Online Appendix

A Data

This section describes in detail the construction of the our data set. Table A.3 and A.2 present summary statistics of the key aggregate variables of our final data set.

Employment and Migration. Our data on migration and employment comes from decadal demographic and economic censuses organized by the Brazilian statistical institute IBGE (*Instituto Brasileiro de Geografia e Estatistica*). For 1970, 1980, 1991, 2000 and 2010, we have micro-level data on migration and employment. For 1950 and 1960, we digitized state-level aggregates from historical census publications. The information in the micro-level data is divided in two questionnaires: one applied to the universe of the population, which asks basic questions about education and the family structure, and one applied to a smaller sample, which asks detailed information on migration, employment and income. We observe both total income, which include transfers from the government, and income in a workers main activity. In our analysis, we use the latter. In 1970 and 1980, 25 percent of the population was sampled for the detailed questionnaire. For 1990, 2000 and 2010, about 25 percent of the population was sampled in smaller municipalities and 10 percent in larger ones. In 2010, the census about 12 million individuals received the detailed questionnaire.

For the census of 1980 onward, we observe the current and the previous municipality of residence of each individual, if they have migrated within the previous 10 years. We use this variable to define migration in the reduced form elasticities that we presented in Section 3, but as we show later in this appendix, our qualitative results are robust to alternative measures of migration.[54] Since less than 0.1 percent of Brazil's population was born abroad, we remove international migrants from our sample. For 1970, we have micro-data with information on the state of birth and the state of residence. For 1950 and 1960, we use information on the total population in each state who were born in each state of Brazil, and also information on total employment per economic activity. In our structural calibration, we use the state of residence and the state of birth to measure the flow of workers between states.

For 1980, 1990, 2000 and 2010, we can use our micro-data to construct the migration flows of workers who were born in state s and live in state s' and work in activity k, which we denote by $L_{ss',kt}$. For 1950, we observe only the migration flows from a state s to a state s', which is given by $L_{ss',t}$. Given that we do not directly observe $L_{ss',kt}$ in 1950, we therefore use entropy methods widely applied in the construction of input-output matrices to obtain $L_{ss',kt}$. Specifically, we apply the algorithm developed in Ireland and Kullback (1968). In our case, the algorithm consists in searching for values of $L_{ss',kt}$ based on a guessed value $\tilde{L}_{ss',kt}$ until $L_{ss',kt}$ are consistent with aggregate data on $L_{s',kt}$ and $L_{s,t}$. In our case, we use data on $L_{ss',k1980}$ from 1980 (adjusting the overall population size to the one in 1950) as guesses of $\tilde{L}_{ss',k1950}$ to construct the values for $L_{ss',k1950}$ that are consistent with observed values of $L_{ss',1950}$.[55]

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

Our data on agricultural revenues come from PAM (*Produção Agrícola Municipal*), which is organized by the Brazilian census bureau. PAM provides municipality level data since 1974 for more than 20 crops and state level data since 1930s for a subset of crops. For cattle, we use data from the agricultural census. We converted the revenues measured in these data sets into value added and computed the share of value added coming from each agricultural activity for each meso-region. We then multiplied these shares of value added by agricultural activity - given at the meso-region level - by the share of value added in agriculture coming

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

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

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

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

Trade Flows. The data on trade flows by agricultural activity come from FAO. The data is disaggregated by good, according to the Harmonized System at the 6 digit level. We classified the trade flows according to the agricultural activities included in our analysis. We focused on the unprocessed versions of each good. For example, for tobacco, we excluded manufactured cigars and, for wheat, we excluded pastry related goods. Since the data from FAO starts in 1960, we extrapolate exports and imports back using data on aggregate exports and imports from IPEA.

For 2010, we use export data by State from Comexstat, a website organized by the Ministry of Development, Industry and Foreign Trade (MDIC). For each State, we observe how much was exported and imported from abroad. According to MDIC, the trade data at the state level is registered according to the location of production. For domestic trade flows, we digitized data on trade flows between states from the annual statistical yearbook reports from the Brazilian government of 1947, 1948, 1949, 1972, 1973 and 1974. For 1999, we use estimates of trade flows between states from Vasconcelos (2001) based on state merchandise and services taxes.

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

B Additional Quantitative Analyses

This section discusses the robustness checks related to Fact 3 in the main body of the paper, an additional motivating fact that complements the ones presented in the main body of the paper, additional inspection of the migration costs and the econometric implications of an alternative model that emphasizes the role of workers sorting.

B.1 Robustness checks for Fact 3

This section discusses three sets of robustness checks for Table 2. In first set of robustness checks, we inspect different sample specifications. In the second one, we analyze a different definition of migration. In the third one, we include a set of variables related to workers' migration history to investigate the mechanisms driving our results.

Tables A.4 and 2 report our results using alternative sample specifications. In the first 6 columns, we experiment with different time lags of $L_{ij,kt-1}$ (10, 20 and 30 years), different official levels of disaggregation of labor markets defined by the Brazilian statistical bureau (meso-region and micro-region) and with the inclusion of additional years of censuses data as available.[56] In our income regressions, our point estimates are statistically the same when we use different lags. When we include additional censuses years, we lose statistical power as we incorporate years with fewer agricultural activities (for example, soybeans are not available as separate activity in 1970). In column 7, we include socioeconomic status (SES) variables (i.e., gender, ethnicity, age and education averaged within each ij, kt). Column 8 shows that our results are similar if we include all workers older than 20 years old in our sample. Panels (c) and (d) select only the cells that are above the bottom quartile in the distribution of farmers in the origin. Lastly, panels a and b in Table A.6 show that if we run the income regressions at the individual level, we recover similar point-estimates relative to the ones in Table 2.

We now turn to an alternative definition of migration, which is based on a workers current state of living and state of birth. In this case, we can only run state-level regressions since we do not observe the municipality of birth of a worker. As expected, Table A.7 shows that our income regressions have point-estimates that are slightly smaller than the ones in our baseline specifications, since we are averaging over workers coming from heterogeneous origins. Table (A.8) shows that our point-estimates are smaller when we include all years between 1980 and 2010, but they are still statistically significant. Our activity choice regressions present larger coefficients. Noticed that, as we moved from micro-regions to meso-regions and states across specifications in different tables, the coefficient for population becomes larger.

Lastly, we inspect whether the results that we found for Table 2 come from workers learning how to produce specific activities in their previous region, or if our results instead reflect a process in which workers are constantly sorting across regions in particular ways that generate a correlation between $L_{ik,t-1}$ and knowledge about how to produce specific activities (as we show formally later in this section, this alternative mechanism, by itself, is inconsistent with the coefficients that we find). To do so, we control - to the best that our data allows us to - for characteristics of the regions where workers come from prior to their previous region. We include in panels (c) and (d) of Table A.9 controls for the share of workers whose last place of living matches their state of birth and also controls for the share of workers who are return migrants. Our results are only slightly affected by the inclusion of such variables. Second, panels (d) and (e) focus only on the sample of workers whose last place of living correspond to their state of birth, which tends to exclude workers who come from third places. In both cases, our results are similar to the ones presented in panels (a) and (b), which present our baseline results without such controls.

B.2 Additional motivating facts

Fact 4: Agricultural revenues increase with immigration from regions employing many farmers in that crop, after controlling for employment. This empirical fact focuses on the aggregate implication of Fact 3. Specifically, we examine how the aggregate production in a region depends on farmers' composition in terms of their region of origin. To do so, we estimate

$$\log(y_{j,kt}) = \iota_{j,t} + \iota_{k,t} + \alpha^A \underbrace{\log workers_{j,kt}}_{Abundance} + \alpha^C \underbrace{\log \sum_i \omega_{ij,kt} workers_{i,kt-1}}_{Composition} + \epsilon_{j,kt}, \tag{19}$$

where $workers_{j,kt}$ is the aggregate number of workers producing activity k in destination j and $\omega_{ij,kt}$ is the share of workers in destination j producing activity k who come from origin i. We estimate equation (19) using two dependent variables: quantity and revenues. The first set of fixed effect on the right hand side captures any level effect such as the size of a region or the overall demand for agricultural goods and the second one captures any crop specific characteristic such as the land intensity. When the composition term is larger for a given destination and crop, then farmers come from origins that are more specialized in the production of this crop.

Table A.11 shows, first, that a 1 percent increase in the abundance of farmers in a region is associated with a 0.9 percent increase in revenues. Second, it shows that a 1 percent increase in the average number of farmers in the origin is associated with a 0.2 percent increase in revenues, even controlling for the abundance

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

of farmers, which indicates that the composition of farmers is strongly associated with total size of the sector in a region. To test the robustness of our results, we add controls for the composition of workers in terms of their socioeconomic characteristics and the share of migrants in a region. To address potential endogeneity in migration patterns, we follow the literature on migration and use predicted values from migration gravity equations where we include only the euclidean distance between regions as a predictor of migration.

Fact 5: Controlling for distance, there is more migration between regions with higher agricultural similarity. To provide additional indirect evidence that part of a farmer's migration decision relates to his skills, we show that crop-specific agricultural similarity is an important driver of migration. Table A.12 presents an alternative to our regressions 3. We exclude origin-destination fixed effects to be able to explore the impact of variation of this kind. We construct agricultural similarity using the GAEZ data set (IIASA/FAO, 2012), based on the index developed in Bazzi, Gaduh, Rothenberg, and Wong (2016), as follows:

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

where $y_{i,k}$ is the potential yield of region i in activity k and $\mathcal{A}_{ij,k}$ is the agricultural similarity between region i and j in activity k. We normalize this index to be between zero and one for our estimation, and then we construct the unweighted average of agricultural similarity across activities k between regions i and j.[57]

Across specifications, our results show that farmers are more likely to migrate when the origin and destination are more similar in their natural conditions. These results hold controlling for distance between origin and destination which, additionally, provides direct evidence that migration decreases with distance.

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

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

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

B.3.1 A "fixed types" model with sorting according to income

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

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

The solution is:

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

where

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

[57]To construct potential yield, we take the average of the maximum attainable yield across pixels in every region according to the high-input use technology. In addition, we normalize potential yields within every crop.

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

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

Hence, the corresponding regression equation is:

$$\log \text{Income}_{ij,kt} = \log w_{j,kt} + \log s_{i,kt} + u^I_{ij,kt}, \tag{20}$$

where we allow again for measurement error in income.

Likewise, the migration equation states that

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

and the corresponding regression equation is:

$$\log L_{ij,kt} = \kappa \log w_{j,kt} + \kappa \log s_{i,kt} - \kappa \log P_{j,t} - \kappa \log \mu_{ij,kt} + \log L_{ikt-1} - \kappa \tilde{\Xi}_{i,kt}. \tag{21}$$

Direct comparison of equation (20) with the results from (12) shows that worker heterogeneity according to origin location is necessary to reproduce them. Moreover, one needs to impose $s_{i,k} = \bar{s}_i L^{\beta}_{i,kt-1}$, to replicate our regression. In turn, using this assumption in (20) shows that the coefficient of $L_{i,kt-1}$ should be $1 + \kappa\beta$, which is at odds with the results from (12). In our specification, in which we essentially use one cohort, one cannot separately identify $L_{i,kt-1}$ from the fixed effect that would capture $\tilde{\Xi}_{i,kt}$.

B.3.2 A "fixed types" model with random sorting

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

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

The standard solution is:

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

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

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

Hence, the corresponding regression equation is:

$$\log \text{Income}_{ij,kt} = \log w_{j,kt} + \log s_{i,kt} + u^I_{ij,kt}, \tag{22}$$

where we allow again for measurement error in income.

Likewise, the migration equation states that

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

and the corresponding regression equation is:

$$\log L_{ij,kt} = \log L_{i,kt-1}. \tag{23}$$

Equation (22) shows that one requires the same conditions as in the previous Section to rationalize the income equation. Moreover, equation (22) shows that the coefficient of $L_{i,kt-1}$ should be 1, which is at odds with the results from (12).

C Model Details

In this section, we first present a formal definition of the equilibrium in Section 4. We then present the full model we use to obtain our quantitative results.

C.1 Equilibrium

Costs. We begin by defining efficiency wages as $w_{j,k} = \pi_{j,k}(s)/s$, i.e., as income per unit of knowledge. When $\beta = 0$, $s_{i,k} = 1$ for all workers, efficiency wages boil down to the standard payments per worker. The cost of a unit of input bundle is $c_{j,kt} = \kappa_\pi w_{j,k}^{1-\gamma_k} r_j^{\gamma_k}$,where $\kappa_\pi \equiv \gamma_k^{\gamma_k}(1-\gamma_k)^{1-\gamma_k}$, and the unit cost of producing a unit of good k in region j is $c_{j,kt}/A_{j,kt}$.

Trade Shares. Utility maximization gives the standard Armington expenditure shares: the share of region j's expenditure in sector k goods produced in region i, $\pi_{ij,kt}$, is given by:

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

Migration Shares and Labor Supply Turning to migration, the definition of the observable component of welfare (4) together with optimal worker sorting gives the share of workers from i choosing to work in region j and sector k, $\lambda_{ij,kt}$

$$\lambda_{ij,kt} = \frac{W_{ij,kt}^{\kappa}}{\Xi_{i,t}^{\kappa}} \tag{24}$$

where $\Xi_{i,t}^{\kappa} \equiv \sum_j \sum_k [w_{j,kt}s_{i,kt}/(\mu_{ij,kt}P_{j,t})]^{\kappa}$. We define the effective units of labor migrating from i to region j, sector k as

$$E_{ij,kt} \equiv s_{i,kt}\lambda_{ij,kt}L_{i,t-1}. \tag{25}$$

Closing the model. We assume that land rents in a region are rebated to workers who live there, proportionally to their labor income. We also assume there are no deficits. Therefore, total expenditure in region j reflects payments to factors there

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

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

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

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

$$w_{j,kt}E_{j,kt} = (1-\gamma_k)\sum_i \pi_{ji,kt}X_{i,kt}. \tag{26}$$

2. Land markets clear in region j:

$$r_{j,t}H_{j,t} = \sum_k \gamma_k \sum_i \pi_{ji,kt} X_{i,kt}, \tag{27}$$

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

$$E_{j,kt} = \sum_i s_{i,kt}\left(L_{i,kt-1}\right) L_{ij,kt}, \tag{28}$$

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

4. Migration flows maximize workers utility

$$L_{ij,kt} = \lambda_{ij,kt} L_{i,t-1}. \tag{29}$$

C.2 Small Open Economy Equilibrium

We start by studying equilibrium prices when the Home economy is small. To do this, we adapt the procedure in Alvarez and Lucas (2007). We use the following assumptions, for each region in Home: (i) $L_i \to 0$, (ii) $A_{i,k}^{\eta-1}/L_i \to \delta_{i,k}$, where $\delta \in (0,\infty)$, (iii) $H_i/L_i \to h_i$, where $h_i \in (0,\infty)$, and (iv) $L_i/L_{i'} \to \ell_{ii'}$

Assuming that, in the limit $w_{i,k} \in (0,\infty)$ and $r_{i,k} \in (0,\infty)$ — which we verify later — the equilibrium price indexes for each region and sector

$$P_{i,k} = \left(\sum_j \left(c_{i,k}\tau_{ji,k}\right)^{1-\eta} A_{i,k}^{\eta-1}\right)^{1-\eta}$$

and assumptions (i) and (ii), imply that

$$P_{i,k} \to \left(\bar{c}_{F,k}\tau_{jF,k}\right)/A_{F,k}, \forall i,k.$$

where $\bar{c}_{F,k} = w_{F,k}^{\alpha(1-\gamma)} r_F^{\gamma\alpha} P_F^{1-\alpha}$ solves the labor and land market clearing conditions for Foreign

$$\begin{aligned} w_{F,k}E_{F,k} &= (1-\gamma)\,\alpha a_k X_F \\ r_F H_F &= \alpha\gamma X_F. \end{aligned}$$

In what follows, we take $w_{F,k}$, r_F and $\{P_{i,k}\}$ as given.

We now characterize the equilibrium wages and rental rates for each region at Home. Using labor market clearing

$$\begin{aligned} w_{i,k}\sum_{i'} \frac{\left(s_{i',k}w_{i,k}\right)^{\kappa}\left(\mu_{i'i,k}P_i\right)^{-\kappa}}{\Xi_{i'}^{\kappa}} s_{i',k}L_{i'} = \alpha\left(1-\gamma\right)\Bigg\{ &\sum_{j\in H}\left(\frac{w_{i,k}^{\alpha(1-\gamma)} r_i^{\alpha\gamma} P_i^{1-\alpha}\tau_{ij,k}/A_{i,k}}{P_{j,k}}\right)^{1-\eta} X_{j,k} \\ &+ \left(\frac{w_{i,k}^{\alpha(1-\gamma)} r_i^{\alpha\gamma} P_i^{1-\alpha}\tau_{ij,k}/A_{i,k}}{P_{F,k}}\right)^{1-\eta} X_{F,k}\Bigg\}, \end{aligned} \tag{30}$$

which implies:

$$w_{i,k}^{1+\kappa-(1-\gamma)\alpha(1-\sigma)}\sum_{i'}\frac{s_{i',k}^{\kappa}\left(\mu_{i'i,k}P_{i}\right)^{-\kappa}}{\sum_{l}\sum_{i''}\left(s_{i',l}w_{i'',l}\right)^{\kappa}\left(\mu_{i''i,l}P_{i''}\right)^{-\kappa}}s_{i',k}\ell_{i'i}=\alpha\left(1-\gamma\right)\delta_{i,k}\left[\left(\frac{r_{i}^{\gamma\alpha}P_{i}^{1-\alpha}\tau_{ij,k}}{P_{F,k}}\right)^{1-\eta}X_{F,k}\right],\tag{31}$$

where the second line divides through by $L_{i,k}$ and uses assumptions (i) through (iv).

Similarly, land market clearing states

$$r_{i}H_{i}=\alpha\left(1-\gamma\right)\left[\sum_{k}\sum_{j\in H}\left(\frac{w_{i,k}^{\alpha(1-\gamma)}r_{i}^{\alpha\gamma}P_{i}^{1-\alpha}\tau_{ij,k}/A_{i,k}}{P_{j,k}}\right)^{1-\eta}X_{j,k}+\left(\frac{w_{i,k}^{\alpha(1-\gamma)}r_{i}^{\alpha\gamma}\tau_{ij,k}P_{i}^{1-\alpha}/A_{i,k}}{P_{F,k}}\right)^{1-\eta}X_{F,k}\right],$$

which implies

$$r_{i}^{1+\alpha\gamma(\eta-1)}=\alpha\left(1-\gamma\right)\sum_{k}w_{i,k}^{\alpha(1-\gamma)(1-\eta)}\delta_{i,k}h_{i}^{-1}\left[\left(\frac{\tau_{ij,k}P_{i}^{1-\alpha}}{P_{F,k}}\right)^{1-\eta}X_{F,k}\right]\tag{32}$$

Taken together, equations (31) and (32) constitute a system of equations for $w_{i,k}$ and r_i jointly for all $i\in H$, which depend only on predetermined constants, parameters, and prices solved above.

C.3 Steady State Equilibrium.

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

$$\begin{aligned}
w_{j,k}E_{j,k} &= \left(1-\gamma_{k}\right)\sum_{i}\pi_{ji,k}X_{i,k}\\
r_{j}H_{j} &= \sum_{k}\gamma_{k}\sum_{i}\pi_{ji,k}X_{i,k}\\
E_{j,k} &= \sum_{i}L_{i,k}^{\beta}L_{ij,k},\\
L_{ij,k} &= \lambda_{ij,k}L_{i}.\\
L_{j} &= \sum_{i}L_{i}\left[\sum_{k}\lambda_{ij,k}\right]
\end{aligned}$$

C.3.1 Closed economy without internal geography

To draw sharp conclusions about how migration and the knowledge externality shape the steady state equilibrium, we focus on a model with two sectors, denoted k and k'. Letting $\omega=w_{k}/w_{k'}$, we can boil down our equilibrium to a labor market clearing equation

$$\omega=\left(\frac{\lambda_{kt}}{\left(1-\lambda_{kt}\right)}\frac{\lambda_{kt-1}^{\beta}}{1-\lambda_{kt-1}^{\beta}}\right)^{1/\sigma}$$

and a migration equation

$$\frac{\lambda_{kt}}{1-\lambda_{kt}}=\left(\frac{\lambda_{kt-1}^{\beta}}{\omega\left(1-\lambda_{kt-1}\right)^{\beta}}\right)^{\kappa}.$$

To study the dynamics of the system, we substitute for ω in the migration equation to obtain

$$\tilde{\lambda}_t = \tilde{\lambda}_{t-1}^{\beta\kappa\frac{\sigma-1}{\kappa+\sigma}},$$

where $\tilde{\lambda}_t = \lambda_{k,t}/\lambda_{k',t}$. We distinguish the following cases. First, if $\sigma < 1$, so sectors are complementary, the unique equilibrium features symmetric labor allocations, $\tilde{\lambda} = 1$. Second, when $\sigma > 1$, there are multiple equilibria: one symmetric, $\tilde{\lambda} = 1$, and two featuring full specialization, $\tilde{\lambda} = 0$ and $\tilde{\lambda}^{-1} = 0$.

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

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

C.3.2 Small open economy

We now turn to the steady state in the case of a small open economy, as described in detail in Appendix Section C.2. For simplicity, set $\gamma = 0$, so labor and intermediate inputs are the only factors of production. From equation (30), which describes labor market clearing for region i and activity k, it is apparent that corners with zeros are not possible if at least one other region i' produced k in the previous period, so that $s_{i',k} > 0$. To see this, note that $L_{i,k} = 0$ only when $w_{i,k} = 0$. If that is the case, the left-hand side of equation (30) equals zero. This is inconsistent with the right-hand side of the same equation, which approaches infinity as $w_{i,k} \to 0$. Therefore, in a small open economy, zero employment is only possible in the trivial case in which $A_{i,k} = 0$.

C.4 Quantitative model

In this section, we give a streamlined description of the model we take to data in Section 5, emphasizing differences with respect to the previous section.

C.4.1 Environment

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

C.4.2 Technology

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

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

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

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

$$P_{j,t}^{1-\sigma} = \sum_{g=1}^{G} \bar{b}_g P_{j,gt}^{1-\sigma},$$

$$P_{j,gt}^{1-\sigma_g} = \sum_{k=1}^{K_g} \bar{a}_k p_{j,kt}^{1-\sigma_s},$$

and

$$p_{j,kt}^{1-\eta_k} = \sum_{i\in\mathcal{W}} \left(\tau_{ij,kt} p_{i,kt}\right)^{1-\eta_k}.$$

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

$$H_{j,t} = g_{j,t} x_{j,t}^{1/\varsigma}. \tag{33}$$

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

C.4.3 Workers

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

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

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

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

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

C.4.4 Expenditure and accounting

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

$$X_{j,t}^F = \frac{\varsigma - 1}{\varsigma} r_{j,t} H_{j,t} + \sum_{g=1}^{G}\sum_{k=1}^{K_g} w_{j,kt} E_{j,kt} + T_{j,t} \tag{34}$$

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

$$X^F_{j,st} = \underbrace{\bar{b}_j \left(\frac{P_{j,st}}{P_{j,t}}\right)^{1-\sigma}}_{\equiv b_{j,st}} X^F_{j,t} \tag{35}$$

$$x^F_{j,kt} = \underbrace{\bar{a}_k \left(\frac{p_{j,kt}}{P_{j,st}}\right)^{1-\sigma_s}}_{\equiv a_{j,kt}} X^F_{j,st} \tag{36}$$

Region j's total expenditure in intermediate inputs

$$X^I_{j,t} = \sum_{g=1}^{G} \sum_{k=1}^{K_g} (1-\alpha_k)\, Y_{j,kt} + \frac{1}{\varsigma} r_{j,t} H_{j,t} \tag{37}$$

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

$$X^I_{j,st} = b_{j,st} X^I_{j,t} \tag{38}$$

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

$$x^I_{j,kt} = a_{j,kt} X^I_{j,st} \tag{39}$$

Region j's Total Expenditure on activity k:

$$x_{j,kt} = x^F_{j,kt} + x^I_{j,kt}$$

Bilateral trade flows

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

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

C.4.5 Endowments and supplies

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

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

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

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

$$\Xi^{\kappa}_{i,t} = \sum_j \sum_k \left[w_{j,kt} s_{i,kt} / \left(\mu_{ij,t} P_{j,t}\right)\right]^{\kappa}$$

C.4.6 Equilibrium definition

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

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

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

2. Land markets clear in region j:

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

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

$$E_{j,kt} = \sum_i s_{i,kt}\,(L_{i,kt-1})\,L_{ij,kt}, \tag{40}$$

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

4. Migration flows maximize workers utility

$$L_{ij,kt} = \lambda_{ij,kt}L_{i,t-1}. \tag{41}$$

where

$$v_{i,kt} = \frac{\bar{v}w_{j,kt}^{1-\rho}}{\bar{v}w_{j,kt}^{1-\rho} + (1-\bar{v})\,r_{j,t}^{1-\rho}}$$
$$Y_{j,kt} = \sum_{j'}\pi_{jj',kt}x_{j',kt}$$

Migration Costs and Comparative Advantage

C.5 Relative Opportunity Costs and Migration at the Regional Level

In this section we provide proofs for the statements in Section 4.3 of the main body of the paper, which relate specialization, comparative advantage and internal migration. We begin by showing under which conditions the following inequality holds

$$\frac{P_{i,k}^A}{P_{i,k'}^A} < \frac{P_{F,k}^A}{P_{F,k'}^A}, \tag{42}$$

where $P_{i,k}^A$ is the price index of sector k when region i is in autarky, which in this model the standard "opportunity-cost" definition of comparative advantage introduced by Haberler (see Deardorff (2005) and French (2017)). We discuss below conditions under which this is also sufficient to predict the patterns of trade specialization. To obtain tractable equations, we adopt the following assumptions: (i) same technology across activities k $\alpha_k = \alpha$ and $\gamma_k = \gamma$; (ii) workers are born of type k and can choose where to live, but not what activity to produce.

Note first that under full trade autarky (i.e., when the regions within Brazil cannot trade), the opportunity cost ratio in region i is given by

$$\begin{aligned}\frac{P_{i,k}^{A}}{P_{i,k'}^{A}} &= \frac{c_{i,k}/A_{i,k}}{c_{i,k'}/A_{i,k'}}\\ &= \frac{\left(w_{i,k}^{1-\gamma}r_i^{\gamma}\right)^{\alpha}P_i^{1-\alpha}/A_{i,k}}{\left(w_{i,k'}^{1-\gamma}r_i^{\gamma}\right)^{\alpha}P_i^{1-\alpha}/A_{i,k'}}\\ &= \frac{w_{i,k}^{(1-\gamma)\alpha}/A_{i,k}}{w_{i,k'}^{(1-\gamma)\alpha}/A_{i,k'}}.\end{aligned}$$

The inequality thus becomes

$$\left(\frac{w_{i,k}}{w_{i,k'}}\right)^{(1-\gamma)\alpha}\frac{A_{i,k'}}{A_{i,k}} < \left(\frac{w_{F,k}}{w_{F,k'}}\right)^{(1-\gamma)\alpha}\frac{A_{F,k'}}{A_{F,k}}. \tag{43}$$

C.5.1 Prohibitive migration costs

When migration costs are prohibitive, so $\mu_{ijk} \to \infty$ for $i \neq j$, then the labor market clearing condition for region i and sector k is given by:

$$w_{i,k}s_{i,k}L_{i,k}^{0} = \alpha\left(1-\gamma\right)a_k X_i,$$

which solving for $w_{i,k}$ gives

$$w_{i,k} = \frac{\alpha\left(1-\gamma\right)a_k X_i}{s_{i,k}L_{i,k}^{0}}.$$

Note that to obtain this expression, we assume region i is in full autarky. Substituting for wages in (43), we show that under prohibitive migration costs, the opportunity-cost definition depends only on productivities and endowments of knowledge and workers.

$$\left(\frac{s_{i,k'}L_{i,k'}^{0}}{s_{i,k}L_{i,k}^{0}}\right)^{(1-\gamma)\alpha}\frac{A_{i,k'}}{A_{i,k}} < \left(\frac{s_{F,k'}L_{F,k'}^{0}}{s_{F,k}L_{F,k}^{0}}\right)^{(1-\gamma)\alpha}\frac{A_{F,k'}}{A_{F,k}} \tag{44}$$

C.5.2 Free migration

With free migration, $\mu_{ij,k} = 1$ for $i \neq j$. Thus labor market in region i, sector k is given by

$$w_{i,k}\sum_{i'} s_{i',k}\lambda_{i',ik}L_{i',k}^{0} = \alpha\left(1-\gamma\right)a_k X_i.$$

Using the migration equations, we obtain

$$\begin{aligned}w_{i,k}\sum_{i'}\frac{\left(w_{i,k}s_{i',k}P_i\right)^{\kappa}}{\sum_h\left(w_{h,k}s_{i',k}P_i\right)^{\kappa}}s_{i',k}L_{i',k}^{0} &= \alpha\left(1-\gamma\right)a_k X_i\\ w_{i,k}^{1+\kappa}\frac{\mathcal{S}_{H,k}}{\sum_h\left(w_{h,k}\right)^{\kappa}} &= \alpha\left(1-\gamma\right)a_k X_i\end{aligned}$$

and solving for wages, we get

$$w_{i,k}^{1+\kappa} = \alpha\left(1-\gamma\right)a_k X_i\sum_h\left(w_{h,k}\right)^{\kappa}\mathcal{S}_{H,k}^{-1}. \tag{45}$$

It will be useful to solve for $w_{i,k}$ here, i.e., eliminate $w_{h,k}$:

$$\frac{w_{i,k}}{w_{h,k}} = \left(\frac{X_i}{X_h}\right)^{\frac{1}{1+\kappa}} \Rightarrow w_{h,k} = w_{i,k}\left(\frac{X_h}{X_i}\right)^{\frac{1}{1+\kappa}},$$

which we use again in (45) to obtain

$$w_{i,k}^1 = \alpha\left(1-\gamma\right)\alpha_k X_i^{1-\frac{\kappa}{1+\kappa}} \mathcal{S}_{H,k}^{-1} \sum_h X_h^{\frac{\kappa}{1+\kappa}}$$

Substituting this expression in (43), we obtain

$$\left(\frac{\mathcal{S}_{H,k'}}{\mathcal{S}_{H,k}}\right)^{(1-\gamma)\alpha} \frac{A_{i,k'}}{A_{i,k}} < \left(\frac{\mathcal{S}_{F,k'}}{\mathcal{S}_{F,k}}\right)^{(1-\gamma)\alpha} \frac{A_{F,k'}}{A_{F,k}}. \tag{46}$$

C.6 Regional Comparative Advantage

In this section, we study how migration relates to region i's export specialization in sector k (relative to sector k' and region F). In particular, we study how migration shapes the inequality

$$\frac{X_{iF,k}}{X_{iF,k'}} > \frac{X_{FF,k}}{X_{FF,k'}}. \tag{47}$$

We then show how the relative opportunity costs in autarky give a sufficient condition for this to hold.

Under the assumptions laid out in the paper, expression (47) is equivalent to

$$\left(\frac{w_{i,k}}{w_{i,k'}}\right)^{(1-\gamma)(1-\eta)\alpha} \left(\frac{A_{i,k'}}{A_{i,k}}\right)^{1-\eta} \left(\frac{\tau_{iF,k}}{\tau_{iF,k'}}\right)^{1-\eta} > \left(\frac{w_{F,k}}{w_{F,k'}}\right)^{(1-\gamma)(1-\eta)\alpha} \left(\frac{A_{F,k'}}{A_{F,k}}\right)^{1-\eta} \tag{48}$$

To obtain conditions relating to exogenous forces in the model, we now solve for wages using labor market clearing. To focus on the sources of comparative advantage related to costs of production, and not of trade, we set $\tau_{iF,k} = \tau_{iF,k'} = 1$, and simplify equation (48) to obtain:

$$\left(\frac{w_{i,k}}{w_{i,k'}}\right)^{(1-\gamma)(1-\eta)\alpha} \left(\frac{A_{i,k'}}{A_{i,k}}\right)^{1-\eta} > \left(\frac{w_{F,k}}{w_{F,k'}}\right)^{(1-\gamma)(1-\eta)\alpha} \left(\frac{A_{F,k'}}{A_{F,k}}\right)^{1-\eta} \tag{49}$$

C.6.1 Prohibitive migration costs

When migration costs are prohibitive, so $\mu_{ijk} \to \infty$ for $i \neq j$, then the labor market clearing condition is

$$w_{i,k} s_{i,k} L_{i,k}^0 = \alpha\left(1-\gamma\right) \sum_j \left(w_{i,k}^{\alpha(1-\gamma)} r_i^{\alpha\gamma} P_i^{1-\alpha} / A_{i,k}\right)^{1-\eta} P_{j,k}^{\eta-1} X_{j,k}$$

where $X_{j,k}$ is expenditure of region j on goods from k. Solving for $w_{i,k}$ gives

$$w_{i,k} = \left(\alpha\left(1-\gamma\right) \frac{\left(r_i^{\alpha\gamma} P_i^{1-\alpha}\right)^{1-\eta}}{A_{i,k}^{1-\eta} s_{i,k} L_{i,k}^0} \sum_j \frac{X_{j,k}}{P_{j,k}^{1-\eta}}\right)^{\frac{1}{1+\alpha(1-\gamma)(\eta-1)}}. \tag{50}$$

Substituting equation (50) into (49), we get

$$\left(\frac{\left(s_{i,k}L^0_{i,k}\right)^{\alpha(1-\gamma)}A_{i,k}}{\left(s_{i,k'}L^0_{i,k'}\right)^{\alpha(1-\gamma)}A_{i,k'}}\right)^{\frac{\eta-1}{1+\alpha(1-\gamma)(\eta-1)}} > \left(\frac{\left(s_{F,k}L^0_{F,k}\right)^{\alpha(1-\gamma)}A_{F,k}}{\left(s_{F,k'}L^0_{F,k'}\right)^{\alpha(1-\gamma)}A_{F,k'}}\right)^{\frac{\eta-1}{1+\alpha(1-\gamma)(\eta-1)}}. \tag{51}$$

C.6.2 Free migration

As will be come clear later, in this section we need to introduce assumptions on land markets. We assume that either land is not a factor of production, so $\gamma = 0$, or that land supply is perfectly elastic, so $r_i = \bar{r}_i$.

With free migration, $\mu_{ij,k} = 1$ for $i \neq j$, and labor market clearing is given by

$$\sum_{i'} w_{i,k}\frac{\left(w_{i,k}s_{i',k}P_i^{-1}\right)^{\kappa}}{\sum_h\left(w_{h,k}s_{i',k}P_h^{-1}\right)^{\kappa}}s_{i',k}L^0_{i',k} = \alpha\left(1-\gamma\right)\sum_j\left(w_{i,k}^{\alpha(1-\gamma)}r_i^{\alpha\gamma}P_i^{1-\alpha}/A_{i,k}\right)^{1-\eta}P_{j,k}^{\eta-1}X_{j,k},$$

which we can use to solve for wages in region i and sector k:

$$w_{i,k} = \left(\frac{\alpha\left(1-\gamma\right)\left(r_i^{\alpha\gamma}P_i^{1-\alpha}\right)^{1-\eta}}{A_{i,k}^{1-\eta}\mathcal{S}_{H,k}}\right)^{\frac{1}{1+\kappa+\alpha(1-\gamma)(\eta-1)}}\left(\left[\sum_h w_{h,k}^{\kappa}\right]\sum_j\frac{X_{j,k}}{P_{j,k}^{1-\eta}}\right)^{\frac{1}{1+\kappa+\alpha(1-\gamma)(\eta-1)}}. \tag{52}$$

Use equation (52), to express wages in i as a function of wages in h, both for sector k:

$$w_{h,k} = w_{i,k}\left(\frac{A_{h,k}^{\eta-1}\left(r_h^{\alpha\gamma}P_h^{1-\alpha}\right)^{1-\eta}}{A_{i,k}^{\eta-1}\left(r_i^{\alpha\gamma}P_i^{1-\alpha}\right)^{1-\eta}}\right)^{\frac{1}{1+\kappa+\alpha(1-\gamma)(\eta-1)}}$$

Under the assumptions of free trade and $\gamma = 0$, this simplifies to

$$w_{h,k} = w_{i,k}\left(\frac{A_{h,k}}{A_{i,k}}\right)^{\frac{\eta-1}{1+\kappa+\alpha(\eta-1)}}$$

and the substitute back into (52) to obtain expression for equilibrium wage

$$w_{i,k} = \left(\frac{\alpha\left(P_i^{1-\alpha}\right)^{1-\eta}}{A_{i,k}^{1-\eta}\mathcal{S}_{H,k}}\right)^{\frac{1}{1+\kappa+\alpha(\eta-1)}}\left(\left[\sum_h w_{i,k}^{\kappa}\left(\frac{A_{h,k}}{A_{i,k}}\right)^{\frac{\kappa(\eta-1)}{1+\kappa+\alpha(\eta-1)}}\right]\sum_j\frac{X_{j,k}}{P_{j,k}^{1-\eta}}\right)^{\frac{1}{1+\kappa+\alpha(\eta-1)}} \Rightarrow$$

$$w_{i,k} = \left\{\frac{\alpha\left(P_i^{1-\alpha}\right)^{1-\eta}}{A_{i,k}^{1-\eta}\mathcal{S}_{H,k}\mathcal{A}_{i,k}}\sum_j\frac{X_{j,k}}{P_{j,k}^{1-\eta}}\right\}^{\frac{1}{1+\alpha(\eta-1)}} \tag{53}$$

where we define $\mathcal{A}_{i,k} \equiv A_{i,k}^{\kappa\frac{\eta-1}{1+\kappa+\alpha(\eta-1)}}/\sum_h A_{h,k}^{\kappa\frac{\eta-1}{1+\kappa+\alpha(\eta-1)}}$.

Substitute the expression above into (49) to obtain

$$\left[\left(\frac{\mathcal{S}_{H,k}\mathcal{A}_{i,k}}{\mathcal{S}_{H,k'}\mathcal{A}_{i,k'}}\right)^{\alpha}\left(\frac{A_{i,k}}{A_{i,k'}}\right)\right]^{\frac{\eta-1}{1+\alpha(\eta-1)}} > \left[\left(\frac{\mathcal{S}_{H,k}\mathcal{A}_{F,k}}{\mathcal{S}_{H,k'}\mathcal{A}_{F,k'}}\right)^{\alpha}\left(\frac{A_{F,k}}{A_{F,k'}}\right)\right]^{\frac{\eta-1}{1+\alpha(\eta-1)}} \tag{54}$$

With this, we complete the proof for regional comparative advantage.

C.6.3 Relation to the opportunity cost definition of CA

For the case in which migration costs are prohibitive, direct comparison of expressions (44) and (51) shows that the opportunity-cost formulation is equivalent to free-trade specialization. For the case of free internal migration, comparison of (46) and (54) gives two results. First, relative opportunity costs in autarky cannot be translated with a constant elasticity into free-trade specialization. Second, in fact, the predictions for specialization coming from relative opportunity costs in autarky can be overturned under free migration. Since the term $\mathcal{A}_{i,k}$ captures the degree of competition for workers and how they are apportioned to i relative to the rest of the economy, in the economy as a whole (as measured by $\sum_h A_{h,k}^{\kappa \frac{(1-\eta)}{1+\kappa+\alpha(\eta-1)}}$), the lack of workers will increase marginal costs under free migration, as to overturn the patterns of specialization.

C.7 Aggregate Comparative Advantage

In this section, we study how migration relates to country H's export specialization in sector k (relative to sector k' and region F). In particular, we study how migration shapes the inequality

$$\frac{X_{HF,k}}{X_{HF,k'}} > \frac{X_{FF,k}}{X_{FF,k'}}. \tag{55}$$

In this section, we focus directly on the case in which land is not a factor of production, or $\gamma = 0$. Together with the assumptions laid out in the paper, (55) is equivalent to

$$\frac{\sum_i \left(w_{i,k}^{\alpha}/A_{i,k}\right)^{1-\eta}}{\sum_i \left(w_{i,k'}^{\alpha}/A_{i,k'}\right)^{1-\eta}} > \frac{w_{F,k}^{\alpha(1-\eta)} A_{F,k}^{\eta-1}}{w_{F,k'}^{\alpha(1-\eta)} A_{F,k'}^{\eta-1}} \tag{56}$$

where we use again the assumption $\tau_{iF,k} = \tau_{iF,k'} = 1$

C.7.1 Prohibitive migration costs

If we substitute equation (50) into the (56), we get

$$\frac{\sum_i \left(s_{i,k} L_{i,k}^0\right)^{\frac{\alpha(\eta-1)}{1+\alpha(\eta-1)}} A_{i,k}^{\frac{\eta-1}{1+\alpha(\eta-1)}}}{\sum_i \left(s_{i,k'} L_{i,k'}^0\right)^{\frac{\alpha(\eta-1)}{1+\alpha(\eta-1)}} A_{i,k'}^{\frac{\eta-1}{1+\alpha(\eta-1)}}} > \frac{\left(s_{F,k} L_{F,k}^0\right)^{\frac{\alpha(\eta-1)}{1+\alpha(\eta-1)}} A_{F,k}^{\frac{\eta-1}{1+\alpha(\eta-1)}}}{\left(s_{F,k'} L_{F,k'}^0\right)^{\frac{\alpha(\eta-1)}{1+\alpha(\eta-1)}} A_{F,k'}^{\frac{\eta-1}{1+\alpha(\eta-1)}}} \tag{57}$$

Expression (57) restates inequality (56) in terms of exogenous forces in the model.

C.7.2 Free migration.

Consider now the case of free labor mobility. In particular, if we substitute equation (53) into the (56), we obtain:

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

A Sufficient Statistic for the Impact of Migration

In this Section we provide a proof for the propositions in Section 4.4 of the paper.

C.8 The impact on the regional direction of trade

We seek to understand the impact of changes in migration costs on the direction of trade

$$\frac{X_{iF,k}/X_{iF,l}}{X_{FF,k}/X_{FF,l}}.$$

Using our model, changes in the direction of trade respond to changes in wages

$$\widehat{\frac{X_{iF,k}/X_{iF,l}}{X_{FF,k}/X_{FF,l}}} = \frac{\hat{w}_{i,k}^{(1-\eta)(1-\gamma)\alpha}/\hat{w}_{i,l}^{(1-\eta)(1-\gamma)\alpha}}{\hat{w}_{F,k}^{(1-\eta)(1-\gamma)\alpha}/\hat{w}_{F,l}^{(1-\eta)(1-\gamma)\alpha}}.$$

Applying hat algebra to equation (31) we obtain:

$$\widehat{w_{i,k}}^{1+\kappa-(1-\gamma)\alpha(1-\eta)}\left[\mathcal{E}_{ii,k}\frac{s_{i,k}^{\kappa}\widehat{P_i^{-\kappa}}}{\sum_l\sum_{i''}\left(s_{i,l}w_{i'',l}\right)^{\kappa}\left(\mu_{ii'',l}P_{i''}\right)^{-\kappa}}s_{i,k}\ell_{ii}\right]+$$

$$\widehat{w_{i,k}}^{1+\kappa-(1-\gamma)\alpha(1-\eta)}\left[\sum_j\mathcal{E}_{ji,k}\frac{s_{j,k}^{\kappa}P_i^{-\kappa}\widehat{\mu_{ji,k}^{-\kappa}}}{\sum_l\sum_{i''}\left(s_{j,l}w_{i'',l}\right)^{\kappa}\left(\mu_{ji'',l}P_{i''}\right)^{-\kappa}}s_{j,k}\ell_{ji}\right]=$$

$$\alpha\left(1-\gamma\right)\delta_{i,k}\left[\left(\frac{\widehat{r_i^{\gamma\alpha}P_i^{1-\alpha}}\tau_{ij,k}}{P_{F,k}}\right)^{1-\eta}X_{F,k}\right]$$

and evaluated at $\mu_{ij,k}\to\infty$ for $i\neq j$, yields

$$\widehat{w_{i,k}} = \left\{\hat{r}_i^{\alpha\gamma(1-\eta)}\left[\mathcal{E}_{ii,k}\frac{\widehat{1}}{\sum_l\sum_{i''}\left(s_{i,l}w_{i'',l}\right)^{\kappa}\left(\mu_{ii'',l}P_{i''}\right)^{-\kappa}}\right]^{-1}\right\}^{\frac{1}{1+\kappa+\alpha(1-\gamma)(\eta-1)}}. \quad (58)$$

Thus, the change in export specialization is given by

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

C.9 The impact on the aggregate direction of trade

The change in the aggregate direction of trade is, when $\gamma=0$:

$$\widehat{\frac{X_{HF,k}/X_{HF,l}}{X_{FF,k}/X_{FF,l}}} = \frac{\sum_{i\in H}S_{i,k}\hat{w}_{i,k}^{(1-\eta)\alpha}}{\sum_{i\in H}S_{i,k}\hat{w}_{i,l}^{(1-\eta)\alpha}}\times\frac{\hat{w}_{F,l}^{(1-\eta)\alpha}}{\hat{w}_{F,k}^{(1-\eta)\alpha}}$$

where $S_{i,k}$ is region $i's$ share in country $H's$ total exports of sector k to country F, in the baseline scenario.

Start by computing changes in welfare as $\mu_{ij,k}\to\infty$:

$$\hat{\Xi}_i^{\kappa} = \left(\sum_k\tilde{\mathcal{E}}_{ii,k}\widehat{w_{i,k}^{\kappa}}\right),$$

where $\tilde{\mathcal{E}}_{ii,k'} = \lambda_{ij,k}/\sum_{k'}\lambda_{ii,k'}$.Substituting $\hat{\Xi}_i$ in equation (58) we obtain

$$\hat{w}_{i,k} = \left\{\mathcal{E}_{ii,k}^{-1}\left(\sum_k \tilde{\mathcal{E}}_{ii,k}\widehat{w_{i,k}^{\kappa}}\right)\right\}^{\frac{1}{1+\kappa+\alpha(\eta-1)}}.$$

Noting that

$$\frac{\hat{w}_{i,k}}{\hat{w}_{i,l}} = \left\{\frac{\mathcal{E}_{ii,l}}{\mathcal{E}_{ii,k}}\right\}^{\frac{1}{1+\kappa+\alpha(1-\gamma)(\eta-1)}},$$

We can rewrite the changes in wages of sector k region i as

$$\hat{w}_{i,k} = \left\{\widehat{w_{i,k}^{\kappa}}\frac{\tilde{\mathcal{E}}_{ii,k}+\sum_{l\neq k}\tilde{\mathcal{E}}_{ii,l}\hat{w}_{i,l}^{\kappa}/\hat{w}_{i,k}^{\kappa}}{\mathcal{E}_{ii,k}}\right\}^{\frac{1}{1+\kappa+\alpha(\eta-1)}}$$

$$\hat{w}_{i,k}^{\frac{1+\alpha(\eta-1)}{1+\kappa+\alpha(\eta-1)}} = \left\{\frac{\tilde{\mathcal{E}}_{ii,k}}{\mathcal{E}_{ii,k}}+\sum_{l\neq k}\frac{\tilde{\mathcal{E}}_{ii,l}}{\mathcal{E}_{ii,k}}\left\{\frac{\mathcal{E}_{ii,l}}{\mathcal{E}_{ii,k}}\right\}^{\frac{\kappa}{1+\kappa+\alpha(1-\gamma)(\eta-1)}}\right\}^{\frac{1}{1+\kappa+\alpha(\eta-1)}}$$

Finally, if we assume that there are no within region distortions to the allocation of labor, nor heterogeneity across workers, this becomes:

$$\hat{w}_{i,k} = \left(\frac{\tilde{\mathcal{E}}_{ii}}{\mathcal{E}_{ii}}\right)^{\frac{1}{1+\kappa+\alpha(\eta-1)}}.$$

Evaluating it in the aggregate bilateral specialization:

$$\frac{\widehat{X_{HF,k}/X_{HF,l}}}{X_{FF,k}/X_{FF,l}} = \frac{\sum_{i\in H}S_{i,k}\left(\frac{\tilde{\mathcal{E}}_{ii}}{\mathcal{E}_{ii}}\right)^{\frac{1-\eta}{1+\kappa+\alpha(\eta-1)}}}{\sum_{i\in H}S_{i,k}\left(\frac{\tilde{\mathcal{E}}_{ii}}{\mathcal{E}_{ii}}\right)^{\frac{1-\eta}{1+\kappa+\alpha(\eta-1)}}}$$

D The Gains from Trade

This Section contains the proofs to the Propositions in Section 7 of the paper. In what follows, we set $\gamma_k = 0$, $\forall k$. Proof of Proposition 2

Let W_i denote the real wage in region i, $W_i = w_i/P_i$. Inverting the domestic trade share, we obtain:

$$W_i = \pi_{ii}^{\frac{1}{\alpha(1-\eta)}} A_i^{\frac{1}{\alpha}},$$

which implies the following changes in real wages in response in changes to fundamentals:

$$\hat{W}_i = (\hat{\pi}_{ii})^{\frac{1}{\alpha(1-\eta)}}\hat{A}_i^{\frac{1}{\alpha}}. \tag{59}$$

Likewise, the implied changes to expected welfare are

$$\hat{\Xi}_i = \left[\sum_j \lambda_{ij}\left(\hat{W}_j\hat{s}_i\hat{\mu}_{ij}^{-1}\right)^{\kappa}\right]^{1/\kappa}, \tag{60}$$

where λ_{ij} are observed migration shares.

We introduce the following notation to indicate four scenarios: (i) B is our baseline with observed trade costs and migration costs, (ii) B,A is the scenario in which, starting from B, we take region i to full trade autarky, (iii) N is the scenario in which, starting from B, we take region i to full migration autarky, and (iv) N, A corresponds to the scenario in which, starting from N, we take region i to full trade autarky.

D.1 Proof of Proposition 2

Observing that $\hat{\pi}_{ii}^{B\to B,A} = \pi_{ii}^{-1}$, i.e., the inverse of the observed trade shares, direct substitution of (59) in (60) yields

$$\hat{\Xi}_i^{B\to B,A} = \left[\sum_j \lambda_{ij} \pi_{ii}^{\frac{\kappa}{\alpha(1-\eta)}}\right]^{1/\kappa}, \tag{61}$$

which completes the proof.

D.2 Proof of Proposition 3

Start by noting that we can write the welfare change from going to autarky, starting from no migration, $N \to N, A$ as

$$\begin{aligned}\hat{\Xi}_i^{N\to N,A} &= \frac{\Xi_i^{N,A}}{\Xi_i^N} \\ &= \frac{\Xi_i^{N,A}}{\Xi_i^B}\frac{\Xi_i^B}{\Xi_i^N} \\ &= \frac{\Xi_i^{N,A}}{\Xi_i^B}\left(\frac{\Xi_i^N}{\Xi_i^B}\right)^{-1}.\end{aligned} \tag{62}$$

We obtain expressions for each of the terms in the last equation.

Applying the same reasoning that led to equation (61), we obtain

$$\left(\frac{\Xi_i^N}{\Xi_i^B}\right)^{-1} = \left(\lambda_{ii}\left(\frac{\pi_{ii}^N}{\pi_{ii}^B}\right)^{\frac{1}{\alpha(1-\eta)}}\right)^{-1},$$

noting that π_{ii}^N is not observed and that π_{ii}^B is simply data. Likewise, we obtain:

$$\frac{\Xi_i^{N,A}}{\Xi_i^B} = \lambda_{ii}\left(\frac{\pi_{ii}^{N,A}}{\pi_{ii}^B}\right)^{\frac{1}{\alpha(1-\sigma)}} = \lambda_{ii}\left(\pi_{ii}^B\right)^{\frac{1}{\alpha(\sigma-1)}}.$$

Substituting the last two expressions in equation (62) we obtain

$$\begin{aligned}\hat{\Xi}_i^{N\to N,A} &= \left(\frac{1}{\pi_{ii}^B}\right)^{\frac{1}{\alpha(1-\sigma)}}\left(\frac{\pi_{ii}^B}{\pi_{ii}^N}\right)^{\frac{1}{\alpha(1-\sigma)}} \\ &= \left(\frac{1}{\pi_{ii}^B}\right)^{\frac{1}{\alpha(1-\sigma)}} T_i,\end{aligned} \tag{63}$$

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

To obtain the result, rewrite equation (61)

$$\hat{\Xi}_i^{B\to B,A} = \left[\lambda_{ij}\left(\pi_{ii}^B\right)^{\frac{\kappa}{\alpha(\sigma-1)}} + \sum_{j\neq i}\lambda_{ij}\left(\pi_{jj}^B\right)^{\frac{\kappa}{\alpha(\sigma-1)}}\right]^{1/\kappa}$$

and use (63) to substitute for π_{ii}^B

$$\hat{\Xi}_i^{B\to B,A} = \left[\lambda_{ii}T^{-\kappa}\left(\hat{\Xi}_i^{N\to N,A}\right)^{\kappa} + \sum_{j\neq i}\lambda_{ij}\left(\pi_{jj}^B\right)^{\frac{\kappa}{\alpha(\sigma-1)}}\right]^{1/\kappa}. \tag{64}$$

D.3 The Multisector Model

The key difficulty in the multi-sector case is that changes trade shares are no longer sufficient statistics for changes in real wages induced by changes in trade costs. Nevertheless, with CES preferences across activities and an elasticity of substitution different from one, one can use changes in observed expenditure shares to proceed.

D.3.1 Gains from Trade

Note first that we can rewrite trade shares as a function of real wages and expenditure shares

$$\pi_{ii,k} = \left(\frac{w_{i,k}^{\alpha_k}P_i^{1-\alpha_k}/A_{i,k}}{P_{i,k}}\right)^{1-\eta}$$
$$\pi_{ij,k} = \left(\frac{W_{i,k}^{\alpha_k}}{A_{i,k}}\frac{P_i}{P_{i,g}}\frac{P_{i,g}}{P_{i,k}}\right)^{1-\eta}$$
$$\pi_{ii,k} = \left(\frac{W_{i,k}^{\alpha_k}}{A_{i,k}}\left(\frac{S_{i,k}}{a_{i,k}}\right)^{\frac{1}{\sigma_g-1}}\left(\frac{S_{i,g}}{b_g}\right)^{\frac{1}{\sigma-1}}\right)^{1-\eta},$$

where we use

$$S_{i,k} = a_{ik}\left(\frac{P_{j,k}}{P_{j,g}}\right)^{1-\sigma_g}$$
$$S_{i,g} = b_g\left(\frac{P_{i,g}}{P_i}\right)^{1-\sigma}$$

We begin by computing the GFT starting from the baseline and going to full trade autarky. Noting that

$$W_{i,k} = \pi_{ii,k}^{\frac{1}{\alpha_k(1-\eta)}}A_{i,k}^{\frac{1}{\alpha_k}}\left(S_{i,k}\right)^{\frac{1}{\alpha_k(1-\sigma_g)}}\left(S_{i,g}\right)^{\frac{1}{\alpha_k(1-\sigma)}},$$

we can compute changes in real wages, $\hat{W}_{i,k}$, and substitute them into the change in expected welfare $\hat{\Xi}_i^{B\to B,A}$:

$$\hat{\Xi}_i^{B\to B,A} = \left[\sum_j\sum_k\lambda_{ij,k}\left(\pi_{ii,k}^{\frac{1}{\alpha_k(\eta-1)}}\hat{S}_{i,k}^{\frac{1}{\alpha_k(1-\sigma_g)}}\hat{S}_{i,g}^{\frac{1}{\alpha_k(1-\sigma)}}\right)^{\kappa}\right]^{1/\kappa}$$

D.3.2 GFT Comparison to the Migration Autarky Scenario

Our goal now is to compare the gains from trade in our baseline scenario to one in which there is no migration. Unfortunately, an exact decomposition such as the one in Proposition 3 is not available. However, we will show that one can cleanly separate the gains arising from migration opportunities, as before.

First, we will see show how real wage changes determine the welfare change going to autarky in a no migration scenario. As before, note that we can decompose the welfare change as

$$\hat{\Xi}_i^{N\to N,A} = \frac{\Xi_i^{N,A}}{\Xi_i^B}\left(\frac{\Xi_i^N}{\Xi_i^B}\right)^{-1}.$$

The first and second terms are given by:

$$\frac{\Xi_i^{N,A}}{\Xi_i^B} = \left(\sum_k \lambda_{ii,k}\left(\frac{s_{i,k}^{N,A}}{s_{i,k}^B}\right)^{\kappa}\left(\frac{1}{\pi_{ii,k}^B}\right)^{\frac{\kappa}{\alpha_k(1-\eta)}}\left(\frac{S_{i,Sk}^{N,A}}{S_{i,Sk}^B}\right)^{\frac{\kappa}{\alpha_k(1-\sigma_A)}}\left(\frac{S_{i,S}^{N,A}}{S_{i,S}^B}\right)^{\frac{\kappa}{\alpha_k(1-\sigma)}}\right)^{1/\kappa}$$

and

$$\left(\frac{\Xi_i^N}{\Xi_i^B}\right)^{-1} = \left(\sum_k \lambda_{ij,k}\left(\frac{s_{i,k}^{N}}{s_{i,k}^B}\right)^{\kappa}\left(\frac{\pi_{ii,k}^N}{\pi_{ii,k}^B}\right)^{\frac{\kappa}{\alpha_k(1-\eta)}}\left(\frac{S_{i,Sk}^{N}}{S_{i,Sk}^B}\right)^{\frac{\kappa}{\alpha_k(1-\sigma_A)}}\left(\frac{S_{i,S}^{N}}{S_{i,S}^B}\right)^{\frac{\kappa}{\alpha_k(1-\sigma)}}\right)^{-1/\kappa}.$$

Putting them together, we obtain

$$\hat{\Xi}_i^{N\to N,A} = \left[\frac{\sum_k \lambda_{ii,k}\left(\frac{1}{\pi_{ii,k}^B}\right)^{\frac{\kappa}{\alpha_k(1-\eta)}}\left(\frac{S_{i,Sk}^{N,A}}{S_{i,Sk}^B}\right)^{\frac{\kappa}{\alpha_k(1-\sigma_A)}}\left(\frac{S_{i,S}^{N,A}}{S_{i,S}^B}\right)^{\frac{\kappa}{\alpha_k(1-\sigma)}}}{\sum_k \lambda_{ii,k}\left(\frac{\pi_{ii,k}^N}{\pi_{ii,k}^B}\right)^{\frac{\kappa}{\alpha_k(1-\eta)}}\left(\frac{S_{i,Sk}^{N}}{S_{i,Sk}^B}\right)^{\frac{\kappa}{\alpha_k(1-\sigma_A)}}\left(\frac{S_{i,S}^{N}}{S_{i,S}^B}\right)^{\frac{\kappa}{\alpha_k(1-\sigma)}}}\right]^{1/\kappa}.$$

Letting

$$\rho_{i,k} = \frac{\lambda_{ii,k}\left(\frac{\pi_{ii,k}^N}{\pi_{ii,k}^B}\right)^{\frac{\kappa}{\alpha_k(1-\eta)}}\left(\frac{S_{i,Sk}^{N}}{S_{i,Sk}^B}\right)^{\frac{\kappa}{\alpha_k(1-\sigma_A)}}\left(\frac{S_{i,S}^{N}}{S_{i,S}^B}\right)^{\frac{\kappa}{\alpha_k(1-\sigma)}}}{\sum_l \lambda_{ii,l}\left(\frac{\pi_{ii,k}^N}{\pi_{ii,k}^B}\right)^{\frac{\kappa}{\alpha_k(1-\eta)}}\left(\frac{S_{i,Sk}^{N}}{S_{i,Sk}^B}\right)^{\frac{\kappa}{\alpha_k(1-\sigma_A)}}\left(\frac{S_{i,S}^{N}}{S_{i,S}^B}\right)^{\frac{\kappa}{\alpha_k(1-\sigma)}}}$$

and

$$\xi_{i,k} = \frac{\left(\frac{1}{\pi_{ii,k}^B}\right)^{\frac{\kappa}{\alpha_k(1-\eta)}}\left(\frac{S_{i,Sk}^{N,A}}{S_{i,Sk}^B}\right)^{\frac{\kappa}{\alpha_k(1-\sigma_A)}}\left(\frac{S_{i,S}^{N,A}}{S_{i,S}^B}\right)^{\frac{\kappa}{\alpha_k(1-\sigma)}}}{\left(\frac{\pi_{ii,k}^N}{\pi_{ii,k}^B}\right)^{\frac{\kappa}{\alpha_k(1-\eta)}}\left(\frac{S_{i,Sk}^{N}}{S_{i,Sk}^B}\right)^{\frac{\kappa}{\alpha_k(1-\sigma_A)}}\left(\frac{S_{i,S}^{N}}{S_{i,S}^B}\right)^{\frac{\kappa}{\alpha_k(1-\sigma)}}}$$

we can rewrite the welfare change as

$$\hat{\Xi}_i^{N\to N,A} = \left[\sum_k \rho_{i,k}\xi_{i,k}\right]^{1/\kappa}.$$

Thus $\{\rho_{i,k}, \xi_{i,k}\}$ and κ fully determine $\hat{\Xi}_i^{N\to N,A}$.

Note that we can write the baseline domestic trade share as

$$\left(\pi_{ii,k}^B\right)^{\frac{\kappa}{\alpha_k(\eta-1)}} = \frac{\xi_{i,k}}{\rho_{i,k}}T_{i,k}.$$

Finally, we can rewrite $\hat{\Xi}_i^{B\to B,A}$ to separate migration opportunities from the components that determine $\hat{\Xi}_i^{N\to N,A}$.

Now we compute $\hat{\Xi}_i^{B\to B,A}$ so

$$\hat{\Xi}_i^{B\to B,A} = \left[\sum_j \sum_k \lambda_{ij,k} \left(\frac{1}{\pi_{jj}^B}\right)^{\frac{\kappa}{\alpha_k(1-\eta)}} \left(\frac{S_{j,Sk}^{B,A}}{S_{j,Sk}^B}\right)^{\frac{\kappa}{\alpha_k(1-\sigma_A)}} \left(\frac{S_{j,S}^{B,A}}{S_{j,S}^B}\right)^{\frac{\kappa}{\alpha_k(1-\sigma)}}\right]^{1/\kappa}$$

$$= \left[\sum_k \lambda_{ii,k} \frac{\xi_{i,k}}{\rho_{i,k}} T_{i,k} \left(\frac{S_{j,Sk}^{B,A}}{S_{j,Sk}^B}\right)^{\frac{\kappa}{\alpha_k(1-\sigma_A)}} \left(\frac{S_{j,S}^{B,A}}{S_{j,S}^B}\right)^{\frac{\kappa}{\alpha_k(1-\sigma)}} + \underbrace{\sum_{j\neq i} \sum_k \lambda_{ij,k} \left(\frac{1}{\pi_{jj}^B}\right)^{\frac{\kappa}{\alpha_k(1-\sigma)}} \left(\frac{S_{j,Sk}^{B,A}}{S_{j,Sk}^B}\right)^{\frac{\kappa}{\alpha_k(1-\sigma_A)}} \left(\frac{S_{j,S}^{B,A}}{S_{j,S}^B}\right)^{\frac{\kappa}{\alpha_k(1-\sigma)}}}_{\text{contribution of migration opportunities}}\right]^{1/\kappa}. \tag{65}$$

Note that this decomposition is analogous to 64, which forms the basis of Proposition 3.

In calculating the contribution of migration opportunities to the GFT in the multi-sector model, we rely on equation 65.

E Calibration Details

This section describes the algorithm that we set up for the calibration of the model. The algorithm can be divided in two steps. In the first step, we calibrate the exogenous parameters related to the goods market equilibrium using data on gross output by sector and region, total exports and imports of Brazil to the rest of the world, and the share of domestic trade of states in Brazil. In the second step, we calibrate the exogenous parameters related to workers' migration using data on migration flows between states and activities and the share of workers living in their region of birth. To simplify notation, we drop time indexes whenever unnecessary for our exposition. In what follows, we denote Brazil as Home and the rest of the world as Foreign.

E.1 Parameters in the calibration algorithm

We begin by discussing a few aspects of the identification of the parameters in our calibration procedure that are important for understanding the algorithm.

Migration costs. We parameterize our matrix of migration costs using a hub-spoke structure (see Ramondo, Rodríguez-Clare, and Saborío-Rodríguez (2016)). In particular, for all regions within a state, workers have to pass through a common location to migrate to other states. This structure allows us aggregate our regional level migration flows to estimate the symmetric component of our migration costs using equation 16 at the aggregate level of states. To see this, write migration cost as $\mu_{ij,k} = \mu_{ss}\mu_i\mu_j\mu_{ss',k}$. With this assumption, migration flows become:

$$L_{ij,k} = \frac{(w_{j,k}s_{i,k}/P_j\mu_{ss}\mu_i\mu_j\mu_{ss',k})^\kappa}{\Xi_i^\kappa} L_i.$$

Let $L_{ss',k} \equiv \sum_{i\in s}\sum_{j\in s'} L_{ij,k}$ be the aggregate flow of workers from state s to state and activity s' and k. Summing the expression above over origins in state s and destinations in state s' gives

$$L_{ss',k} = \sum_{i\in s}\sum_{j\in s'} \frac{(w_{j,k}s_{i,k}/P_i\mu_{ss}\mu_i\mu_j\mu_{ss',k})^\kappa}{\Xi_i^\kappa} L_i.$$

Straightforward manipulations give

$$\log L_{ss',k} = \alpha_{s,k} + \alpha_{s',k} - \kappa \log \mu_{ss'} - \kappa \log \mu_{ss',k},$$

where $\alpha_{s,k} \equiv \sum_{i \in s} \frac{(s_{i,k}/\mu_i)^\kappa}{\Xi_i^\kappa}$ and $\alpha_{s',k} \equiv \sum_{j \in s'} (w_{j,k}/P_j \mu_j)^\kappa$.

Prices and natural advantages. We calibrate natural advantages ($A_{j,k}$) using data on gross output per region and activity. The intuition for the calibration of natural advantages is the same as the one described in detail in Allen and Arkolakis (2014): if two regions have the same distance to other markets, but the model predicts that one region sells more than the other in a given activity, this difference is attributed to differences in natural advantages. In practice, we first search for prices $p_{j,k}$, which has a direct relationship with natural advantages ($A_{j,k}$) given that $p_{j,k} = c_{j,k}/A_{j,k}$, where $c_{j,k}$ is the unit cost of production and $A_{j,k}$ is the absolute advantage. With calibrated values for $p_{j,k}$, in the end of our algorithm, we can construct $c_{j,k}$ using model-implied values of $w_{j,k}$, r_i and P_i and recover $A_{j,k}$.

Trade costs and international trade. We calibrate trade costs τ_k and relative productivities between Home and Foreign as to exactly match trade shares between Home and Foreign in the absence of trade deficits. We illustrate this point using a simple 2 by 2 case (without internal geography). The system of equations defining trade shares is given by

$$\frac{X_{HH,k}}{X_{H,k}} = \frac{(p_{H,k})^{1-\sigma}}{(p_{H,k})^{1-\sigma} + (\tau_k p_{F,k})^{1-\sigma}}$$
$$\frac{X_{HF,k}}{X_{H,k}} = \frac{(\tau_k p_{H,k})^{1-\sigma}}{(p_{H,k})^{1-\sigma} + (\tau_k p_{F,k})^{1-\sigma}}$$
$$\frac{X_{FF,k}}{X_{F,k}} = \frac{(p_{H,k})^{1-\sigma}}{(p_{H,k})^{1-\sigma} + (\tau_k p_{F,k})^{1-\sigma}}$$
$$\frac{X_{HH,k}}{X_{H,k}} = \frac{(p_{H,k})^{1-\sigma}}{(p_{H,k})^{1-\sigma} + (\tau_k p_{F,k})^{1-\sigma}}$$

where $X_{jj',k}$ is the trade flow of country j to j' and $X_{j,k}$ is the total consumption of country j. We have four equations and three variables ($p_{H,k}$, $p_{F,k}$ and τ_k). Two of these equations are dependent and, therefore, satisfying 2 out of the four equations of this system provides a perfect match of the model with the data. We therefore use such equations to recover trade costs τ_k and relative prices $p_{H,k}/p_{F,k}$. Noticed that we do not have any degree of freedom left to calibrate an asymmetric trade cost in the example given above. Alternatively, we could assume that prices are the same across countries $p_{H,k} = p_{F,k}$, but that trade costs are asymmetric.

E.2 Calibration algorithm

Our algorithm calibrates preference shifters ($a_{H,kt}$ and $a_{F,kt}$), productivities ($A_{j,kt}$ and $b_{j,kt}$), migration costs (μ_t^0, $\mu_{ss',kt}$), trade costs (δ_t^0, δ_t^1 and $\delta_{ij,kt}$), taking the following parameters as given: preference parameters (σ and σ_S), production technology parameters (α_k, γ_k, ρ and ς), worker heterogeneity (β and κ), the elasticity of migration cost with respect to distance (μ^1) and the symmetric component of migration costs between states ($\mu_{ss,t}$). As described below, our algorithm consists of three major. In what follows, we drop index t to save on notation unless necessary.

Step 1: Trade equilibrium. In the first step of our calibration, we search for values of prices ($p_{j,k}$), preference shifters ($a_{H,kt}$ and $a_{F,kt}$), international trade costs (δ_k), the level of domestic trade costs (δ^0) and the elasticity of trade cost with respect to distance (δ^1) such that different moments of the data, detailed below, are consistent with a goods market equilibrium in the model. Our calibration requires measures of expenditure X_j by region, which we construct by summing the gross output across activities and distributing Brazil's trade deficit across regions proportionally to X_j.[58]

We define $p_{j,k} \equiv \tilde{p}_{j,k} p_{H,k}$, where $p_{H,k}$ is the average price of goods in Home and $\sum_{j \in H} \tilde{p}_{j,k} = 1$, use superscripts g for guessed values and let $\widehat{x}$ be model-implied value for observed value x. Using these definitions, the algorithm consists of the following steps:

1. Guess values for $\left(\delta^0\right)^g$ and $\left(\delta^1\right)^g$
2. Guess values for $(\tilde{p}_{j,k})^g$, $(b_{j,k})^g$, $(a_{H,k})^g$ and $(a_{F,k})^g$
3. Guess values for $(\delta_k)^g$ and $(p_{H,k})^g$
4. Compute τ_{ijk} using the guesses of $\left(\delta^0\right)^g$, $\left(\delta^1\right)^g$ and $(\delta_k)^g$
5. Compute prices $p_{j,k}$, $P_{j,k}$, $P_{j,s}$ and P_j
6. Compute model-implied trade flows $\widehat{X}_{HF,k}$ and $\widehat{X}_{FH,k}$, revenues $\widehat{Y}_{F,k}$, $\widehat{Y}_{H,k}$, and $\widehat{Y}_{j,k}$ and share of trade between states within Home $\sum_k \sum_s \widehat{X}_{ss,k} / \sum_k \widehat{X}_{HH,k}$
7. Compute the differences between model-implied trade flows $\widehat{X}_{HF,k}$ and $\widehat{X}_{FH,k}$ with data $X_{HF,k}$ and $X_{FH,k}$. If the differences are smaller than a tolerance level, move to next step, otherwise, generate new values for $(\delta_k)^g$ and $(p_{H,k})^g$ with an updating rule and return to step 3
8. Compute the differences between model-implied revenues $\widehat{Y}_{F,k}$, $\widehat{Y}_{H,k}$ and $\widehat{Y}_{j,k}$ with data $Y_{F,k}$, $Y_{H,k}$ and $Y_{j,k}$. If the differences are smaller than a tolerance level, move to the next step, otherwise, generate new values for $(\tilde{p}_{j,k})^g$, $(b_{j,k})^g$ and $(a_{j,k})^g$ using an updating rule and return to step 3.
9. Compute the difference between model-implied $\sum_k \sum_s \frac{\widehat{X}_{ss,k}}{\sum_k \widehat{X}_{HH,k}}$ and data $\sum_k \sum_s \frac{X_{ss,k}}{\sum_k X_{HH,k}}$, as well as the difference between the model-implied elasticity of trade with respect to distance and the corresponding elasticity in the data. If the differences are smaller than a tolerance level, finish the calibration. Otherwise, generate new values for $\left(\delta^0\right)^g$ and $\left(\delta^1\right)^g$ using an updating rule and return to step 1

Step 2: Migration equilibrium In the second step of our calibration, we search for values of migration costs (μ_0), migration shocks ($\tilde{\mu}_{ss',k}$), land intensity γ_k such that different moments of the data are consistent with workers migration choices in the model. We use the price index of every region generated in the previous calibration step and construct the supply of efficiency labor in every origin region $E_{j,k}$ using $s_{i,kt} = L^{\beta}_{i,kt-1}$ if $L_{i,kt-1} > 0$ and $s_{i,kt} = \min\{L^{\beta}_{i,kt-1}\}$ if $L_{i,kt-1} = 0$.

The steps for the calibration of migration costs are as follows:

1. Guess values for $(\mu_0)^g$ and $(\tilde{\mu}_{ss',k})^g$
2. Compute $\mu_{ss,k}$ using the guesses of $(\mu_0)^g$ and $(\tilde{\mu}_{ss',k})^g$
3. Guess values of $(\gamma_k)^g$
4. Guess values of $(w_{j,k})^g$ and $(r_j)^g$
5. Compute real income in each region using $w^0_{j,k}$ and P_j
6. Compute model-implied migration $\widehat{L}_{ss',k}$ and $\widehat{L}_{ii}$, efficiency labor $\widehat{E}_{j,k}$, land use $\widehat{H}_j$ and cost share of labor in Home

[58]In the exercises in which we study the gains from trade, we move the economy to autarky. To avoid endogeneizing the deficits or keeping them constant when there is no trade (or more generally, taking a stance on how they change in our counterfactuals), we set deficits to zero in all simulations of our model.

7. Compute predicted demand for efficiency labor $\widehat{E}_{j,k}$ and for land $\widehat{H}_j$

8. Compute the differences between model-implied $\widehat{E}_{j,k}$ and $\widehat{H}_j$ and $E_{j,k}$ and H_j. If the differences are smaller than a tolerance level, move to the next step, otherwise, generate new values for $(w_{j,k})^g$ and $(r_j)^g$ using an updating rule and return to step 5.

9. Compute the difference between model-implied aggregate cost share of labor in Home and the difference with the data. If the difference is smaller than a tolerance level, move to the next step, otherwise, generate new values for $(\gamma_k)^g$ using an updating rule and return to step 3.

10. Compute the differences between $\widehat{L}_{ss',k}$ and the average of $\widehat{L}_{ii}$ in Home and their corresponding values in the data. If the differences are smaller than a tolerance level, move to the next step, otherwise, generate new values for $(\mu_0)^g$ and $(\tilde{\mu}_{ss',k})^g$ using an updating rule and return to step 1

Step 3: Natural advantages and land supply productivity

Once with values for model-implied wages $(w_{j,k})$, land rents (r_j) and price indexes (P_i), we can construct the unit cost of production $c_{j,kt}$ and recover the natural advantage of a region from $A_{j,k} = c_{j,k}/A_{j,k}$. For the land supply, we recover land productivity using $b_{j,t} = (\varsigma P_{j,t}/r_{j,t})^{1/\varsigma} (H_{j,t})^{(\varsigma-1)/\varsigma}$.

E.3 Additional analysis of migration costs

To validate our migration costs, we proceed with two exercises. First, we compare the migration costs estimated in our regressions to a common approach used by the literature, which is based on the Head and Ries index. Specifically, to quantify migration costs between the west and the east in our model, we estimated equation 16 and recovered $\mu_{ss',t}$ from the fixed costs. An alternative approach is to disregard the sector dimension of the data, aggregate up all the migration fluxes to the state level, and compute the Head and Ries index given by:

$$HR_{ss',t} = \left(\frac{L_{ss',t}}{L_{ss,t}} \times \frac{L_{s's,t}}{L_{s's',t}} \right)^{-\frac{1}{2\kappa}}.$$

Unreported results show that migration costs between regions based on the formula above give results that are extremely close to the ones obtained from the fixed effect approach that we adopted, with a correlation of 0.98. It shows that the results that we obtain are similar to the ones that would be obtained using a common approach used in the literature.

F Quantitative Drivers of International GFT

In Section 7.3, we discussed the impact of limiting East-West migration on the international gains from trade. Here, we report additional details on the mechanisms behind those results. First, in the main body of the paper, we computed the ratio of the GFT in a model with one activity relative to a model with multiple ones. We used this statistic to examine the role of comparative advantage in shaping the GFT. The correlation between the changes in the international GFT (between the baseline and the counterfactual) and changes in this statistic (also between the baseline and the counterfactual) equals 0.51. Figure A.11 studies the relationships between wages, employment, and trade. Panel (a) shows that agricultural wages (relative to manufacturing) increase substantially when we limit migration, up to 60 percent in some regions, revealing that Eastern migrants tend to work in agriculture when they migrate to the West. In the East, agricultural wages drop only modestly. As a consequence, regions in the West experience a drop in the share of revenue coming from exports, but accompanied by an increase in the share of manufacturing in total exports in the West (as shown in Panel (b)). These figures show that the East-West migration deepened the specialization of the West in activities in agriculture, in which it has a strong comparative advantage.

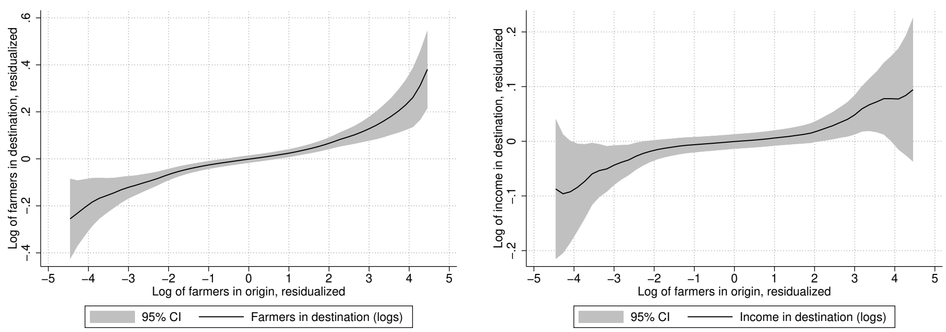
G Additional Figures and Tables

Figure A.1: Federal Government Propaganda about the March to the West, 1940s

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

Figure A.2: Local Polynomial Regressions of the Influence of the Region of Origin on Crop Choice and Income of Farmers in their Destination Region

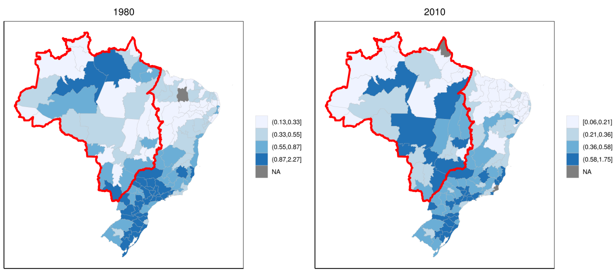
(a) Farmers in destination, residualized (b) Income in destination, residualized



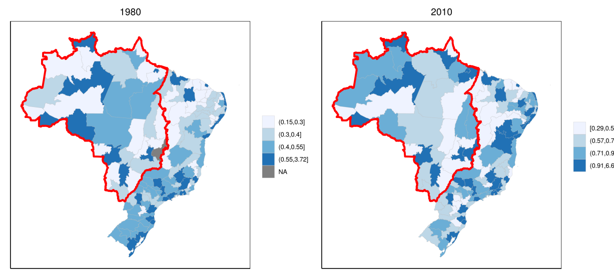
Notes: To construct this figure, we first absorb origin-destination-year and destination-crop-year fixed effects from dependent and independent variables of interest in equations 2 and 3. Using the residuals from these variables, we run a non-parametric regression using a local polynomial smooth.

Figure A.3: Calibrated Wedges: Productivities Relative to Foreign I

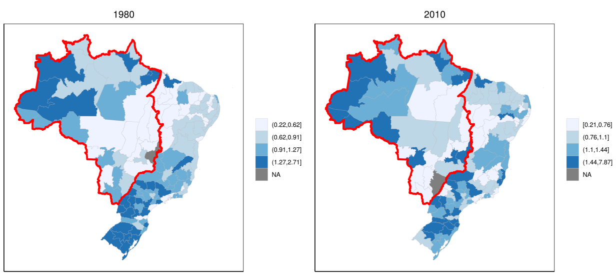
(a) Services



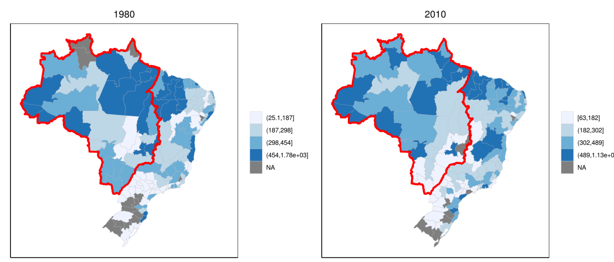
(b) Manufacturing



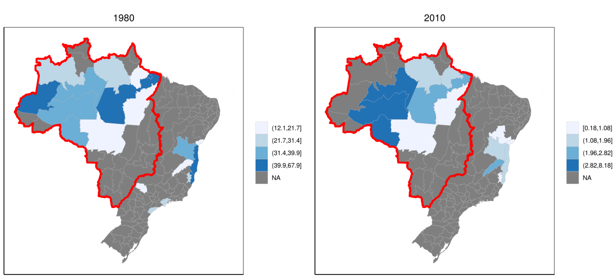
(c) Rest of Agriculture



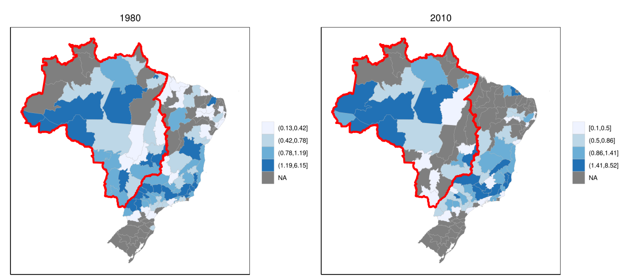
(d) Banana



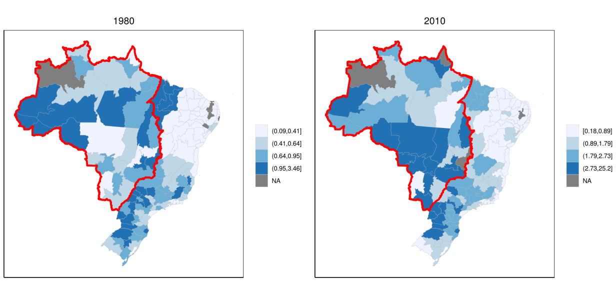
(e) Cacao



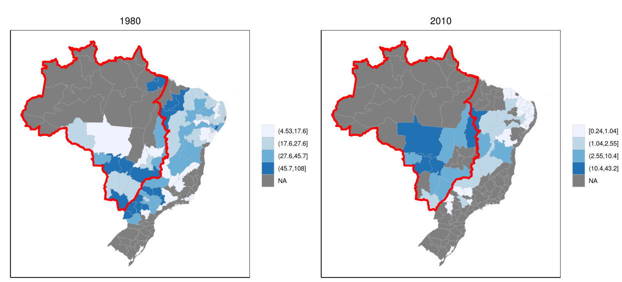
(f) Coffee



(g) Corn



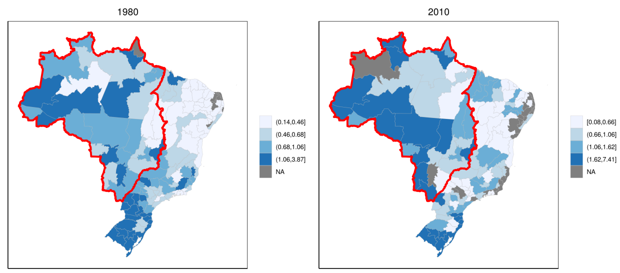
(h) Cotton



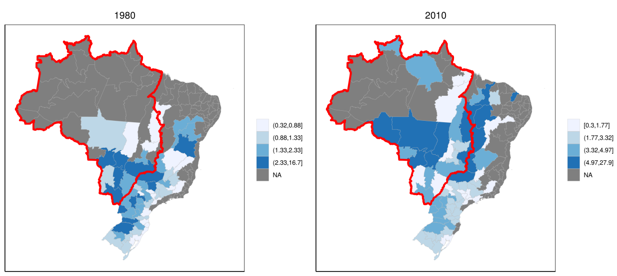
Notes: The figure plots calibrated values of exogenous productivity, $A_{i,k}$, in each region relative to that of Foreign, in a given year.

Figure A.4: Calibrated Wedges: Productivities Relative to Foreign II

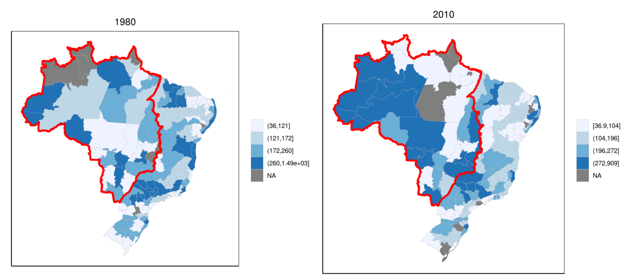
(a) Rice



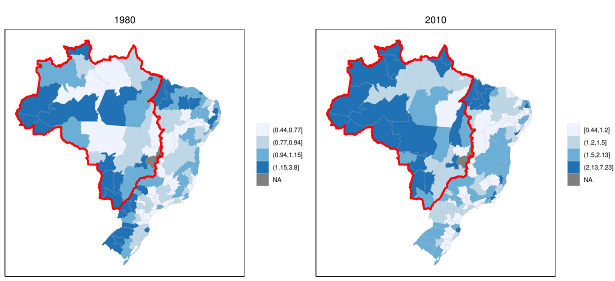
(b) Soy



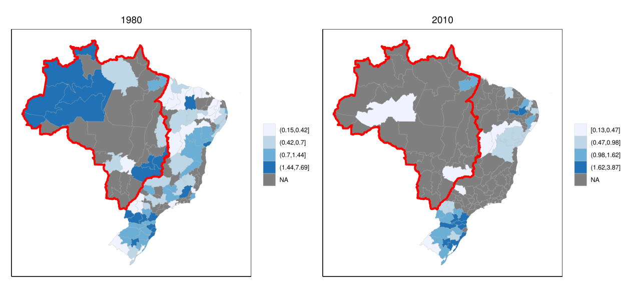
(c) Sugarcane



(d) Beef



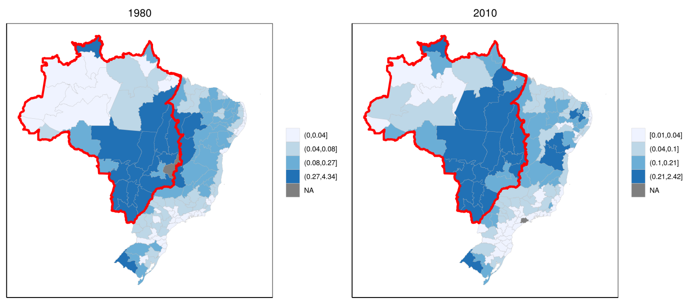
(e) Tobacco



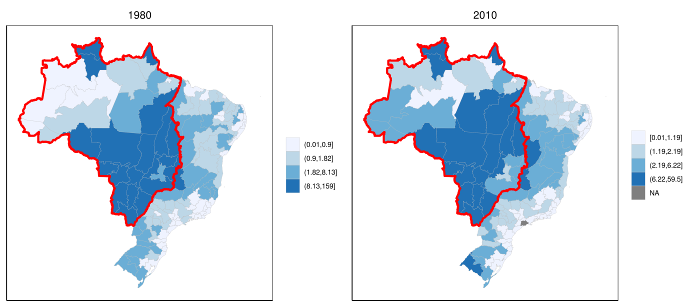
Notes: The figure plots calibrated values of exogenous productivity, $A_{i,k}$, in each region relative to that of Foreign, in a given year.

Figure A.5: Calibrated Wedges: Land Production

(a) Land Production Productivity relative to Foreign



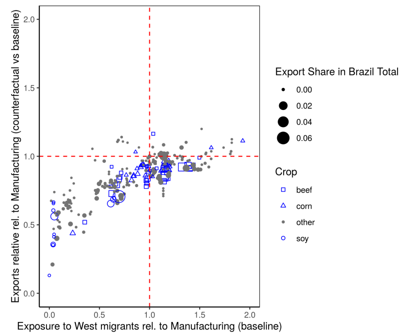
(b) Land to Labor relative to Foreign



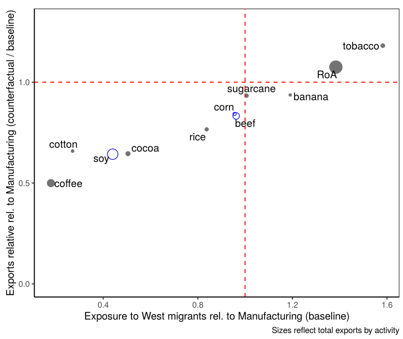
Notes: Panel (a) plots calibrated values of exogenous land production productivity, g_i, in each region relative to that of Foreign, in a given year. Panel (b) plots the observed land to labor ratio in each region relative to that of Foreign.

Figure A.6: Counterfactual Changes for 1980

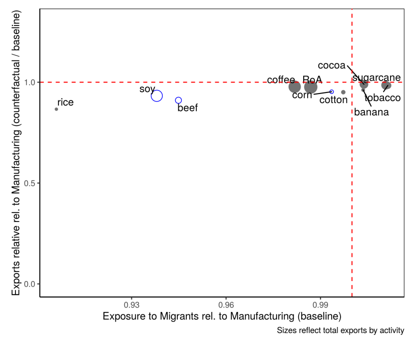
(a) Regional, 1980



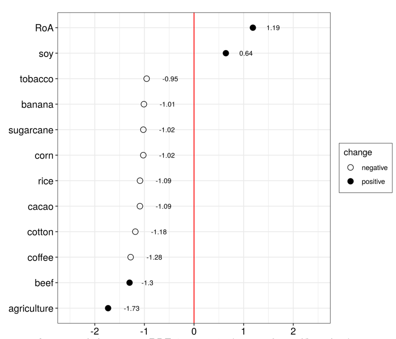
(b) West, 1980



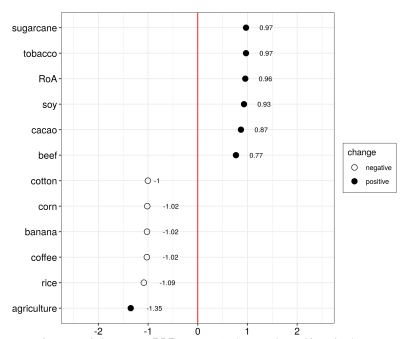
(c) Brazil, 1980



(d) West, 1980



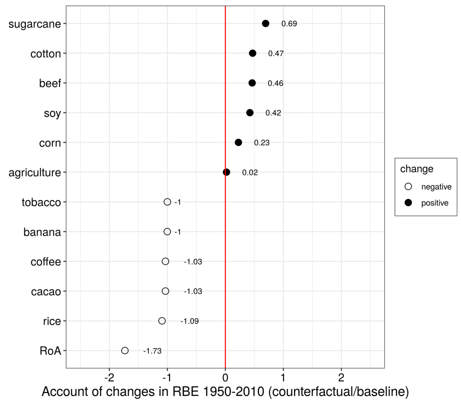
(e) Brazil, 1980



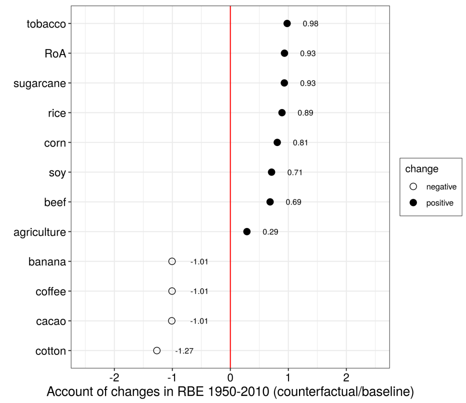
Notes: See the notes of Figures 5, 6, and 7.

Figure A.7: Accounting for Observed Changes in Specialization

(a) West



(b) Brazil



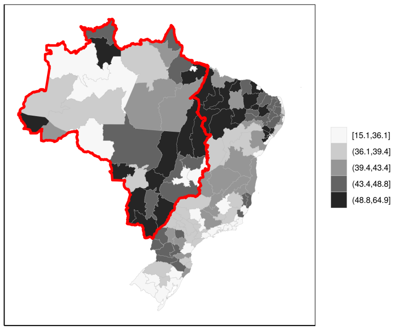
Notes: These figures show the counterfactual change in specialization for each activity, relative to the that in the data, between 1950 and 2010. Specifically, each panel plots, for each activity k

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

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

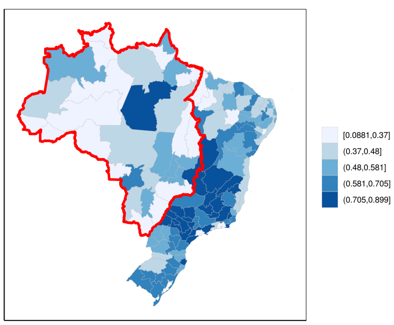
Figure A.8: Comparative Advantage, Migration, and the Losses from Autarky, 1980

(a) The Losses from Autarky

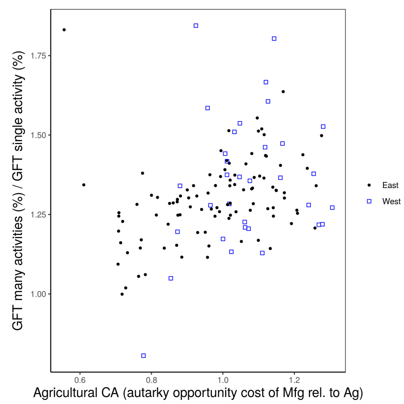


(b) The Contribution of Migration Opportunities

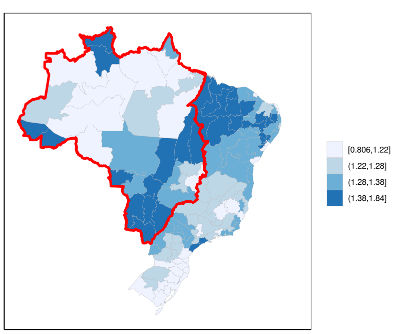
(c) The Contribution of Migration Opportunities



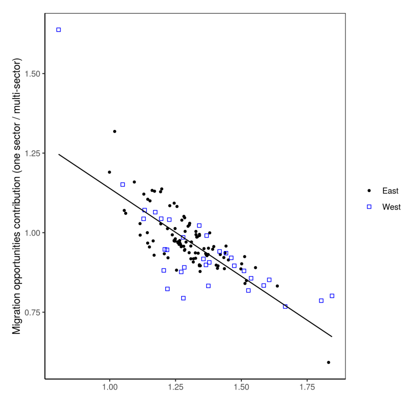
(d) The Contribution of Comparative Advantage



(e) Mapping the Contribution of Comparative Advantage



(f) Mapping the Contribution of Comparative Advantage



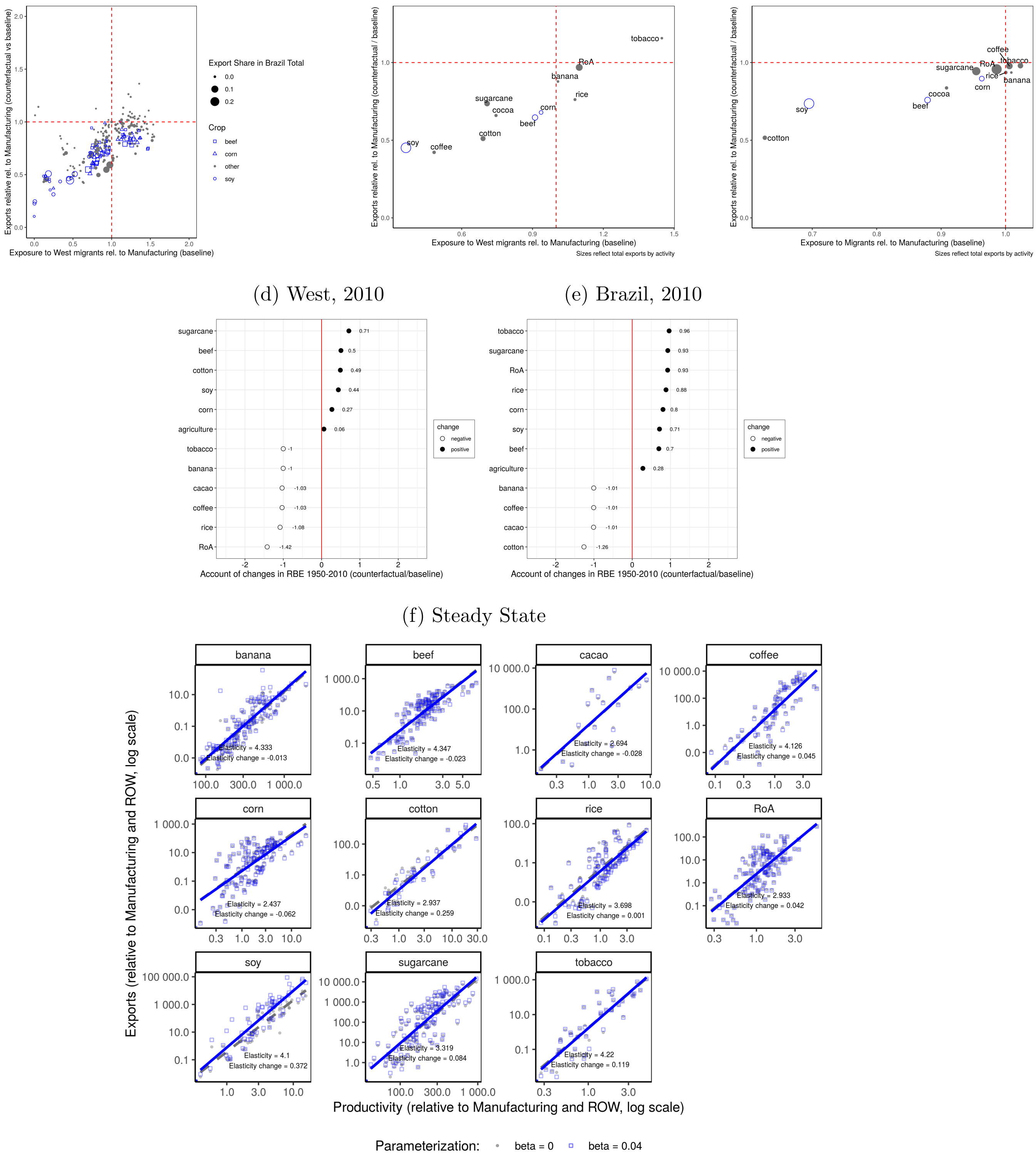
Notes: All simulations are for 1980 and each observation is a region. Panel (a) shows the welfare losses from letting each region go to full trade autarky. Panel (b): The horizontal axis the fraction of workers leaving that region in the baseline simulation. The vertical axis measures the ratio of the welfare cost that results solely from migration opportunities (i.e. setting the domestic contribution to zero) to the total costs. Panel (c): See description for Panel (b). Panel (d): The vertical axis plots the ratio of the losses from autarky in a single-sector version of our model to those in our multi-sector model, showing how the losses increase due to intersectoral heterogeneity. The horizontal axis shows the autarky opportunity cost of manufacturing relative to agriculture. Panel (e) shows the horizontal axis in Panel (d). Panel (f): The vertical axis is the same as in in Panel (d). The horizontal axis shows the change in the contribution of migration opportunities in going from a one-sector to a multi-sector version of our model.

Figure A.9: Knowledge as a function of Employment Shares

(a) Regional, 2010 (b) West, 2010 (c) Brazil, 2010

(d) West, 2010 (e) Brazil, 2010

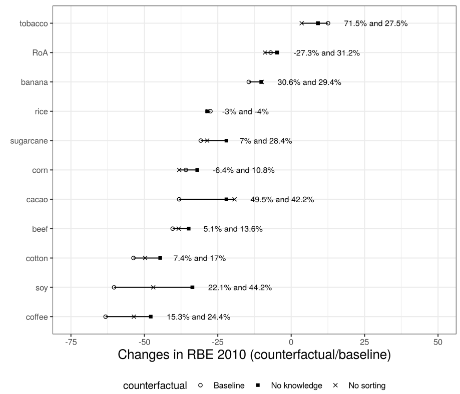
(f) Steady State



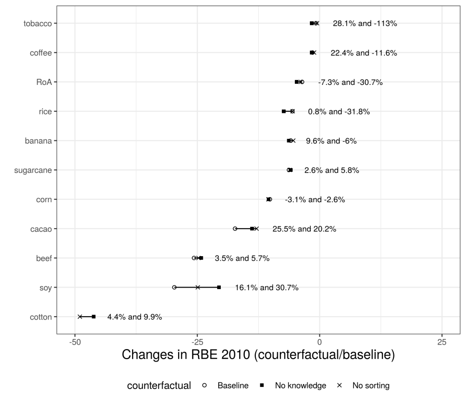
Notes: The figures present our main results in a calibration where $s_{i,k} = \bar{s}\left(L_{i,kt-1}/L_{iAg,t-1}\right)^{\beta}$

Figure A.10: Measuring the Contribution of Knowledge 2010

(a) In the West



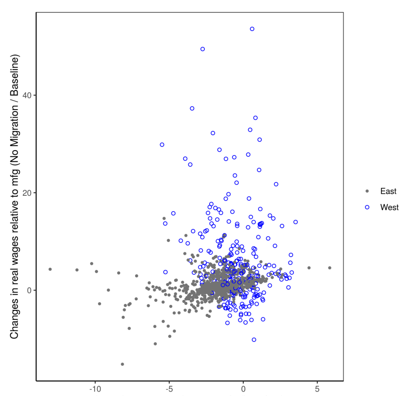
(b) In Brazil



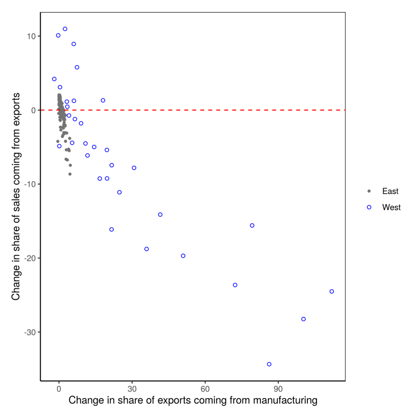
Notes. Each row is an activity aggregate. The hollow circle presents the counterfactual change in export specialization in our baseline calibration, which targets state-state-employment flows. The cross presents the counterfactual change in specialization in a calibration that targets state-state migration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

Figure A.11: The Impact of the March on Real Wages, Employment and Exports

(a) Real Wages



(b) Exports/Sales vs Mfg Exports/Exports



Notes: All simulations are for 2010. This panel reports the impact of limiting East-West migration. In panels (a) each observation is a pair of agricultural activity and region. In panels (b) each observation is a region. Panel (a) shows changes in agricultural wages relative to manufacturing between the counterfactual economy and the baseline on the y-axis and the baseline log of wages in agricultural activities relative to manufacturing on the x-axis. Panel (d) shows changes in foreign exports relative to total sales on the y-axis and changes in the share of manufacturing in total exports in the x-axis (correlation = -0.83).

Table A.1: Evolution of Revealed Comparative Advantage

	Brazil			East	West
	1950	1980	2010	2010	2010
	(1)	(2)	(3)	(4)	(5)
banana	1.92	1.29	0.36	0.43	0.00
cacao	17.08	14.98	1.26	1.51	0.01
coffee	35.42	17.58	15.34	18.33	0.29
corn	1.53	0.34	6.70	2.23	29.22
cotton	2.96	0.50	5.72	3.93	14.76
beef	2.43	3.17	9.53	5.95	27.59
rice	0.59	0.27	1.35	1.59	0.15
soy	0.00	15.33	22.56	15.14	59.96
sugarcane	3.35	6.08	29.03	33.21	7.97
tobacco	1.48	4.07	6.37	7.63	0.00
rest of agriculture	0.71	1.66	1.69	1.77	1.30
agriculture	3.37	3.71	4.47	4.06	6.55
manufacturing	0.27	0.63	0.73	0.76	0.57

Notes: This table shows the evolution of revealed comparative advantage as measured by the Balassa index:

$$RCA_{i,k} = \frac{X_{i,k} / \sum_k X_{i,k}}{\sum_i X_{i,k} / \sum_i \sum_k X_{i,k}},$$

where $X_{i,k}$ are exports from country i in activity k.

Table A.2: Aggregate Summary Statistics

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

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

Table A.3: Summary Statistics by Activity (in percentages)

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

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

Table A.4: The Relationship between Farmers' Choices and Earning and their Region of Origin (OLS - Robustness)

Geographic Unit	Meso	Meso	Meso	Meso	Meso	Micro	Meso	Meso
Lag (years)	30	20	10	20	10	30	30	30
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
a. Farmers in destination (logs)								
Farmers in origin	0.075***	0.086***	0.099***	0.097***	0.112***	0.072***	0.076***	0.081***
	(0.014)	(0.011)	(0.012)	(0.008)	(0.006)	(0.013)	(0.014)	(0.010)
R^2	0.751	0.738	0.738	0.771	0.779	0.775	0.752	0.759
Obs	7375	8443	8393	14449	24604	15437	7375	9597
b: Income (logs)								
Farmers in origin	0.023**	0.016	0.023**	0.003	0.013***	0.023*	0.026**	0.024***
	(0.010)	(0.010)	(0.010)	(0.007)	(0.004)	(0.012)	(0.010)	(0.008)
R^2	0.702	0.677	0.682	0.665	0.659	0.730	0.729	0.688
Obs	6794	7727	7685	13640	23530	14132	6794	8844
c. Farmers in destination (logs) - Above Q1								
Farmers in origin	0.101***	0.144***	0.166***	0.140***	0.167***	0.097***	0.103***	0.085***
	(0.022)	(0.018)	(0.019)	(0.014)	(0.010)	(0.021)	(0.022)	(0.017)
R^2	0.774	0.770	0.774	0.800	0.811	0.791	0.777	0.779
Obs	5609	6422	6395	10942	18614	11787	5609	7271
b: Income (logs) - Above Q1								
Farmers in origin	0.047***	0.051***	0.051***	0.039***	0.024***	0.037*	0.049***	0.039***
	(0.016)	(0.017)	(0.016)	(0.013)	(0.008)	(0.019)	(0.016)	(0.013)
R^2	0.729	0.704	0.712	0.687	0.682	0.746	0.755	0.715
Obs	5180	5903	5883	10365	17860	10872	5180	6712
Dest-Act-Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Dest-Orig-Year FE	Y	Y	Y	Y	Y	Y	Y	Y
SES							Y	
Years: 2000-2010	Y	Y	Y			Y	Y	Y
Years: 1990-2010				Y				
Years: 1980-2010					Y			
Age: 30-60	Y	Y	Y	Y	Y	Y	Y	Y
Age: 20-								Y

Notes: * / ** / + denotes significance at the 10 / 5 / 1 percent level. Standard errors clustered at the destination-crop-year level in parenthesis. This table replicates Panels (a) and (b) of Table 2 using several alternative specifications relative to our main results in the paper based on OLS. Our sample exclude non-migrants.

Table A.5: The Relationship between Farmers' Choices and Earning and their Region of Origin (PPML - Robustness)

Geographic Unit	Meso	Meso	Meso	Meso	Meso	Micro	Meso	Meso
Lag (years)	30	20	10	20	10	30	30	30
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
a. Farmers in destination (logs)								
Farmers in origin	0.120***	0.129***	0.148***	0.147***	0.168***	0.074***	0.121***	0.129***
	(0.013)	(0.011)	(0.012)	(0.009)	(0.007)	(0.012)	(0.013)	(0.011)
Obs	7375	8443	8393	14449	24604	15437	7375	9597
b: Income (logs)								
Farmers in origin	0.045***	0.043***	0.036***	0.018*	-0.003	0.029*	0.044***	0.023*
	(0.012)	(0.013)	(0.012)	(0.010)	(0.010)	(0.016)	(0.012)	(0.013)
Obs	6794	7727	7685	13640	23530	14132	6794	8844
c. Farmers in destination (logs) - Above Q1								
Farmers in origin	0.131***	0.182***	0.186***	0.181***	0.215***	0.104***	0.127***	0.133***
	(0.023)	(0.017)	(0.019)	(0.013)	(0.010)	(0.019)	(0.022)	(0.017)
Obs	5609	6422	6395	10942	18614	11787	5609	7271
b: Income (logs) - Above Q1								
Farmers in origin	0.083***	0.081***	0.059***	0.056***	0.006	0.042**	0.083***	0.044***
	(0.020)	(0.019)	(0.017)	(0.016)	(0.020)	(0.018)	(0.022)	(0.017)
Obs	5180	5903	5883	10365	17860	10872	5180	6712
Dest-Act-Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Dest-Orig-Year FE	Y	Y	Y	Y	Y	Y	Y	Y
SES							Y	
Years: 2000-2010	Y	Y	Y			Y	Y	Y
Years: 1990-2010				Y				
Years: 1980-2010					Y			
Age: 30-60	Y	Y	Y	Y	Y	Y	Y	Y
Age: 20-								Y

Notes: * / ** / $^{+}$ denotes significance at the 10 / 5 / 1 percent level. Standard errors clustered at the destination-crop-year level in parenthesis. This table replicates Panels (a) and (b) of Table 2 using several alternative specifications relative to our main results in the paper based on PPML. Our sample exclude non-migrants.

Table A.6: The Relationship between Farmers' Income and their Region of Origin (Individual Level Regressions)

	OLS	OLS	OLS	PPML	PPML
	(1)	(2)	(3)	(4)	(5)
a. Income (logs)					
Farmers in origin	0.006	0.014**	0.043***	0.045***	0.082***
	(0.004)	(0.006)	(0.013)	(0.010)	(0.019)
R^2	0.257	0.386	0.389	-	-
Obs	18913	18913	13841	18913	13841
b. Income (logs) - SES controls					
Farmers in origin	0.006*	0.012**	0.033**	0.038***	0.070***
	(0.003)	(0.006)	(0.013)	(0.010)	(0.020)
R^2	0.366	0.460	0.459	-	-
Obs	18913	18913	13841	18913	13841
c. Income (logs) - Controls for previous migration					
Farmers in origin	0.006*	0.012**	0.033**	0.038***	0.070***
	(0.003)	(0.006)	(0.013)	(0.010)	(0.020)
R^2	0.366	0.460	0.459	-	-
Obs	18913	18913	13841	18913	13841
d. Income (logs) - Migrants from state of birth					
Farmers in origin	0.009	0.005	0.047***	0.033***	0.087***
	(0.005)	(0.008)	(0.017)	(0.012)	(0.022)
R^2	0.280	0.411	0.410	-	-
Obs	11964	11964	9340	11964	9340
Dest-Act-Year FE	Y	Y	Y	Y	Y
Dest-Orig-Year FE		Y	Y	Y	Y
Above Q1			Y		Y

Notes: See table 2 for a description of columns. This table replicates Panel (a) of Table 2 using individual level data. Our sample excludes non-migrants.

Table A.7: The Relationship between Farmers' Choices and Income and their Region of Origin - State Level Regressions

	OLS	OLS	OLS	PPML	PPML	PPML
	(1)	(2)	(3)	(4)	(5)	(6)
a. Farmers in destination (logs)						
Farmers in origin	0.289***	0.157***	0.210***	0.210***	0.236***	0.213***
	(0.017)	(0.015)	(0.021)	(0.019)	(0.017)	(0.018)
R^2	0.462	0.849	0.883	-	-	-
Obs	2750	2750	2018	2750	2018	7948
b: Income (logs)						
Farmers in origin	0.061***	0.032***	0.033**	0.032**	0.039**	-
	(0.008)	(0.009)	(0.014)	(0.013)	(0.018)	-
R^2	0.361	0.612	0.639	-	-	-
Obs	2573	2573	1900	2573	1900	
Dest-Act-Year FE	Y	Y	Y	Y	Y	Y
Dest-Orig-Year FE		Y	Y	Y	Y	Y
Above Q1			Y		Y	
Include zeros						Y

Notes: * / ** / *** denotes significance at the 10 / 5 / 1 percent level. Standard errors clustered at the destination-crop-year level in parenthesis. This table replicates Panels (a) and (b) of Table 2 using state-level variation. Farmers in origin is defined by the state of birth. Here we include only 2000 and 2010, which are the years included in our main analysis.

Table A.8: The Relationship between Farmers' Choices and Income and their Region of Origin - State Level Regressions - All Years

	OLS	OLS	OLS	PPML	PPML	PPML
	(1)	(2)	(3)	(4)	(5)	(6)
a. Farmers in destination (logs)						
Farmers in origin	0.271***	0.126***	0.156***	0.162***	0.178***	0.164***
	(0.015)	(0.012)	(0.017)	(0.014)	(0.021)	(0.014)
R^2	0.445	0.871	0.901	-	-	-
Obs	4492	4492	3282	4492	3282	12080
b: Income (logs)						
Farmers in origin	0.048***	0.023***	0.029***	0.015*	0.024**	-
	(0.006)	(0.006)	(0.009)	(0.009)	(0.011)	-
R^2	0.404	0.622	0.662	-	-	-
Obs	4286	4286	3146	4286	3146	
Dest-Act-Year FE	Y	Y	Y	Y	Y	Y
Dest-Orig-Year FE		Y	Y	Y	Y	Y
Above Q1			Y		Y	
Include zeros						Y

Notes: * / ** / *** denotes significance at the 10 / 5 / 1 percent level. Standard errors clustered at the destination-crop-year level in parenthesis. This table replicates Panels (a) and (b) of Table 2 using state-level variation. Farmers in origin is defined by the state of birth. Here, we include 1980, 1990, 2000 and 2010, which are the years with available data.

Table A.9: The Relationship between Farmers' Choices and Income and their Region of Origin with Controls for previous Migration

	OLS	OLS	OLS	PPML	PPML
	(1)	(2)	(3)	(4)	(5)
a. Farmers in destination (logs)					
Farmers in origin	0.074***	0.075***	0.096***	0.120***	0.118***
	(0.007)	(0.014)	(0.024)	(0.013)	(0.023)
R^2	0.183	0.751	0.775	-	-
Obs	7375	7375	5478	7375	5478
b. Income (logs)					
Farmers in origin	0.016***	0.023**	0.076***	0.045***	0.076***
	(0.005)	(0.010)	(0.019)	(0.012)	(0.019)
R^2	0.342	0.702	-	-	-
Obs	6794	6794	5056	6794	5056
c. Farmers in destination (logs) - Controls for previous migration					
Farmers in origin	0.066***	0.074***	0.097***	0.120***	0.119***
	(0.007)	(0.013)	(0.024)	(0.013)	(0.022)
R^2	0.192	0.751	0.776	-	-
Obs	7375	7375	5478	7375	5478
d. Income (logs) - Controls for previous migration					
Farmers in origin	0.016***	0.023**	0.079***	0.045***	0.079***
	(0.005)	(0.010)	(0.019)	(0.012)	(0.019)
R^2	0.342	0.703	-	-	-
Obs	6794	6794	5056	6794	5056
e. Farmers in destination (logs) - Migrants from state of birth					
Farmers in origin	0.088***	0.117***	0.101***	0.161***	0.136***
	(0.009)	(0.019)	(0.032)	(0.019)	(0.027)
R^2	0.237	0.752	0.766	-	-
Obs	4794	4794	3560	4794	3560
f. Income (logs) - Migrants from state of birth					
Farmers in origin	0.020***	0.019	0.063***	0.036***	0.063***
	(0.006)	(0.013)	(0.024)	(0.011)	(0.024)
R^2	0.360	0.706	-	-	-
Obs	4462	4462	3326	4462	3326
Dest-Act-Year FE	Y	Y	Y	Y	Y
Dest-Orig-Year FE		Y	Y	Y	Y
Above Q1			Y		Y
Include zeros					

Notes: * / ** / *** denotes significance at the 10 / 5 / 1 percent level. Standard errors clustered at the destination-crop-year level in parenthesis. This table replicates Panels (a) and (b) of Table 2 adding controls for previous migration history. In Panels (c) and (d), we include the share of return migrants and the share of farmers who come from their region of origin, which control for third migration of workers between regions. In Panels (e) and (f), we run regressions only with workers who come from their state of birth, who represent 40 percent of our sample.

Table A.10: The Relationship between Workers' Choices and Income and their Region of Origin - Manufacturing Activities

	OLS	OLS	OLS	PPML	PPML	PPML
	(1)	(2)	(3)	(4)	(5)	(6)
a. Income (logs)						
Workers in origin	0.075***	0.051***	0.063***	0.060***	0.059***	-
	(0.006)	(0.016)	(0.021)	(0.018)	(0.021)	-
R^2	0.380	0.673	0.682	-	-	-
Obs	6921	6921	5410	6921	5410	
b. Workers in destination (logs)						
Workers in origin	0.077***	0.111***	0.142***	0.152***	0.188***	0.195***
	(0.006)	(0.016)	(0.022)	(0.015)	(0.018)	(0.013)
R^2	0.284	0.686	0.708	-	-	-
Obs	6939	6939	5427	6939	5427	166153
Dest-Act-Year FE	Y	Y	Y	Y	Y	Y
Dest-Orig-Year FE		Y	Y	Y	Y	Y
Above Q1			Y		Y	
Include zeros						Y

Notes: * / ** / *** denotes significance at the 10 / 5 / 1 percent level. Standard errors clustered at the destination-crop-year level in parenthesis. This table replicates Panels (a) and (b) of Table 2 using data for manufacturing workers. We include the years of 2000 and 2010. The manufacturing activities included are: automotive, leather, furniture, processed tobacco, oil, paper, clothing, textile, perfume, and wood.

Table A.11: Relationship between the Composition of Farmers and Agricultural Output

	Revenues			Quantity		
Explanatory Variable	(1)	(2)	(3)	(4)	(5)	(6)
Abundance	0.911***	0.874***	0.887***	0.966***	0.900***	0.914***
	(0.040)	(0.066)	(0.064)	(0.042)	(0.068)	(0.067)
Composition	0.198***	0.192***	0.183***	0.248***	0.241***	0.232***
	(0.046)	(0.046)	(0.047)	(0.051)	(0.051)	(0.052)
R^2	0.858	0.859	0.859	0.831	0.832	0.832
Obs	1371	1371	1371	1417	1417	1417
Region-Year	Y	Y	Y	Y	Y	Y
Activity-Year	Y	Y	Y	Y	Y	Y
Controls: total migration		Y	Y		Y	Y
Controls: SES		Y	Y		Y	Y
Predicted composition			Y			Y

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

Table A.12: Migration of Agricultural Workers and Agricultural Similarity

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

Notes: * / ** / *** denotes significance at the 10 / 5 / 1 percent level. Robust standard errors clustered at the destination-year level. Agricultural similarity is based on the formulation presented in Bazzi, Gaduh, Rothenberg, and Wong (2016) and is computed based on potential yield data from FAO-GAEZ.