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

Heitor S. Pellegrina
University of Notre Dame

Sebastian Sotelo
University of Michigan*†

First Version: January 2021
This Version: January 2024

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. Migration allowed workers to sort according to their own comparative advantage, reallocating knowledge and raw labor to high-productivity regions. In 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 25 percent of the observed changes in specialization. Road expansions were key drivers of migration. Migration opportunities, moreover, account for a substantial share of the gains from trade.

Keywords: International Trade, Migration, Comparative Advantage

*We thank the editor, three anonymous referees, and Dominick Bartelme, Jonathan Eaton, Andrei Levchenko, Parag Mahajan, 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 many institutions. Andrés Fajardo, Natalia Martinez Camelo, Brian Cevallos Fujiy, Elliot Kang, Lucas Correa Dias, and Razi Iqbal 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: hpelleg3@nd.edu and ssotelo@umich.edu. All errors are our own.

†An Online 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 Rodriguez-Clare, 2015). But while recent work has also documented that comparative advantage itself evolves over time (e.g. Levchenko and Zhang, 2016), 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 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 focus on two channels. First, migration determines the allocation of labor across regions and activities that differ in their local productivity. Second, heterogeneous workers sort according to their own comparative advantage, which determines the size and composition of the labor supply across economic activities and regions. To measure the strength of these mechanisms, we incorporate them into a dynamic 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 three empirical facts that guide our modeling approach. 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 specialization, 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 (Fact 1). 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).

We then document two facts that jointly demonstrate that the migration of workers from different origins shapes regional productivity and specialization. We first show that average labor productivity in a region-crop pair is higher when immigrants to that region come from origins with high exogenous productivity in that same crop (Fact 2). That is, migrants from high-productivity origins increase labor productivity in their destinations. This fact remains agnostic, however, on how a worker's own productivity in a crop is shaped by his origin region. We next establish that higher sectoral employment in a migrant's origin increases his income and employment (Fact 3). Comparing farmers who emigrate from different origins but end up in the same destination and working in the same crop, those coming from a region with 1 percent larger employment in that crop earn 0.03 to 0.11 percent larger incomes and have a 0.08 to 0.23 percent larger employment in that crop. In both Facts 2 and 3, we complement OLS estimates with instrumental variable strategies to address endogeneity concerns.[1] Our conclusions are robust to a comprehensive battery of tests, including the use of different samples and data sets, alternative measurements, controls for selection, and verifying other implications of worker heterogeneity in the data.

After establishing that migrants' knowledge is shaped by employment in their origin and that they impact productivity in their destination, we develop an overlapping generations model of trade with many activities and forward-looking, heterogeneous workers. Workers acquire activity-specific knowledge through exposure to employment 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 and knowledge of the labor force.

Using the model, we obtain new analytical results relating migration to productivity and comparative advantage. We first describe how geography shapes access to knowledge from other regions and how it mediates the impact of population growth on welfare and real wages. Second, we show that in a special case of our model, the non-migrant employment share in a region-activity pair summarizes the impact of migrants on comparative advantage. Although valid only under special circumstances, this result is useful to understand why the impact of

[1] We exploit historical coincidences in push and pull factors over past decades, following recent work by Burchardi et al. (2019). For Fact 3, in addition, we build an instrument that shifts employment based purely on geographical proximity to other regions, adapting the approach in Harris (1954) and Donaldson and Hornbeck (2016).

migration was spatially heterogeneous during the March.

We map the model's key parameters to the reduced-form elasticities from Fact 3, and calibrate the rest of parameters by inverting the model so that it matches the evolution of trade flows, gross output, and migration flows between 1950 and 2010. Having calibrated the model, we establish the quantitative importance of worker knowledge in our setting. We first show that that limiting the portability of knowledge is a powerful deterrent to migration and specialization. In a counterfactual scenario in which, upon migration, eastern workers lose their knowledge advantage relative to western workers, Brazil's specialization in the new comparative-advantage crops would be 13 percent smaller and agricultural migration to the West 37 percent smaller. We then show that the steady state of our baseline economy features stronger specialization in soy, reflecting the increased sorting of workers according to their comparative advantage.

Next, we show that the evolution of migration costs was critical in settling the West. In a counterfactual scenario in which migration costs remain at their 1950 level, we find that migration to the West would have been 60 percent lower by 2010. Trade specialization in agricultural goods, moreover, would have been pervasively lower across Western regions, and the largest effects would occur in crop-region pairs where observed migrant shares are largest. Aggregating these regional changes yields large swings in specialization in new commodities for the West as a whole: Our comparative advantage index drops by 47 percent for soy and by about 32 percent for beef and corn. For Brazil as a whole, soy and beef specialization are 29 and 25 percent lower. In addition, the drop in migration costs accounts for up to 25 percent of the *observed* evolution of Brazil's specialization in the new comparative advantage commodities, between 1950 and 2010 (or 32 percent of the export share growth of these commodities). Studying the role of specific policies that fostered migration, we find that the expansion of highways to the West was a key driver of Brazil's comparative advantage, unlike land grants, whose effect was concentrated in the cattle industry.

We close our analysis by studying welfare. We show that migration opportunities account for up to 60 percent of the full gains from trade (i.e. welfare costs of not being able to trade with any other region). In addition, heterogeneity in the patterns of comparative advantage across regions has a large quantitative impact on how migration opportunities shape 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.[2]

Our paper contributes to a literature that measures the sources and evolution of com-

[2]We do not include the costs of expanding agriculture into the Amazon. Recent papers explore the link between trade and deforestation (e.g. Dominguez-Iino, 2023; Farrokhi et al., 2023; Hsiao, 2023).

parative 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 relative productivity and factor abundance differences are key drivers of comparative advantage.[3] Buera and Oberfield (2020); Cai et al. (2022); Lind and Ramondo (2018), among others, study the diffusion of ideas and comparative advantage in open economies. We contribute to this literature by establishing that domestic migration determines the evolution of regional and national comparative advantage, especially 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, 2022, Porteous, 2019, Pellegrina, 2022, Sotelo, 2020), development and structural change (Tombe, 2015; Farrokhi and Pellegrina, 2020; Porteous, 2020), and the implications of climate change (Costinot et al., 2016 and Gouel and Laborde, 2021). In most of this literature, 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.

Our results thus complement a recent literature studying the development of soybean in Brazil. Focusing on the 2000s, Bustos et al. (2016) and Bustos et al. (2019) show that regions more exposed to the arrival of GMO soy saw faster structural transformation, and that these increases in industrialization came without substantial innovation gains. Coinciding with our analysis period, Pellegrina (2022) studies the adaptation of soybean to tropical regions starting in the 1970s. Relative to these studies, we emphasize the impact that knowledge mobility had on shaping comparative advantage within the agricultural sector, taking the evolution of productivity as given. Our results also indicate that soybeans are particularly affected by the transfer of knowledge from the East to the West.

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 (2022), 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.

[3] 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).

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 (2022), Artuç et al. (2010), and Caliendo et al. (2019), we also study a new source of complementarity between trade and migration: the differential propensity of workers—driven by ability or geography—to migrate to different regions and pursue different activities.[5] Desmet et al. (2018) characterize the growth properties of a spatial model of innovation and growth. We add to their results by showing that geography shapes access to knowledge from other regions and that learning from others, in our context, introduces weak scale effects in growth.

Lastly, we contribute to a literature documenting how migrants' knowledge shape their host economy. 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. Arkolakis et al. (2020) study the impact of migrants on the technological frontier in the U.S. in the 19th century. de la Roca and Puga (2017) show that a migrant's past environment shapes his learning and future productivity. Our contribution to this literature is twofold. First, relying on detailed individual-level surveys, we provide new evidence 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.

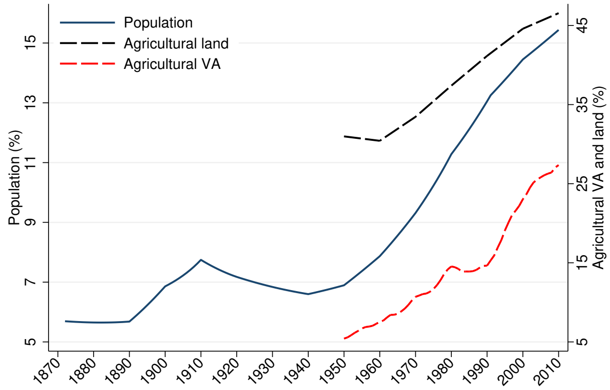
2 The March to the West

This section presents a brief history of the March and discusses how these developments inform the rest of the paper.

[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 drives FDI across US counties. Cardoso and Ramanarayanan (2022) 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.

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



Notes: This figure shows the evolution of the West's shares of population, agricultural land, and agricultural value added in Brazil's total. 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.

2.1 Background on the March

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.[6]

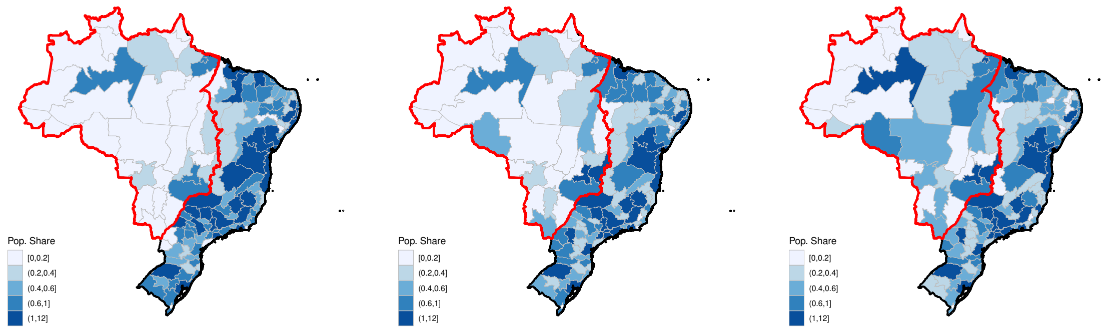
The onset of the March. Reflecting the history of Brazil's settlement, until the 1950s population was concentrated along the Atlantic shore. The economy was organized mostly around export-oriented commodities such as sugarcane, coffee, and cotton, which benefited from easy access to ports. With the exceptions of gold extraction in Minas Gerais during the 18th century and rubber exploitation in the Amazon in the late 19th century, poor access to the interior limited economic activity there (Baer, 2001).

The March to the West began in the 1940s, when urbanization and population growth took off in Brazil. Concerned with food security and population pressure in the urban centers of the southeast, president Getulio Vargas launched a large-scale project, named the

[6]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.

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 Brazil's population that live in each meso-region, which is the geographic unit in our analysis. The red contour shows the meso-regions classified as the West.

"March to the West", to occupy the Central-West region and boosted this initiative through several policies and institutions. Eight National Agricultural Colonies (Colônias Agrícolas Nacionais or CAN) were created in the 1940s to orderly populate the Center-West and the North, and which allocated land lots of 20 to 50 hectares to migrants from other regions (Brasil, 1941). In addition to granting land, in 1943 the government created the Fundação Brasil Central (FBC), an institution charged with channeling private and public capital to lay the infrastructure for making the expansion economically viable. The FBC continued the work started with Roncador-Xingu, the first publicly funded expedition to the West in 1941, whose goal was to chart unpopulated regions in the interior that were amenable to building new cities (unlike previous expeditions, which were primarily concerned with the study of nature). The FBC created new population nuclei in modern day Goias and Mato Grosso; it coordinated the construction of the Tocantins rail line, which connected the new agricultural frontier, South to North; and it created regional commerce centers with basic infrastructure for itinerant traders (Villas Bôas and Villas Bôas, 1994). In parallel, the government launched an information campaign that proclaimed, e.g., that "the true sense of Brazilianness is the March to the West" (Appendix Figure O.2 and Vargas, 1938), with the intention of improving perceptions about the region.

While Getulio Vargas' government gave the occupation of Brazil's interior its initial thrust, it was not until the 1960s, when successive governments undertook larger investments in infrastructure, 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 president Juscelino Kubitschek moved the Brazilian capital from the coastal city of Rio de

Janeiro to Brasília, a newly constructed city in the Central-West region. Complementing this political decision, the government built highways to connect Brasília to the rest of the country (Morten and Oliveira, 2016; Bird and Straub, 2020).

Between the 1960s and the 1980s, a military dictatorship in Brazil expanded Kubitschek's projects so as to further integrate the North of Brazil. The government invested in new roads under a national transportation plan (*Plano Nacional de Viação*) and created a free economic zone for assembly plants in Manaus, next to the Amazon river. In 1970 the newly-created National Institute for Colonization and Agrarian Reform (INCRA) became the primary means of distributing land in Brazil. In the West, INCRA distributed land under federal ownership—including 60 percent of the land in the State of Mato Grosso—to private and public colonization companies, with the goal of establishing new agricultural centers.

Alongside these investments in infrastructure and land reforms, the Brazilian agricultural research institute EMBRAPA, founded in 1973, expanded its research on crop adaptation to regions closer to the tropics. An emblematic result of this effort was the adaptation of soybeans to tropical areas during the 1970s (Nehring, 2016; Sabel et al., 2012; Amann et al., 2018). These scientific advances continued well into the 2000s, partly due to private R&D, when new seeds increased the productivity of soybeans (Bustos et al., 2016).

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.

2.2 Economic Mechanisms

We now discuss briefly the role of knowledge in this episode and how we will map key historical developments into the quantitative model we develop later.

Knowledge and Migration during the March.

The knowledge of migrants has contributed to the expansion of new economic activities in several historical episodes.[7] The Brazilian experience was no different. To give an example, case studies suggest that *gauchos*, migrants from the South of Brazil, led the expansion of

[7]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).

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." (Sabel et al., 2012, p.181). To give another example, migrants also shaped the expansion of coffee "[...] the commercial production in the region only consolidated in the mid-1970s in the colonization projects, particularly in Rondonia. Colonizers, originating from coffee producing regions, brought the production to the Amazon" (Marcolan and Espindula, 2015). Focusing on the expansion of commercial crops in the West of Brazil, Sanders and Bein (1976) posit an important role for knowledge carried by migrants: "Crop shifts generally require new production and marketing technologies. How do farmers on the frontier obtain this information? One hypothesis is that the continuing waves of immigration make possible these crop shifts. According to this hypothesis, the immigrants bring in the production knowledge of the particular crop from other areas."

The structure of Brazilian farms, which are usually managed by small teams of workers, is also suggestive of the importance of knowledge in the organization of production. According to the agricultural census of 2006, the average number of workers per farm in the West is 3.5, with small variation across municipalities, despite substantial variation in landholdings. Most farmers in the region, moreover, own and manage their farms. In addition, about 90 percent of farm managers are land holders, as opposed to externally hired managers.

Knowledge, which makes workers heterogeneous and leads them to sort according to their own comparative advantage, is a central mechanism in our model. After carefully documenting that this heterogeneity matters to the impact of migrants in their destinations and that it is tied to employment in each worker's origin, we incorporate it into a quantitative OLG framework.

Incentives to Labor Mobility, Trade Barrier Reductions, and Productivity.

First, the government policies discussed above encouraged migration to the West. The FBC and, later on, the *Plano Nacional de Viação* reduced substantially travel times to the West. The government's information campaign also seems to have been effective in shifting preferences for living in the West.[8] Land grants from INCRA and the CANs, in addition, provided economic incentives for farmers to leave crowded areas in the East for the promise of

[8] As recounted later on by demographer Neiva (1985): "There, on the radio, it was night or day and they talked about that government propaganda, talking about the agricultural colonies and such. Anyone who didn't go there was lazy, because the government was giving land there."

abundant land—and high earnings—in the West. We capture infrastructure and information as a reduction in migration costs and, separately, we capture land grants as profit rebates from new land development to workers, quantifying their impact separately.

Second, government policy also facilitated domestic and international trade by lowering the physical cost of shipping goods through new highway infrastructure and assembly zones. Our model will capture these policies as reductions in trade costs, which we measure separately from reductions to migration costs.

Third, the government boosted agricultural productivity in the West, combining technical support with initiatives such as EMBRAPA. As documented by Sabel et al. (2012), migrants' needs interacted with government investments: "In the very beginning, growers were unsuccessful with their crops, as the soybean varieties that they planted matured prematurely [...] After one or two rice harvests, growers started to gravitate toward soybeans, putting pressure on research institutions to help develop varieties suited to the region's environment, as well as techniques to improve the soil. [...] The creation of EMBRAPA in 1973 was especially useful in bringing together research that was already available in other public research institutions and universities, and transferring this knowledge to farmers. Its creation also proved to be valuable in showing growers that production was possible by means already available." We recover changes in productivity by inverting our model and matching revenue and employment data. Our model allows productivity to be endogenous via local external economies, which captures the spirit of EMBRAPAs response to agricultural workers' demands.

3 Data

We collect data to construct a panel on employment, migration, gross output, and trade for Brazil between 1950 and 2010. Our data contains 133 meso-regions—which can be aggregated into 26 States—,[9] two countries (Brazil and the rest of the world), and 13 economic activities (11 agricultural activities, manufacturing, and services).[10] We briefly describe the data here and relegate details to Online Appendix OA.

[9] 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.

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

To measure migration flows and incomes, we use state-level migration and employment variables coming from decadal Brazilian demographic censuses from 1950 until 2010. For 1970 onward, we also observe responses to a detailed questionnaire, applied to 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, including total earnings and hundreds of activities (in particular, the crop a farmer grows). We complement our main data set with a panel on wages and activities for all formal workers, 2005-2015, using administrative data from Relação Anual de Informações Sociais (RAIS).

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. 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). Data on land use and total labor employment for Brazil are from IPEA and for the rest of the world from FAO.

Lastly, we utilize government data on the expansion of highways in Brazil since the 1950s and employ the Fast Marching Method to compute travel distances between any two meso-regions, incorporating the variable speeds associated with different road types. We also gather data from INCRA on all land settlements assigned in Brazil over the same period.

4 Facts about Migration and Comparative Advantage

This section documents three facts about migration and comparative advantage in Brazil. Fact 1 describes Brazil's rising specialization in soy, corn, and cattle, relative to the world, between 1950 and 2010. Next, we proceed in two steps to demonstrate that the settlement of the West by new migrants with specific knowledge was one of the drivers of this evolution. As a first step, Fact 2 shows that the composition of migrants impacts labor productivity in a destination: Labor productivity is higher in a crop-destination pair whose migrant composition favors migrants from more productive locations. This exercise remains agnostic about the precise channel through which origin differences shape workers differences and, ultimately, destination productivity. In our second step, Fact 3 shows how employment in the origin region impacts each worker's productivity and employment choices. In both steps we develop empirical designs to give our estimates a causal interpretation.

4.1 Fact 1: The Evolution of Brazil's Comparative Advantage

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 activity k' and region i') in year t as follows:

$$RBE_{ii',kk',t} \equiv \frac{X_{iFk,t}/X_{i'Fk,t}}{X_{iFk',t}/X_{i'Fk',t}}, \tag{1}$$

where $X_{ijk,t}$ are $i's$ sales to j in activity k at date t. This index is a useful indicator of comparative advantage for two 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.[11]

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)
soy	0.01	60.88	69.74	44.85	237.02	0.00	29.43
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
sugarcane	18.82	27.91	52.28	57.32	18.36	5.70	16.13
cotton	9.41	0.72	3.28	2.16	10.83	6.23	1.79
coffee	655.55	339.56	166.11	190.24	3.95	59.22	8.01
rest of agriculture	1.01	2.19	2.51	2.52	2.47	13.66	28.40
cacao	273.44	299.89	15.00	17.22	0.11	8.98	0.40
rice	0.19	0.07	0.43	0.48	0.06	0.44	0.56
banana	5.09	1.81	0.38	0.44	0.00	0.49	0.06
tobacco	6.70	19.92	64.27	73.84	0.00	1.76	4.39
agriculture	4.98	5.08	6.14	5.35	11.50	79.77	31.96
manufacturing	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. Crops are ordered according to the West's relative bilateral exports in 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. The table shows that by 2010, Brazil is intensively specialized in several commodities relative to ROW, among them coffee and

[11] 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. Online Appendix Table O.3 shows that using instead Balassa's Revealed Comparative Advantage index provides similar conclusions. French (2017) and Costinot et al. (2012) discuss different empirical measures comparative of advantage.

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, rising from close to zero in 1950.

4.2 Fact 2: Productivity and the Composition of Migrants

We now show that migrants coming from high-productivity regions increase aggregate productivity in their destination, using the following specification:

$$\ln\left(\frac{Q_{sk,t}}{\ell_{sk,t}}\right) = \delta_{k,t} + \delta_{s,t} + \kappa \underbrace{\ln\left(\sum_{s'\neq s} \frac{M_{s's,t}}{M_{s,t}} S_{s'k}\right)}_{\text{Migrants' Composition}} + X'_{sk,t}\beta + u_{sk,t}, \tag{2}$$

where s, k, and t denote state, crop, and year, and $Q_{sk,t}$ is physical output (in tons), $\ell_{sk,t}$ is employment, $S_{s',k}$ is an exogenous measure of land productivity based on FAO-GAEZ, $M_{s's,t}$ is the stock of migrants from s' living in s, and $M_{s,t}$ is the total stock of migrants in s. Across specifications, the controls $X_{sk,t}$ include combinations of sectoral employment, land suitability, the nonmigrant share, and lagged labor productivity by region and sector. We also control for crop-broad region fixed effects, to capture variation across regions and sectors coming purely from geography. Note, moreover, that because we use shares, our regression ignores variation purely due to scale. To exploit long-run variation in migration flows, we use data coming from the census, for $t = 1980$, 1990, 2000 and 2010, and because this strategy requires us to build on historical migration flows going back as far as possible, our main specification estimates these regressions at the state level.

The goal of this regression is to show that average worker productivity in a given activity and state responds to the composition of its immigrant labor force. The sum over migrants' origins, therefore, does not include destination s itself. Because we want to ensure we exploit exogenous variation to estimate the impact of migrants from different origins on productivity, our primary specification uses $S_{s'k}$ as an exogenous characteristic of the origin, but our results hold when we use alternative measures—including employment in the origin—as we show below. The fixed effects $\delta_{k,t}$ control for country wide shocks that affect each crop (such as the arrival of a new variety) and $\delta_{s,t}$ control for destination-specific shocks that might increase local productivity or draw in better workers on average. With the inclusion of $\delta_{s,t}$, κ measures the impact on *relative* labor productivity, as the theory of comparative advantage would suggest. In equation (2), any unobserved shock that shifts the composition of interstate migrants, while simultaneously shifting worker productivity in the destination, poses a threat to identification. To guard against such concerns, we develop an IV strategy that removes any origin– and destination-specific factors driving migration from s to s', while also controlling for bilateral variables such as distance and agricultural similarity. Doing so ensures that the predicted migration flows are not systematically correlated with distance and other bilateral factors, year by year. Following Burchardi et al. (2019), we generate a predicted stock of migrants that reflects exclusively historical coincidences in push and pull forces at the state level since the 1950s. Specifically, we instrument the composition term in equation (2) with $\ln\left(\sum_{s'\neq s}\left(\widehat{M}_{s's,t}/\widehat{M}_{s,t}\right)S_{s'k}\right)$, where $\widehat{M}_{s's,t}$ and $\widehat{M}_{s,t}$ are constructed according to their method, as explained in Appendix A.[12]

Table 2: The Impact of Migrants's Composition on Average Labor Productivity

	DV: Log of Output per Worker						
	OLS	OLS	OLS	OLS	IV	IV	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Composition	0.205***	0.189***	0.275***	0.059	0.226**	0.344***	0.276***
	(0.059)	(0.055)	(0.056)	(0.074)	(0.109)	(0.121)	(0.090)
R2 or K-P	0.885	0.917	0.929	0.969	29.106	22.115	31.020
Obs	751	751	751	751	751	751	751
P-value (Ho = to col 3)				0.265			
Controls							
-Crop suitability	Y	Y	Y	-	Y	Y	Y
-State/Year FE	Y	Y	Y	Y	Y	Y	Y
-Crop/Year FE	Y	Y	Y	Y	Y	Y	Y
-Broad Region/Crop FE	-	Y	Y	-	-	Y	Y
-State/Crop FE	-	-	-	Y	-	-	-
-Lagged productivity	-	-	Y	Y	-	-	Y
-Labor employment	-	-	Y	Y	-	-	Y
-Share of migrants	-	-	Y	Y	-	-	Y

Notes: * / ** / *** denotes significance at the 10 / 5 / 1 percent level. Regressions at the state and crop level. Standard errors clustered at the state-crop level are in parenthesis. Columns (5) to (7) report Kleibergen-Paap (K-P) statistic instead of R2. Column (1) to (7) control for the destination's crop suitability based on our FAO-GAEZ measure. Column (2), (3), (6), and (7) add broad region fixed effects for 5 statistical zones of Brazil, i.e., South, North, Southeast Northeast, and Midwest. Column (4) adds state-crop fixed effects. Columns (3), (4) and (7) control for lagged worker productivity by 30 years, the total labor employment in a given year for that same destination and activity pair, and the total share of migrants producing that same activity. Regressions are weighted by states' initial population size. The sample only includes region-crop pairs with positive production.

Table 2 reports our estimates of equation (2). Across specifications, the point estimate is between 0.06 and 0.35. IV estimates tend to be larger than their OLS counterparts, but are indistinguishable statistically. Our interpretation of a coefficient of 0.2 is as follows: If the

[12] Using meso-region level data, which are available for a shorter span, we examined a similar strategy and obtain qualitatively similar conclusions (see Appendix OB).

composition of migrants changes, raising average land productivity of the migrants' origins by 1 percent, then labor productivity in the destination rises by 0.2 percent.[13]

Details, robustness, and complementary results. Online Appendix OB shows that our results change little under the following alternative specifications. First, our results remain if we construct our instrument excluding controls for bilateral distance and agricultural similarity when we predict the stocks of migrants (Online Appendix Table O.6). Second, our results hold when we include more or fewer census years in the regression, which suggests particular years do not drive the results (Online Appendix Table O.6). Third, we show that no individual crop drives our results (Online Appendix Table O.7). Fourth, our results are quite similar when we measure origin heterogeneity using employment, total output, or total revenue at the origin, instead of crop suitability—although we prefer our main specification, which uses exogenous suitability $S_{s'k}$, to avoid endogeneity (Online Appendix Table O.8). Fourth, we confirm that other regional outcomes also respond to composition in the direction we would expect: The elasticity of revenues per worker is about 0.2, while that of prices is about -0.09 (Online Appendix Table O.9).[14]

A final concern about our results is that they could be driven by the confluence of (i) predicted migration flows $\hat{M}_{s's,t}$ that systematically correlate with distance and (ii) spatial correlation in crop-specific suitability, S_{sk}. As we mentioned earlier, however, our implementation of the Burchardi et al. (2019) instrument ensures predicted flows are not correlated with distance. To mitigate concerns about spatial correlation in productivity, we take three steps. First, we add controls that target specifically the average suitability of neighbor regions. Second, we exclude the nearest state from the computation of migrant composition. Third, we add controls in the construction of the predicted migration flows that control for the difference in suitability between origin and destination regions. Online Appendix Table O.10 shows that our results remain similar with these new specifications.

These robustness tests and specifications indicate that our IV, which exploits cross-sectional variation, provides a valid identification strategy. An alternative is to include finer, state-crop fixed effects (δ_{sk}), so as to control for any characteristic that is fixed over time and focus purely on within-group variation (Column 4). In our context, that strat-

[13]Alternatively, imagine a change in the origin of all migrants, from a region in the 25 percentile of the distribution of $S_{s',k}$ to one at the 75 percentile—which in the data entails a 22-fold increase in land productivity—, then labor productivity of the destination region would be 0.8 times higher. To put this number in perspective, within crops, labor productivity in the 75 percentile relative to the 25 percentile in the distribution of labor productivity is 4.8 times higher.

[14]These results hold even when we look within coffee, across varieties (robusta vs. arabica). When the share of migrants coming from, e.g., arabica producing regions increases, so does the share of arabica in total coffee production in the destination (Online Appendix Table O.11).

egy comes with two potential limitations. First, it removes cross-sectional variation that is useful for estimation. Our main regressor changes slowly over time, so that in column (4) the standard errors increase by one-third, and the estimates in columns (3) and (4) are statistically indistinguishable.[15] Moreover, although the point estimate falls to 0.059 and we cannot reject the null of no effect in Column (4), the coefficient remains economically significant (see interpretation in footnote 13).[16] Second, state-crop FEs do not control for omitted characteristics that change over time, such as local productivity growth that induces migration between particular regions, thus changing the migrant composition. Our IV strategy, instead, controls for a number of characteristics that would be absorbed by the fixed effects, while also removing biases arising from migration between particular regions, because, year by year, we add bilateral controls and exclude bilateral flows from the push and pull variables. Based on these considerations, we conclude the evidence indicates that migration composition causes productivity shifts in receiving regions.

We have documented that the migrant labor force shapes relative labor productivity in its destination. We have remained agnostic, however, about the specific mechanism through which origin characteristics translate into individual heterogeneity. We turn next to this question.

4.3 Fact 3: The Impact of Origin on Worker Income and Occupation

This section shows that, upon migration, farmers who originate 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. The core of our exercise is to compare migrant farmers working in the same activity and destination—therefore, under the same agro-climatic conditions and institutions—but who come from origins with different patterns of agricultural specialization. Specifically, we estimate the following regressions:

$$\ln \text{income}_{ijk,t} = \iota^{I}_{jk,t} + \iota^{I}_{ij,t} + \alpha^{I} \ln \text{workers}_{ik,t-1} + \epsilon^{I}_{ijk,t}, \tag{3}$$

$$\ln \text{workers}_{ijk,t} = \iota^{W}_{jk,t} + \iota^{W}_{ij,t} + \alpha^{W} \ln \text{workers}_{ik,t-1} + \epsilon^{W}_{ijk,t}, \tag{4}$$

[15] Further, fixed effect specifications that exploit longitudinal variation can magnify the attenuation bias coming from measurement error, when the explanatory variable is strongly correlated over time (Griliches and Hausman, 1986; Bound and Krueger, 1991).

[16] If we include state-crop fixed effects into our IV strategy, the first-stage F-statistic drops to 2.8, which renders 2SLS estimates meaningless. This is because the instrument is also quite correlated over time, not because exclusion restriction is violated. In fact, separately regressing migrant composition and our instrument on state-crop fixed effects has an R2 of 0.7 and 0.95.

where i indexes the origin region, j the destination region, k the agricultural activity and t the year, while $\text{income}_{ijk,t}$ and $\text{workers}_{ijk,t}$ denote average income and number of workers in each cell. We use meso-regions in the regressions, and to gain precision, we stack data for the years $t = 2000$ and $t = 2010$. With an eye toward our modeling strategy later, we use a thirty-year lag as our measure of $\text{workers}_{ik,t-1}$.[17] To avoid including the same farmer in both sides of the equation, we exclude non-migrants from our sample.

Table 3: The Relation between Farmers' Income, Choices, and Employment in the Origin

	OLS (1)	OLS (2)	OLS (3)	PPML (4)	PPML (5)	2SLS (6)	2SLS (7)	2SLS (8)
a. Income (logs)								
Farmers in origin	0.026**	0.048***	0.070	0.047***	0.073***	0.051***	0.106**	0.059***
	(0.011)	(0.017)	(0.085)	(0.016)	(0.025)	(0.015)	(0.051)	(0.016)
R^2 or K-P	0.699	0.724	0.839	0.338	0.257	24.393	21.202	14.250
Overid. p								0.277
Obs	6778	5262	5262	6778	5262	6778	6778	6778
b. Farmers in destination (logs)								
Farmers in origin	0.084***	0.100***	0.095	0.129***	0.130***	0.133***	0.232**	0.147***
	(0.016)	(0.022)	(0.077)	(0.021)	(0.032)	(0.034)	(0.094)	(0.033)
R^2 or K-P	0.755	0.776	0.935	0.784	0.801	24.393	21.202	14.250
Overid. p								0.269
Obs	6778	5262	5262	6778	5262	6778	6778	6778
Dest/Act/Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Dest/Orig/Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Dest/Orig/Act FE	-	-	Y	-	-	-	-	-
Above Q1	-	Y	Y	-	Y	-	-	-
Crop Suitability	-	-	-	-	-	Y	Y	Y
Harris's IV	-	-	-	-	-	Y	-	Y
Mig. Comp. IV	-	-	-	-	-	-	Y	Y

Notes: * / ** / *** denotes significance at the 10 / 5 / 1 percent level. Regressions at the meso-region level. Standard errors, multiway clustered at the destination-crop-year and origin-year level, are in parenthesis. The unit of observation is a origin-destination-activity-year cell. Columns (6) to (8) control for suitability measures based on FAO-GAEZ at the origin and activity. Column (6) uses as an instrument the access of an origin-activity pair to workers producing that same activity based on $\ln \sum_{j \neq i} \mu_{ji,t-1}^{-1} \text{workers}_{jk,t-1}$, where $\mu_{ij,t-1}$ is the migration cost between regions i and j. Column (7) uses as an instrument the access of a meso-region to origin-activity of state-migrants $\ln \sum_{s' \neq s} \left(\widehat{M}_{s's,t} / \widehat{M}_{s,t} \right) S_{s',k}$. Column (8) uses both instruments and presents the Hansen over-identification p-value. All regressions exclude non-migrants.

OLS results. Panel A in Table 3 shows estimates of α^I using our detailed census sample. In our baseline specification, 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 0.026 percent (Column 1). Since origin-destination-

[17] 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 worker's current and previous meso-region of residence. We discuss alternative measures of migration in Appendix Section OC, where we run the same regressions at the state level, since we do observe the state of birth.

crop cells containing zero workers are frequent in our sample, we explore three additional specifications. In Column (2) we drop observations in the bottom quartile of the distribution of workers$_{ijk,t-1}$. As one would expect if low values of our regressor make sampling zeros more likely, this specification increases our coefficients to 0.048. We repeat these estimations with a PPML estimator, to avoid the bias due to heteroskedasticity (Silva and Tenreyro, 2006). Columns (4) and (5) show that doing so increases our estimates to 0.047 and 0.073.

In Panel B, the elasticity of migrants' employment with respect to the number of farmers in the origin, α^W, rises from 0.084 to 0.130 across the same specifications. 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.

Our chosen specification will allow us to give α^W and α^I precise structural interpretations through the lens of our model. The rest of this section argues that we can also give our estimates a causal interpretation, and it shows that our conclusions are largely robust to alternative econometric specifications and robustness checks. We also explore empirically other observable implications of worker sorting according to comparative advantage. Appendices B and OC collect the details of each exercise below.

The role of fixed effects in identification. In equation (3), α^I measures the association between a farmer's comparative advantage and the size of the workforce in his origin region. In equation (4), α^W captures how this comparative advantage translates into activity choice. By including destination-activity fixed effects ($\iota^W_{jk,t}$ and $\iota^I_{jk,t}$), 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, any local Marshallian externality, such as knowledge spillovers from other workers, and any local investments in technology and machinery (which may follow from increased employment). By including origin-destination fixed effects ($\iota^W_{ij,t}$ and $\iota^I_{ij,t}$), we control for common factors that are origin-destination but not activity-specific, such as access to education and financial capital, bilateral migration costs, total agricultural activity in the origin, and weather similarity, as well as average differences in human capital or other sources of productivity for all workers from the same origin. To give an example, we identify worker heterogeneity from the fact that, facing the same local conditions in j = Mato Grosso, high-earnings workers in k = soy come disproportionately from i = Rio Grande do Sul where, 30 years earlier, soy employment was high relative to other regions and other crops.

In column (3), we add origin-destination-crop fixed effects to estimate our coefficient using only within-group variation. Because our point estimates remain quite similar, but our standard errors increase by a factor of five (so we are unable to reject that the coefficient is different from zero), we conclude that the more demanding fixed-effects specification eliminates variation that is useful to gain precision, making our other specifications preferable. As with regression (2), the new fixed effects preclude us from deploying our IV strategy below, which incorporates cross-sectional variation to achieve identification.

IV results. While our fixed-effects strategy controls already for several confounding factors, a potential concern is that other unobserved, origin-activity factors might simultaneously increase both employment in certain origin-activity pairs (thirty years ago) and the productivity of workers born in those regions in their new destinations. This would lead to biased estimates of α^I and α^W.

We therefore construct two instruments that generate plausible exogenous variation in employment at the origin-activity pair. Our first instrument adapts Harris's measure of goods market access, to measure instead access to workers from other regions (Harris, 1954; Donaldson and Hornbeck, 2016). We construct $IV^1_{ik,t-1} = \ln \sum_{j \neq i} \left[\mu^{-1}_{ji,t-1} \text{workers}_{jk,t-1}\right]$, where $\mu_{ij,t-1}$ measures migration frictions from i to j.[18] This instrument shifts the number of workers in an origin-activity (i,k) based exclusively on i's proximity to workers k in other regions. To ensure that this shifter is uncorrelated with productivity shocks at (i,k), we add controls for crop suitability at (i,k) and we exclude from the instrument regions j at different distances from i.

Our second instrument adapts the same approach we used for Fact 2. We construct a measure of migrants' composition $IV^2_{ik,t-1} = \ln \left(\sum_{s' \neq s(i)} \frac{\widehat{M}_{s's(i),t-1}}{\widehat{M}_{s(i),t-1}} S_{s'k} \right)$, where $s(i)$ denotes meso-region i's state. This instrument shifts employment in origin-activity (i,k) using historical coincidences in push factors that are unrelated to destination region i's productivity (as captured in $\hat{M}_{s's,t}$ and $\hat{M}_{s,t}$, constructed following Burchardi et al., 2019). This employment shift is larger for an activity k in which origin s' is exogenously more productive, consistent with employment being larger there.

Column (6) in Table 3 shows that the Harris instrument yields similar results to OLS. Column (7) shows that the push-pull instrument yields larger, albeit substantially more imprecise, coefficients. In Column (8) we combine the instruments and obtain similar results, with a coefficient of 0.060. Moreover, we cannot reject the hypothesis that both instruments are valid. Across samples, the OLS and IV coefficients are quite similar, which leads us to conclude that, to the extent that endogeneity exists, its impact on our estimation is modest.

Appendix Table C.3 presents a set of robustness tests, all of which are associated with

[18] We set $\mu_{ij,t} = \delta_0 d^{\delta_1}_{ij,t}$ where $d_{ij,t}$ is the travel distance in time given by the network of highways available at time t and we set $\delta_1 = 0.05$ based on Bryan and Morten (2019). The choice of δ_0 is immaterial due to the log specification. We experiment with larger values of δ_1 in Appendix B and report results in Table C.3.

the construction of the Harris instrument. One potential concern is that unobserved shocks are correlated geographically between s and s', which would violate our identification assumption. We address this concern by evaluating three other approaches to the construction of the instrument: (i) we include meso-regions within 1 hour drive from the origin (excluded in our baseline), (ii) we exclude meso-regions that are within 3 hours drive from the origin, and (iii) we construct an alternative measure of $\mu_{ij,t}$ with smaller migration frictions, which gives a more prominent role to faraway regions. Our point estimates are comparable across these alternative specifications.

Alternative measures of economic activity in the origin. We next provide results strongly indicative that employment is the best measure, in a statistical sense, of origin characteristics. Together with origin-activity employment, we include, one at a time, total revenues, total incomes, total land use, and an aggregation of workers in broad categories of crops, as alternative measures of economic activity, and output per worker as a measure of worker productivity in the origin. Appendix Tables C.1 and C.2 show that the association with origin-activity employment persists when we add these other measures as independent variables. Our estimates of the impact of employment are similar across specifications, whereas the point estimates for the impact of each of these alternative measures is always close to zero. In our quantitative model, we build on these results and posit that origin employment drives the learning of migrants.

RAIS data. To bolster our results, Online Appendix Table O.13 shows estimates of equations (3) and (4) using the RAIS data set for the same specifications reported in Table 3. Across the board, we obtain similar results to those using the Census, although the new estimates of α^I are smaller, and those of α^W larger. It is reassuring that we find similar OLS and IV results, even though direct comparisons are difficult because the timespan of RAIS is shorter and it includes only formal workers, its crop coverage is different, and its measure of income includes only wages. Online Appendix Table O.14 shows that results using this sample are also robust to alternative construction of the Harris instrument and Online Appendix Tables O.17 to O.20 include alternative measures of economic activity in the origin.

Robustness to measurement, sample choices, and alternative interpretations After the large battery of tests summarized next and discussed in Online Appendix OC, our conclusions about the relationship between migrants' heterogeneity and activity employment at the origin remain unchanged.

First, we use alternative definitions of migration based on the state of birth, rather than the previous meso-region, to ensure that neither the geographic unit of analysis nor the definition of migration drive our results (Online Appendix Table O.21).

Second, we find that no single crop drives our results, as we estimate our equations dropping crops one at a time (Online Appendix Table O.22). This is true, in particular, even if we drop the most capital-intensive activities, which rules out the hypothesis that we are capturing differential access to capital. Importantly, the results are unchanged when we drop soy, a crop that responds strongly to migration in our quantitative exercises.

Third, to address the possibility of selection bias, we conduct two tests using the data at the worker level (Appendix Table C.4). First, we control for a host of socio-economic and demographic characteristics using our individual level data to capture differences in the composition of workers. Second, following Dahl's (2002) approach to control for selection bias, Appendix Table C.4 shows our earnings equation using that same data at the individual level, but augmenting it to include a subset of the choice probabilities of individuals from origin i across options (j, k).

Fourth, we control for proxies of previous networks of migrants and show that our results are not driven by workers inserted in social networks that allow them to gain access to better wages (Online Appendix Table O.23). Alternatively, to study the role of first-mover advantage, we include a full set of fixed effects for the years of residence in the destination in our individual-level regressions (Online Appendix Table C.4). Our results remain. We emphasize, however, that if workers were fundamentally homogeneous and only differed in terms of their access to better wages, then we would not have found an impact of workers' composition on productivity, as Fact 2 shows.[19]

Fifth, using RAIS data, where we can track workers over time, our results remain when we control for previous experience in activity k in the previous ten years (Online Appendix Table O.24).

Sixth, we examine whether the results that we find are also present in the other sectors of the economy and we find that, albeit at a coarser level of aggregation, our results also carry over to manufacturing sectors (Online Appendix Table O.25).[20]

[19] This evidence is also inconsistent with a model where workers are heterogeneous but cannot between sectors. Online Appendix OH.2 shows that a version of our model in which workers cannot switch jobs tends to generate α^W equal to 1 in equation (4) and α^I equal to 0 in equation (3).

[20] Throughout the paper, we focus on knowledge heterogeneity within the agricultural sector. We do so for several reasons, including the outsize importance of Brazil's agriculture in the world. Most importantly, one key advantage of the agricultural sector is that we can construct quite granular sectoral definitions within agriculture, which allows us to compare workers within narrowly defined employment categories, as well as the evolution of revenues for these subsectors across space. For manufacturing, on the contrary, the sectoral classification is quite broad, especially if one is to create consistent categories across years and across space.

Eighth, we show non-parametrically the variation that identifies α^I and α^W. In both cases, a log-linear relationship provides a reasonable approximation (Online Appendix Figure O.1).

Additional evidence consistent with worker heterogeneity. We conclude this section with two additional empirical facts that are consistent with workers' knowledge driving their comparative advantage. First, we observe larger flows of migrants between origins and destination whose agricultural suitability is more similar, according to an index akin to that in Bazzi et al. (2016), even after controlling for bilateral distance. This fact suggests that workers from origins that are specialized in a given activity are more knowledgeable about that activity and migrate accordingly (see Appendix B and Appendix Table C.5).

Second, we provide suggestive evidence consistent with heterogeneous workers responding differently to international trade shocks. To do so, we track the income of workers born before the 1980s in regions that were specialized in soybeans—a new comparative-advantage crop—and coffee—a more traditional crop—, and we study how such workers fared following Brazil's trade liberalization in the 1990s (Dix-Carneiro and Kovak, 2017) and the rise of China (Costa et al., 2016). After the 1990s, workers' earnings in soy grew at a higher rate if those workers were born in soy producing regions. Conversely, the incomes of workers born in regions specialized in coffee fell relative to that of other coffee workers. These results suggest that skill heterogeneity, linked to the region of birth, has lifelong consequences for workers and induces differential exposure to international trade (Appendix Figure C.1).

In what follows, 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 that depends on region-activity employment.

5 A Model of Migration and Comparative Advantage

We develop a model of trade and migration, in which comparative advantage is driven by regional productivity and the activity-specific knowledge of workers. We use the model to obtain, first, analytical results about the relation between geography and knowledge accumulation and, second, to derive a simple expression based on observables that will guide our quantitative analysis. To simplify our exposition, we present a stripped down version of the model. Online Appendices OE-OG contain details and proofs.

5.1 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 world composite, denoted by F. There are $k = 1, \ldots, K$ goods (or activities), and each region produces a unique variety of each. At each time, the geography of the economy, Γ_t, is given by a set of exogenous natural advantages, iceberg trade costs, migration costs, and a land endowment: $\Gamma_t = \left\{\bar{A}_{jk,t}, \tau_{ijk,t}, f_{ijk,t}, H_{j,t}\right\}$.

Demography and preferences. People live two periods, young and old. An adult at time t, upon observing his knowledge, chooses an activity-location pair and then has one child. Let $L_{j,t}$ denote the adult population at time t in region j.

An adult born in $t-1$ obtains utility from his household's consumption and the future welfare of his children. In particular, the welfare of a parent who moved from i to work on j, k (net of moving costs) is:

$$V_{ijk,t} = u\left(W_{ijk,t}\right) + \delta \mathbb{E}\left[\max_{j'k'}\left\{V_{jj'k',t+1} - f_{jj'k',t+1} + \varepsilon_{jj'k',t+1}/\theta\right\}\right], \tag{5}$$

In equation (5), the first term denotes the adult's indirect utility (which incorporates children's consumption), and it is given by

$$W_{ijk,t} = \frac{w_{jk,t}s_{ik,t}}{P_{j,t}}, \tag{6}$$

where $w_{jk,t}$ is the wage per efficiency unit of labor, and $P_{j,t}$ is the price index of final output consumption in destination j (defined below). Reflecting our empirical findings, $s_{ik,t}$ is the worker's knowledge to produce in activity k, which depends on the region he comes from. Workers can migrate within the Home country, but not between Home and Foreign. The second term in equation (5) denotes the expected welfare of the child born in j, where $f_{jj'k',t+1}$ is the child's utility migration cost.[21] The parameter δ measures the relative importance of the child's utility, and the preference shocks ε are drawn i.i.d across locations and activities from a Type-I Extreme Value distribution with zero mean (see Allen and Donaldson, 2022)[22]

[21] We adopt this two-period OLG formulation for two main reasons. First, using RAIS data Online Appendix Table O.29 shows that Brazilian workers spend more than 90 percent of their previous 10 years of employment working on the same crop. Second, introducing multiple periods of employment would increase substantially the state space, because a workers' migration and occupation history determines his knowledge. We thus simplify by assuming knowledge is only acquired once, in the location of birth.

[22] We introduce preference shocks, as opposed to productivity shocks, since they readily deliver the estimating equations (3) and (4). This formulation also recognizes that workers migrate for many reasons beyond productivity differentials.

The adult born in i chooses destination j and activity k to solve

$$\max_{j,k} \{V_{ijk,t} - f_{ijk,t} + \varepsilon_{ijk,t}/\theta\}.$$

Further specializing the period return function to logarithms, the value function becomes

$$V_{ijk,t} = \ln W_{ijk,t} + \delta \Upsilon_{j,t+1} \tag{7}$$

and, using our distributional assumptions, the expectation in (5) becomes

$$\Upsilon_{i,t} = \frac{1}{\theta} \ln \left(\sum_{j,k} \exp\left(\theta\left(\ln W_{ijk,t} - f_{ijk,t} + \delta \Upsilon_{j,t+1}\right)\right) \right). \tag{8}$$

Knowledge endowment. Motivated by Fact 3, we assume a worker's knowledge depends on the employment structure of the region where he is born. Specifically, a child born at $t-1$ is characterized by a vector of activity-specific productivities, $s_{ik,t}$, according to:[23]

$$s_{ik,t} = \bar{s}_k \ell_{ik,t-1}^{\beta}, \tag{9}$$

where $\ell_{ik,t-1}$ is the number of workers in (i,k) in the previous period and $\bar{s}_k$ is a crop-specific knowledge term.[24]

Technology. The cost of producing one unit of good k in region j is given by

$$\frac{c_{jk,t}}{A_{jk,t}}. \tag{10}$$

[23]This formulation captures in a transparent way the relation between worker heterogeneity and migrant origin from Fact 3. Since our regressions include origin-destination fixed effects, we cannot determine if our parameterization of $s(\ell_{ik,t-1})$ 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 (Ramondo et al., 2016). Online Appendix Figure C.3 shows that our quantitative results remain almost unchanged when considering an alternative formulation in which knowledge depends on agricultural employment shares, $s_{i,k} = (\ell_{ik,t-1}/L_{i,t-1})^{\beta}$. Online Appendix OI offers a microeconomic foundation for this formulation, based on a meeting and learning process between young and old.

[24]Online Appendix OJ discusses an extension of our model in which workers learn both via externalities and through their parents. We find that, without the learning externality, the parental-learning model can replicate our reduced-form results only by generating patterns of intergenerational occupational mobility that are inconsistent with previous findings in the literature. The reason is that, to rationalize the empirical coefficients in equation (3), the model gives the children of parents in activity k too large a productivity premium. In addition, the parental-learning model suggests that our reduced-form results might be driven by the composition of workers in the origin—i.e., a larger proportion of high-knowledge parents. We find, however, that our results are unaffected by the inclusion of controls for worker productivity in the origin (Online Appendix Table O.26).

In expression (10), $c_{jk,t}$ is the unit cost of the input bundle:

$$c_{jk,t} = \bar{c}_k \left(w_{jk,t}^{1-\gamma_k} r_{j,t}^{\gamma_k} \right)^{\alpha_k} P_{j,t}^{(1-\alpha)}, \tag{11}$$

where $w_{j,kt}$ are efficiency wages, $r_{j,t}$ are land rents, $\bar{c}_k$ is a technology constant, and γ_k and α_k are the share of land and the share of value added. Regional productivity $A_{jk,t}$ reflects exogenous natural advantage and agglomeration forces,

$$A_{jk,t} = \bar{A}_{jk,t} \ell_{jk,t}^{\xi}, \tag{12}$$

where $\ell_{jk,t}$ is the total employment in region j, good k.

In each region j, a representative firm aggregates all regional varieties of good k into aggregate output, with constant elasticity of substitution η_k. Good k output is likewise aggregated into final output, with constant expenditure shares a_k. Only varieties are traded. Therefore, the price index for good k in region j is given by

$$p_{jk,t}^{1-\eta_k} = \sum_{i \in \mathcal{W}} \left(\tau_{ijk,t} c_{ik,t} / A_{ik,t} \right)^{1-\eta_k}, \tag{13}$$

while that of final output is

$$P_{j,t} = \prod_{k=1}^{K} p_{jk,t}^{a_k}. \tag{14}$$

5.2 Equilibrium

Optimal worker sorting gives the share of workers from i choosing to work in region j and activity k:

$$\lambda_{ijk,t} = \left(\frac{W_{ijk,t} / \mu_{ijk,t} \cdot \upsilon_{j,t+1}^{\delta}}{\upsilon_{i,t}} \right)^{\theta}, \tag{15}$$

where we define $\mu_{ijk,t} \equiv \exp\left(f_{ijk,t}\right)$ and $\upsilon_{i,t} \equiv \exp\left(\Upsilon_{i,t}\right)$ to simplify notation. We define the effective supply of labor to region j, activity k as

$$E_{jk,t} \equiv \sum_{i} s_{ik,t} \lambda_{ijk,t} L_{i,t-1}, \tag{16}$$

whereas the supply of raw labor is given by

$$\ell_{jk,t} = \sum_{i} \lambda_{ijk,t} L_{i,t-1}. \tag{17}$$

To close the model, we assume that land rents are paid to local landowners.

The share of region $j's$ sector k expenditure going to origin i is given by

$$\pi_{ijk,t} = \left(\frac{c_{ik,t} \tau_{ijk,t} / A_{ik,t}}{P_{jk,t}} \right)^{1-\eta_k} \tag{18}$$

Total expenditure in region j, $X_{j,t}$, reflects final and intermediate expenditure:

$$X_{j,t} = \sum_k w_{jk,t} E_{jk,t} + r_{j,t} H_{j,t} + \sum_{k=1}^{K} (1 - \alpha_k) Y_{jk,t}, \tag{19}$$

where we impose trade balance and $Y_{jk,t}$ denotes revenues in activity k,

$$Y_{jk,t} = \sum_i \pi_{jik,t} a_k X_{i,t}. \tag{20}$$

Labor market clearing requires

$$w_{ik,t} E_{ik,t} = \alpha_k (1 - \gamma_k) Y_{ik,t}, \tag{21}$$

while land market clearing requires

$$r_{i,t} H_{i,t} = \sum_k \alpha_k \gamma_k Y_{ik,t} \tag{22}$$

Definition 1. (Equilibrium) Given a sequence of geographies $\{\Gamma_t\}$ and initial labor allocations, $\{\ell_{ik,0}\}_{ik}$, a general equilibrium is a sequence of factor prices and labor allocations $\{r_{j,t}, w_{jk,t}, E_{jk,t}\}_{jk,t}$, and expected welfare $\{\Upsilon_{j,t}\}$, such that, for each region j and activity k and time t: (i) Workers choosing to migrate to (j, k) satisfy (15) given (9) and factor prices, (ii) the market for efficiency units of labor clears, (21) and (16), (iii) land markets clear for region j, (22).

The characterization of dynamic spatial equilibria is a current area of work (see e.g. Allen and Donaldson, 2022; Desmet et al., 2018; Kleinman et al., 2023), and to the best of our knowledge, there are no general sufficient conditions that guarantee uniqueness in models with many sectors, like ours.

Nevertheless, building on existing work and focusing on the steady state, we offer a partial characterization. A difference relative to existing work is that in our model knowledge externalities increase the effective supply of workers and are, therefore, distinct from Marshallian externalities or scale effects in models of imperfect competition, which instead raise the demand for labor (Bartelme et al., 2019b; Kucheryavyy et al., 2023). Online Appendix OF.1 shows this clearly, by mapping a one-sector version of our model to the framework of Allen and Donaldson (2022). We also show that in this one-sector setting, absent migration costs, our knowledge mechanism does not affect the multiplicity properties of the model.

Online Appendix OF.2 characterizes instead the steady state in a two-sector, one-region model. With more than one sector, multiplicity can arise either because of agglomeration forces (as usual) or knowledge externalities but, once again, operating differently through supply and demand of labor. We also construct a simple example in which the knowledge mechanism produces multiple equilibria, where workers could specialize in an activity in which exogenous productivity is lower, so that the economy might fall into a poverty trap.

5.3 Geography, skill accumulation, and comparative advantage

This section briefly summarizes two key results whose full exposition we relegate to Online Appendix OF.3. Our first result is that population growth can act as an engine of growth in our model. In the presence of our knowledge mechanism and agglomeration externalities, a steady state with larger population can feature higher levels of real wages and higher welfare. As Appendix Proposition 2 shows, the strength of these effects increases in β, which governs the magnitude of the knowledge mechanism.[25] The result is reminiscent of endogenous growth models with weak scale effects, in which population growth is necessary to sustain economic growth. Importantly, geography has no impact on the distribution of these gains, which are equal across regions.

Our second result shows that, at any date t, the inflow of knowledge through migration into region i is independent of region i's characteristics, such as real wages and welfare, but does depend on bilateral migration costs: Proximity to regions with abundant production in an activity k will improve the average knowledge of region i in that activity. In fact, absent migration costs, all regions have exactly the same access to knowledge through migration.

5.4 A Guide to our Quantitative Results

We now introduce a result to aid the interpretation of our quantitative results in Section 7.

Proposition 1. *Assume that Home is a small open economy, and that $\xi = 0$, $\eta_k = \eta$, $\gamma_k = \gamma$, $\alpha_k = \alpha$, $\forall k$. Suppose migration to region i becomes prohibitively costly, $\mu'_{ijk,\tau} \to \infty$, $\forall j \neq i$,*

[25] Because land supply is not perfectly elastic, a larger population can instead lead to lower welfare. Proposition 2 shows this happens if land intensity in production is large enough.

at all dates $\tau \geq t$. Then the change in specialization at time t is given by:

$$\frac{\widehat{X_{iFk,t}/X_{iFm,t}}}{X_{FFk,t}/X_{FFm,t}} = \left(\mathcal{E}_{iik,t}/\mathcal{E}_{iim,t}\right)^{\frac{\alpha(1-\gamma)(\eta-1)}{1+\theta+\alpha(1-\gamma)(\eta-1)}}, \tag{23}$$

where $\mathcal{E}_{ijk,t}$ is the share of i's workers in j's effective labor force in activity k, at date t, i.e., $\mathcal{E}_{ijk,t} \equiv E_{ijk,t}/E_{jk,t}$, and $\hat{z} \equiv z'/z$ for variable z.

The proposition shows that, under some simplifying assumptions, the relative share of domestic workers in total effective labor supply contains information on how relative marginal costs respond to limiting migration.

There are two reasons why exposure $\mathcal{E}_{ijk,t}$ varies across crops and regions in our setting. The first is worker heterogeneity, $s_{ik,t}$, which makes workers from i sort differentially across activities k within destination j. 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.

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 in Section 7, which quantitatively evaluates the impact of observed changes in migration costs.

Online Appendix OG extends this result to the case where agglomeration economies are present, $\xi > 0$. Two key differences emerge. First, the wage increase to a migration shock is somewhat mitigated, because labor demand is flatter in the presence of externalities. This is the same channel discussed in Proposition 1. Second, the same migration shock directly affects comparative advantage because agglomeration forces depend directly on the number of workers. This effect is additional to what Proposition 1 states.

6 Taking the Model to Brazilian Data

We now take our model to Brazilian data, to simulate the impact of westward migration and the policies that drove it. We map the model to Brazil's economy in 1950, 1980, and 2010, 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. Online Appendix OL describes our quantitative model and Appendix OM details our calibration procedure.

6.1 The Quantitative Model

Relative to the model presented in the last section, we add five 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; (iii) land endogenously supplied by local governments that use a decreasing returns technology that requires final output—with a productivity $g_{j,t}$ and an elasticity of land supply to land rents ζ—, where profits from land development are rebated to farmers proportionally to their wages; (iv) endogenous amenities given by $u_{i,t} = \bar{u}_{i,t} L_{i,t}^{-\chi}$; and (v) Brazil's population growth relative to the rest of the world.

6.2 Calibration

We provide below an overview of how we calibrate (i) worker heterogeneity, (ii) trade costs, (iii) technology and preferences, and (iv) migration costs.

Worker heterogeneity. The preference dispersion parameter, θ, and the worker productivity parameter, β, map to the reduced-form elasticities we estimated in Fact 3. Using equation (9), our model relates the income of migrants, $w_{jk,t} \times s_{ik,t}$, to the employment in the origin as follows:

$$\ln\left(\text{income}_{ijk,t}\right) = \iota^I_{jk,t} + \iota^I_{ij,t} + \beta \ln \ell_{ik,t-1} + u^I_{ijk,t}. \tag{24}$$

To motivate the error $u^I_{ijk,t}$, we posit measurement error in income.

Analogously, we substitute equation (24) into (15) to obtain our regression for activity choice:

$$\ln \ell_{ijk,t} = \iota^L_{jk,t} + \iota^L_{ij,t} + \theta\beta \ln \ell_{ik,t-1} + u^L_{ijk,t}, \tag{25}$$

where $u^L_{ijk,t}$ 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). Appendix OH.1 contains the derivations of equations (24) and (25), and it explains what each fixed effect captures. The appendix also relates this formulation to the assumptions needed to ensure identification both via OLS and IV.

Our model thus gives a structural interpretation to Fact 3 and shows that the estimated coefficient α^I in regression (3) measures β, while coefficient α^W in regression (4) measures $\theta\beta$. A model in which workers' knowledge is independent of their origin would be unable to speak to the reduced-form coefficients in Fact 3.

Our estimates for β range between 0.026 and 0.105, and the implied values for θ range between 2.08 and 3.23. Using different strategies, a few recent papers have estimated similar values of θ, which controls the elasticity of migration with respect to the value of different region-activity pairs. 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.[26] Since β is new to our theory, there is no direct benchmark with which to compare it. But our results are similar to those of sectoral scale economies (Antweiler and Trefler, 2002; Bartelme et al., 2019b) and of the effect of city size on productivity (de la Roca and Puga, 2017). Given our range of estimates, we set $\theta = 2.5$ and $\beta = 0.05$, to be conservative about our conclusions.

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

$$\tau_{ijk,t} = \delta_t^0 \left[\left(dist_{ij,t} \right)^{\delta_t^1} \right]^{\iota_{ij}^T} \left[\delta_{kt} \left(dport_{i,t} \right)^{\delta_t^1} \left(dport_{j,t} \right)^{\delta_t^1} \right]^{1-\iota_{ij}^T}, \tag{26}$$

where ι_{ij}^T is a dummy variable that equals one if i and j belong to same country and zero otherwise, $dist_{ij,t}$ is the travel distance in hours between i and j, and $dport_{i,t}$ is the minimum travel distance to the nearest port (for $j = F$, we set $dport_j = 1$). All distances are measured according to the highway network available at time t using the Fast Marching Method (Online Appendix OA).

We calibrate δ_t^0 to match the observed share of intra-regional trade in total domestic trade in Brazil. Specifically, 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 ratio of 0.7 for 1950, 0.60 for 1980, and 0.50 for 2010, which gives δ_t^0 of 3.71, 1.90 and 1.77 respectively. We pick δ_t^1 to match 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 gives δ_t^1 of 0.08, 0.07, and 0.05 respectively. Lastly, we calibrate δ_{kt} to match Brazil's trade with the rest of the world.

[26] We effectively impose the restriction that the elasticity of migration across sectors is the same as across regions. As observed by Borusyak et al. (2022), papers using variation across sectors, such as Artuç et al. (2010), find similar elasticities to those in papers using variation across regions, such as Caliendo et al. (2019). In our case, we exploit variation across both sectors and regions, and our preferred point estimates are in line with the previous literature.

Technology, preferences, and amenities. 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 the land share to zero. 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). We set the cost share of land to 0.21 for all agricultural activities based on Dias Avila and Evenson (2010). Lastly, we set the land supply elasticity ζ to 1.5 and, as in Desmet et al. (2018), we pick $\xi = 0.06$ and $\chi = 0.32$.[27]

Next, we set σ to 0.4 based on Comin et al. (2021), σ_A to 2.5 according to Sotelo (2020), η_k in agriculture to 9.5, and 5.5 for other sectors following Caliendo and Parro (2015). These elasticities allow for structural transformation between broad sectors in response to changes in sectoral prices. We set the discount rate $\delta = 0.172 = (1/1.0603)^{30}$, consistent with Allen and Donaldson (2022).

Having calibrated trade costs, technology, and preferences, we follow the model-inversion logic laid out by Allen and Arkolakis (2014) and calibrate $\bar{A}_{jk,t}$ to match observed gross output, $\bar{u}_{i,t}$ to match the total population in every meso-region, $g_{j,t}$ to match observed agricultural land use, and country-level sectoral preference shifters to match country-level apparent consumption by activity.

Migration Costs. With an eye toward counterfactual analysis, we extract a state-state component from migration costs, $\mu_{ss',t}$, and break down the remainder into a state-state-activity component, $\mu_{ss'k,t}$ and a geographic component that operates across meso-regions. Formally, we assume that $\mu_{ijk,t} = 1$ for $i = j$ and, for $i \neq j$, that

$$\mu_{ijk,t} = \mu_t^0 \left[(dist_{ij,t})^{\mu^1} \right]^{\iota_{ij}^M} \left[\mu_{ss',t}\mu_{ss'k,t} \left(dcap_{i,t}\right)^{\mu^1} \left(dcap_{j,t}\right)^{\mu^1} \right]^{1-\iota_{ij}^M} \tag{27}$$

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 $dcap_{i,t}$ is the travel distance between region i and the state capital of i. We thus assume workers travel through the capital to access meso-regions in other states.

To recover $\mu_{ss',t}$, based on our parameterization in equation (27), we use OLS to estimate:

$$\ln\left(L_{ss'k,t}\right) = \alpha_{s,t} + \alpha_{s'k,t} + \tilde{\mu}_{ss',t} + \epsilon_{ss'k,t}, \tag{28}$$

[27] 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 (2021).

where $\alpha_{s,t}$ and $\alpha_{s'k,t}$ are origin and destination-activity fixed effects and $\epsilon_{ss'k,t}$ is an error term. We recover $\mu_{ss',t}$ from $\tilde{\mu}_{ss',t} \equiv -\theta \ln(\mu_{ss',t})$.[28] Given our estimates of $\mu_{ss',t}$, we set μ_t^0 to match the share of workers living in their meso-region of birth and $\mu_{ss'k,t}$ to match the migration of workers between states and activities. The k dimension in $\mu_{ss'k,t}$ captures differences in labor allocations across activities driven by factors other than wages, such as misallocation. We set μ^1, the elasticity of migration cost with respect to distance, to 0.05 based on Bryan and Morten (2019).

6.3 The March to the West as seen Through our Model

Table 4 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 investments in transportation infrastructure. Migration costs from the East to the West, in particular, declined by a half. To benchmark our migration costs with previous literature, we note that our estimates are comparable to the ones obtained by Tombe and Zhu (2019), who find an overall migration cost between provinces in China of 25 circa 2000.[29]

Panel (b) shows that productivities, $\bar{A}_{ik,t}$, in the West rose sharply relative to the East, especially in agriculture. For soybeans, specifically, the upward trend in relative productivity is in line with the EMBRAPA's and other research efforts, started in the 1970s, to increase soybean productivity in the West, as discussed in Bustos et al. (2016) and Pellegrina (2022). Panel (b) also indicates that the productivity of the land supply sector $g_{j,t}$ increased in the West relative to the East, consistent with the government's increasing efforts to facilitate land settlement and acquisition in the West via the CANs and INCRA.

Panel (c) shows that our calibration captures Brazil's increasing trade openness through a reduction of international trade costs, consistent with the evidence reported in Dix-Carneiro and Kovak (2017). Domestic trade costs also declined, in line with the transportation policies that fostered East-West trade integration.

[28] Online Appendix OM.1 shows that the hub and spoke formulation in equation (27) allows us to aggregate migration flows and run gravity equations at the state level in a theoretically consistent manner. Since the level of $\tilde{\mu}_{ss',t}$ is not identified from equation (28), we normalize $\tilde{\mu}_{ss',t} = 1$ for $s = s'$. Note also that, because we use data at a higher level of aggregation, we are less likely to incorrectly infer infinite migration costs from sampling zeros.

[29] We also regress employment shares, at the meso-region level, in region j activity k which are not targeted in our calibration procedure, against their model-implied counterpart and find a coefficient of 0.98 (and a R^2 of 0.92) in 2010.

Table 4: Description of the Parameters Recovered in the Calibration

	1950 (1)	1980 (2)	2010 (3)
a. Migration costs			
Average migration costs	9.54	7.15	6.52
Migration costs between states in the East and in the West	20.27	11.33	9.79
Elast. of migration costs w.r.t. travel distance	0.44	0.34	0.29
b. Productivity			
Productivity in manufacturing in the West relative to the East	0.58	0.85	0.88
Productivity in agriculture in the West relative to the East	0.95	1.10	1.44
- Soybeans	0.18	2.32	1.96
- Livestock	0.91	1.23	1.36
- Corn	1.25	1.18	2.59
Productivity of land supply in the West relative to the East	0.45	0.80	2.97
c. Trade costs			
Trade cost between Brazil and RoW - manufacturing	5.36	3.70	2.98
Trade cost between Brazil and RoW - agriculture	12.89	4.76	3.40
Elast. of trade cost w.r.t. travel distance	0.07	0.07	0.09

Notes: This table shows results from the calibration of the model. Migration costs are the harmonic average across state pairs. 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.

Put together, these results paint a clear picture of the forces that produced the March. Starting in the 1950s, a series of shocks and government interventions dramatically increased the West's agricultural productivity across the board, especially in new activities such as soybean, livestock, 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.

6.4 The Quantitative Importance of Knowledge

Having calibrated our model, we close this section by establishing the quantitative importance of worker knowledge heterogeneity in generating migration and comparative advantage, both in the short and the long run.

Reducing the portability of knowledge. In our model, the ability to bring knowledge fosters migration. To make this point clear, we study two counterfactual scenarios in which we reduce the portability of knowledge from the East to the West. We set knowledge according to $\tilde{s}_{ijk,t} = \phi_{ijk} \times s_{ik,t}$, where $\phi_{ijk} = \bar{\phi}_k < 1$ when i is in the East and j in the West, and $\phi_{ijk} = 1$ otherwise (for $t > 1950$ and k in the agricultural sector). We pick the values of $\bar{\phi}_k$ to ensure that, if they decide to migrate, Eastern workers lose their knowledge advantage relative to Western workers.

Table 5: The Quantitative Impact of Knowledge Portability

	Portability to the West		
	25%	Crop-specific	
	Δ RBE	Δ RBE	$\bar{\phi}_k$
Crop	(1)	(2)	(3)
soy	-15.0	-29.1	0.90
corn	-4.0	-0.8	0.68
beef	-8.9	0.0	0.24
sugarcane	-3.6	-2.4	0.17
cotton	-18.7	-34.6	0.87
coffee	-1.3	-1.5	1.00
rest of agriculture	-2.1	-0.8	1.00
cacao	-6.7	0.9	0.64
rice	-2.3	0.1	0.95
banana	-2.1	-2.1	0.91
tobacco	-0.9	-0.5	0.82

Notes: This table shows the aggregate impact of limiting the portability of knowledge from the East to the West of the country, between 1950 and 2010, on RBE of crop k relative to manufacturing in 2010. Column (1) sets $\bar{\phi}_k = 0.75$. Column (2) uses the knowledge reduction factors stated in Column (3). Column (3) shows the knowledge of the average Western worker relative to one in the East.

Table 5, Column (1) shows the results reducing East to West portability by 25 percent, which is the average knowledge difference across all crops, between East and West. Brazil's relative bilateral exports index in 2010 would shrink about 15 percent in soybean, 9 percent in livestock, and 4 percent in corn, which demonstrates the ability to bring knowledge through migration is quantitatively important in the aggregate.

An alternative exercise is shown in Column (2) in which we adjust ϕ_{ijk} to reflect differences in the average value of s_{ik} between East and West, this time crop by crop. In addition to limiting knowledge portability, this exercise removes the comparative advantage across crops of Eastern migrants. In this scenario the impact on soy is largest (30 percent), even though this crop does not face the largest mobility penalty.[30]

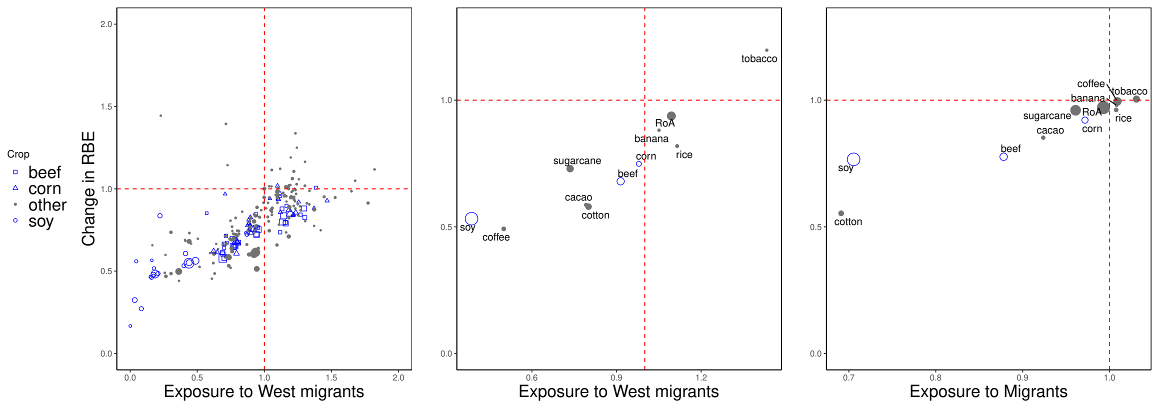
These crop responses are driven by limited migration: In the first scenario, agricultural migration to the West drops by 37 percent, while in the second, it does so by 16.

Knowledge and migration in the long run. In Appendix C.1, we study the impact of worker knowledge heterogeneity in the long run. We find that Brazil's agricultural transformation continues over time, as soybeans become a dominant activity within agriculture and the economy more broadly. Because of migration costs, knowledge and employment rise disproportionately in regions with high employment in soy, improving the sorting of workers to

[30] In our quantitative exercises, cotton specialization responds strongly to migration. As Table 1 shows, however, this crop is unimportant in the aggregate, so we do not discuss it.

Figure 3: Counterfactual Change in Specialization 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 RBE relative to manufacturing (as a percentage of the baseline RBE). 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 RBE 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 RBE relative to manufacturing (as a percentage of the baseline value). The size of the markers represents the magnitude of exports in the baseline. Online Appendix Figure (O.8) reports results for 1980.

natural productivity.

7 Migration Costs, Policy, and Trade during the March

We now evaluate the impact of migration costs and of specific policies that fostered migration during the March. We first examine a counterfactual economy in which the East-West migration costs we backed out remain at their 1950 levels. Next, we isolate the impact of highways and land grants on migration and specialization.

7.1 The Impact of Migration Costs on Specialization

In this exercise, we keep the state-to-state component we recover from our gravity regressions, $\mu_{ss',t}$, constant at its 1950 level, while letting all other exogenous shocks evolve as in the baseline economy, including the residual component $\mu_{ss'k,t}$.[31] We also allow all other exogenous shocks to evolve as in the baseline economy.

We first note that the measured reduction in migration costs accounts for 59 percent of Westbound migration: In our counterfactual economy, the share of employment in the West rises from 6.9 to 10.3 percent (as opposed to 15.4 percent in the data). The rest is due to other factors, including productivity, land supply, and trade costs shocks.

Panel (a) in Figure 3 plots, on the vertical axis, counterfactual changes in the year 2010's RBE specialization across regions in the West, relative to the rest of the world, using manufacturing as a reference. 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. We highlight three patterns from this figure. First, most region-activity pairs fall below one 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, traditional activities. Third, the intuition from Proposition 1 carries over to this richer setting, in that there is a strong relation between exposure to migration and changes in specialization.

Panels (b) and (c) 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 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. Counterfactual specialization relative to the baseline is 1/2 lower for soy and about 1/3 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. In contrast, aggregate drops in soy and cattle specialization are large (25 and 23 percent) and follow, to a large extent, the evolution of the West's agriculture. The exposures in the East to migrants from the West are much lower in general, which means that the majority of our aggregate results follow from the changes that occur in the West.

Lastly, we benchmark these counterfactual results to data, asking how much migration costs account for the *observed* evolution of specialization. Appendix Figure O.10 shows that

[31] The discussion that follows focuses on our cumulative results for 2010. Online Appendix Figure O.8 displays results for 1980.

reductions in migration costs account for up to 25 percent of the country-wide evolution of specialization in soy, cattle, and corn that we observe in the data between 1950 and 2010 (and 32 percent of the evolution of the export shares). For the West as a whole, migration accounts for almost twice as much. These are the results we expect given the prominent role of the West in the exports of these crops by 2010, and given that these crops were disproportionate receivers of migrants in the previous six decades.

7.2 The Role of Policy

We now isolate the effects of two government policies—the expansion of highways towards the West and the land settlement initiatives led by INCRA—and compare how important they were relative to the full migration cost reductions we measure in Section 7.1. Online Appendix OD describes the construction of the counterfactual scenarios we examine.

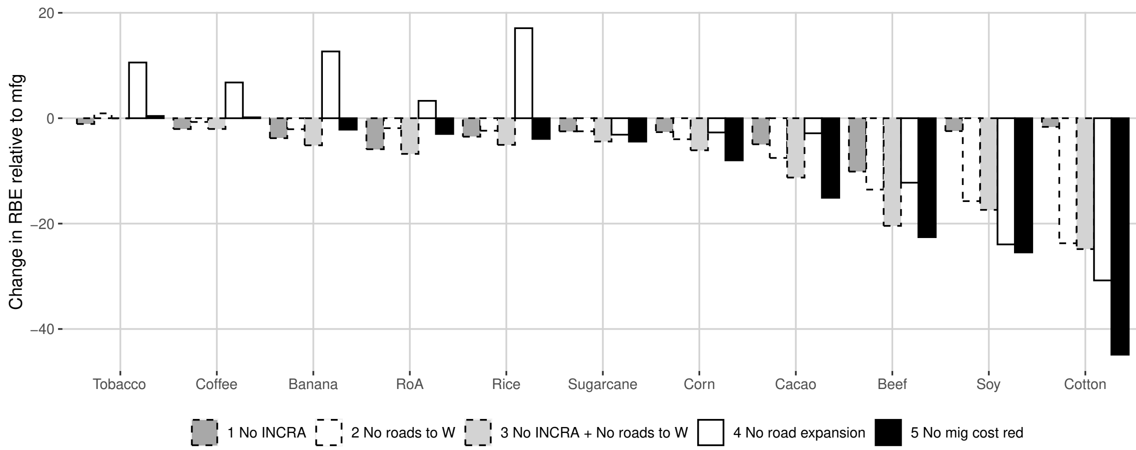
To gauge the impact of highways, we regress the model-implied migration costs between states—$\tilde{\mu}_{ss',t}$ in equation (28)—, against travel distance between states by road, applying the instrumental variables in Morten and Oliveira (2016). Using the resulting coefficients, we construct counterfactual migration costs for 1980 and 2010, assuming highways not expanded since the 1950s (i) to the West and (ii) at all. We keep migration costs fixed at those levels, while allowing the rest of the economy to evolve as in the baseline.

To gauge the role of land settlements, we project our estimated land supply productivity $g_{i,t}$, 1950-2010, on the share of land in meso-region i that was assigned via INCRA's land settlements. We then construct counterfactual productivity $g_{i,t}^{CF}$ distributions by aligning the average expansion of land settlement projects in the West to that of the East (in the data, 12 percent of the land in the West was allocated via INCRA against 4 percent in the East). Everything else, including migration costs, evolves as in the baseline calibration.

The expansion of highways to the West accounts for approximately 30 percent of all the migration to the West, and 50 percent of the migration that we can attribute to reductions in migration cost (Appendix Table O.31). If we add the expansion of land settlements, these numbers rise to 33 percent and 52 percent. If we remove all highways that were constructed after the 1950s (including those constructed in the East), we account for 60 percent of all the migration to the West, and 85 percent of the migration induced by migration costs. The expansion of highways therefore accounts for a sizable share of the migration costs we backed out in Section 7.1.

Figure 4 reports changes in the Brazil's 2010 specialization under several scenarios (Online Appendix Figure O.9 shows the counterpart for the West). The solid black bars repeat the full migration cost exercise from Section 7.1 to provide a benchmark. The expansion of

Figure 4: The Impact of Integration Policies on Brazil's Comparative Advantage



Notes: This figure shows the impact of different integration policies adopted by the government on Brazil's aggregate patterns of specialization, as measured by relative bilateral exports (RBE) (Section 7.2). It examines (1) the land-redistribution policies by INCRA in the West, (2) the expansion of highways in the West, (3) the combination of land-redistribution and highways, (4) the overall expansion of highways in Brazil since the 1950s, and (5), to benchmark these policy results, the overall reductions in migration cost as discussed in Section 7.1.

highways to the West accounts for roughly one third of the changes in specialization in soy, cattle, and corn from migration cost reductions. The impact of INCRA land grants is more heterogeneous: While it accounts for half of the expansion of cattle, its impact is muted on soy. Putting both policies together accounts for 76 percent of the change specialization in soy induced by the full migration costs. The corresponding numbers for cattle and corn are 90 percent and 68 percent.

8 Migration, Comparative Advantage, and the GFT

To close the paper, we summarize theoretical and quantitative results that explain the relation between migration, comparative advantage, and the gains from trade (in the case of $\delta = 0$). We discuss first how migration shapes the "full" losses from trade, i.e. the loss for a region that stops trading altogether. Then we discuss how the March shaped the gains from trading with the rest of the world. Online Appendix OK develops these results in more detail.

A basic result that holds regardless of comparative advantage is that migration links regional gains from trade across all regions within a country. As we show in Proposition 3, the loss in expected utility from full autarky for a region in Brazil is the weighted average of such losses in every other region. The reason is that workers' expected utility incorporates the possibility of migrating to other destinations within the country. In addition, as we show in Proposition 4, the losses from autarky can be broken down into a "domestic" contribution, which are the welfare losses absent migration, and the contribution of migration opportunities, which measure how the absence of trade affects all other regions in Brazil.

Additionally, we show how these results explain how the gains from *international* trade were affected by the March to the West and worker comparative advantage. Comparing the baseline gains from international trade to those under no migration cost reduction, aggregate international gains from trade are unaffected, but the effect of migration costs differs greatly across regions. In particular, since migrants to the West sort disproportionately into agriculture, migration reinforces that region's comparative advantage relative to Foreign, exacerbating the losses from international autarky.

9 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 productivity growth in new sectors in the West, by deploying their knowledge upon migrating there.

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. (2022). Volatility and the Gains From Trade. *Econometrica*, 90(5):2053–2092.

Allen, T. and Donaldson, D. (2022). Persistence and path dependence in the spatial economy.

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., Lee, S. K., and Peters, M. (2020). 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 Rodriguez-Clare, A. (2019b). 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.

Borusyak, K., Dix-Carneiro, R., and Kovak, B. (2022). Understanding migration responses to local shocks. *Available at SSRN 4086847*.

Bound, J. and Krueger, A. B. (1991). The extent of measurement error in longitudinal earnings data: Do two wrongs make a right? *Journal of labor economics*, 9(1):1–24.

Brasil (1941). Decreto-lei n 3.059, de 14 de fevereiro de 1941. Diario Oficial da Uniao.

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. (2020). The Global Diffusion of Ideas. *Econometrica*, 88(1):83–114.

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. (2019). Industrialization without Innovation. NBER Working Papers 25871, National Bureau of Economic Research, Inc.

Cai, J., Li, N., and Santacreu, A. M. (2022). Knowledge Diffusion, Trade, and Innovation across Countries and Sectors. *American Economic Journal: Macroeconomics*, 14(1):104–145.

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. (2022). Immigrants and exports: Firm - level evidence from canada. *Canadian Journal of Economics/Revue canadienne d'Ã©conomique*, 55(3):1250–1293.

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

Comin, D., Lashkari, D., and Mestieri, M. (2021). Structural change with long run income and price effects. *Econometrica*, 89(1):311–374.

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

Costa, F., Garred, J., and Pessoa, J. P. (2016). Winners and losers from a commodities-for-manufactures trade boom. *Journal of International Economics*, 102:50–69.

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 Rodriguez-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.

Desmet, K., Nagy, D. K., and Rossi-Hansberg, E. (2018). The geography of development. *Journal of Political Economy*, 126(3):903–983.

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.

Dominguez-Iino, T. (2023). Efficiency and redistribution in environmental policy: An equilibrium analysis of agricultural supply chains.

Donaldson, D. and Hornbeck, R. (2016). Railroads and American Economic Growth: A "Market Access" Approach. *The Quarterly Journal of Economics*, 131(2):799–858.

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. J. (2022). Trade, structural transformation, and development: Evidence from argentina 1869 - 1914. *Journal of Political Economy*, 130(5):1249–1318.

Farrokhi, F., Kang, E., Pellegrina, H., and Sotelo, S. (2023). Deforestation: A global and dynamic approach.

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.

Gouel, C. and Laborde, D. (2021). The crucial role of domestic and international market-mediated adaptation to climate change. *Journal of Environmental Economics and Management*, 106(C).

Griliches, Z. and Hausman, J. A. (1986). Errors in variables in panel data. *Journal of econometrics*, 31(1):93–118.

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.

Harris, C. D. (1954). The market as a factor in the localization of industry in the united states. *Annals of the Association of American Geographers*, 44(4):315–348.

Hsiao, A. (2023). Coordination and commitment in international climate action: Evidence from palm oil.

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

Kleinman, B., Liu, E., and Redding, S. J. (2023). Dynamic spatial general equilibrium. *Econometrica*, 91(2):385–424.

Kucheryavyy, K., Lyn, G., and Rodriguez-Clare, A. (2023). Grounded by gravity: A well-behaved trade model with industry-level economies of scale. *American Economic Journal: Macroeconomics*, 15(2):372–412.

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.

Neiva, I. C. (1985). O outro lado da colônia: contradições e formas de resistência popular na colônia agrícola de goiás. *Revista Brasileira de Estudos de População*, 2(1):181–186.

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. (2022). Trade, productivity, and the spatial organization of agriculture: Evidence from brazil. *Journal of Development Economics*, 156.

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., Rodriguez-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*.

Sanders, J. H. and Bein, F. L. (1976). Agricultural development on the brazilian frontier: southern mato grosso. *Economic Development and Cultural Change*, 24(3):593–610.

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.

A Appendix to Section 4.2

A.1 The Instrumental Variable Approach

This section explains the construction of the instrument we use in Section 4.2. Our goal is to isolate variation in migration coming strictly from historical coincidences in push and pull factors, building on the approach first developed by Burchardi et al. (2019), and more recently applied in Burchardi et al. (2020). To do so, we first estimate a zero-th state regression to identify these push and pull factors, and then we use them to construct the instrument.

Our "zero-th stage" regression as follows:

$$M_{s'sk,t} = \delta^0_{s',t} + \delta^0_{sk,t} + \sum_{\tau=1950}^{t} \alpha^0_{\tau,t} \times \left(I^{-r(s)}_{s',\tau} \frac{I^{-r(s')}_{s,\tau}}{I^{-r(s')}_{\tau}} \right) + X'_{s's}\beta^0_t + u^0_{s'sk,t}, \tag{A.1}$$

where $I^{-r(s)}_{s',\tau}$, which measures *push* factors, is the observed flow of migrants who left origin state s' at time τ excluding those who migrated from s' to any destination state that belongs to the same region r of state s; $I^{-r(s')}_{s,\tau}$, which measures *pull* factors, is the flow of migrants who arrived at destination s at time τ excluding those who migrated from any origin that belongs to the same region of s'; $I^{-r(s')}_{\tau}$ is the total flow of migrants in τ in Brazil excluding those who come from origins in the same region as s'; and $X_{s's,t}$ is a vector of controls that includes the agricultural similarity between s and s' and the travel distance.

We then construct the predicted flow of migrants between s and s' based on

$$\hat{M}_{s's,t} = \sum_{\tau=1950}^{t} \hat{\alpha}_{\tau,t} \times \left(I^{-r(s)}_{s',\tau} \frac{I^{-r(s')}_{s,\tau}}{I^{-r(s')}_{\tau}} \right)^{\perp},$$

where $\hat{\alpha}_{\tau,t}$ denotes the estimate from equation (A.1) and $\perp$ indicates that we use only variation of that variable residualized against all controls in equation (A.1).[32]

A key feature of this approach is that we control for any origin-time ($\delta^0_{s',t}$) and destination-activity-time ($\delta^0_{sk,t}$) fixed effects inducing migration from s' to s, since we only exploit the residualized variation to generate predictions for the stock of migrants. The interaction between push and pull forces is therefore independent of any unobserved destination-activity shock inducing more migration from all origins to that destination-activity pair (such as

[32] Burchardi et al. (2019) provide a micro-foundation for this instrument based on a dynamic model of migration in which the stock of migrants (in their case ancestry) evolves recursively, drawing from the new flow of migrants being added to the existing stock of migrants. Our method differs from theirs as we only observe the first ancestry link (based on the region of birth), but not previous ancestry links (for example, given by the region of birth of an individual's parents).

suitability for the production of a crop), and any unobserved origin specific shock inducing out-migration (such as low market access). Additionally, we note that by excluding from the push and pull variables the bilateral migration between s and s', we remove the influence of unobserved factors (such as climate) driving that bilateral relationship. Consider, for example, a destination state s that is suitable for coffee and workers from an origin s' who are skilled in coffee production. Our pull forces are based only on the total flow of workers who migrated into s who do not come from origin s' (or any other origin state in the same region) and our push forces are based *only* on the flow of workers who left that origin s and migrated to destinations other than s' (or any other destination in the same region). Then if coffee suitability were the only reason why workers from origin s' were migrating to destination s, our instrument would predict no migration. This procedure thus removes spurious correlations between $M_{s's,t}$ and the error term $u_{sk,t}$ in equation (2) in the main body of the paper.

A final concern is that the historical coincidences between push and pull forces might still be correlated with other bilateral factors, such as distance. Following Burchardi et al. (2020), we address this concern by adding controls for bilateral distance and agricultural similarity between destination and origin.

A.2 Complementary Evidence: Arabica vs. Robusta.

To complement the evidence in Section 4.2, we exploit data from the agricultural census of 2010, which allows us to measure the share of production in each meso-region of each of the two sub-types of coffee: arabica, considered a higher quality type of coffee, and robusta coffee. According to Marcolan and Espindula (2015), migrants in the north of Brazil learned how to produce arabica coffee with coffee growers from the region of Espírito Santo, where there is a larger concentration of producers of arabica. Online Appendix Table O.11 reports on this exercise, based on equation 2, in which we use as the dependent variable the share of coffee output of type arabica, and for the composition term we substitute the suitability term by the share of output of coffee of type arabica in the region of origin. It indicates that a region tends to increase its share of production of arabica coffee when it has a larger share of migrants coming from regions that are specialized in arabica.

B Appendix to Section 4.3

Alternative construction of the Harris' instrument. In our baseline results, we construct $IV^1_{ik,t-1} = \sum_{j \neq i} \mu^{-1}_{ji,t-1} \text{workers}_{jk,t-1}$ removing any region that is within 1 hour drive from the origin i, where $\mu_{ji,t} = \delta_0 d_{ji,t}^{\delta_1}$ and $\delta_1 = 0.05$. Appendix Table C.3 reports results using alternative formulations of this instrument, using the Census data (results using RAIS are in Table O.14). Columns (2) and (6) construct $\mu_{ji,t}$ using a lower values for δ_1 of 0.025. Columns (3) and (7) include meso-regions within 1 hour drive from the origin. Columns (4) and (8) exclude all regions that are within 3 hours drive from the origin.

Controlling for selection. Column 2 in Appendix Table C.4 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 (3).[33] To deal with this potential bias, we follow Dahl (2002) and assume that the migration probabilities $\lambda_{ijk,t}$ 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 $\hat{p}_{ijk,t}$ are consistent estimates of the migration probabilities $\lambda_{ijk,t}$ 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:

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

where $f(\widehat{p}_{ijk,t})$ 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., $\widehat{p}_{ijk,t}$), 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.

Anecdotal Evidence on Migrant Heterogeneity. Additional anecdotal evidence highlights the role of knowledge in the expansion of special varieties of 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

[33]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.

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. 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.

Supporting Evidence on 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, $\mathcal{A}_{ijk}$:

$$\mathcal{A}_{ijk} = -\left| y_{ik}^{FAO} - y_{jk}^{FAO} \right|$$

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

Appendix Table C.5, 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.

Supporting Evidence on the Evolution of Workers' Income and Changes in the International Environment. We study farmers' income for the cohort born between 1950 and 1980. We first track farmers born in regions specialized in the production of soybeans, who benefited from the changes in terms of trade after the 1990s. We then track farmers born in regions specialized in coffee, which has experienced a decline in its relative importance.

Using a triple-difference specification, Appendix Table C.6 shows that the earnings of workers producing soy who were born in top producing regions increase substantially after the 2000s, when the economy experience trade liberalization and the rise of China. Using a more flexible specification, Appendix Figure C.1 shows no differential trends before the 2000s for such workers, but an uptake after then. In contrast, after the 2000s there is a reduction in incomes of coffee growers who where born in coffee-specialized regions. The figure C.1 shows that trend accelerated since the 1990s, which is consistent with the continuous drop in Brazil's comparative advantage in coffee.

C Quantitative Results

C.1 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. We find that, according to our baseline calibration of $\beta = 0.05$, Brazil's agricultural transformation is far from complete. In the model, 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. The aggregate export and revenue shares of soybeans continue to grow and settle about 60 percentage points and 14 percentage points (about 6 times as large as their values in 2010). This transformation is reflected in employment by activity, where we see that the soybean sector increases its employment shares from 0.1 percent to 7 percent. In contrast, the steady state that obtains under $\beta = 0$ is quite close to the observed equilibrium of 2010.

The endogenous adjustment of knowledge, $s_{i,k}$, underlies these results: In the presence of migration costs, knowledge rises disproportionately in high-employment regions, because workers are more likely to stay there. As Appendix Figure C.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 0.44 percentage points (12 percent). In other words, besides the increase in total employment, the sorting of workers is substantially strengthened. Thus, the reallocation of knowledge quantitatively reinforces natural advantage differences.

While these results suggest that the intergenerational knowledge spillovers have potentially large impacts on long-run development, one should be cautious when interpreting these results. Part of the growth of the soybean sector occurs at the expense of services and manufacturing, for which $\beta = 0$ in all our parameterizations quantitative model.

C.2 Alternative Parametrizations

Appendix Figure C.3 shows the impact of reductions in migration costs under different parameterizations. We consider models in which: (1) there are no differences in land intensity between manufacturing and agriculture, (2) there are no agglomeration forces or congestion costs; (3) knowledge depends on the share of workers in the origin producing a certain activity, $s_{ik} = \bar{s}\left(L_{ik,t-1}/L_{iA,t-1}\right)^{\beta}$; and (4) agents are myopic. Across specifications, the impact of migration cost reductions on RBE is similar. The reason is that we re-calibrate the model in each specification, so as to match the exposure of a region to migrants (see 5.4).

Table C.1: Alternative RHS - Earnings Regression - PPML

	(1)	(2)	(3)	(4)	(5)	(6)
Farmers in origin	0.047***	0.038**	0.054**	0.053***	0.050***	0.051***
	(0.016)	(0.017)	(0.025)	(0.020)	(0.018)	(0.019)
Revenues		0.008				
		(0.010)				
Quantity			-0.006			
			(0.016)			
Land				-0.008		
				(0.015)		
Agr-group					-0.008	
					(0.025)	
Worker productivity in origin						-0.006
						(0.018)
R^2	0.338	0.338	0.338	0.338	0.338	0.338
Obs	6778	6778	6778	6778	6778	6778

Notes: * / ** / *** denotes significance at the 10 / 5 / 1 percent level. Multiway clustered standard errors clustered at destination-activity-year and origin-year level. All specifications include destination-activity-year and destination-origin fixed effects. In column 5, we aggregate crops into 4 groups: fruits (banana, cacao, and coffee), grains (corn, soy, and rice), and other (cotton, livestock, tobacco and sugarcane). Online Appendix Table (O.15) shows results for OLS.

Table C.2: Alternative RHS - Population Regression - PPML

	(1)	(2)	(3)	(4)	(5)	(6)
Farmers in origin	0.129***	0.126***	0.123***	0.121***	0.148***	0.124***
	(0.021)	(0.023)	(0.024)	(0.026)	(0.028)	(0.024)
Revenues		0.003				
		(0.011)				
Quantity			0.005			
			(0.016)			
Land				0.010		
				(0.021)		
Agr-group					-0.045	
					(0.038)	
Worker productivity in origin						0.010
						(0.021)
R^2	0.784	0.784	0.784	0.784	0.784	0.784
Obs	6778	6778	6778	6778	6778	6778

Notes: * / ** / *** denotes significance at the 10 / 5 / 1 percent level. Multiway clustered standard errors clustered at destination-activity-year and origin-year level. For explanation, see notes in Appendix Table (C.1). Online Appendix Table (O.16) shows results for OLS.

Table C.3: The Relation between Farmers' Income, Choices, and Region of Origin - Alternative Specifications for the Instrument

	2SLS	2SLS	2SLS	2SLS
		Smaller	Include	Exclude
	Bline	Cost	Dist. < 1h	Dist. < 3h
	(1)	(2)	(3)	(4)
a. Income (logs)				
Farmers in origin	0.048***	0.050***	0.052***	0.045**
	(0.019)	(0.014)	(0.015)	(0.022)
R^2 or K-P	94.211	19.762	23.552	132.482
Obs	6778	6778	6778	6778
b. Farmers in destination (logs)				
Farmers in origin	0.075**	0.131***	0.134***	0.101***
	(0.032)	(0.032)	(0.035)	(0.030)
R^2 or K-P	94.211	19.762	23.552	132.482
Obs	6778	6778	6778	6778

Notes: * / ** / *** denotes significance at the 10 / 5 / 1 percent level. Multiway clustered standard errors clustered at destination-activity-year and origin-year level. All specifications include destination-activity-year and destination-origin fixed effects. Column 1 presents our baseline results, in which we exclude all regions that are within 1 hour of distance of the meso-region. Column 2 constructs the Harris IV using a lower migration cost between regions, instead of $\delta_1 = 0.05$ we use $\delta_1 = 0.025$, which gives a higher weight to regions farther away from an origin. Column 3 include the regions that are within 1 hour and column 4 presents a more strict specification in which we exclude all meso-region within 3 hours drive.

Table C.4: The Relation between Farmers' Income, Choices, and Region of Origin - Individual Level Regressions

	PPML	PPML	PPML	PPML	2SLS	2SLS	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Farmers in origin	0.047***	0.022*	0.042***	0.047***	0.020*	0.037	0.022**
	(0.011)	(0.017)	(0.013)	(0.015)	(0.010)	(0.031)	(0.010)
R^2 or K-P	0.185	0.185	0.236	0.185	26.573	19.508	16.134
Overid. p							0.577
Obs	20002	20002	20002	20002	19335	19335	19335
Dest/Act/Year FE	Y	Y	Y	Y	Y	Y	Y
Dest/Orig/Year FE	Y	Y	Y	Y	Y	Y	Y
Controls selection	-	Y	-	-	-	-	-
SES variables.	-	-	Y	-	-	-	-
Time in meso-region	-	-	-	Y	-	-	-
Harris's IV	-	-	-	-	Y	-	Y
Mig. Comp. IV	-	-	-	-	-	Y	Y

Notes: * / ** / *** denotes significance at the 10 / 5 / 1 percent level. Multiway clustered standard errors clustered at destination-activity-year and origin-year level. This table shows results using individual level regressions. Column 2 follows Dahl's strategy to control for selection bias and includes a exible polynomial of probability choices. Column 3 include socio-economic status variables: sex, ethnicity, age, age squared, education and education squared. Column 4 include a set of dummy variables for the years of living in the current municipality. Columns 5 to 8 estimate the regressions using the instruments discussed in Section 4.

Table C.5: Migration of Agricultural Workers and Agricultural Similarity

	DV: Log of migration flows	
	(1)	(2)
$\log(\mathcal{A}_{ij})$	0.963***	0.367***
	(0.065)	(0.043)
$\log(dist_{ij})$		-1.231***
		(0.026)
R^2	0.185	0.471
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 $\mathcal{A}_{ijk} = -\left|y_{ik}^{FAO} - y_{jk}^{FAO}\right|$ across crops, where y_{ik}^{FAO} averages potential yields in crop k for all cells contained in region i, using data from FAO-GAEZ.

Table C.6: Impact of 1990s on Income of Farmers who were Born between 1950 and 1980 in Specific Regions

	Soy		Coffee	
	(1)	(2)	(3)	(4)
After 1990 gain	0.233**	0.134*	-0.141**	-0.214***
	(0.0934)	(0.0799)	(0.0605)	(0.0529)
R^2	0.032	0.175	0.012	0.172
Obs	2122720	2122720	2122720	2122720
Interaction terms	Y	Y	Y	Y
Socio Economic Status	-	Y	-	Y

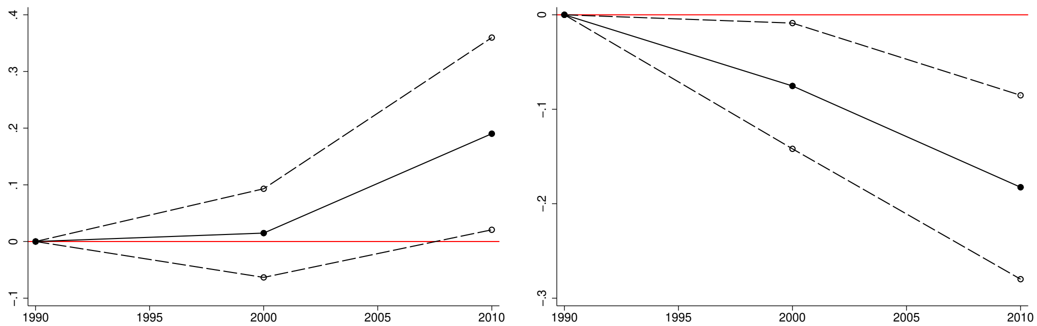
Notes: * / ** / *** denotes significance at the 10 / 5 / 1 percent level. Robust standard errors clustered at the meso-region in parenthesis. To produce this table, we estimate

$$\begin{aligned} \ln earning_{ikt} = {} & \beta_0 \left(T_t \times C_k \times S_i\right) + \beta_1 \left(T_t \times C_k\right) + \beta_2 \left(C_k \times S_i\right) + \beta_3 \left(T_t \times S_i\right) \\ & + \beta_4 T_t + \beta_5 C_k + \beta_6 S_i + \epsilon_{ikt} \end{aligned} \tag{C.2}$$

where $T_t = 1\left(t > 2000\right)$, C_k indicates if the worker produces soy, and S_s indicates if the worker was born in a state that is specialized in soy, where we define as being specialized in soy as the top three states in terms of the employment in soy. Conversely we run the same specification for coffee producers, a crop in which Brazil has been losing comparative advantage since the 1950s.

Figure C.1: Evolution of Income, 1990-2010, for Farmers born in Specialized Regions

(a) Soy (b) Coffee

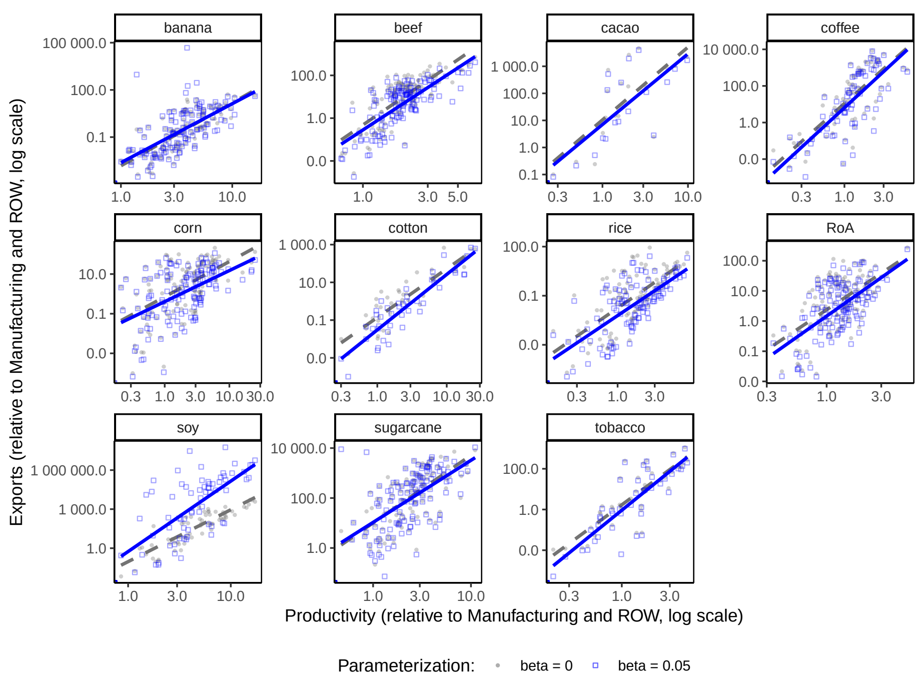


Notes: This figure shows the evolution of income of workers born in states specialized in soybeans when producing soybeans in panel (a). Specifically, it shows the results from the estimation of the following specification:

$$\ln\left(\text{income}_{iskt}\right) = \sum_{t>1990}\left[\rho_t\left(S_s \times C_k \times T_t\right) + \gamma_t\left(C_k \times T_t\right) + \psi_t\left(S_s \times T_t\right) + \omega_t T_t\right] + \phi\left(S_s \times C_k\right) + \lambda S_s + \omega C_k + \epsilon_{iskt}$$

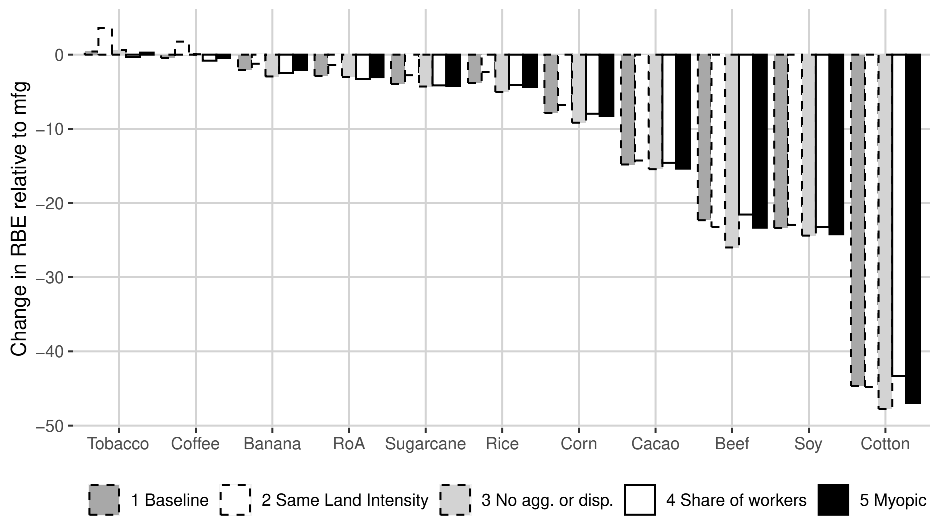
where S_s indicates if the worker was born in a state that is specialized in soy, C_k indicates if the worker produces soy, T_t is an indicator variable for the year t, and ϵ_{iskt} is the error term. We estimate the equation above via OLS using data for $t=$1990, 2000 and 2010. We report $\hat{\rho}_t$ and the 90 percent confidence interval with errors clustered at the meso-region level. Panel (b) shows the same statistic, but in the case of coffee.

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



Notes: Each observation is a region-crop pair, under $\beta = 0.05$ and $\beta = 0$. The x-axis is the productivity of that crop (relative to manufacturing and relative to ROW). The y-axis shows exports in that crop (relative to manufacturing and relative to ROW). For each calibration we also report regression lines.

Figure C.3: Alternative Parametrizations of the Model - Impact of No Migration to the West on Brazil's specialization



Notes: This figure shows the impact on Brazil's specialization of no reductions in migration costs between the East and the West of Brazil under alternative parametrizations of the model. It shows results for (1) our baseline calibration and calibrations in which (2) the land-intensity is the same between manufacturing, services and agricultural activities ($\gamma_k = \gamma$), (3) there are no agglomeration or congestion forces ($\xi = \chi = 0$), (4) knowledge externality is based on the share of workers instead of the scale ($s_{ik,t} = \bar{s}_k L_{ik,t-1}/L^{\beta}_{i,t-1}$), and (5) agents are myopic ($\delta = 0$). See Online Appendix Figure O.11 for the effects on the West.

References

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.

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

Burchardi, K. B., Chaney, T., Hassan, T. A., Tarquinio, L., and Terry, S. J. (2020). Immigration, innovation, and growth. Technical report, National Bureau of Economic Research.

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

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

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