to the other observables and fixed effects (e.g. as in Eaton and Kortum, 2002). Our model thus gives a structural interpretation to Fact 3, and show that the estimated coefficient α^I in regression (3) measures β, while coefficient α^W in regression (2) measures $\kappa\beta$.

The estimates for β range between 0.023 and 0.083, and the implied values for κ range between 1.58 and 3.26. Using quite different strategies, a few recent papers have estimated similar values of κ, which controls the elasticity of migration with respect to real wages. For example, using migration data for Brazil, Morten and Oliveira (2016) estimate a value of 1.9 and, using migration data from China, Tombe and Zhu (2019) find values between 1.2 and 1.6. Since β is new to our theory, there is no direct benchmark with which to compare it. But our results are comparable to those of sectoral scale economies (Antweiler and Trefler, 2002; Bartelme et al., 2019) and of the effect of city size on productivity (de la Roca and Puga, 2017). Henceforth, we set $\kappa = 2$ and $\beta = 0.04$. In robustness analysis, we also consider $\beta \in \{0, 0.083\}$.

Trade Costs. We need to calibrate trade costs between regions for each activity and period, which gives a total of $134 \times 134 \times 13 \times 3$ parameters. Given the trade data we have available, this requires us to parameterize trade costs. Akin to Ramondo et al. (2016), we assume for $i = j$, that $\tau_{ij,kt} = 1$ and, for $i \neq j$, that

$$\tau_{ij,kt} = \delta^0_t \left[(dist_{ij})^{\delta^1_t}\right]^{\iota^T_{ij}} \left[\delta_{kt} \left(dport_i\right)^{\delta^1_t} \left(dport_j\right)^{\delta^1_t}\right]^{1-\iota^T_{ij}}, \tag{16}$$

where ι^T_{ij} is a dummy variable that equals one if i and j belong to same country and zero otherwise, $dist_{ij}$ is the travel distance between i and j, and $dport_{ij}$ is the minimum travel distance to the nearest port (for $j = F$, we set $dport_j = 1$).

We calibrate δ^0_t to match the observed share of intra-regional trade in total domestic trade in Brazil. In particular, we construct $\sum_{s \in H} X_{ss}/X_{HH}$, where X_{ss} are sales of state s to itself, and X_{HH} are sales of Brazil to itself. We target a domestic trade ratios of 0.7 for 1950, 0.65 for 1980, and 0.60 for 2010, which gives δ^0_t of 0.08, 0.42 and 0.58. We choose δ^1_t so that the model matches the empirical elasticity of trade flows between states with respect to distance. For 2010, the OLS estimate of this elasticity is 1.05; for 1980, 1.25, and for 1950, 2.5, which translates into δ_t^1 of 0.22, 0.25, and 0.52 respectively. Lastly, we calibrate δ_{kt} to match Brazil's trade with the rest of the world.

Technology and Preferences. We set the share of value added α_k to 0.30 for manufacturing, 0.55 for agriculture and 0.6 for services according to the aggregate share of value added in the World Input-Output Database. For non-agricultural activities, we set $\gamma_k = 0$. For agriculture, we set $\rho = 0.5$, which is the mid-value between Costinot et al. (2016), who assume perfect complementarity between land and labor ($\rho = 0$), and a Cobb-Douglas production function ($\rho = 1$), as in Restuccia et al. (2008). Our main specification sets $\gamma_k = 0.21$ for all agricultural activities based on Dias Avila and Evenson (2010), thus allowing for Heckscher-Ohlin forces only between agriculture and manufacturing. Lastly, we set the land supply elasticity ζ to 1.5.[37]

As for preferences, we set σ to 0.4 based on Comin et al. (2015), σ_a to 2.5 according to Sotelo (2020), η_k in agriculture to 9.5, and η_k for other sectors to 5.5 following Caliendo and Parro (2015).

Having calibrated trade costs, technology and preferences, we follow the model-inversion logic laid out by Allen and Arkolakis (2014) and calibrate $A_{j,kt}$ to match observed gross output, $b_{j,t}$ to match observed agricultural land use, and $a_{H,kt}$ and $a_{F,kt}$ to match country-level apparent consumption by activity.

Migration Costs. With an eye toward counterfactual analyses, we extract a state-state component from migration costs, and break down the remainder into a state-state-activity component and geographic component that operates across meso-regions. Formally, we assume, for $i = j$, that $\mu_{ij,kt} = 1$ and, for $i \neq j$, that

$$\mu_{ij,kt} = \mu_t^0 \left[(dist_{ij})^{\mu^1} \right]^{\iota_{ij}^M} \left[\mu_{ss',t} \mu_{ss',kt} \left(dcap_i \right)^{\mu^1} \left(dcap_j \right)^{\mu^1} \right]^{1-\iota_{ij}^M} \tag{17}$$

where ι_{ij}^M is an indicator for whether i and j belong to the same state, s and s' denote states, $\mu_{ss',t}$ is a symmetric interstate migration cost (i.e., $\mu_{ss',t} = \mu_{s's,t}$), and $dcap_i$ is the travel distance between region i and the state capital of i. As such, workers have to pass through the capital to access meso-regions in other states.

To recover $\mu_{ss',t}$, based on our parameterization in equation (17), we use OLS to estimate the following equation:

$$\log \left(L_{ss',kt} \right) = \alpha_{s,t} + \alpha_{s',kt} + \tilde{\mu}_{ss',t} + \epsilon_{ss',kt}, \tag{18}$$

[37] There is some discussion in the literature about the right elasticity of land supply over long periods of time. We pick a number in the range between Costinot et al. (2016) and Gouel and Laborde (2018).

where $\alpha_{s,t}$ and $\alpha_{s',kt}$ are destination-activity and origin fixed effects and $\epsilon_{ss',kt}$ is an error term. We recover $\mu_{ss',t}$ from $\tilde{\mu}_{ss',t} \equiv -\kappa \log(\mu_{ss',t})$.[38] Given our estimates of $\mu_{ss',t}$, we adjust μ_t^0 in the model to match the share of workers living in their meso-region of birth and $\mu_{ss',kt}$ to match the migration of workers between states and activities. The k dimension in $\mu_{ss',kt}$ captures differences in labor allocations across activities driven by factors other than wages, such as misallocation. We set μ^1 to 0.05, which controls the elasticity of migration costs with respect to distance, based on Bryan and Morten (2019).[39]

5.3 The March to the West as seen Through our Model

Table 5 presents selected descriptive statistics of our calibration. Each panel highlights a different exogenous driver of migration in our model: migration costs, productivity, and trade costs.

Panel (a) shows that domestic migration costs declined sharply between 1950 and 1980, in line with what we expect, given Brazil's large-scale public investments in transportation infrastructure.[40] Migration costs from the East to the West declined by two-thirds. Appendix Figure G.6 correlates our estimates of the symmetric component of migration cost ($\mu_{ss',t}$), recovered from the fixed effects from equation (18), with distance between states. Reassuringly, migration costs are strongly correlated with distance, even though we do not use any measure of distance to estimate $\mu_{ss,t}$. Consistent with increased government investment to deepen ties between the East and the West, this correlation drops over time as shown in Panel (a). To benchmark our migration costs with previous literature, we note that our migration cost estimates are comparable to the ones obtained in Tombe and Zhu (2019), who find an overall migration cost between provinces in China of 25 circa 2000.[41]

Panel (b) shows the evolution of productivities, $A_{i,kt}$, in the West relative to the East. The West's agricultural productivity caught up and surpassed that of the East—a reason

[38] Supplementary Appendix SH.1 shows that the hub and spoke formulation allows us to aggregate migration flows and run gravity equations in a theoretically consistent manner. Also, Supplementary Appendix Figure S.7 reports that our estimates of $\mu_{ss',t}$ are strongly correlated with a Head and Ries index applied to aggregate migration flows between states. Note that we target state-state-activity flows in this calibration, instead of meso-regions. The reason is these data have a higher level of aggregation, which makes it less likely that we wrongly infer infinite migration costs from sampling zeros.

[39] Using the calibrated value for μ^1, our model-implied migration elasticity at the state level is comparable to the elasticity estimated in Morten and Oliveira (2016). Specifically, the model-implied elasticity of state to state migration flow across years with respect to distance here is -1.36 (1950-2010), whereas and in Morten and Oliveira (2016) is -1.32 (1940-2010).

[40] To compute migration cost in this section, we remove $\mu_{ss',kt}$ from $\mu_{ij,kt}$, since $\mu_{ss',kt}$ can be interpreted as sectoral wedges that are not related to geography. Results are qualitatively the same if we incorporate $\mu_{ss',kt}$ into $\mu_{ij,kt}$.

[41] We also regress employment in region j activity k (normalized by the level in each state), which is not targeted in our calibration procedure, against its model-implied counterpart and find a coefficient of 0.94 (and a R^2 of 0.51) in 2010.

Table 5: Description of the Parameters Recovered in the Calibration

	1950 (1)	1980 (2)	2010 (3)
a. Migration costs			
Average migration costs	33.62	20.49	18.38
Migration costs within states	2.21	1.91	1.78
Migration costs between states: East versus West	103.04	42.61	37.84
Elast. of migration costs w.r.t. travel distance	1.00	0.80	0.72
b. Productivity			
Productivity in manufacturing in the West relative to the East	0.48	0.75	0.82
Productivity in agriculture in the West relative to the East	0.95	1.25	1.56
- Soybeans	0.40	1.86	1.67
- Livestock	1.34	0.87	0.80
- Corn	1.01	0.97	2.04
Productivity of land supply in the West relative to the East	0.31	0.52	1.31
c. Trade costs			
Trade cost between Brazil and RoW - manufacturing	8.83	5.13	3.96
Trade cost between Brazil and RoW - agriculture	9.70	5.42	4.06
Elast. of trade cost w.r.t. travel distance	0.15	0.14	0.12

Notes: This table shows results from the calibration of the model. Migration costs are presented in terms of its harmonic average. Productivity is averaged using employment weights. International trade costs are averaged using trade-flow weights. The elasticity of migration cost with respect to travel distance is the slope of a regression of the log of estimated migration costs between states against the log of travel distance.

for the mismatch between labor allocation and productivity that we discussed in Section 4.3. For soybeans, specifically, the upward trend in relative productivity is in line with the research efforts from EMBRAPA to increase soybean productivity in the West in the 1970s. Panel (b) also indicates that the productivity of the land supply sector $b_{j,t}$ increased in the West relative to the East, consistent with the government's increasing efforts to facilitate land settlement and acquisition in the West.

Panel (c) shows that our calibration also captures Brazil's increasing trade openness through a reduction of international trade costs. Domestic trade costs also declined and became less sensitive to distance, in line with the transportation policies that fostered East-West trade integration.

Put together, these results paint a clear picture of the forces that produced the March. Starting in the 1950s, a series of shocks dramatically increased the West's agricultural productivity across the board, especially in new crops such as soybean and corn. In response to concomitant reductions in East-West migration costs, and taking advantage of these productivity shocks and of the West's relative abundance of land, Eastern migrants sorted throughout the West and fueled its expansion into domestic and foreign markets.

6 The Quantitative Impact of Migration on Trade

Motivated by the Brazilian government's domestic integration policies, we ask: How would Brazil's specialization have evolved if these policies, as they pertain to migration costs, had never been enacted? We examine a counterfactual economy in which migration costs between the East and the West remain at their 1950 levels, while all other exogenous shocks evolve as in the baseline economy. Specifically, we keep the state-to-state component we recover from our gravity regressions, $\mu_{ss',t}$, constant at its 1950 level, while letting the residual component $\mu_{ss',kt}$ evolve.[42]

Before discussing the effects of migration on trade, we note that the measured reductions in migration costs accounts for 56 percent of Westbound migration: In our counterfactual economy, the share of employment in the West rises from 6.9 to 10.6 percent (as opposed to 15 percent in the data). The rest is due to other factors, including productivity, land supply, and trade costs shocks.[43]

6.1 The Impact of Migration Costs on Specialization

Panel (a) in Figure 4 plots, on the vertical axis, counterfactual changes in specialization across regions in the West, relative to the rest of the world, using manufacturing as a reference (i.e., we use our RBE index).[44] The horizontal axis shows the baseline share of workers born in the West in total effective labor supply of each region and activity, relative to manufacturing. With Proposition 1 in mind, note that the smaller this baseline share, the more exposed is the region-activity pair to a reduction of immigration from the East, compared to manufacturing. We highlight three patterns from this figure. First, most region-activity pairs fall below zero on the vertical axis, meaning that reductions in migration costs shifted trade towards agriculture and away from manufacturing. Second, although soybean production is the most affected activity with specialization shifting often by more than 50 percent, the impact is also large for other, more traditional activities. Third, there is a strong relation between exposure to migration and changes in specialization—suggesting that the intuitions from Proposition 1 carry over to this more general counterfactual.

In panels (b) and (c), we aggregate these regional changes for the West and for Brazil. Panel (b) confirms that the decline of migration costs led to a marked expansion of the West's

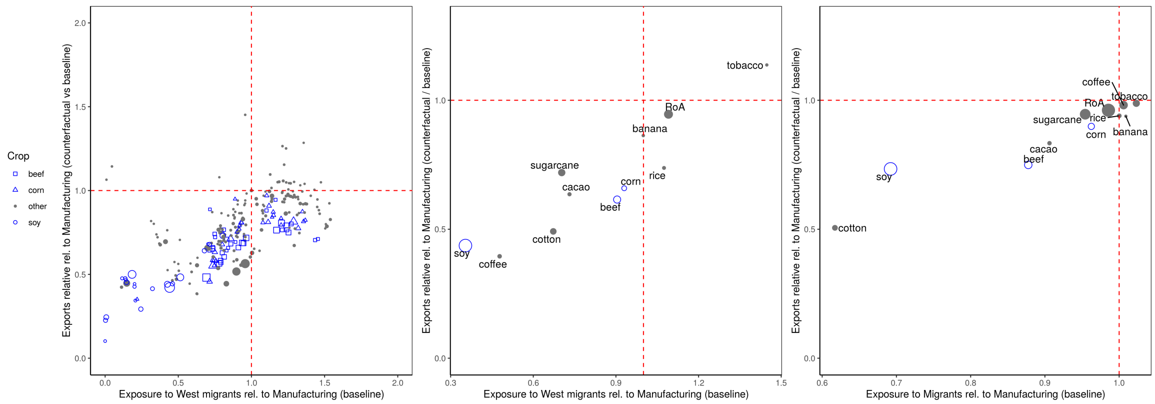
[42] The discussion that follows focuses on our cumulative results for 2010. Supplementary Appendix Figure S.3 displays results for 1980.

[43] In the absence of any exogenous shocks, the population in the West drops slightly from 6.9 to 6.7, which indicates that the observed migration is almost entirely driven by new shocks, rather than the 1950 allocation being way from the steady state.

[44] The exposures in the East to migrants from the West are much lower in general, as suggested by Appendix Figure G.5. This means that the vast majority of our aggregate results follow from the changes that occur in the West.

Figure 4: Counterfactual Change in Exports relative to Manufacturing (2010), for the West and Brazil as a whole

(a) Meso-regions - West (b) West (c) Brazil



Notes: Panel (a): Each observation is a region-activity pair in the West. The horizontal axis measures the fraction of effective employment comprised by workers born in the West itself. The vertical axis shows the counterfactual change in exports relative to manufacturing (as a percentage of the baseline exports). The size of the markers represents the magnitude of exports in the baseline as a fraction of Brazil's total in that activity. Panel (b): Each observation is an activity aggregate. The horizontal axis measures the fraction of effective employment in the West comprised by workers born in the West. The vertical axis measures the counterfactual change in West's exports relative to manufacturing (as a percentage the baseline value). The size of the markers represents the magnitude of exports in the baseline. Panel (c): Each observation is an activity aggregate. The horizontal axis measures the fraction of effective employment born in the same region where they work, weighted by baseline exports. The vertical axis shows the counterfactual change in Brazil's exports relative to manufacturing (as a percentage of the baseline value). The size of the markers represents the magnitude of exports in the baseline.

agriculture relative to manufacturing. It also shows that the agricultural activities that are more exposed to Eastern migrants are the ones that expanded the most due to migration (recall Fact 2). Counterfactual specialization relative to the baseline is 58 percent lower for soy and about 35 percent lower for beef and corn. In addition, among the activities that benefited most from migration in the West are some traditional crops, such as coffee and cacao. However, as Panel (c) shows, despite a large impact on the specialization patterns of the West itself, the expansion of traditional crops had little impact on the specialization of Brazil as a whole. The reason is that, although these traditional activities grew in the West due to migration, the West's share in the aggregate production of such activities is minimal.[45] In contrast, aggregate counterfactual drops in soy and cattle specialization are large (29 and 25 percent) and follow, to a large extent, the evolution of the West's agriculture.

To benchmark our results to data, Appendix Figure G.7 shows that reductions in migration costs account for up to 30 percent of the country-wide *observed* evolution of specialization

[45]Cotton responds strongly to migration costs shocks both in the West and in the aggregate. As Table 1 shows, however, it is relatively unimportant in the aggregate.

in soy, cattle, and corn between 1950 and 2010 (or 27 percent of the observed evolution of the export shares over the same period). For the West as a whole, migration accounts for almost twice as much.

6.2 Quantifying the Margins of Comparative Advantage

We now measure the contribution of worker heterogeneity and factor intensity to our results. To do so, we recalibrate the model under different specifications and compute the same counterfactual scenarios in which migration costs remain at their 1950 levels.

6.2.1 Worker Heterogeneity

Proposition 1 suggests that migration operates through variation in effective-labor exposure across regions and activities. With the exception of sector-specific migration costs, $\mu_{ss',kt}$, in our model the only force generating sorting is origin-specific worker heterogeneity, $s_{i,kt}$. To assess the importance of worker heterogeneity for the impact of migration, we recalibrate our model in steps.[46] First, we modify our base calibration to target origin-destination labor flows and, separately, employment by activity and region. This first step removes the sorting across activities that comes purely from $\mu_{ss',kt}$ (i.e., we replace $\mu_{ss',kt}$ with $\mu_{ss',t}$ and $\mu_{s',kt}$). In the second step, we recalibrate the model by additionally setting $\beta = 0$, which shuts down sorting completely (yielding exposure $\mathcal{E}_{ii,k}/\mathcal{E}_{ii,l} = 1$ for any k and l).

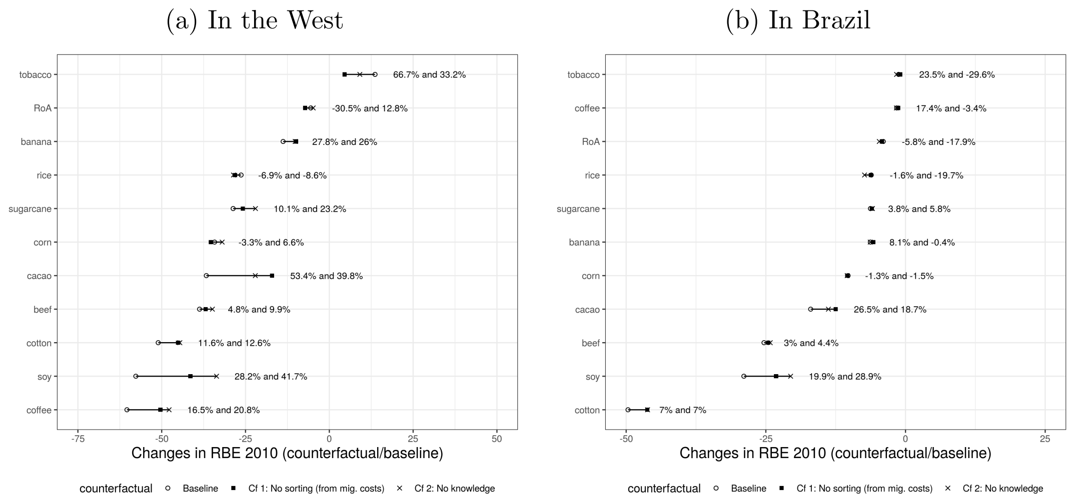
Starting with the West, Figure 5 (a) plots, for each activity, the counterfactual change in specialization in our baseline calibration and compares it to that in these two new calibrations. Completely eliminating sorting dramatically affects specialization across the board. Among crops in which the baseline impact of migration is large, we observe the largest effects on soybean specialization, which drops by about 42 percent relative to our baseline, but the effect is also sizable for cacao, coffee, and beef (40, 21, and 10 percent). Focusing again on soy, about one-third of the effect of sorting comes from worker heterogeneity, as opposed to migration costs. We conclude that knowledge, and sorting more generally, was critical in the differential expansion of agricultural activities in the West. Turning to Brazil as a whole, the role of knowledge is most quantitatively important for soybean: absent sorting, the impact of migration on specialization would be about a 29 percent smaller and, again, about one-third of this reduction is due to knowledge.[47]

In Appendix E, we study how worker heterogeneity shapes specialization in the steady

[46] An indication of worker heterogeneity is that, in our baseline calibration, a regression of effective labor flows across origins and destination-activity pairs, $L_{ij,k}$, against skill heterogeneity, $s_{i,k}$, accounts for 55 percent of the variation in 2010 and 32 percent in 1980.

[47] Supplementary Appendix Figure S.6 shows that setting $\beta = 0.083$ retains the total role of sorting unchanged, but increases the importance of knowledge from one third to one half.

Figure 5: Measuring the Contribution of Knowledge (2010)



Notes. Each row is an activity aggregate. The hollow circle presents the counterfactual change in export specialization in our baseline calibration, which targets state-state-employment flows. The square presents the counterfactual change in specialization in a calibration that targets state-state migration and state-activity employment, separately. The cross corresponds to a calibration in which, additionally, $\beta = 0$. For each activity we present the difference in the impact of migration relative to the baseline, as a percentage of the latter. Panel (a) presents results for the West; Panel (b), for the Brazil as a whole.

state. We find that Brazil's agricultural transformation continues over time, as soybean exports continue to grow, while other crops shrink. The reason is that knowledge rises disproportionately in regions with high employment in soy.

6.2.2 Factor Proportions

In our baseline calibration, agriculture is the only activity that uses land in production, so Heckscher-Ohlin forces operate only between manufacturing and agriculture. Appendix Figure G.8 compares our baseline results to those in a model that recognizes that land intensity varies within agriculture between cattle and crops (0.33 vs 0.17 value-added share). Doing so mitigates the impact of migration on cattle specialization—almost 1.3 p.p. in the aggregate, or 5.3 percent lower. For most other agricultural activities, the impacts are modest. The reason is that, in this new calibration, land commands a larger value-added share in cattle, compared to other agricultural activities (0.33 vs 0.17). Hence the direct impact of a labor shortfall on marginal costs is lower for cattle than for crops. Because land supply is not perfectly elastic, in equilibrium the price of land declines, which also favors land-intensive activities the most.

7 Migration, Comparative Advantage, and the GFT

To complete our evaluation of the impact of migration on international trade, we assess how it affects the gains from trade (hereafter, GFT). In our model, a measure of welfare is the expected utility attained by a person born in region i, Ξ_i. Denoting by $\hat{\Xi}_i$ the change in welfare from going to autarky, the GFT are $1 - \hat{\Xi}_i$.[48]

We first show analytically that, in a one-activity model, regional terms of trade and migration opportunities generate new sources of gains from trade. We then show that migration and comparative advantage are key drivers of the GFT in our multi-sector environment.

7.1 Gains from Trade with One Activity

We start with a result that, following directly from the definitions of Ξ_i and of trade shares π_{ii}, highlights that changes in expected utility depend on changes in real wages in all regions to which workers can migrate; these changes, in turn, can be computed using observed regional trade shares.

Proposition 3. *Using observed trade shares, one can compute the losses from full trade autarky, i.e.* $\tau_{ij} \to \infty, \forall i \neq j$, *as*

$$\hat{\Xi}_i^{B \to B,A} = \left(\sum_j \lambda_{ij} \pi_{ii}^{\frac{\kappa}{\alpha(\eta-1)}} \right)^{1/\kappa}, \tag{19}$$

where λ_{ij} *are observed migration shares and* π_{ii} *are observed domestic expenditure shares, and "B" denotes baseline and "B, A" denotes trade autarky starting from B.*

Absent migration, i.e. when $\lambda_{ii} = 1$ and $\lambda_{ij} = 0$ for any $j \neq i$, equation (19) collapses to the canonical formula for the GFT (see Arkolakis et al., 2012). The next proposition relates the GFT in the baseline to those in a situation where migration across regions is not allowed.[49]

Proposition 4. *For region i, the autarky losses in the baseline economy* $\hat{\Xi}_i^{B \to B,A}$ *and the*

[48]To obtain the analytic result below, we return to a formulation in which land is not a productive factor. We focus again on 2010, but Supplementary Appendix Figure S.4 shows that our conclusions hold also for 1980. Our results complement the approach in Galle et al. (2017) by bringing in geographic mobility frictions, which do not play a prominent role in their analysis.

[49]Bonadio (2020) studies how international migration impacts market access for receiving regions. We focus on how migration interacts with the gains from trade.

no-migration economy $\hat{\Xi}_i^{N\to N,A}$ *are related by the following equation*

$$\left(\hat{\Xi}_i^{B\to B,A}\right)^{\kappa} = \underbrace{\lambda_{ii} T_i^{-\kappa} \left(\hat{\Xi}_i^{N\to N,A}\right)^{\kappa}}_{\text{domestic contribution}} + \underbrace{\sum_{j\neq i} \lambda_{ij} \left(\pi_{jj}^{B}\right)^{\frac{\kappa}{\alpha(\eta-1)}}}_{\text{migration opportunities contribution}}, \tag{20}$$

where $T_i = \left(\pi_{ii}^B / \pi_{ii}^N\right)^{1/\alpha(1-\sigma)}$.

Note first that two components contribute to the baseline losses from full autarky.[50] The first component is the losses from autarky that would occur without migration, $\hat{\Xi}_i^{N\to N,A}$, whose weight is given by the fraction of workers who stay in i, λ_{ii}. The coefficient T_i corrects for the fact that migration, by itself, worsens the terms of trade for regions that receive workers. The second component measures the contribution of migration opportunities: Additional welfare losses also arise from migration destinations in which real wages drop when there is no trade. Note that if region i is a large receiver of migrants — and hence T_i is small — the losses from autarky tend to be smaller in the baseline economy than in the economy without migration, and so migration attenuates the losses from autarky (i.e., migration reduces the gains from trade).

7.2 Gains from Trade with Multiple Activities

The forces we have uncovered in Section 7.1—which state how migration opportunities shape the full GFT with one activity—carry over to a model with multiple activities. Namely, (i) changes in real wages in other regions contribute to the GFT, and (ii) migration by itself induces changes in local real wages via changes in the terms of trade.[51] We now examine their quantitative importance in our baseline model.

But with multiple activities, these migration-related forces also interact with comparative advantage. For one thing, larger proportions of workers sort into region-activity combinations with high efficiency relative to the rest of the world, governing the initial $\lambda_{ij,k}$ shares. For another, comparative advantage activities tend to experience larger reductions in real wages from going to trade autarky.

Panel (a) in Appendix Figure G.9 shows that the full GFT are large (26 percent on average) but vary substantially across regions. Using equation (20), we can compute the share of the full GFT accounted for exclusively by migration opportunities (i.e., the welfare loss from

[50] We continue to write changes in real wages as a function of domestic trade shares π_{ii} although the first term is not directly observable, to both keep the symmetry with equation (19) and emphasize that some of these components are directly observable.

[51] Dix-Carneiro and Kovak (2017) show that exposure to a trade liberalization shock has protracted, negative effects across Brazilian labor markets. Our results show that negative outcomes in a local labor market also impact welfare for workers who could potentially migrate to that labor market.

real wage losses in migration destinations).[52] The contribution of migration opportunities ranges from minimal to almost the full losses from trade, with an average of 61 percent. As shown in panel (b), the migration-opportunity contribution is correlated with the fraction of people who leave when migration is available, $1 - \sum_k \lambda_{ii,k}$, as one expects from Equation (20). But the correlation is far from perfect, which highlights that comparative advantage and regional heterogeneity also shape the importance the migration-opportunity component. Since the West tends to receive workers from the East, the migration-opportunity contribution is smaller on average for the former than the latter (49 percent and 62 percent; see Panel (c)). This share is particularly low in large urban centers in the East and expanding agricultural regions in the West.

To assess the impact of comparative advantage on the full GFT, we compute the ratio of the GFT in a model with one activity relative to a model with multiple ones.[53] As panels (d) and (e) in Appendix Figure G.9 show, the contribution of multiplicity of activities to the GFT is strongly associated with comparative advantage in agriculture relative to the rest of the world. For many regions in the West, including multiple activities almost doubles the GFT. Since comparative advantage is inherently a local characteristic of region i,it operates chiefly by strengthening the response of $i's$ wages to trade and less so by shifting migration opportunities (i.e., it tends to shift the "domestic" component in equation (20)).

Lastly, we study in Panel (f) how migration and comparative advantage interact to determine the GFT. On the y-axis, we measure how the contribution of migration opportunities changes, when going from a single-activity model to one with multiple activities. On the x-axis we measure again the GFT with many activities relative to the GFT with one activity. The figure shows a clear negative relation, with a slope of -0.51, which means that the larger the role of comparative advantage in the GFT, the smaller the share of migration opportunities in the GFT (relative to a single-activity model). Hence, although migration opportunities create further gains from trade in a many-activity model (relative to a single-activity one), these gains do not rise as fast as the additional gains coming from domestic markets.

Appendix F details how the gains from *international* trade were affected by the March

[52] That is, we set $T_i^{-\kappa}\left(\hat{\Xi}_i^{N\to N,A}\right)^{\kappa}$ equal to 1 in Equation (20) and divide by the full GFT. Supplementary Appendix SF shows the equivalent of expression (20) in a model with many activities, which we use to make these calculations. Although a decomposition such as (20)—which exactly links welfare losses in the two scenarios—is not available in a model with multiple activities, one can still easily separate the contribution of migration opportunities to the total gains from trade, as we do in this exercise. The correction term T adds little quantitatively to our results. The average $T_i^{-\kappa}$ across meso-regions is approximately 1.01 with a standard deviation of 0.06.

[53] We calculate the one-sector losses using equation (19). Levchenko and Zhang (2016) perform a similar comparison underscoring the quantitative importance of comparative advantage.

to the West. Migration cost reductions leave aggregate international gains from trade unaffected. But the effect differs across regions. For example, since migrants to the West sort disproportionately into agriculture, migration reinforces that region's comparative advantage relative to Foreign, exacerbating the losses from autarky.

8 Conclusions

Domestic labor allocation shapes regional and aggregate comparative advantage. In the Brazilian experience, the decline in East-West migration costs that began in 1950 encouraged the development of new industries, such as soy, cattle, and corn, ultimately helping to transform Brazil's agriculture. Key to these new developments was that workers from the East, where these commodities were already being produced, took advantage of rapid productivity growth in new sectors in the West through migration.

Previous research—including ours—often takes comparative advantage as an unchanging feature of the world. This paper shows that comparative advantage evolves to reflect the way migration interacts with, and sometimes amplifies, natural advantage. We have also shown how policies to improve spatial labor mobility complement other policies whose impact is localized, such as boosting regional productivity, or country-wide, such as tariffs, to determine their ultimate impact.

References

Adão, R. (2015). Worker heterogeneity, wage inequality, and international trade: Theory and evidence from brazil. *Unpublished paper, MIT.*

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

Allen, T. and Atkin, D. (2016). Volatility and the gains from trade. Working Paper 22276, National Bureau of Economic Research.

Allen, T. and Donaldson, D. (2020). Persistence and path dependence in the spatial economy.

Alvarez, F. and Lucas, R. J. (2007). General equilibrium analysis of the eaton-kortum model of international trade. *Journal of Monetary Economics*, 54(6):1726–1768.

Amann, E., Azzoni, C., and Baer, W. (2018). *The Oxford handbook of the Brazilian economy.* Oxford University Press.

Antweiler, W. and Trefler, D. (2002). Increasing Returns and All That: A View from Trade. *American Economic Review*, 92(1):93–119.

Arkolakis, C., Costinot, A., and Rodriguez-Clare, A. (2012). New trade models, same old gains? *American Economic Review*, 102(1):94–130.

Arkolakis, C., Lee, S. K., and Peters, M. (2018). European immigrants and the united states' rise to thetechnological frontier in the 19th century.

Artuç, E., Chaudhuri, S., and McLaren, J. (2010). Trade shocks and labor adjustment: A structural empirical approach. *American Economic Review*, 100(3):1008–45.

Baer, W. (2001). *Brazilian Economy, The: Growth and Development: Growth and Development*. ABC-CLIO.

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

Bartelme, D. G., Costinot, A., Donaldson, D., and Rodríguez-Clare, A. (2019). The textbook case for industrial policy: Theory meets data. Technical report, National Bureau of Economic Research.

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

Bird, J. and Straub, S. (2020). The brasília experiment: The heterogeneous impact of road access on spatial development in brazil. *World Development*, 127:104739.

Bombardini, M., Gallipoli, G., and Pupato, G. (2012). Skill Dispersion and Trade Flows. *American Economic Review*, 102(5):2327–2348.

Bonadio, B. (2020). Migrants, trade and market access.

Browder, J. O., Pedlowski, M. A., Walker, R., Wynne, R. H., Summers, P. M., Abad, A., Becerra-Cordoba, N., and Mil-Homens, J. (2008). Revisiting theories of frontier expansion in the brazilian amazon: a survey of the colonist farming population in rondonia's post-frontier, 1992–2002. *World Development*, 36(8):1469–1492.

Bryan, G. and Morten, M. (2019). The aggregate productivity effects of internal migration: Evidence from indonesia. *Journal of Political Economy*, 127(5):2229–2268.

Buera, F. J. and Oberfield, E. (2016). The global diffusion of ideas. Working Paper 21844, National Bureau of Economic Research.

Burchardi, K. B., Chaney, T., and Hassan, T. A. (2019). Migrants, Ancestors, and Foreign Investments. *Review of Economic Studies*, 86(4):1448–1486.

Bustos, P., Caprettini, B., and Ponticelli, J. (2016). Agricultural productivity and structural transformation: Evidence from brazil. *American Economic Review*, 106(6):1320–65.

Bustos, P., Castro Vincenzi, J. M., Monras, J., and Ponticelli, J. (2020). Industrialization without innovation.

Cai, J., Li, N., and Santacreu, A. M. (2019). Knowledge diffusion, trade, and innovation across countries and sectors.

Caliendo, L., Dvorkin, M., and Parro, F. (2019). Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Trade Shock. *Econometrica*, 87(3):741–835.

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

Cardoso, M. and Ramanarayanan, A. (2019). Immigrants and exports: Firm-level evidence from canada.

Chor, D. (2010). Unpacking sources of comparative advantage: A quantitative approach. *Journal of International Economics*, 82(2):152–167.

Comin, D. A., Lashkari, D., and Mestieri, M. (2015). Structural change with long-run income and price effects. Technical report, National Bureau of Economic Research.

Cosar, A. K. and Fajgelbaum, P. D. (2016). Internal Geography, International Trade, and Regional Specialization. *American Economic Journal: Microeconomics*, 8(1):24–56.

Costinot, A. and Donaldson, D. (2014). How large are the gains from economic integration? theory and evidence from u.s. agriculture, 1880-1997.

Costinot, A., Donaldson, D., and Komunjer, I. (2012). What goods do countries trade? a quantitative exploration of ricardo's ideas. *Review of Economic Studies*, 79(2):581–608.

Costinot, A., Donaldson, D., and Smith, C. (2016). Evolving comparative advantage and the impact of climate change in agricultural markets: Evidence from 1.7 million fields around the world. *Journal of Political Economy*, 124:205–248.

Costinot, A. and Rodríguez-Clare, A. (2015). Chapter 4 - trade theory with numbers: Quantifying the consequences of globalization. In Elhanan Helpman, K. R. and Gopinath, G., editors, *Handbook of International Economics*, volume 4 of *Handbook of International Economics*, pages 197 – 261. Elsevier.

Courant, P. N. and Deardorff, A. V. (1992). International Trade with Lumpy Countries. *Journal of Political Economy*, 100(1):198–210.

Crosby, A. W. (1973). *The Columbian Exchange. Biological and Cultural Consequences of 1492*. Number 2 in Contributions in American Studies. Greenwood.

Dahl, G. B. (2002). Mobility and the return to education: Testing a roy model with multiple markets. *Econometrica*, 70(6):2367–2420.

de la Roca, J. and Puga, D. (2017). Learning by working in big cities. *The Review of Economic Studies*, 84(1):106–142.

Deardorff, A. V. (2005). How robust is comparative advantage? *Review of International Economics*, 13(5):1004–1016.

Di Giovanni, J., Levchenko, A. A., and Ortega, F. (2015). A global view of cross-border migration. *Journal of the European Economic Association*, 13(1):168–202.

Dias Avila, A. F. and Evenson, R. E. (2010). *Total Factor Productivity Growth in Agriculture: The Role of Technological Capital*, volume 4 of *Handbook of Agricultural Economics*, chapter 72, pages 3769–3822. Elsevier.

Dix-Carneiro, R. and Kovak, B. K. (2017). Trade liberalization and regional dynamics. *American Economic Review*, 107(10):2908–46.

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

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

Fajgelbaum, P. and Redding, S. (2014). External integration and internal development: Evidence from argentina, 1870-1914. NBER Working Papers 20217, National Bureau of Economic Research, Inc.

Farrokhi, F. and Pellegrina, H. S. (2020). Global trade and margins of productivity in agriculture. *Working Paper.*

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

Fujiwara, T., Morales, E., and Porcher, C. (2020). A revealed-preference approach to measuring information frictions in migration.

Galle, S., Rodríguez-Clare, A., and Yi, M. (2017). Slicing the pie: Quantifying the aggregate and distributional effects of trade. Working Paper 23737, National Bureau of Economic Research.

Gouel, C. and Laborde, D. (2018). The crucial role of international trade in adaptation to climate change. Working Paper 25221, National Bureau of Economic Research.

Grossman, G. M. and Maggi, G. (2000). Diversity and trade. *American Economic Review*, 90(5):1255–1275.

Hanson, G. H., Lind, N., and Muendler, M.-A. (2015). The dynamics of comparative advantage. Working Paper 21753, National Bureau of Economic Research.

IIASA/FAO (2012). *Global Agro-ecological Zones (GAEZ v3.0).* IIASA, Laxenburg, Austria and FAO, Rome, Italy.

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

Klein, P. and Ventura, G. (2009). Productivity differences and the dynamic effects of labor movements. *Journal of Monetary Economics*, 56(8):1059–1073.

Kucheryavyy, K., Lyn, G., and Rodríguez-Clare, A. (2016). Grounded by gravity: A well-behaved trade model with industry-level economies of scale. Working Paper 22484, National Bureau of Economic Research.

Levchenko, A. A. (2007). Institutional Quality and International Trade. *Review of Economic Studies*, 74(3):791–819.

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

Lind, N. and Ramondo, N. (2018). Innovation, knowledge diffusion, and globalization. Working Paper 25071, National Bureau of Economic Research.

Manova, K. (2013). Credit Constraints, Heterogeneous Firms, and International Trade. *Review of Economic Studies*, 80(2):711–744.

Marcolan, A. L. and Espindula, M. C. (2015). *Café na Amazônia.* Brasília, DF: Embrapa, 2015.

Morrow, P. M. (2010). Ricardian-Heckscher-Ohlin comparative advantage: Theory and evidence. *Journal of International Economics*, 82(2):137–151.

Morten, M. and Oliveira, J. (2016). The effects of roads on trade and migration: Evidence from a planned capital city. *NBER Working Paper*, 22158.

Nagy, D. (2020). Hinterlands, city formation and growth: Evidence from the u.s. westward expansion.

Nehring, R. (2016). Yield of dreams: Marching west and the politics of scientific knowledge in the brazilian agricultural research corporation (embrapa). *Geoforum*, 77:206–217.

Nunn, N. (2007). Relationship-Specificity, Incomplete Contracts, and the Pattern of Trade. *The Quarterly Journal of Economics*, 122(2):569–600.

Ohnsorge, F. and Trefler, D. (2007). Sorting It Out: International Trade with Heterogeneous Workers. *Journal of Political Economy*, 115(5):868–892.

Olmstead, A. L. and Rhode, P. W. (2008). Creating abundance. *Cambridge Books*.

Pellegrina, H. S. (2020). Trade, productivity, and the spatial organization of agriculture: Evidence from brazil. *Working paper*.

Porcher, C. (2020). Migration with costly information.

Porteous, O. (2019). High trade costs and their consequences: An estimated dynamic model of african agricultural storage and trade. *American Economic Journal: Applied Economics*, 11(4):327–66.

Porteous, O. (2020). Trade and agricultural technology adoption: Evidence from africa. *Journal of Development Economics*.

Ramondo, N., Rodríguez-Clare, A., and Saborío-Rodríguez, M. (2016). Trade, domestic frictions, and scale effects. *American Economic Review*, 106(10):3159–84.

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

Redding, S. J. and Rossi-Hansberg, E. (2017). Quantitative spatial economics. *Annual Review of Economics*, 9(1):21–58.

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

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

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

Silva, J. S. and Tenreyro, S. (2006). The log of gravity. *The Review of Economics and statistics*, 88(4):641–658.

Sotelo, S. (2020). Domestic trade frictions and agriculture. *Forthcoming Journal of Political Economy.*

Tombe, T. (2015). The missing food problem. *American Economic Journal: Macroeconomics*, 7(3):1–33.

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

Vargas, G. (1938). *A nova politica do Brasil - Vol V.*

Vasconcelos, J. R. d. (2001). Matriz do fluxo de comércio interestadual de bens e serviços no brasil-1999. Technical report.

Villas Bôas, O. and Villas Bôas, C. (1994). A marcha para o oeste. *A epopéia da Expedição Roncador-Xingu. São Paulo: Editora Globo.*

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

Online Appendix

A Data

This section describes how we construct our data set. Appendix Tables A.1 and A.2 present summary statistics of the key aggregate variables.

Employment and Migration. Our data on migration and employment comes from decadal demographic and economic censuses administered by the Brazilian statistical institute IBGE (*Instituto Brasileiro de Geografia e Estatistica*). For 1970, 1980, 1991, 2000 and 2010, we have micro-level data on migration and employment. For 1950 and 1960, we digitized state-level aggregates from historical census publications. The information in the micro-level data is divided in two questionnaires: one applied to the universe of the population, which asks basic questions about education and family structure, and one applied to a sample of approximately 25 percent of households, which asks detailed information on migration, employment and income.[54] We observe both total income, which includes transfers from the government, and income from a worker's main activity, which includes any source of income, in addition to wages. In our analysis, we use the latter.

For the census of 1980 onward, we observe the current and the previous municipality of residence of each individual, if they have migrated within the previous 10 years. We use this variable to define migration in the reduced form elasticities that we presented in Section 3, but as we show later in this appendix, our results are robust to alternative measures of migration.[55] Since less than 0.1 percent of Brazil's population was born abroad, we remove international migrants from our sample. For 1970, we have micro-data with information on the state of birth and the state of residence. For 1950 and 1960, we use information on the total population in each state who were born in each state of Brazil, and also information on total employment per economic activity. In our structural calibration, we use the state of residence and the state of birth to measure the flow of workers between states.

[54] In 1970 and 1980, 25 percent of the population was sampled for the detailed questionnaire. For 1990, 2000 and 2010, about 25 percent of the population was sampled in smaller municipalities and 10 percent in larger ones. In 2010, the census about 12 million individuals received the detailed questionnaire.

[55] The exception is the census of 2000, which asks individuals their previous state of residence, their municipality of residence in 1995, but not their previous municipality of residence.

Table A.1: Aggregate Summary Statistics

	1950 (1)	1960 (2)	1970 (3)	1980 (4)	1990 (5)	2000 (6)	2010 (7)
a. Migration							
- East to East	0.874	0.845	0.825	0.759	0.735	0.719	0.687
- East to West	0.085	0.105	0.137	0.191	0.205	0.205	0.218
- East to West + West to East	0.107	0.129	0.152	0.212	0.229	0.235	0.255
- West to East	0.022	0.023	0.015	0.021	0.024	0.030	0.038
- West to West	0.019	0.026	0.024	0.029	0.036	0.046	0.058
b. Economic Aggregates							
- Brazil's GDP	1.000	1.000	1.000	1.000	1.000	1.000	1.000
- Exports	0.071	0.109	0.062	0.080	0.075	0.086	0.117
- Imports	0.071	0.117	0.071	0.093	0.047	0.088	0.125
- World's GDP	99.415	94.235	68.613	43.943	50.187	51.973	31.772

Notes: Panel (a) shows different the share of different categories of migrants. We define migration based on a workers state of living and state of birth. It shows that East to West migration accounted for 8 percent of all interstate migrants in 1950. Panel (b) presents values normalized by Brazil's GDP in a given year.

For 1980, 1990, 2000 and 2010, we can use our micro-data to construct the migration flows of workers who were born in state s and live in state s' and work in activity k, which we denote by $L_{ss',kt}$. For 1950, we observe only the migration flows from a state s to a state s', which is given by $L_{ss',t}$. Given that we do not directly observe $L_{ss',kt}$ in 1950, we therefore use entropy methods widely applied in the construction of input-output matrices to obtain $L_{ss',kt}$. Specifically, we apply the algorithm developed in Ireland and Kullback (1968). In our case, the algorithm consists in searching for values of $L_{ss',kt}$ based on a guessed value $\tilde{L}_{ss',kt}$ until $L_{ss',kt}$ are consistent with aggregate data on $L_{s',kt}$ and $L_{s,t}$. We use data on $L_{ss',k1980}$ (adjusting the overall population size to the one in 1950) as guesses of $\tilde{L}_{ss',k1950}$ to construct the values of $L_{ss',k1950}$.[56]

Gross Output. To construct gross output and value added per meso-region and activity, we apply a procedure that ensures that our aggregates are consistent with the ones measured in dollars by FAO and United Nations. We compute shares of value added by meso-region and activity and multiply such values by aggregate values from FAO and UN. When necessary, we apply the algorithm developed in Ireland and Kullback (1968) to construct more disaggregated values.

Our data on agricultural revenues come from PAM (*Produção Agrícola Municipal*), which is maintained by the Brazilian census bureau. PAM provides municipality level data since 1974 for more than 20 crops and state level data since 1930s for a subset of crops. For cattle, we use data from the agricultural census. We converted the revenues measured in

[56] In 1950 we do not observe labor employment in soybeans. We therefore complemented our numbers with special reports from EMBRAPA on historical production of soybeans.

these data sets into value added and computed the share of value added coming from each agricultural activity for each meso-region. We then multiplied these shares of value added by agricultural activity—given at the meso-region level—by the share of value added in agriculture coming from each meso-region relative to the total value added in agriculture in Brazil, which is measured by IPEA. Lastly, we multiply the share of value added by meso-region and agricultural activity by the aggregate value added in agriculture measured in dollars from the UN. For the manufacturing and service sectors, we use the share of value added for each meso-region measured by IPEA, we then multiply such shares by the aggregate value added measured by UN. Lastly, with data on value added, we computed gross output using the share of value added in the World Input-Output Database.

For 1950, we do not observe value added by economic activity at the meso-region level, but only aggregates at the state-level. As with migration, we use the entropy method developed in Ireland and Kullback (1968) to adjust our values. We search for values of value added $VA_{j,kt}$ based on guesses of value added $\tilde{VA}_{j,kt}$ until they are consistent with observed value added by state and activity $VA_{s,kt}$ and value added by meso-region $VA_{j,t}$. We use the values from 1980 as guesses of $\tilde{VA}_{j,kt}$ for 1950.

Trade Flows. The data on trade flows by agricultural activity come from FAO. The data is disaggregated by good, according to the Harmonized System at the 6 digit level. We classified the trade flows according to the agricultural activities included in our analysis. 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. Since the data from FAO starts in 1960, we extrapolate exports and imports back using data on aggregate exports and imports from IPEA.

For 2010, we use export data by state from 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 registered according to the location of production. For domestic trade flows, we digitized data on trade flows between states from the annual statistical yearbook reports from the Brazilian government of 1947, 1948, 1949, 1972, 1973 and 1974. For 1999, we use estimates of trade flows between states from Vasconcelos (2001) based on state merchandise and services taxes.

Table A.2: Summary Statistics by Activity (in percentages)

	services	mfg	agriculture	rest of agri	banana	cocoa	coffee	corn	cotton	livestock	rice	soy	sugarcane	tobacco
				Percentage within Agriculture										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
a. Value Added														
-1950	53.5	21.6	24.8	27.7	1.3	1.3	16.0	7.3	6.9	27.3	6.8	0.2	4.1	1.0
-1980	56.5	33.8	9.7	24.2	1.6	2.3	7.1	9.9	2.8	27.3	7.1	9.2	7.2	1.2
-2010	74.0	21.0	5.0	22.8	2.4	0.6	6.5	7.3	1.7	25.9	3.8	15.7	10.5	2.7
b. Workers														
-1950	65.3	7.5	27.2	31.5	0.8	0.8	16.2	10.0	3.7	27.0	5.2	0.1	3.7	0.8
-1980	68.8	14.0	17.3	20.6	0.7	1.9	8.0	14.0	6.1	27.0	11.3	3.9	4.7	1.8
-2010	84.5	7.5	8.0	27.8	1.3	1.2	12.3	7.6	0.1	27.7	2.9	3.0	13.0	3.1
c. Exports from Brazil to the ROW														
-1950	0.0	20.2	79.8	13.7	0.5	9.0	59.2	1.3	6.2	2.2	0.4	0.0	5.7	1.8
-1980	0.0	55.6	44.4	28.2	0.2	8.6	23.9	0.5	0.5	3.5	0.2	21.4	9.0	4.1
-2010	0.0	68.0	32.0	28.4	0.1	0.4	8.0	3.9	1.8	6.9	0.6	29.4	16.1	4.4
d. Imports of Brazil from the ROW														
-1950	0.0	82.9	17.1	99.4	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0
-1980	0.0	90.1	9.9	73.7	0.0	0.0	1.9	8.0	0.1	4.2	4.5	7.6	0.0	0.1
-2010	0.0	95.4	4.6	87.9	0.0	2.2	0.3	1.6	1.2	2.1	3.3	0.8	0.0	0.6

Notes: This table shows the distribution of economic activity that we match in our calibrated model. Data for workers per activity comes from Brazilian censuses. Data for value added per sector comes from United Nations since the 1970s and extrapolated back to 1950 based on data from IPEA-DATA. Data on trade comes from COMTRADE and IPEA-DATA.

Labor and Land in the Rest of the World. We normalize Brazil's population to 1 in each period, and to avoid dealing with population growth we set Foreign's population at 37, it's magnitude relative to Brazil in 1980. To preserve Heckscher-Ohlin forces as they occur in the data, we impute the rest of the world's land endowment each year (1950, 1980, 2010) so that the relative land-labor ratio between Brazil and the rest of the world, i.e., $(H_{BR,t}/L_{BR,t})\,/\,(H_{F,t}/L_{F,t})$ is reproduced by the model in each year. Within the rest of the world, we allocate labor proportionally to value added across sectors.

B Anecdotal Evidence on Migrant Heterogeneity

In Section (3), we present anecdotal evidence on how migrants' knowledge aided in the expansion of soybeans in the West. Additional anecdotal evidence points to this mechanism also being present in the westward expansion of other crops, such as coffee. For example: "The new amazonian experience with the "black gold" is the result of the entrepreneurship of migrants coming from Paraná, Minas Gerais and Espírito Santo [...] Farmers from Paraná and Minas Gerais brought arabica coffee to the region and farmers from Espírito Santo brought canephora coffee (i.e., robusta), which they cultivated in their region of origin." Marcolan and Espindula (2015), p. 13.

A cursory inspection of our migration data suggests that migration patterns are consistent with this idea. For instance, in the municipality of Sorriso, the largest producer of soybean in the West today, 26 percent of the labor force employed in soy comes from from Rio Grande do Sul, the state with the highest soybean output per farmer in the East. The data also bears this out for coffee, as Marcolan and Espindula (2015) suggest. In the State of Rondônia, the West's leading producer of robusta, this higher-quality bean accounts for 20 percent of production. The main origin of immigrant farmers in Rondônia is the State of Espírito Santo, where robusta accounts for 30 percent of coffee output. A much smaller share of immigrants in Rondônia comes from other Eastern states specialized in arabica beans. We find similar patterns for cacao and sugarcane. In Fact 3 in Section 3 in the main body of the paper, we examine this relation rigorously.

C Additional Empirical Evidence on Worker Heterogeneity

This section discusses in more detail the robustness checks related to Facts 3 and 4 that we present in the main body of the paper and discusses the econometric implications of an alternative model that emphasizes the role of workers sorting.

C.1 Robustness checks for Fact 3

C.1.1 Alternative specifications and interpretations

Appendix Table C.3 reports our main results using alternative sample specifications. As indicated in the table, we experiment with different time lags of $L_{ij,kt-1}$ (20 and 30 years), different official levels of disaggregation of labor markets defined by the Brazilian statistical bureau (meso-region and micro-region), and with the inclusion of additional years of censuses data as available.[57] In addition, we use an alternative definition of migration, which is based on a workers current state of living and state of birth. In this case, we can only run state-level regressions since we do not observe the municipality of birth of a worker. We include results for this alternative explanatory variable in the same Appendix Table C.3 in rows in which regions are states.

Appendix Figure C.1 presents local polynomial regressions of farmers and income that correspond to regressions 2 and 3. We first absorb origin-destination-year and destination-crop-year fixed effects; we then run a polynomial regressions on the residuals of the dependent and independent variables of interest in equations 2 and 3.

[57] Micro-region is a finer-level of labor market aggregation relative to the meso-region. Specifically, we have 137 official meso-regions in Brazil and 558 micro-regions.

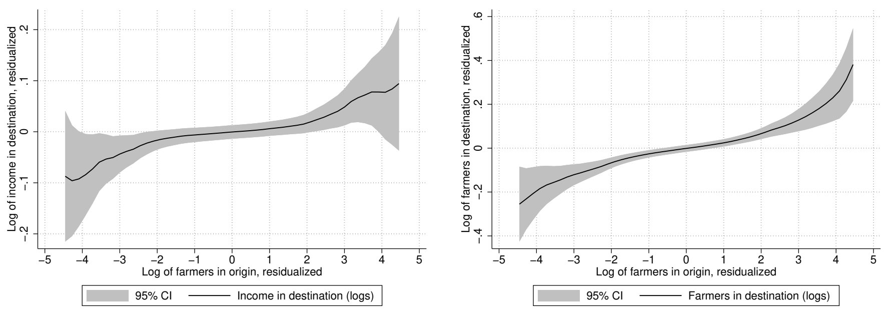
Table C.3: The Relationship between Farmers' Choices and Earning and their Region of Origin - Alternative Sample Specifications

Specification				Dep Var: Income			Dep Var: Farmers		
Geog Unit	Lag	Year	Age	Coef (1)	SE (2)	Obs (3)	Coef (4)	SE (5)	Obs (6)
a. OLS									
Meso	30	2000-2010	30-60	0.023**	(0.010)	6794	0.075***	(0.014)	7375
Meso	20	2000-2010	30-60	0.016	(0.010)	7727	0.086***	(0.011)	8443
Meso	20	1990-2010	30-60	0.003	(0.007)	13640	0.097***	(0.008)	14449
State	30	2000-2010	30-60	0.023*	(0.012)	14132	0.072***	(0.013)	15437
Micro	30	2000-2010	30-60	0.032***	(0.009)	2573	0.157***	(0.015)	2750
Meso	30	2000-2010	20-	0.024***	(0.008)	8844	0.081***	(0.010)	9597
b. OLS (Above Q1)									
Meso	30	2000-2010	30-60	0.047***	(0.016)	5180	0.101***	(0.022)	5609
Meso	20	2000-2010	30-60	0.051***	(0.017)	5903	0.144***	(0.018)	6422
Meso	20	1990-2010	30-60	0.039***	(0.013)	10365	0.140***	(0.014)	10942
State	30	2000-2010	30-60	0.037*	(0.019)	10872	0.097***	(0.021)	11787
Micro	30	2000-2010	30-60	0.033**	(0.014)	1900	0.210***	(0.021)	2018
Meso	30	2000-2010	20-	0.039***	(0.013)	6712	0.085***	(0.017)	7271
c. PPML									
Meso	30	2000-2010	30-60	0.045***	(0.012)	6794	0.120***	(0.013)	7375
Meso	20	2000-2010	30-60	0.043***	(0.013)	7727	0.129***	(0.011)	8443
Meso	20	1990-2010	30-60	0.018*	(0.010)	13640	0.147***	(0.009)	14449
State	30	2000-2010	30-60	0.029*	(0.016)	14132	0.074***	(0.012)	15437
Micro	30	2000-2010	30-60	0.032**	(0.013)	2573	0.213***	(0.018)	7948
Meso	30	2000-2010	20-	0.023*	(0.013)	8844	0.129***	(0.011)	9597
d. PPML (Above Q1)									
Meso	30	2000-2010	30-60	0.083***	(0.020)	5180	0.131***	(0.023)	5609
Meso	20	2000-2010	30-60	0.081***	(0.019)	5903	0.182***	(0.017)	6422
Meso	20	1990-2010	30-60	0.056***	(0.016)	10365	0.181***	(0.013)	10942
State	30	2000-2010	30-60	0.042**	(0.018)	10872	0.104***	(0.019)	11787
Micro	30	2000-2010	30-60	0.039**	(0.018)	1900	0.236***	(0.017)	2018
Meso	30	2000-2010	20-	0.044***	(0.017)	6712	0.133***	(0.017)	7271

Notes: This table replicates Table 2 using alternative specifications. Regressions at the meso- and micro-region level use previous region of living as the previous location of workers. Regressions at the state level use instead the state of birth as the previous location of workers. Our sample exclude non-migrants.

Figure C.1: Local Polynomial Regressions of the Influence of the Region of Origin on Crop Choice and Income of Farmers in their Destination Region

(a) Income in destination, residualized (b) Farmers in destination, residualized



Notes: To construct this figure, we first absorb origin-destination-year and destination-crop-year fixed effects from dependent and independent variables of interest in equations 2 and 3. Using the residuals from these variables, we run a non-parametric regression using a local polynomial smooth.

Appendix Table C.4 reports results with alternative explanatory variables. First, we use total income of workers lagged by 30 years as reported in the demographic census. In addition, based on data from the agricultural census of 1970, we also experiment with total revenues and total land use. Table C.5 shows our results are not driven by the inclusion of any particular crop.

We check if the relationships that we uncovered for the case of agriculture are also found more generally in other sectors of the economy. Appendix Table C.6 shows that if we run the same regressions with workers employed in manufacturing activities, we find similar patterns in the data.

Appendix Table C.7 shows the checks related to alternative explanations that we discuss in the main body of the paper.

Table C.4: The Relationship between Farmers' Choices and Earning and their Region of Origin - Different RHS

	RHS: Total Income			RHS: Total Revenues			RHS: Total Land Use		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
a. Dep Var: Income									
Column 1	0.024***	(0.004)	6788	0.018***	(0.005)	6898	0.008*	(0.004)	7522
Column 2	0.023**	(0.010)	6788	0.027***	(0.009)	6898	0.014*	(0.008)	7522
Column 3	0.047***	(0.017)	5212	0.038***	(0.010)	4755	0.024**	(0.012)	5131
Column 4	0.044***	(0.012)	6788	0.051***	(0.013)	6898	0.035***	(0.011)	7522
Column 5	0.079***	(0.019)	5212	0.057***	(0.012)	4755	0.040***	(0.013)	5131
b. Dep Var: Farmers									
Column 1	0.080***	(0.007)	7369	0.038***	(0.006)	7588	0.050***	(0.005)	8227
Column 2	0.077***	(0.014)	7369	0.075***	(0.011)	7588	0.065***	(0.010)	8227
Column 3	0.087***	(0.024)	5622	0.055***	(0.016)	5152	0.045***	(0.016)	5534
Column 4	0.122***	(0.013)	7369	0.082***	(0.012)	7588	0.095***	(0.009)	8227
Column 5	0.114***	(0.023)	5622	0.043**	(0.017)	5152	0.062***	(0.015)	5534
Column 6	0.165***	(0.011)	127706	0.150***	(0.010)	144218	0.149***	(0.008)	154036

Notes: This table replicates Table 2 using different explanatory variables. Total earnings come from the demographic census and is 30 years lagged. Total revenues and total land use come from the agricultural census of 1970.

Table C.5: The Relationship between Farmers' Choices and Earning and their Region of Origin - Dropping Activities

	Dep Var: Income					Dep Var: Farmers				
Dropped	OLS		PPML			OLS		PPML		
Activity	Coef	SE	Coef	SE	Obs	Coef	SE	Coef	SE	Obs
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
none	0.047***	(0.016)	0.083***	(0.020)	5180	0.101***	(0.022)	0.131***	(0.023)	5609
soy	0.035*	(0.018)	0.067***	(0.024)	4985	0.083***	(0.024)	0.114***	(0.025)	5404
sugarcane	0.036*	(0.019)	0.068***	(0.018)	4404	0.086***	(0.027)	0.116***	(0.026)	4817
tobacco	0.051***	(0.016)	0.087***	(0.020)	5052	0.096***	(0.022)	0.127***	(0.024)	5472
cotton	0.049***	(0.016)	0.086***	(0.020)	5091	0.103***	(0.022)	0.131***	(0.023)	5513
coffee	0.044**	(0.021)	0.084***	(0.026)	4260	0.087***	(0.026)	0.112***	(0.029)	4663
banana	0.050***	(0.016)	0.085***	(0.020)	4970	0.108***	(0.023)	0.135***	(0.023)	5378
rice	0.057***	(0.017)	0.093***	(0.020)	4802	0.110***	(0.024)	0.135***	(0.025)	5118
corn	0.051***	(0.016)	0.082***	(0.021)	4905	0.124***	(0.023)	0.143***	(0.024)	5209
cotton	0.049***	(0.016)	0.086***	(0.020)	5091	0.103***	(0.022)	0.131***	(0.023)	5513
cotton	0.049***	(0.016)	0.086***	(0.020)	5091	0.103***	(0.022)	0.131***	(0.023)	5513

Notes: This table replicates columns 3 and 5 from Table 2. In each row of the table here we drop a different activity from our sample. The row in the top replicates the baseline results from Table 2 for comparison.

Table C.6: The Relationship between Workers' Choices and Income and their Region of Origin - Manufacturing Activities

	(1)	(2)	(3)
a. Dep Var: Income			
Column 1	0.075***	(0.006)	6921
Column 2	0.051***	(0.016)	6921
Column 3	0.063***	(0.021)	5410
Column 4	0.060***	(0.018)	6921
Column 5	0.059***	(0.021)	5410
b. Dep Var: Farmers			
Column 1	0.077***	(0.006)	6939
Column 2	0.111***	(0.016)	6939
Column 3	0.142***	(0.022)	5427
Column 4	0.152***	(0.015)	6939
Column 5	0.188***	(0.018)	5427
Column 6	0.195***	(0.013)	166153

Notes: See Table 2 for description of columns. The manufacturing activities included are: automotive, leather, furniture, processed tobacco, oil, paper, clothing, textile, perfume, and wood.

Table C.7: The Relation between Farmers' Income, Choices, and Region of Origin - Inspecting Alternative Interpretations

Unit	Cell	Ind	Ind	Ind	Ind	Ind	Ind
	No Cap Intensive		SES	Time in Region	Previous Migrants	Control for Selection	State of Origin
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
a. OLS							
Farmers in origin	0.041*	0.040***	0.040***	0.034**	0.032**	0.034*	0.048**
	(0.021)	(0.013)	(0.014)	(0.014)	(0.016)	(0.022)	(0.019)
R^2	0.731	0.389	0.389	0.456	0.371	0.389	0.392
Obs	4768	13784	13784	13784	12075	13784	13784
b. PPML							
Farmers in origin	0.075**	0.077***	0.078***	0.072***	0.068**	0.049*	0.088***
	(0.030)	(0.024)	(0.024)	(0.026)	(0.031)	(0.035)	(0.031)
Obs	4768	13784	13784	13784	12075	13784	13784

Notes: Bootstrapped standard errors in column 6. For the remaining columns, see description in Table 2. Column 1 presents a regression at the cell level and columns 2 to 7 at the individual level. All columns include destination-activity-year and destination-origin fixed effects. All columns drop bottom quartile in the distribution of the dependent variable. Column 1 drop three capital intensive activities: soy, cotton, and sugarcane. Column 3 controls for socio-economic status: gender, age, age squared, indicator variables for ethnicity, education, and education squared. Column 4 includes fixed effects for every year the worker has been in the region. Column 5 controls for the log of worker who migrated to the same region to produce the same activity in the previous census (i.e., for $L_{ijk,t-1}$). Column 6 follows Dahl's strategy to control for selection bias and includes a flexible polynomial of probability choices. Column 7 includes fixed effects for state of origin-activity-year. We exclude return migrants and non-migrants.

C.1.2 Controlling for selectivity bias

Column 6 in Table C.7 controls for selection into region j and activity k. The reason for this check is that one might worry that the sample of individuals is selected into option j, k in a way that induces a correlation between the error and our main regressor in equation

2.[58] To deal with this potential bias, we follow Dahl (2002) and assume that the migration probabilities $\lambda_{i,jkt}$ are summary statistics of the bias induced by selection into a particular option j, k. Our first step is to compute migration probabilities for each triplet (i, j, k) using all of our micro data. These shares, which we denote $\hat{p}_{i,jkt}$ are consistent estimates of the migration probabilities $\lambda_{i,jkt}$ in our model. Our second step is to estimate our income regression, augmenting it with a subset of these probabilities for each individual, as in:

$$\log \text{income}_{\omega ij,kt} = \iota^{I}_{j,kt} + \iota^{I}_{ij,t} + \alpha^{I} \log \text{workers}_{i,kt-1} + f\left(\widehat{p}_{ij,kt}\right) + \epsilon^{I}_{\omega ij,kt},$$

where $f(\widehat{p}_{ij,kt})$ is a flexible polynomial function of the vector of probability choices and ω is an individual. In addition to the probability of the realized choice (i.e., $\widehat{p}_{ij,kt}$), we also have to include in this function cross probabilities related to unrealized choices. Because of the large dimensionality of our choice set, we follow Dahl (2002) and include a subset of probability choices in our earnings regression: (1) the probability of the realized choice, (2) the highest probability of an individual from a given origin i producing in sector k choosing any other pair of region and crop, and (3) the highest probability of an individual from a origin state and activity k choosing any other pair of region and crop. We also include the square and cubic terms of each of the probabilities in our regressions.

C.2 State-level analysis of Fact 4

Using state-level data on exports by state and activity, Appendix Table C.8 shows that states with a larger share of agricultural workers originating from states with higher employment in that crop export more of that activity.

C.3 Econometric Implications of a "Fixed Types" Model

This Section explores the implications of a model in which individuals have assigned "types" (i.e. they can only work on one activity) at the time of their location and occupation decision. We show that, if workers sort according to income, or if they sort at random, such a model yields econometric implications, that are inconsistent with the results found in Section 3.

In both sections below, we consider the following common setup. Workers are born with a type k— and associated productivity $s_{i,kt}$— in a number proportional to their parents, $L_{i,kt-1}$. Idiosyncratic shocks are drawn from the same Fréchet distribution as before, with dispersion parameter κ

A "fixed types" model with sorting according to income

[58]Note that this is not a concern if selection is entirely driven by preference shocks, since these do not show up in the wage regression. This is, in fact, the formulation we adopt in the theory section of our paper.

Table C.8: Relationship between the Composition of Farmers and Exports - State Level

	Exports			
Method	OLS	OLS	PPML	OLS-IV
Explanatory Variable	(1)	(2)	(3)	(4)
Abund.	1.023***	0.706***	0.532***	0.486*
	(0.231)	(0.242)	(0.136)	(0.263)
Comp.	0.094	0.514**	0.299*	0.993***
	(0.289)	(0.243)	(0.159)	(0.370)
R^2	0.684	0.752		0.396
Obs	193	193	321	193
Region-Year FE	Y	Y	Y	Y
Activity-Year FE	Y	Y	Y	Y
Controls: Migrants		Y	Y	Y
Controls: SES + Farm size		Y	Y	Y

Notes: * / ** / *** denotes significance at the 10 / 5 / 1 percent level. Standard errors clustered at the state-region level in parenthesis. The composition of farmers is the log of the number of farmers in the origin as weighted by the share of farmers in a destination. Regressions include the years of 2000 and 2010.

Migration Probability. Migration probabilities $\tilde{\lambda}_{ij,kt}$ reflect the following problem

$$\max_j \left\{ \frac{w_{j,kt} s_{i,kt}}{P_{j,t} \mu_{ij,kt}} \varepsilon_j \right\}.$$

The solution is:

$$\tilde{\lambda}_{ij,kt} = \frac{W^{\kappa}_{ij,kt}}{\tilde{\Xi}^{\kappa}_{i,kt}}$$

where $\tilde{\Xi}^{\kappa}_{i,kt} \equiv \sum_{j'} W^{\kappa}_{ij',kt}$.

Econometric specification. The income of a worker of type k from i going to region j are given by

$$\text{Income}_{ijkt} = w_{j,kt} s_{i,kt},$$

Hence, the corresponding regression equation is:

$$\log \text{Income}_{ij,kt} = \log w_{j,kt} + \log s_{i,kt} + u^I_{ij,kt}, \tag{21}$$

where we allow again for measurement error in income.

Likewise, the migration equation states that

$$L_{ij,kt} = \tilde{\lambda}_{ij,kt} L_{i,kt-1},$$

and the corresponding regression equation is:

$$\log L_{ij,kt} = \kappa \log w_{j,kt} + \kappa \log s_{i,kt} - \kappa \log P_{j,t} - \kappa \log \mu_{ij,kt} + \log L_{ikt-1} - \kappa \tilde{\Xi}_{i,kt}. \qquad (22)$$

Direct comparison of equation (21) with the results from (14) shows that worker heterogeneity according to origin location is necessary to reproduce them. Moreover, one needs to impose $s_{i,k} = \bar{s}_i L^{\beta}_{i,kt-1}$, to replicate our regression. In turn, using this assumption in (21) shows that the coefficient of $L_{i,kt-1}$ should be $1 + \kappa\beta$, which is at odds with the results from (15). In our specification, in which we essentially use one cohort, one cannot separately identify $L_{i,kt-1}$ from the fixed effect that would capture $\tilde{\Xi}_{i,kt}$.

A "fixed types" model with random sorting

Migration Probability. Migration probabilities $\tilde{\lambda}_{ij,kt}$ reflect the following problem

$$\max_j \{\varepsilon_j\}.$$

The standard solution is:

$$\tilde{\lambda}_{ij,kt} = \frac{1}{I}$$

Econometric specification. The income of a worker of type k from i going to j is

$$\text{Income}_{ij,kt} = w_{j,kt} s_{i,kt},$$

Hence, the corresponding regression equation is:

$$\log \text{Income}_{ij,kt} = \log w_{j,kt} + \log s_{i,kt} + u^I_{ij,kt}, \qquad (23)$$

where we allow again for measurement error in income.

Likewise, the migration equation states that

$$L_{ij,kt} = \frac{1}{I} L_{i,kt-1},$$

and the corresponding regression equation is:

$$\log L_{ij,kt} = \log L_{i,kt-1}. \qquad (24)$$

Equation (23) shows that one requires the same conditions as in the previous Section to rationalize the income equation. Moreover, equation (23) shows that the coefficient of $L_{i,kt-1}$ should be 1, which is at odds with the results from (14).

D The Quantitative Model

This section presents the full quantitative model. In Supplementary Appendix Section SA, we provide a full description of the simplified, non-quantitative model we introduced in Section 4 of the main body of the paper.

D.1 Quantitative model

D.1.1 Environment

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. We denote all regions in the world by $\mathcal{W}$. There are $g = 1, ..., G$ sectors, which corresponds to manufacturing, services and agriculture. Within each sector we have $k = 1, \ldots, K_g$ industries (or economic activities) and each region produces an unique variety of each good. Time is discrete and indexed by t. Iceberg trade and migration costs deter the flux of agents and goods across space. At each time, the geography of the economy is given by a set of natural advantages, a matrix of bilateral trade costs, a matrix of bilateral migration costs, and a land productivity term: $\{A_{j,kt}, \tau_{ij,kt}, \mu_{ij,kt}, g_{j,t}\}$. We omit time indexes whenever unnecessary for our presentation.

D.1.2 Technology

Goods production. An agricultural worker with knowledge s rents land and produces according to

$$q_{j,k}(s) = A_{j,k}\left(\left(1-\nu_k\right)s^{1-\rho} + \nu_k l^{1-\rho}\right)^{\frac{\alpha_k}{1-\rho}} m^{1-\alpha_k},$$

where s is the worker's knowledge, l is land, and m is the use of intermediate inputs. The parameters ρ and α_k measure the elasticity of substitution of land and labor, and the value-added share.

Final good aggregator. In each region, a competitive firms aggregate goods according to a nested CES technology and transforms them into sectoral and final output. The corresponding price indexes are

$$P_{j,t}^{1-\sigma} = \sum_{g=1}^{G} \bar{b}_g P_{j,gt}^{1-\sigma}, \quad P_{j,gt}^{1-\sigma_g} = \sum_{k=1}^{K_g} \bar{a}_k p_{j,kt}^{1-\sigma_s}, \text{ and } p_{j,kt}^{1-\eta_k} = \sum_{i\in\mathcal{W}} \left(\tau_{ij,kt} p_{i,kt}\right)^{1-\eta_k}.$$

Land supply. To capture adjustments in the quantity of land, we introduce a government that develops farmland ($H_{j,t}$) using the following technology:

$$H_{j,t} = g_{j,t} x_{j,t}^{1/\varsigma}. \tag{25}$$

where $g_{j,t}$ is the productivity of the land technology and $x_{j,t}$ is a final output requirement. The government prices land competitively and rebates profits to farmers proportionately to their land use. With this formulation, the elasticity of land supply with respect to land rent is $\zeta = 1/(\varsigma - 1)$.

D.1.3 Workers

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

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

where ε is drawn i.i.d from $G(\varepsilon) = \exp(-\varepsilon^{-\kappa})$ and W_{ijkt} is given by

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

where $P_{j,t}$ is the price of a unit of final goods in region j.

D.1.4 Expenditure and accounting

Total final expenditure in region j comes from payments to factors and net transfers, $T_{j,t}$:

$$X_{j,t}^F = \frac{\varsigma - 1}{\varsigma} r_{j,t} H_{j,t} + \sum_{g=1}^{G} \sum_{k=1}^{K_g} w_{j,kt} E_{j,kt} + T_{j,t} \tag{26}$$

Region $j's$ Final expenditure on sector g, and good k within sector g is given by:

$$X_{j,st}^F = \underbrace{\bar{b}_j \left(\frac{P_{j,st}}{P_{j,t}} \right)^{1-\sigma}}_{\equiv b_{j,st}} X_{j,t}^F \tag{27}$$

$$x_{j,kt}^F = \underbrace{\bar{a}_k \left(\frac{p_{j,kt}}{P_{j,st}} \right)^{1-\sigma_s}}_{\equiv a_{j,kt}} X_{j,st}^F \tag{28}$$

Region j's total expenditure in intermediate inputs

$$X_{j,t}^{I}=\sum_{g=1}^{G}\sum_{k=1}^{K_g}\left(1-\alpha_k\right)Y_{j,kt}+\frac{1}{\varsigma}r_{j,t}H_{j,t} \tag{29}$$

where $Y_{j,kt}$ is region j's revenues of activity k. Region j's Expenditure on intermediate inputs from sector s

$$X_{j,st}^{I}=b_{j,st}X_{j,t}^{I} \tag{30}$$

Region $j's$ Expenditure on intermediate inputs from activity k, sector s

$$x_{j,kt}^{I}=a_{j,kt}X_{j,st}^{I} \tag{31}$$

Region j's Total Expenditure on activity k:

$$x_{j,kt}=x_{j,kt}^{F}+x_{j,kt}^{I}$$

Bilateral trade flows

$$x_{ij,kt}=\pi_{ij,kt}x_{j,kt}$$

$$\pi_{ij,kt}=\frac{\left(c_{i,kt}\tau_{ij,kt}\right)^{1-\eta_k}}{p_{j,kt}^{1-\eta_k}}$$

D.1.5 Endowments and supplies

$$E_{j,kt}=\sum_{i=1}s_{i,kt}L_{ij,kt}$$

$$H_{j,t}=g_{j,t}\left(\frac{g_{j,t}}{\varsigma}\right)^{\frac{1}{\varsigma-1}}\left(\frac{r_{j,t}}{P_{j,t}}\right)^{\frac{1}{\varsigma-1}}$$

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

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

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

D.1.6 Equilibrium definition

Given a geography for $t = 1, \ldots, \infty$ and initial labor allocations in period 0, $\{L_{i,k0}\}_{i,k}$, a competitive equilibrium is a sequence of migration flows, efficient labor allocations, and prices, for each origin i, destination j and good k, $\{L_{ij,kt}, E_{i,kt}, w_{i,kt}, r_{i,t}\}$, that satisfy

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

$$w_{i,kt}E_{i,kt} = v_{i,kt}\alpha_k Y_{i,kt}$$

2. Land markets clear in region j:

$$r_{i,t}H_{i,t} = \sum_{g=1}^{G}\sum_{k=1}^{K_g}\left(1 - v_{i,kt}\right)\alpha_k Y_{i,kt}$$

3. Total immigration determines the effective supply of labor in region j, good k:

$$E_{j,kt} = \sum_{i} s_{i,kt}\left(L_{i,kt-1}\right)L_{ij,kt}, \tag{32}$$

where the function $s_{i,kt}$ is defined in equation (6).

4. Migration flows maximize workers utility

$$L_{ij,kt} = \lambda_{ij,kt}L_{i,t-1}. \tag{33}$$

where

$$v_{i,kt} = \frac{\bar{v}w_{j,kt}^{1-\rho}}{\bar{v}w_{j,kt}^{1-\rho} + \left(1 - \bar{v}\right)r_{j,t}^{1-\rho}}$$

$$Y_{j,kt} = \sum_{j'}\pi_{jj',kt}x_{j',kt}$$

Migration Costs and Comparative Advantage

E Quantitative Results for the Steady State

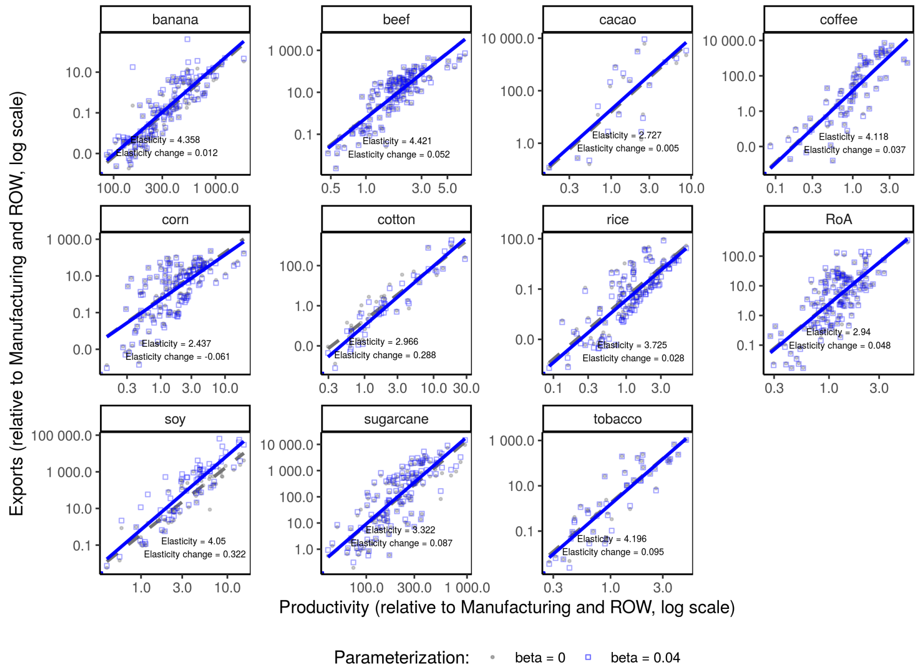
This section exploits the dynamic structure of our model to examine how, matching exactly the same data in 1950, 1980, and 2010, the paths of the economy would diverge in the long run with different values of β.[59] We find that Brazil's agricultural transformation is far from complete. Export specialization in soybean continues to intensify, as do this crop's export and revenue shares; meanwhile all other crops—among them other new crops, such as cattle and corn—become less important. Aggregate export and revenue shares of soybeans continue to grow and settle about 15 p.p. and 1 p.p. higher in the presence of worker heterogeneity, i.e., slightly more than double their 2010 levels.

The endogenous adjustment of knowledge, $s_{i,k}$, underlies these results: On the one hand, migration makes worker knowledge available to all regions; on the other hand, in the presence of migration costs, knowledge rises disproportionately in high-employment regions, because workers are more likely to stay there. As Figure E.2 shows, these forces play out most prominently in the case of soy, for which the elasticity of international specialization to relative productivity increases by 0.32 percentage points (8 percent). This outcome is explained by the fact that soy is a large-employment activity and, moreover, this employment is concentrated in a few regions. Thus, the reallocation of knowledge quantitatively reinforces natural advantage differences.[60]

[59] To obtain the steady state, we simulate the economy forward, keeping all exogenous productivity, trade, migration, and land wedges constant at their 2010 levels, until period to period changes in wages and employment are negligible. Although dynamics are long-lived, the largest part of the steady-state transformation occurs rapidly: the half-life of the the differential between the $\beta = 0.04$ and $\beta = 0$ calibrations is only 2 periods, or 60 years.

[60] Supplementary Appendix Figure S.5 collects all our results for changes in specialization at the regional and aggregate levels, both in 2010 and the steady state, for a specification that removes the scale effects of knowledge, that is, $s_{i,k} = \bar{s}\,(L_{i,kt-1}/L_{i,At-1})^{\beta}$ and shows that this choice makes little quantitative difference.

Figure E.2: Measuring the Contribution of Knowledge in the Steady State



Notes. Each panel corresponds to a crop and each observation is a region in Brazil. In each panel, the horizontal axis is the productivity of that crop (relative to manufacturing and relative to ROW, in logs) and the vertical axis shows exports in that crop (relative to manufacturing and relative to ROW,in logs). For each calibration we also report regression lines.

F International Gains from Trade and the March

Having established the contribution of different economic mechanisms in the case of full GFT and full migration autarky in the main body of the paper, this section returns to our main counterfactual and studies how the March to the West interacted with international GFT.

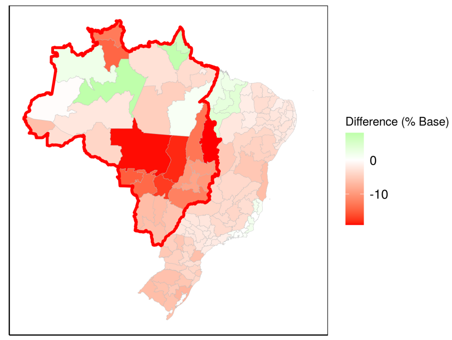
Appendix Figure F.3 (a) maps the international GFT across regions. For Brazil as a whole, the international GFT are 5.0 percent, reflecting that it is a relatively closed economy. Within the country, nevertheless, there are regions for which international trade is crucial, and limiting it can cut down welfare by as much as 11 percent. Panel (b) presents the impact of limiting East-West migration on the gains from international trade. The average international GFT drops by a modest 3.1 percent (0.15 percentage points out of the 5.0 percent baseline), but again, across regions, the differentials range from -19 to 7 percent. The impact of migration is particularly large for the Central-West region, which hosts a large production share of Brazil's new export activities, and which also received the majority of Eastern migrants. We now proceed to disentangle the forces behind these international GFT differentials.

Echoing our Section 7.2 results, we begin by computing the contribution of East-West

Figure F.3: The March to the West and the International Gains from Trade (2010)

(a) The Gains from International Trade

(b) Counterfactual Changes in International GFT



Notes: Panel (a) shows for each region the gains from trade with the rest world, defined as the welfare cost of prohibiting foreign trade only (but allowing domestic trade). Panel (b) subtracts the baseline gains from trade from the gains from trade in the counterfactual scenario (no East-West migration).

migration to international GFT in our baseline economy. Across regions, on average 22 percent of the international GFT are associated with East-West migration opportunities. In the counterfactual economy, these opportunities are not available to workers, which tends to lower the international GFT across the board.

To understand the regional variation in Panel (b), the interaction between migration and comparative advantage is key. Consider first what happens to real wages in each region. From Fact 2, we know that Eastern workers sort disproportionately into agriculture when they migrate to the West and, especially, the Central-West. This means that Eastern migrants sort according to the West's international comparative advantage, which makes Western sales more reliant on international markets, rather than domestic ones. In the West, therefore, the drop in real wages from going to international trade autarky is larger in the baseline, when migration is allowed. The exact opposite happens (i) in a few regions in the northeast—which also have a comparative advantage in agriculture relative to ROW, but instead receive the Eastern agricultural workers in the no-migration counterfactual—and (ii) in the manufacturing regions in the Amazon, such as Manaus. In the rest of the East, because changes in labor supply are small, these effects are quite muted. Finally, note that an additional consequence of limiting migration is to make local real wages, relative to the ones associated with migration opportunities, have a larger weight in the expected welfare of workers born in each region.

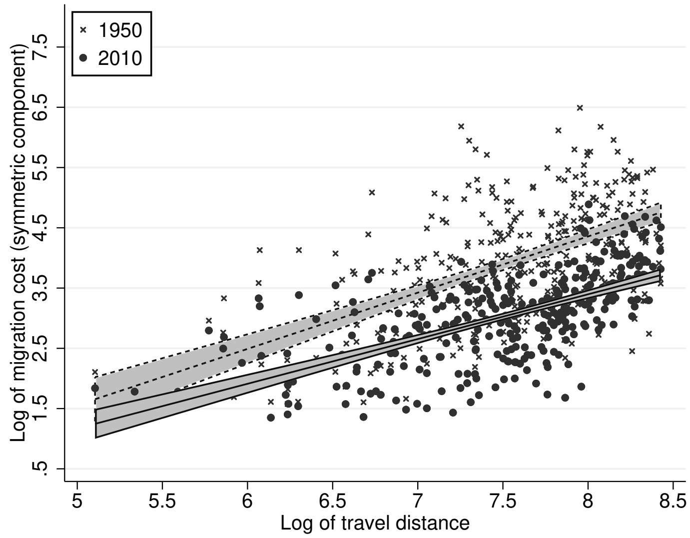
Putting these forces together, we conclude that the international GFT in high-employment regions in the East were not greatly affected by migration, which explains why aggregate GFT are insensitive to it. But the large heterogeneity we observe across other regions is driven by how migration interacts with the forces of comparative advantage.

G Appendix Figures and Tables

Figure G.4: Federal Government Propaganda about the March to the West, 1940s

Notes: Poster features Getulio Vargas, (president 1930-1945 and 1951-1954). The quote in the bottom translates to "The true meaning of Brazilianness is the March to the West". This quote comes from one of his famous speeches, later named the "The speech at midnight" ("*O discurso da meia noite*") given at midnight on December 31st 1937 from Guanabara Palace - Getulio's official residence - and transmitted via the national radio (Vargas, 1938).

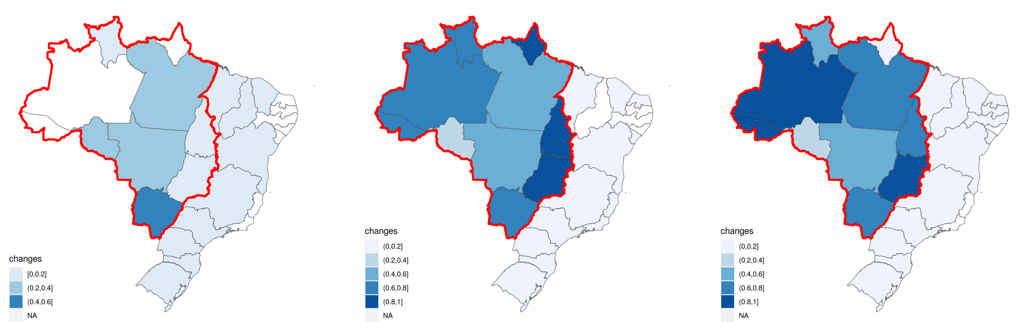
Figure G.6: Correlation between Migration Costs and Distances



Notes: This figure plots the log of the symmetric component of migration costs between states ($\mu_{ss',t}$) against travel distance between states. We estimate migration costs based on the fixed effects of gravity equations of migration in which we use no information on distance. Lines in the figure show the best linear fit and the 95 percent confidence interval.

Figure G.5: The Share of Workers born in the West across all States in 2010, for Selected Activities

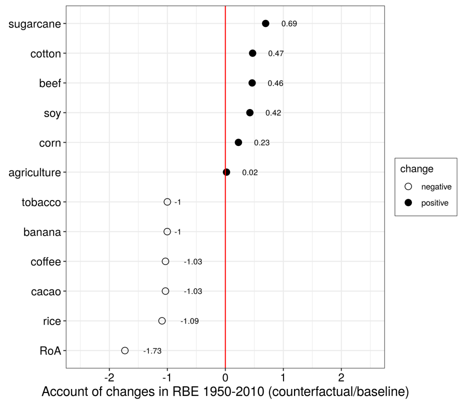
(a) soy (b) cattle (c) corn



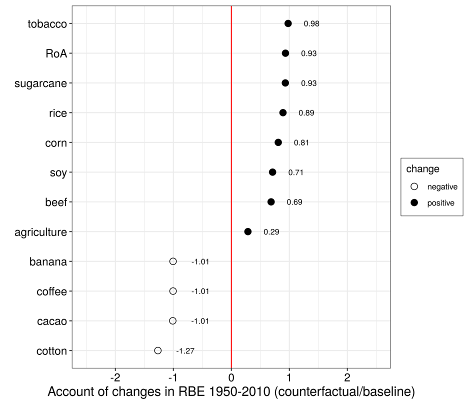
Notes: Each region is a state. For each state we present the fraction of the total employment in 2010 comprised by workers born in the West. The red contour shows the states we classify as the West.

Figure G.7: Accounting for Observed Changes in Specialization

(a) West



(b) Brazil

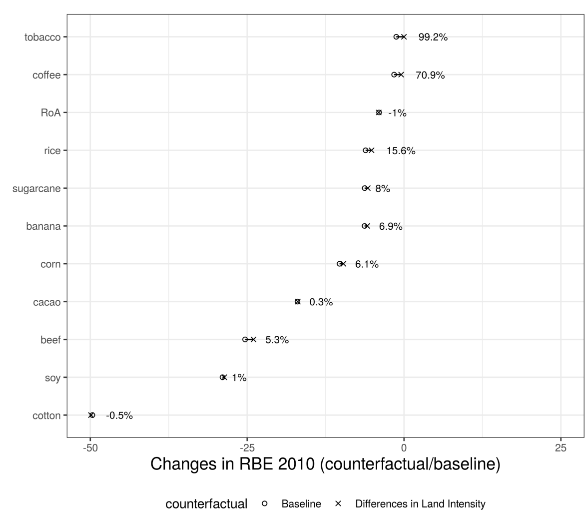


Notes: These figures show the counterfactual change in specialization for each activity, relative to the that in the data, between 1950 and 2010. Specifically, each panel plots, for each activity k

$$\frac{RBE^{\text{counterfactual}}_{BF,kMfg,2010}/RBE^{\text{baseline}}_{BF,kMfg,1950} - 1}{|RBE^{\text{baseline}}_{BF,kMfg,2010}/RBE^{\text{baseline}}_{BF,kMfg,1950} - 1|}.$$

Activities in black circles are those in which RBE grew over this period, while those in white circles shrunk.

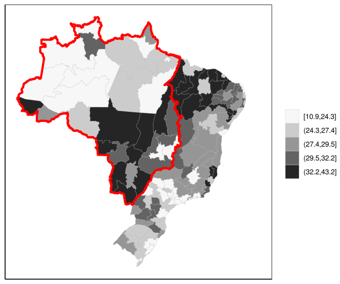
Figure G.8: Measuring the Contribution of Factor Intensity (2010)



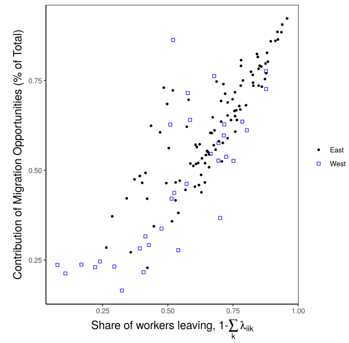
Notes. Each row is an activity aggregate at the country level. The hollow circle presents the counterfactual change in export specialization in our baseline calibration. The cross presents the counterfactual change in specialization in a calibration that gives different land intensities in production to crops and cattle. The numbers represent the drop in the second counterfactual relative to the baseline, in percentage terms.

Figure G.9: Comparative Advantage, Migration, and the Losses from Autarky (2010)

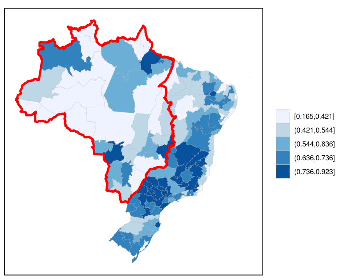
(a) The Full Gains from Trade



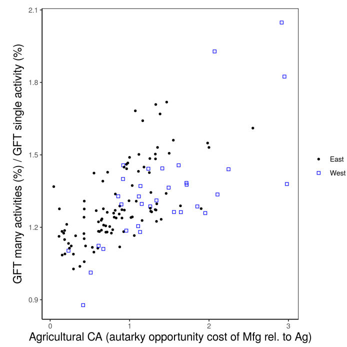
(b) The Contribution of Migration Opportunities



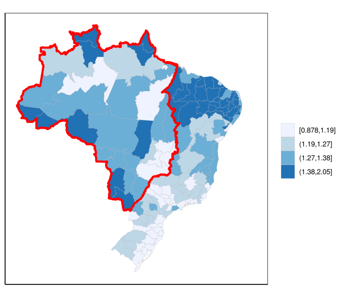
(c) Mapping the Contribution of Migration Opportunities



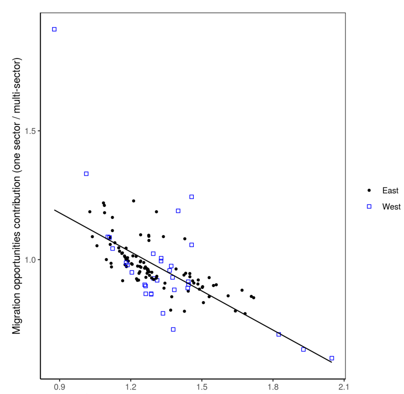
(d) Comparative Advantage and the GFT



(e) Mapping the Contribution of Comparative Advantage



(f) The Interaction of Migration and Comparative Advantage



Notes: All simulations are for 2010 and each observation is a region. Panel (a) shows the welfare losses from letting each region go to full trade autarky, as a percent of baseline welfare. Panel (b): The horizontal axis the fraction of workers leaving that region in the baseline simulation. The vertical axis measures the ratio of the welfare cost that results solely from migration opportunities (i.e. setting the domestic contribution to zero) to the total costs and taking the ratio of $\hat{\Xi}_i$. Panel (c): shows the vertical axis of Panel (b). Panel (d): The vertical axis plots the ratio of the GFT with many activities to the GFT with one activity, both as percent of baseline. The horizontal axis shows the autarky opportunity cost of manufacturing relative to agriculture (measured as relative price indexes in autarky). Panel (e) shows the horizontal axis in Panel (d). Panel (f): The vertical axis is the same as in in Panel (d). The horizontal axis shows the change in the contribution of migration opportunities in going from a one-activity to a many-activity version of our model.