WPS7264

Policy Research Working Paper 7264

Does Access to Foreign Markets Shape Internal Migration?

Evidence from Brazil

Laura Hering
Rodrigo Paillacar

WORLD BANK GROUP
Development Economics Vice Presidency
Development Policy Department
May 2015

Abstract

This paper investigates how internal migration is affected by Brazil's increased integration into the world economy. It analyzes the impact of regional differences in access to foreign demand on sector-specific bilateral migration rates between the Brazilian states for the years 1995 to 2003. Using international trade data, a foreign market access measure is computed at the sectoral level, which is exogenous to domestic migration. A higher foreign market access is associated with a higher local labor demand and attracts workers via two potential channels: higher wages and new job opportunities. Results show that both channels play a significant role in internal migration. Further, we find a heterogeneous impact across industries, according to their comparative advantage on the world market. However, the observed impact is driven by the strong reaction of low-educated workers to changes in market access. This finding is consistent with the fact that Brazil is exporting mainly goods that are intensive in unskilled labor.

This paper is a product of the Development Policy Department, Development Economics Vice Presidency. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://econ.worldbank.org. The authors may be contacted at laura.hering@gmail.com and paillacar@gmail.com.

Does Access to Foreign Markets Shape Internal Migration? Evidence from Brazil

Laura Hering and Rodrigo Paillacar

JEL codes: F16, F66, R12, R23.

Keywords: Migration, international trade, economic geography, Brazil

Laura Hering (corresponding author) is an assistant professor at the Erasmus School of Economics (Department of Economics) and a Tinbergen Research Fellow; her email address is laura.hering@gmail.com. Rodrigo Paillacar is an assistant professor at the University of Cergy-Pontoise (Laboratoire THEMA); his email address is rodrigo.paillacar@gmail.com. This research has been conducted as part of the project Labex MME-DII (ANR11-LBX-0023-01). We thank the editor, two anonymous referees, Maarten Bosker, Matthieu Couttenier, Fabian Gouret, Philippe Martin, Sandra Poncet, Loriane Py, Cristina Terra, Vincent Rebeyrol and Gonzague Vannoorenberghe for their helpful suggestions.

INTRODUCTION

A considerable amount of literature provides evidence that a country generally benefits from opening up to international trade. However, within the country, these benefits are often unevenly distributed. This can cause a rise in regional wage disparities, both across and within industries, which may lead to changes in the spatial distribution of the domestic economic activity.

In this paper, we investigate how internal migration is affected by Brazil's increased integration into the world economy. More specifically, we analyze the impact of changes in foreign demand for Brazilian goods on sector-specific bilateral migration rates between the 27 Brazilian states for the years 1995 to 2003.

In order to identify the effect of international trade on the local labor market in a specific sector, we compute a region-sector specific measure of foreign demand, which is derived from a standard gravity equation that can be obtained from various trade models. The location of the region with respect to its potential trading partners plays a key role in determining a region's market access. Firms located in regions closer to large consumer markets have a higher market access due to lower trade costs, thereby giving them a competitive advantage in these markets. An increase in a region's market access therefore reflects a higher demand for its products and consequently a higher labor demand.

We show in this paper that an increase in a region's access to foreign markets attracts migrants via two channels: i) an indirect effect via an increase in the local wage premium and ii) a direct effect resulting from the creation of new job opportunities.

The positive effect of foreign market access on wages is already well documented for various countries, including Brazil (Fally et al. 2010).[1] In this paper we focus on the second channel, which captures the impact of market access on migration beyond its effect via a change in local wages.

[1] The impact of market access on wages is by now well studied empirically. See, among others, Hanson (2005) for the United States, Head and Mayer (2006) for Europe, and Hering and Poncet (2010) for China. The theoretical link is modeled explicitly in the so-called "New Economic Geography wage equation" (Fujita et al. 1999), but Head and Mayer (2011) point out that such wage equations can be established in numerous trade models.

Higher market access is expected to also have a direct effect on migration essentially due to a higher number of vacancies, which increases the probability of employment. Alternatively, the type of jobs created as a result of an increased foreign demand can be considered to be of better quality. In Brazil, as in many emerging countries, firms in the export industry are preferred employers.[2] Next to a higher employment probability, an increase in the market access variable can thus also capture long-term considerations in the migration decision. These aspects are typically excluded when migration is modeled as depending only on spot wages, which themselves cannot capture the workers' wage profile or nonpecuniary aspects linked to the job (Aguayo-Tellez et al. 2010).

Our sector-specific foreign market access measure identifies the net effect of foreign demand on the local labor market. Note that a positive shock to foreign market access does not necessarily mean that only jobs in exporting firms will be created. Due to spillovers or an increase in connected activities (e.g., outsourced tasks), the increase in demand for exported goods may also lead to a change in labor demand in non-exporting local firms in the same sector.

The main advantage of our market access measure is that it is by construction exogenous to domestic factors, such as local labor market regulations or a region's comparative advantage in the supply of goods in a specific sector. Thus, we do not risk confounding the role of foreign demand with local characteristics, in particular the local export capacity, which may be affected by domestic migration.[3]

Performing the analysis of bilateral migration at the sectoral level is motivated by some recent studies on Brazil's labor market, which present evidence for a very low sectoral mobility of Brazilian workers (Menezes-Filho and Muendler, 2011; Muendler, 2008). Therefore, in this paper we focus on labor

[2] Exporters are likely to offer more long-term employments, propose a steeper wage gradient and better working conditions (see e.g., Wagner (2012) for an overview).

[3] This is possible because our approach allows us to separate the foreign demand from a region's production and export capacity. By excluding all supply side factors from our market access measure, we eliminate the possibility of reverse causality between internal migration and international market access.

migration that takes place within sectors.[4]

The sectoral approach has two important advantages, which we exploit in our identification strategy. First, in contrast to our sectoral measure, an aggregated market access variable would be potentially correlated with the evolution of other unobserved migration determinants that vary over time and across states (i.e., amenities, price levels, institutional quality). Constructing migration rates and market access by sector allows us to include year-location fixed effects, which control for these unobserved location characteristics. Second, this allows us to study the heterogeneous effect of market access across industries.

Our results show that regional differences in access to international markets indeed affect internal migration patterns. Foreign demand impacts migration also directly and not only by means of an increased wage level. These findings suggest that new job opportunities created by higher foreign demand are important location determinants.

Further, our results indicate that the effect of market access is generally stronger, the higher the industry's comparative advantage is on the world market.

Moreover, we find that the impact of market access on sectoral migration rates is driven by the low-educated workers. This could be explained by Brazil's relative abundance of low-skilled labor. A higher market access represents a stronger increase in demand for goods intensive in low-skilled labor, in which Brazil has a comparative advantage on the world market (Muriel and Terra, 2009). Thus, these workers are more likely to be affected by a change in the foreign demand.

Although several studies explore the link between trade and migration, they have mostly focused on international migration patterns (cf., for example, Ortega and Peri, 2013, and Letouzè et al. 2009). Yet, internal migration flows have a far greater magnitude than international flows and hence may modify a country's development path much more sensibly. This is of particular relevance in fast urbanizing developing countries like Brazil.

[4] Appendix S2, provides additional results on the issue of potential sectoral relocation.

Closest to our work is the paper by Aguayo-Tellez et al. (2010), which also applies to Brazil. These authors show that workers in formal employments are attracted to states with a higher concentration of foreign owned establishments. We differ from that paper in that we we focus only on employment opportunities that are created by a change in foreign demand. However, as explained above, these new vacancies can also be in non-exporting and domestically owned firms.[5] Further, our analysis also includes informal workers, who account for at least 38 percent of the Brazilian workforce (Henley et al. 2009).

A few papers have studied the role of imports in the location choice of individuals and can be considered as complementary to our work. Kovak (2011; 2013) studies the effect of import competition on internal migration patterns in Brazil. He finds that regions specialized in industries experiencing larger tariff cuts see their wages decrease, which in turn triggers outmigration. In the same spirit, Autor et al. (2013) show how import competition from China affects local labor markets in the United States. They find that stronger import competition is associated with a higher reduction in manufacturing employment. However, their setting requires internal migration in reaction to trade shocks being negligible.[6]

EMPIRICAL METHODOLOGY

The empirical specification of our migration equation is based on an additive random utility model.[7] Every individual *k* from location *i* maximizes the indirect utility V_{kij} across all possible destinations *j*. In a general utility differential approach, the individual location choice M_{kij} can then be written as:

$$M_{kij} = 1 \quad \textit{if and only if } V_{kij} = max(V_{ki1}, \ldots, V_{kiJ}), \quad = 0 \textit{ otherwise.}$$

The indirect utility V_{kij} can be decomposed as follows:

$$V_{kij} = X_{ij}\beta + \xi_{ij} + e_{kij} \tag{1}$$

[5] In our empirical analysis, the presence of exporters and foreign owned firms is controlled for via location-year fixed effects.
[6] Note also that their proxy of trade exposure is only region-time specific. Since we exploit the sectoral dimension and control for location-time fixed effects, we automatically account for this measure.
[7] This model choice is standard in the recent migration literature and is used, for example, in Grogger and Hanson (2011) and Kovak (2011). For a detailed description on the derivation of the empirical specification see Bertoli and Fernández-Huertas Moraga (2013).

where X_{ij} are the characteristics of location *j*. The subscript *i* is included, as characteristics of *j* can vary across original locations *i* (e.g., bilateral distance). β is a vector of marginal utilities and ξ_{ij} represents unobserved location characteristics. The idiosyncratic error term e_{kij} is included to allow individuals from the same origin to choose different locations. We make the standard assumption that this error term follows an i.i.d. Type I extreme value distribution.

Given that individuals select the location that maximizes their utility, the probability that an individual from *i* will choose destination *j* is defined by

$$Pr(V_{kij} > V_{kim}) \quad \forall j \neq m \tag{2}$$

Replacing the indirect utilities by their definitions of equation 1 and rearranging terms, the probability that individual *k* will move from *i* to *j* is given by:

$$Pr(e_{kij} - e_{kim} > X_{im}\beta - X_{ij}\beta + \xi_{im} - \xi_{ij}) \;\; \forall j \neq m \tag{3}$$

McFadden (1974) shows that under the assumption of an i.i.d. extreme value distribution of the individual error term, migration probabilities can be expressed as

$$Pr(M_{kij} = 1) = \frac{exp(X_{ij}\beta + \xi_{ij})}{\Sigma_{j=1}^{J} exp(X_{ij}\beta + \xi_{ij})} = s_{ij} \tag{4}$$

Following Berry (1994), this individual migration probability can be interpreted as the share of individuals from *i* migrating to *j*, s_{ij}. Similarly, the share of stayers of region *i*, s_{ii}, can be written as

$$Pr(M_{kii} = 1) = \frac{exp(X_{ii}\beta + \xi_{ii})}{\Sigma_{j=1}^{J} exp(X_{ij}\beta + \xi_{ij})} = s_{ii} \tag{5}$$

Dividing equation 4 by equation 5 and taking the log yields

$$ln(\frac{s_{ij}}{s_{ii}}) = ln\left(\frac{exp(X_{ij}\beta + \xi_{ij})}{exp(X_{ii}\beta + \xi_{ii})}\right) = \beta(X_{ij} - X_{ii}) + \xi_{ij} - \xi_{ii} \tag{6}$$

We now have an aggregate discrete choice model that accounts for unobserved location characteristics ξ and whose parameters can be estimated using conventional linear estimation techniques.

To obtain our empirical specification, we add the time dimension *t* and the sectoral dimension *s* and replace the vector *X* with our location-sector specific variables of interest.[8] This gives us our first benchmark specification:

$$\ln m_{ijst} = \ln \frac{s_{ijst}}{s_{iist}} = \alpha + \beta_1 \Delta MA_{ijs\tilde{t}} + \beta_2 \Delta w_{ijs\tilde{t}} + FE_{ij} + FE_{st} + FE_{it} + FE_{jt} + \varepsilon_{ijst} \tag{7}$$

m_{ijst} is the observed migration rate between state *i* and *j* for sector *s* in the household survey of year *t*. It is simply defined as the number of migrants going from *i* to *j* divided by the number of stayers. Individuals are considered as migrants when they declare having lived five years ago (*t* – 5) in a different state than their current state of residence. Since we do not know the exact moment of migration, all independent variables are constructed as means over the years *t*–4 to *t*–1. This is indicated by the index $\tilde{t}$.[9]

Our main variable of interest is the market access gap between states *i* and *j* for sector *s*, $\Delta MA_{ijs\tilde{t}} = ln \frac{MA_{js\tilde{t}}}{MA_{is\tilde{t}}}$. An increase in this variable makes state *j* relatively more attractive, either because of i) a higher wage level or ii) new job opportunities (more or better jobs). We can isolate the second channel by including the wage gap, $\Delta w_{ijs\tilde{t}}$, in our benchmark specification. Adding the wage variable has an additional important advantage: it also captures other sector and time varying characteristics of the local labor market that we cannot observe but which are potentially correlated with foreign market access (e.g., sector-specific productivity differentials).

A lower number of available jobs typically also corresponds to a higher unemployment rate. But a higher unemployment rate can also reflect limitations on the labor supply side or a mismatch on the local labor market between vacancies and job seekers. While in some specifications we explicitly include regional differences in unemployment rates, our benchmark estimation includes FE_{it} and

[8] Here we make the implicit assumption that workers do not switch sectors, and thus their migration decision depends only on state characteristics (e.g., price level) or the characteristics of their own sector (e.g., sectoral market access).

[9] Our benchmark results hold also when specifying our independent variables as four-year lags instead of the mean over the previous four years.

FE_{jt}, which correspond to origin-year and destination-year dummies. These account for time-varying differences across states, including the unemployment rate, amenities or price levels, which are also considered to be important determinants of migration.

Bilateral fixed effects FE_{ij} take into account time-invariant specificities concerning migration between two particular states (e.g., moving costs, migration networks). FE_{st} represents sector-year fixed effects.

In the presence of these numerous sets of fixed effects, we identify β_1 by exploiting the variation of market access within the same pair of states over time and across industries. The exact ranking of market access across states or sectors is therefore not of importance.

By definition, ε_{ijst} is a i.i.d. bilateral error term. However, using equation 6 it can be shown that all ε_{ijst} from the same origin *i* depend on the same ξ_{ii}. This leads to a non-zero covariance of ε_{ijst} for observations with the same origin *i* in year *t*. In all our regressions, we therefore cluster our standard errors by the state of origin-year level. Appendix S3 discusses the assumption of the independence of irrelevant alternatives (IIA) that is underlying our model.

MARKET ACCESS: DERIVATION AND CONSTRUCTION

Theoretical Derivation of Market Access

In this subsection, we provide the formal definition of market access and how it can be derived from a standard gravity model of trade.[10]

According to structural gravity models, exports EX_{ijs} in sector *s* from region *i* to partner *j* can be written as

$$EX_{ijs} = \phi_{ijs} S_{is} M_{js} = \phi_{ijs} \underbrace{\frac{Y_{is}}{\Pi_{is}}}_{S_{is}} \underbrace{\frac{E_{js}}{P_{js}}}_{M_{js}} \tag{8}$$

[10] This subsection borrows from the presentation of the general framework in Head and Mayer (2013). Although initially derived from a trade model of monopolistic competition, these authors show how market access can be obtained also in other market structures, notably in a setting with perfect competition and technology differences (Eaton and Kortum 2002), or in trade models accounting for firm heterogeneity (Chaney, 2008).

with $0 \le \phi_{ijs} \le 1$. This equation decomposes exports into three components: The term ϕ_{ijs} reflects the accessibility of market *j* for the exporters from location *i* in sector *s*. A ϕ_{ijs} of 1 indicates free trade and $\phi_{ijs} = 0$ refers to prohibitively high trade costs and thus zero exports.

The terms S_{is} and M_{js} are often referred to in the literature as the supply and market capacity. They capture all the considerations that make exporter *i* a competitive exporter and partner *j* an attractive destination in sector *s*. More precisely, the supply capacity depends on the total output $Y_{is} = \sum_j EX_{ijs}$ of sector *s* in location *i*, as well as the local firms' price competitiveness, Π_{is}. The market capacity of *j* in sector *s* depends on location *j*'s total expenditure on goods from sector *s*, $E_{js} = \sum_i EX_{ijs}$, and the prevailing price index in sector *s* on market *j*, P_{js}.

The terms Π_{is} and P_{js} are the so-called outward and inward "multilateral resistance terms" (Anderson and van Wincoop 2003). These terms take into account that bilateral trade relationships are affected by competition from third countries.

Given equation 8, region *i*'s relative access to every individual market *j* for sector *s* is defined by $\frac{E_{js}\phi_{ijs}}{P_{js}}$. Region *i*'s total market access in sector *s* can be obtained by summing over all destinations *j*:

$$MA_{is} = \sum_j \frac{E_{js}\phi_{ijs}}{P_{js}} = \sum_j \phi_{ijs} M_{js} \tag{9}$$

MA_{is} measures the overall ease for firms in location *i* to access all domestic and foreign markets *j* in sector *s*. It represents an expenditure-weighted average of relative access, as it weights the market capacity of each potential destination *j* by their accessibility from region *i*.

By summing only over foreign countries, we obtain an international market access measure, which solely captures the demand for goods from location *i* coming from abroad.

Market Access Calculation

We estimate the market access measure presented in equation 9 via a gravity trade regression, following Redding and Venables (2004). This methodology is rarely applied in regional studies because of data limitations: bilateral trade flows are often unavailable at the subnational level, particularly for developing countries. Brazil is a fortunate exception since it provides information on international trade flows at the sectoral level for each of its twenty-seven states.

Our trade data set covers the years 1991 to 2002 and eleven sectors.[11] It contains international trade flows between the twenty-seven Brazilian states and 170 partner countries and flows among the 170 foreign countries.

The empirical specification of the trade equation follows from Equation 8. After taking the logs, we obtain

$$\ln EX_{ijs} = \ln \phi_{ijs} + \ln S_{is} + \ln M_{js} \tag{10}$$

For the calculation of a sector-state specific market access variable that varies over time, we estimate equation 10 separately for every sector-year pair.

In the regressions, sector-specific market capacity (M_{js}) and supply capacity (S_{is}) of every trading partner are captured by sector-importer (FM_{js}) and sector-exporter (FX_{is}) fixed effects. ϕ_{ijs} can be specified using different measures of trade costs. Specifically, we consider bilateral distance (d_{ij}), whether partners share a common border (B_{ij}), the presence of a free trade agreement between the two trading partners (RTA_{ij}) and whether the two are members of the WTO or its predecessor GATT (WTO_{ij}). Since we estimate the trade equation separately for every sector-year pair, we can drop the subscript *s*. Our empirical specification of the trade equation can then be written as

$$\ln EX_{ij} = \delta \ln d_{ij} + \lambda_1 B_{ij} + \lambda_2 RTA_{ij} + \lambda_3 WTO_{ij} + FX_i + FM_j + v_{ij} \tag{11}$$

where v_{ij} is a random bilateral error term.

In total, we run 132 regressions (12 years $\times$ 11 sectors). Given that all coefficients and fixed effects are allowed to vary over time and across sectors, this

[11] For details on sectoral classification and data sources for the variables used in this section see appendix S4.1 and S4.2.

enables us to build a time-varying market access specific for each state-sector combination.

Market access for state i in sector s in year t is built by weighting each predicted market capacity, $\hat{M}_{jst}$, by the estimates of the corresponding bilateral trade costs, $\hat{\phi}_{ijst}$. These weighted market capacities are then summed up to one single variable per state-sector pair:

$$\begin{aligned} MA_{ist} &= \sum_j \hat{\phi}_{ijst} \hat{M}_{jst} \\ &= \sum_j^R \exp(\hat{\delta}_{st} \ln d_{ij} + \hat{\lambda}_{1st} B_{ij} + \hat{\lambda}_{2st} RTA_{ijt} + \hat{\lambda}_{3st} WTO_{ijt} + \hat{FM}_{jst}) \end{aligned} \tag{12}$$

We sum over R countries, where R includes only foreign countries and not the Brazilian states. This way, market access exclusively captures the *foreign* demand addressed to each Brazilian state.[12] Market access thus differs from predicted exports as it excludes the local supply capacity.

Our measure can be considered exogenous to bilateral migration rates since all effects of internal migration on the states' exports (imports) are captured by the estimates of the export (import) capacities of the Brazilian states. These are however not included in our measure. By excluding the exporter fixed effects, we ensure that our measure is exogenous to all domestic factors that affect the state's export supply capacity, such as its comparative advantage in sector s, the local infrastructure or changes in the labor force.[13]

By focusing on foreign market access we eliminate the possible reverse causality that can arise when immigrants raise local consumption and hence the local market capacity: a local shock inducing the arrival of additional migrants may increase consumption in the host region and thus domestic market access but does not affect the access to foreign markets.

[12] To be consistent across sectors and years, each MA_{is} is constructed using the estimated market capacities and trade costs of always the exact same one hundred countries. These are the countries that import goods from all sectors in all years and thus provides us for all sector-year combinations with the necessary estimates for trade costs and importer fixed effects.

[13] We present also robustness checks including the difference in the states' exporter fixed effects as control variable ($\Delta supply_{ijs\tilde{t}}$), to verify that our market access coefficient is not correlated to supply factors.

Finally, also the variables to proxy trade costs can all safely be regarded as exogenous to internal migration within Brazil (at least for the time horizon under study).

Table A-1 summarizes by industry the coefficients obtained from the trade regressions (equation 11). Coefficients on the trade cost variables have the expected sign, and magnitudes are in line with the literature (cf. Head and Mayer, 2013). However, there are some important differences across sectors, in particular in the distance coefficient. The last column summarizes the time varying importer fixed effect, representing the sector-specific market capacity of each destination country. Appendix S1.1 provides some descriptive statistics. Appendix S1.2 calculates various alternative market access measures.

HOUSEHOLD SURVEY DATA

Our main data set is the yearly household survey *Pesquisa Nacional por Amostra de Domicilios* (PNAD) collected by the Brazilian Institute of Geography and Statistics (IBGE). The PNAD does not follow individuals but interviews a different random and representative sample of residents each year (between 310,000 and 390,000 per year). We use the PNAD for the years 1992 to 2003 (with data missing for 1994 and 2000).[14]

Migration Rates

We identify an individual as a migrant when the answer given to the question “In which Brazilian state did you live five years ago?” differs from the actual state of residence. Our sample is limited to individuals who declare having a job in a tradeable sector, earning a positive wage, having lived in Brazil five years ago and being between twenty and sixty-five at the time of the interview.

We distinguish eleven tradeable sectors that can be matched with the trade data and construct bilateral migration rates separately for each sector.[15] We do not have any information about the individual’s work five years ago. Nevertheless, as argued above, we can make the reasonable assumption that individuals already

[14] In 1994 the PNAD was not conducted because of a strike. 1991 and 2000 were years of the population census.
[15] See appendix S4.1 for details on the industrial classification.

worked in the same sector as in the year of the survey. Bilateral migration rates are then defined as the number of migrants from state i to j over the number of workers that stayed in state i and declare working in sector s at the time of the interview. InTable 3, we construct sectoral migration rates separately by educational attainment. The workers are treated as highly educated if they attended high school for at least one year; otherwise they are regarded as low educated.[16]

Despite the presence of a relatively high number of zero migration flows among the states, the PNAD is considered to be representative of overall migration rates and thus adequate for studying migration patterns within Brazil (Fiess and Verner, 2003; Cunha, 2002). In robustness checks, we will also address the problem of unobserved flows by running Poisson-Maximum-Likelihood estimations including zero-flows.[17]

In our final data set, close to 3 percent of the individuals have moved states at least once within five years prior to the interview. Even though most of the migrants are low qualified in absolute terms, the highly educated individuals are the more mobile throughout the years (2.75 percent versus 3.53 percent). Table A-2 compares interstate mobility across sectors. Whereas less than 3 percent of the workers in basic metals, machinery, textile, and agriculture migrated within the last five years, this percentage is above 4 percent in the wood industry.

Sector-State Specific Wages

Our key control variable is the sector-specific wage gap, $\Delta w_{ijs\tilde{t}}$. This variable accounts for most sector-state specific characteristics of the Brazilian labor market (such as sectoral and regional variations in employment regulations and labor productivity). Moreover, it controls for the indirect impact of market access on migration.

However, due to endogeneity concerns, we do not rely on the observed

[16] In our empirical analysis, we exclude migration rates that are constructed with less than six observations. Results are robust when maintaining all observed flows and when omitting the top five and bottom five percent of migration rates. Also, using a sample limited to household heads yields overall very similar results (available upon request).

[17] We have 7722 potential origin-destination-sector cells ($27 \times 26 \times 11$) but observe at least one positive migration rate for only 1748 cells. In the Poisson estimations we replace all missing values with zeros for these 1748 sector-origin-destination combinations.

average wage levels. The main potential source of endogeneity in our case stems from self-selected migration. The personal characteristics (e.g., education, age) that drive the location choice are also major wage determinants. Thus, the observed wage level in a region depends on the composition of the local labor force, including the immigrants. We treat this issue by correcting for self-selected migration, following the methodology developed by Dahl (2002).[18]

$\Delta w_{ijs\tilde{t}}$ is constructed from estimates of a modified Mincerian wage equation that is run separately for every state-year combination *jt*.[19] The obtained parameters on the individual characteristics are then used to predict wages that each individual *k* would potentially earn in each of the twenty-seven states in year *t*. The effect of sector-specific market access on the wage level is accounted for by sector fixed effects.

The final wage gap for year *t* is defined as

$$\Delta w_{ijst} = \hat{w}_{ijst} - \hat{w}_{iist} \tag{13}$$

where $\hat{w}_{iist}$ is the average of the predicted wages that all individuals *k* in sector *s* who actually lived in *i* five years ago would earn in state *i* in year *t*. $\hat{w}_{ijst}$ uses the same set of individuals *k* and is defined as the average of the predicted wages in year *t* that all workers in sector *s* coming from state *i* would have potentially earned in state *j* (regardless of whether in year *t* they actually live in *j* or not).

This aggregation method keeps the composition of the labor force constant across the states, since the same individuals are used for computing the regional wage at the origin and at the destination. Thus, differences in regional wage levels are only due to variations in the estimated parameters of the wage equations and

[18] This approach has become standard in the recent migration literature. For a most recent study see Bertoli et al. (2013). For a detailed description of the methodology see Dahl (2002).

[19] We regress individual hourly wages over the standard wage determinants age, age squared, education, gender, ethnic group, and sector dummies plus an individual correction term. The correction terms are the individuals' migration probabilities as proposed by Dahl (2002). The individual probability of moving from *i* to *j* is constructed using only observed personal characteristics (educational attainment, age group, gender, family status, and state of origin). By adding a polynomial of these migration probabilities to the wage equations, we get consistent estimates of the coefficients on the wage determinants. Estimation results of the wage equations are available upon request.

not to the composition of the labor force.[20]

There is however one remaining source of potential reverse causality, which results from the possibility that with more sizeable immigration levels, migrants may exert a negative impact on the local wage level. But so far, studies concerning the impact of migration on wages are not conclusive and indicate either a weak positive or neutral effect.[21] Moreover, bilateral flows, compared to total immigration, can be considered of small magnitudes, which justifies the assumption that general equilibrium effects are of second order. Therefore, we are confident that our wage variable is not subject to important endogeneity concerns, even though it is not directly addressing all general equilibrium issues.

In Tabole 3, we use migration rates that are constructed separately for highly educated and low-educated workers. Here, the wage variable takes different values for the different educational groups *e*. Δw_{ijst}^{e} is constructed as in equation 13 but takes the average of the predicted wages only for the relevant group of workers.

Main Results

Sector-Specific Market Access and Migration Rates

In column 1 of table 1, we start by estimating a standard model of migration with a reduced set of fixed effects. Instead, next to sector-specific wage gaps, we also take into account the regional differences in unemployment rates, $\Delta u_{ij\tilde{t}}$, population size, $\Delta pop_{ij\tilde{t}}$, and homicide rates, $\Delta death_{ij\tilde{t}}$.[22]

Homicide rates are considered as a proxy for crime and security. For both the unemployment gap and the difference in homicide rates, we expect a negative

[20] Note that Δw_{ijst} is constructed using predicted wages in levels and not in log, as do Grogger and Hanson (2011). When repeating our main estimations with the wage variables in log, wages are not significant and market access shows a higher coefficient. Overall, this would not affect our general conclusions on market access. However, given the highly significant results for wages in levels, we believe that wages in this form are the relevant variable for the estimation of the location decision of workers in Brazil.

[21] In order to explain these findings, more complex models have been proposed that take into account investment reactions or other adjustment channels to migration (Dustmann et al. 2013; Moretti 2011). Accounting for all of these general equilibrium effects would require a careful treatment of the potential interactions between wages, the housing sector, and investment, among other potential outcomes. Yet, preliminary work by Morten and Oliveira (2014) indicates that these alternative adjustment channels are of little overall importance for Brazil.

[22] See appendix S4.3 for the sources of the additional control variables and the construction of the unemployment rate.

impact. The expected sign of population is ambiguous. Although there are more jobs available in large states, there are also possible congestion costs. In column 1, all coefficients have the expected sign and are significant, except for the population variable.[23]

Table 1: Sectoral Market Access and Bilateral Migration

Dep. variable:	$\ln(\text{migrants}_{ijst}/\text{stayers}_{iist})$					
	(1)	(2) *benchmark I*	(3)	(4) PPML	(5)	(6)
$\Delta MA_{ijs\tilde{t}}$	0.617^{a}	0.571^{a}	0.745^{a}	0.983^{a}	0.846^{a}	0.544^{a}
	(0.086)	(0.097)	(0.116)	(0.259)	(0.280)	(0.114)
$\Delta w_{ijs\tilde{t}}$	0.251^{a}	0.311^{a}		0.170^{a}	0.041	0.294^{a}
	(0.044)	(0.048)		(0.056)	(0.033)	(0.050)
$\Delta u_{ij\tilde{t}}$	-0.262^{a}				-0.268^{a}	
	(0.077)				(0.066)	
$\Delta pop_{ij\tilde{t}}$	-0.031				-0.368	
	(0.745)				(0.621)	
$\Delta death_{ij\tilde{t}}$	-0.129^{c}				-0.078	
	(0.074)				(0.061)	
$\Delta supply_{ijs\tilde{t}}$						0.034^{a}
						(0.009)
FE_{ij}	yes	yes	yes	yes	yes	yes
FE_{st}	yes	yes	yes	yes	yes	yes
FE_{it} & FE_{jt}		yes	yes	yes		yes

[23] We do not adjust standard errors for the fact that the market access and wage variables are themselves estimated. Bootstrapping standard errors is prohibitive given the already considerable computational requirements for the construction of each of these variables.

FE_{is} & FE_{js}					yes	
Observations	4183	4183	4183	13927	4183	3798

Heteroskedasticity-robust standard errors clustered at the state of origin-year level appear in parentheses. [a], [b] and [c] indicate significance at the 1%, 5% and 10% confidence levels. Source: Authors' analysis based on data described in the text.

Column 2 contains our preferred specification described in equation 7. Here we include destination-year and origin-year fixed effects to control for time and state varying variables like the price index or the presence of foreign owned firms. Despite the addition of these controls, the magnitude of the coefficient of market access decreases only slightly and remains significant at the 1 percent level. The observed effect here corresponds to the impact that market access has on migration beyond its indirect impact via the wage gap.

This direct effect can be interpreted as the consequence of improved job opportunities generated through several mechanisms. Notably, this direct effect of international demand could be the result of the growth in the number of vacancies, an increase in the tightness of the labor market or more "high quality" jobs.[24] Due to the lack of more detailed data, we cannot identify which is the exact channel, but all of these would increase the utility of workers in this state and thus attract more migrants.

In column 3, we repeat our benchmark estimate but exclude the wage variable. This specification captures the joint effect market access has on migration via the two possible channels: higher wages and more job opportunities. As expected, the coefficient of market access is higher and remains highly significant, when wages are excluded.[25]

[24] Helpman et al. (2010) develop a model that may lead to another possible explanation for our finding of a significant market access coefficient next to a significant wage gap: When firms do not react to an increase in market access by opening more positions but with screening more intensively to obtain a better match, this may attract suitable candidates from other regions. When this mechanism is not fully capitalized into wages, our market access coefficient could represent the better matching between employers and employees provoked by deeper trade integration.

[25] All main results hold also when using destination-origin-year fixed effect instead of destination-year, origin-year, and origin-destination dummies (results available upon request). Table S1.3 presents some sensitivity analyses of our benchmark equation on our market access measure, with overall similar coefficients. Table S2.1 shows that all main results hold for a subsample of workers in sector-specific occupations only.

Since our empirical specification derives from an aggregate discrete choice model (grouped logit model), the estimated coefficients cannot be directly interpreted as marginal effects. To find the partial effect of a change in a location characteristic on the migration probability between two states, we need to differentiate equation 4 with respect to the X_{ij} of interest, which can be written as:

$$\frac{\partial s_{ijst}}{\partial X_{ijst}} = \beta s_{ijst}(1 - s_{ijst}) \tag{14}$$

To evaluate the importance of the direct effect of market access on domestic migration, we replace β with the estimated coefficient of market access and s_{ijst} with the observed migration probabilities. Equation 14 then tells us how the probability of migrating from state *i* to any state *j* in sector *s* in year *t* is affected by a change of 1 percent in the sectoral market access gap.

The values of the elasticities for the 4183 observations in our benchmark specification (column 2) range from 0.0003 to 0.14, with an average elasticity of 0.012. For an increase of 1 percent in the market access gap, this translates into a substantial growth of 34 percent to 57 percent in the number of migrants for each observation. Using the estimates from column 3, which consider the joint effect via both channels, this increase reaches 44 percent to 74 percent.

The last three columns of table 1 provide robustness checks. Column 4 replicates our benchmark estimation using the Poisson Pseudo-Maximum Likelihood estimator (PPML) to deal with the high number of zero-migration flows. The coefficient of our key variable of interest remains highly significant, confirming the positive impact of market access on migration rates. The large standard error in column 4 indicates that even though the coefficient is higher than in the previous estimates, the magnitude is not significantly different from the one in the benchmark equation.[26]

Columns 5 and 6 address the concern that the positive coefficient on market

[26] The data set in column 4 consists of all the 1748 sector-origin-destination combinations for which we observe a positive migration flow for at least one year. The panel is not entirely balanced since we exclude fifty-seven migration rates because i) they are constructed with less than six individual observations; or ii) we do not have wage data for the origin-sector combination.

access could reflect state *j*'s comparative advantage in the export supply of a particular sector *s* if these two are correlated. To make sure that our variable of interest is indeed capturing regional differences in access to foreign markets, column 5 includes sector-destination and sector-origin dummies, which account for sector-region specific characteristics, such as a potential comparative advantage of state *i* in sector *s*.[27] Our market access coefficient remains comparable to the previous estimates. However, the parameter of the wage gap becomes very small and turns insignificant. This suggests that even though wages vary a lot between sectors and states, the yearly variation within sector-state combinations is relatively low, which makes it difficult to identify the effect of wages on migration in the presence of these additional fixed effects.

In column 6, we add an additional variable ($\Delta supply_{ijs\bar{t}}$), which captures the difference between regions in their capacity to supply goods in sector *s*. This variable is the four years average of the estimated exporter fixed effect for each Brazilian state in sector *s* from the gravity trade equation (equation 11) and captures the supply capacity of each exporting region. The higher the comparative advantage of a state in sector *s*, the higher its supply capacity. Even though the coefficient of this variable is positive, we do not want to give it a strong causal interpretation, as this measure is likely to be endogenous to domestic migration.[28]

A highly significant coefficient of market access also in these last two specifications gives us further confidence that the spatial structure of foreign demand matters and that our results are not driven by any local comparative advantage in a specific industry correlated with our market access variable.

Heterogeneous Impact by Industries

Workers in different industries might react differently to changes in market access. This could arise, for example, from a different degree of dependence of the industries on foreign demand or different labor market structures across industries

[27] We exclude here the origin-year and destination-year fixed effects to reduce the number of fixed effects. Including all sets of dummies would substantially reduce the variation left to explain.

[28] The number of observations in column 6 is reduced since not all Brazilian states have been exporting in all sectors during our sample period. As a consequence, we cannot estimate all the sector-year specific exporter fixed effects for each state.

affecting the mobility of workers. To test empirically for heterogeneity in the role of market access in the migration pattern, we allow the coefficient of market access to vary across all eleven industries.

In column 1 of table 2, all sectors, except *Electrical & Electronics*, exhibit a positive and significant coefficient. This shows that the positive effect of market access that we found before is not driven by any particular sector. Column 2 also allows the coefficient on the sector-specific wage variable to vary by industry. Although this decreases the magnitudes of the market access coefficients, these estimates confirm the findings of column 1.

Magnitudes of the market access coefficient vary substantially, leading to important differences in marginal effects across sectors (from on average 0.005 for *Electrical & Electronics* to 0.1 for *Wood*). A first indication for a possible source of such a variation across sectors lies in the sector's comparative advantage on the world market. After Brazil opened itself to foreign trade, certain sectors started to flourish, whereas others experienced a substantial decline.

The industries in table 2 are categorized into three groups (*high*, *medium*, and *low*) according to their comparative advantage on the world market.[29] Sectors with an international comparative advantage have on average higher and more significant coefficients for market access. Columns 3 and 4 repeat the estimations from the first two columns, but restrict the coefficients so as to be the same for all industries within a group. The t-test in the bottom line of the table rejects the hypothesis of equality between the market access coefficient of the group with comparative advantage and that with a comparative disadvantage.

These results suggest that workers in more international competitive industries are moving to higher market access regions and taking full advantage of the positive economic prospects linked to increased exposure to exports. Our findings can thus help to explain the concentration of certain industries in specific regions.

[29] This classification of industries is based on the measure of revealed comparative advantage for Brazilian industries proposed by Muendler (2007) (for details see appendix S4.4.)

In contrast, workers in disadvantaged industries seem less sensitive to changes in foreign market access. Since international demand for their goods is generally low, better access to foreign markets will have less additional value for workers in these industries. As a consequence, market access is expected to play a less important role in the location decision of these workers.

Table 2: Market Access Impact by Sector

Dependent variable:	ln(migrants$_{ijst}$/stayers$_{iist}$) (1)	(2)	(3)	(4)
High: comparative advantage industries				
$\Delta MA_{ijs\tilde{t}} \times$ Agriculture	2.949^a	-0.018		
	(0.435)	(0.298)		
$\Delta MA_{ijs\tilde{t}} \times$ Food	1.377^a	0.924^b		
	(0.451)	(0.424)		
$\Delta MA_{ijs\tilde{t}} \times$ Wood	2.334^a	1.474^a		
	(0.433)	(0.383)		
$\Delta MA_{ijs\tilde{t}} \times$ Plastic & non-metallic	0.522^a	0.234^c		
	(0.136)	(0.134)		
$\Delta MA_{ijs\tilde{t}} \times$ Basic metals	1.028^a	0.529^a		
	(0.197)	(0.183)		
(β_H) $\Delta MA_{ijs\tilde{t}} \times$ *Strong Adv*			0.810^a	0.829^a
			(0.150)	(0.152)
Medium: no comparative advantage				
$\Delta MA_{ijs\tilde{t}} \times$ Mining	1.062^b	0.915^a		
	(0.467)	(0.340)		
$\Delta MA_{ijs\tilde{t}} \times$ Textiles	2.008^a	1.564^a		
	(0.241)	(0.218)		
$\Delta MA_{ijs\tilde{t}} \times$	0.440^b	0.263		
	(0.184)	(0.182)		

Chemical & Pharmaceuticals				
$\Delta MA_{ijs\tilde{t}} \times$ Machinery and others	0.785[a] (0.153)	0.378[b] (0.161)		
(β_M) $\Delta MA_{ijs\tilde{t}} \times$ *Medium Adv*			0.570[a] (0.118)	0.596[a] (0.118)
Low: comparative disadvantage				
$\Delta MA_{ijs\tilde{t}} \times$ Paper & Printing	0.658[a] (0.119)	0.442[a] (0.113)		
$\Delta MA_{ijs\tilde{t}} \times$ Electrical & Electronics	0.226 (0.331)	-0.383 (0.331)		
(β_L) $\Delta MA_{ijs\tilde{t}} \times$ *Low Adv*			0.340[a] (0.106)	0.302[a] (0.104)
Observations	4183	4183	4183	4183
$H_0 : \beta_H = \beta_L$ (p-value)			0.001	0.000

Heteroskedasticity-robust standard errors clustered at the state of origin-year level appear in parentheses. [a], [b] and [c] indicate significance at the 1%, 5% and 10% confidence levels. All regressions include the fixed effects FE_{ij}, FE_{st} and FE_{it} & FE_{jt}. Columns 1 and 3 restrict the coefficient of sector-specific wage gaps to be the same across all industries. Column 2 and 4 allow the coefficient of the wage gap to vary across industries in the same way as market access. Wage coefficients are not reported for the sake of brevity. They are mostly positive and significant. For details on the industry classification see appendix S4.4. Source: Authors' analysis based on data described in the text.

Empirical Results by Sector and Education

In this section, we distinguish between highly educated and low-educated workers. Figure A-1 displays differences in migrant shares between the two educational groups for each state for the years 1995 and 2003. Over the sample period, highly educated migrants were more likely to move to the South and Northeast, while the Center region has become a more popular destination for low-educated migrants. These differences in the location choices suggest that the utility of migrating to a specific state might vary across educational levels.

We thus investigate whether the observed differences in migration patterns can be explained partly by a heterogeneous impact of sectoral market access, depending on the educational attainment of the individuals.

However, there is no clear theoretical prediction on whether the effect of market access on migration rates should be stronger for highly educated or low-educated workers. On the one hand, a more pronounced reaction of highly qualified workers to a change in market access would be in line with the New Economic Geography model by Redding and Schott (2003). Their model predicts that higher market access leads to a higher wage premium for skilled workers. Thus, we could expect that highly educated workers have a stronger incentive to go to states with high market access to benefit from the additional wage premium or a steeper wage gradient in these regions.

On the other hand, numerous theoretical and empirical studies have suggested that highly educated workers are more sensible to certain region-specific amenities.[30] At the same time, highly educated workers might have better access to well-paid jobs. From this perspective, higher wages and career opportunities created by a higher foreign demand could play a minor role in the migration decision of these individuals.

Fally et al. (2010) show that in Brazil, the states with higher foreign market access pay low qualified workers relatively more than highly qualified workers. This finding is in line with traditional trade theory. The Stolper-Samuelson mechanism predicts that in the case of trade liberalization, there should be an increase in the relative returns of the production factor, which is relatively more

---

[30] For example, Levy and Wadycki (1974) have shown that in Venezuela educated individuals tend to value amenities much more than low-qualified individuals. More recently, Adamson et al. (2004) find that returns to education for the higher educated workers fall with the population size in US metropolitan areas, which is also consistent with a skill-biased effect of amenities.

abundant in the country. Thus, in the case for Brazil, we could expect a strong effect of market access on migration for low-educated workers via the indirect wage channel.

Menezes-Filho and Muendler (2011) and Corseuil et al. (2013) provide a first indication that trade liberalization could also lead to a strong adjustment via the direct channel for low-educated workers. Both studies document for Brazil that higher educational attainment contributes to increased employment durations. Low-educated workers are thus more likely to be laid off and obliged to move for new employment.

To test for a heterogeneous role of foreign demand depending on educational attainment, we adapt equation 7 to allow the coefficient of the independent variables to be different for highly educated and low-educated workers. Our second benchmark specification can then be written as

$$\begin{aligned} \ln m_{ijst}^{e} &= \alpha + \beta_H \Delta MA_{ijs\tilde{t}} \times High^{e} + \beta_L \Delta MA_{ijs\tilde{t}} \times Low^{e} \\ &\quad + \beta_3 \Delta w_{ijs\tilde{t}}^{e} \times High^{e} + \beta_4 \Delta w_{ijs\tilde{t}}^{e} \times Low^{e} + FE_{st}^{e} + FE_{ij}^{e} + FE_{it}^{e} + FE_{jt}^{e} + \varepsilon_{ijst}^{e} \end{aligned} \tag{15}$$

where m_{ijst}^{e} is defined as the number of migrants in sector *s* belonging to educational group *e* in year *t* moving from *i* to *j* divided by the number of stayers. The dummy *High* (*Low*) takes the value one when the migration rate is constructed with high (low)-educated workers. The wage gap, $\Delta w_{ijs\tilde{t}}^{e}$, is calculated using means of predicted wages that vary across states, sectors and skill groups. As before, $\tilde{t}$ indicates that independent variables are constructed as means over the years *t*–4 to *t*–1. To take into account that other migration determinants might also vary according to educational attainment, all included fixed effects (FE^{e}) are allowed to differ between the two groups.[31]

Table 3: Bilateral Migration by Education

Dependent	$\ln(\text{migrants}_{ijst}^{e} / \text{stayers}_{iist}^{e})$

[31] This specification corresponds to splitting the sample between high and low qualified workers. Migration rates of highly educated workers represent 34 percent of our final sample.

variable:

	(1)	(2)	(3)	(4)
		benchmark II		
(β_H) $\Delta MA_{ijs\tilde{t}}$ *times* *High edu*	0.058	0.041	0.071	-0.006
	(0.062)	(0.078)	(0.087)	(0.091)
(β_L) $\Delta MA_{ijs\tilde{t}} \times$ *Low edu*	0.871[a]	0.917[a]	1.069[a]	0.890[a]
	(0.132)	(0.151)	(0.161)	(0.170)
$\Delta w^e_{ijs\tilde{t}} \times$ *High edu*	0.188[a]	0.233[a]		0.220[a]
	(0.025)	(0.028)		(0.030)
$\Delta w^e_{ijs\tilde{t}} \times$ *Low edu*	0.149[b]	0.202[b]		0.179[c]
	(0.072)	(0.090)		(0.099)
$\Delta u^e_{ij\tilde{t}} \times$ *High edu*	-0.148[c]			
	(0.083)			
$\Delta u^e_{ij\tilde{t}} \times$ *Low edu*	-0.167[c]			
	(0.095)			
$\Delta pop_{ij\tilde{t}} \times$ *High edu*	1.073			
	(0.937)			
$\Delta pop_{ij\tilde{t}} \times$ *Low edu*	-0.184			
	(0.996)			
$\Delta death_{ij\tilde{t}} \times$ *High edu*	-0.169			
	(0.118)			
$\Delta death_{ij\tilde{t}} \times$ *Low edu*	-0.043			
	(0.069)			
$\Delta supply_{ijs\tilde{t}} \times$ *High edu*				0.034[a]
				(0.009)

$\Delta supply_{ijs\tilde{t}} \times$ *Low edu*				0.023^b (0.011)
FE_{ij}^e	yes	yes	yes	yes
FE_{st}^e	yes	yes	yes	yes
FE_{it}^e & FE_{jt}^e		yes	yes	yes
Observations	4614	4614	4614	4209
$H_0 : \beta_H = \beta_L$ (p-value)	0.000	0.000	0.000	0.000

Heteroskedasticity-robust standard errors clustered at the state of origin-year level appear in parentheses. ^{a}, ^{b} and ^{c} indicate significance at the 1%, 5% and 10% confidence levels. Source: Authors' analysis based on data described in the text.

Table 3 reports results on the heterogeneous impact of market access across educational groups. As in table 1, we display first estimation results for a less restrictive specification. Column 1 does not include the state-year fixed effects, but instead the relative population size, the unemployment gap, and the difference in homicide rates. Column 2 contains our second benchmark specification (equation 15) and column 3 excludes the wage variable to obtain the joint effect of market access on migration via both channels. Column 4 adds the gap of the sector-state specific export supply capacity.

In all specifications, the coefficients of the control variables, including wages, are similar across educational groups. However, the coefficient of market access is significant at conventional levels only for low-qualified workers. The t-tests reported at the bottom of the table clearly reject the hypothesis of a uniform impact of market access across educational groups.

In light of our findings, the differences in the observed migration patterns across educational levels can be partly explained by a different sensitivity to foreign market access.

Economic opportunities associated with international trade seem most important for the location choice of low educated individuals. The strong impact of market access for low-qualified workers can be explained by the fact that Brazil is exporting mainly goods that are intensive in unskilled labor. The industries in which Brazil has a high comparative advantage on the world market exhibit a higher share of low-skilled workers.[32] Consequently, an increase in the demand for exported goods signifies a higher demand and more jobs for low-educated workers. Also when controlling for wage differentials, β_L remains highly significant. This indicates that new employment opportunities created by a stronger local export activity are indeed important for the location choice of this group.

For the highly educated workers, market access remains insignificant in all specifications, even if we exclude wages in order to estimate the joint impact of foreign market access via both channels. The interpretation of this result is less straightforward since it might be driven by various forces, as explained above. One possible explanation is that these individuals have, in general, easier access to "high quality" jobs with good working conditions and career prospects. The alternative explanation of a predominant role of amenities with respect to economic considerations is, however, at odds with the fact that highly educated workers are also responsive to wage differentials.

SIMULATIONS

Before we conclude, we use the estimated coefficients of column 2 of table 3 to simulate the implied change in each observed migration rate in response to a positive shock in the foreign demand for Brazilian goods. This provides more intuition for our results and allows to identify the regions that are particularly affected by a specific demand shock.

Table 4: Effects of Changes in MA

State	Immigrants	Positive demand shock in	Decrease

[32] In the PNAD of 1992, the share of high-skilled workers is highest in the two industries with comparative disadvantage. The share of skilled workers was lowest in agriculture and wood, which are in the highest comparative advantage group.

		(share in local pop.) (in %)	Mercosur (in %)	EU (in %)	NAFTA (in %)	in internal distance (in %)
		(1)	(2)	(3)	(4)	(5)
North	Rondonia	.8	-6.5	1.7	5	-.2
	Acre	1.4	-.8	.8	.3	1.3
	Amazonas	.7	-1.7	-.6	.6	.9
	Roraima	4.2	-1.1	-1.1	1.5	1.2
	Para	1	-1.3	-.6	2.2	-.3
	Amapa	.4	-.6	-1.4	1.8	-.2
	Tocantins	6.7	-1.8	.9	1.2	1
Northeast	Maranhao	.3	-1.6	.8	1.3	-.2
	Piaui	1.3	-3	1.7	2	.1
	Ceara	1	-3.4	2.9	1.1	-.1
	Rio Grande do N.	1.2	-4.4	2.9	1.8	-.6
	Paraiba	.8	-4.1	3.5	.6	0
	Pernambuco	.6	-3.6	3.1	.4	0
	Alagoas	.4	-1	.8	.1	-.1
	Sergipe	1.5	-3.2	1.9	1.1	.1
	Bahia	.5	-2.7	1.8	.6	-.3
Southeast	Minas Gerais	.8	-.7	.3	0	-.2
	Espirito Santo	1.2	-.7	.5	-.2	-.2
	Rio de	.2	1.3	-.9	-.6	-.4

	Janeiro					
	Sao Paulo	.6	3.3	-1.9	-1.4	-1
South	Parana	1.2	5.6	-2.1	-2	6.4
	Santa Catarina	1.1	2.8	-.9	-1.2	-3.9
	Rio Grande do S.	.7	11.4	-5.1	-5.1	-2.2
Center	Mato Grosso do S.	3.8	-1.9	.4	.4	-.3
	Mato Grosso	3.9	-1.2	.1	-.5	.3
	Goias	1.6	.8	-.5	-1.2	.2
	Distrito Federal	4	1.4	-.8	-1.5	.6

Source: own calculations. Immigrant shares in column 1 are the observed shares in the PNAD, constructed based on the sectors included in our analysis. The changes in the immigration shares in columns 2 to 5 are obtained with help of the estimates of column 2 of table 3. Column 2 to 4 simulate the consequences of an increase by 3% in the market capacity of the corresponding group of countries. Column 5 assumes a decrease in the internal distance by 10%. Source: Authors' analysis based on data described in the text.

In table 4, we simulate the effects of four different shocks to $\Delta MA_{ijs\tilde{t}}$. Column 1 reports the average share of immigrants of each state over the sample period. Columns 2 to 5 show for each state how this share would be affected by these different demand shocks.

Marginal effects of market access for the highly educated workers being very low, the implied change in the number of migrants is driven by the low-educated workers. Note that the numbers presented in this table correspond to partial equilibrium effects since our simulation rules out any impact of market access on migration going through the indirect effect of wage differentials. Also, the model does not incorporate any potential impacts of migration on housing costs or other congestion costs.[33]

The first scenario supposes an increase of 3 percent in the demand coming from the Mercosur members (Argentina, Uruguay and Paraguay).[34] This increase in the relative importance of the Mercosur countries affects states differently. Notably the increase in market access of the Southern states which are closer to the Mercosur partners will be higher than for the Northern states. The resulting change in the market access gap between two states impacts directly the bilateral migration rates. When summing over all sending states, sectors, and the two educational groups, we can calculate the total number of additional immigrants a state will receive. Column 2 shows that states in the South will see the most important increase in their share of immigrants, whereas the North and Northeast are relatively less well connected to these markets and will attract less migrants.

Columns 3 and 4 repeat the exercise for an increase in 3 percent of the demand coming from one of the other two main destinations of Brazil's exports, respectively the European Union and the NAFTA countries.

Results are different here: for these two scenarios it is the Southern states that will see the strongest decrease in the share of immigrants. In contrast, the geographic proximity of the Northern and the Northeastern states to the European Union and NAFTA countries leads to an important increase in their market access and hence in the immigration share of these states.

These changes in migration patterns in response to a change in the access to foreign demand illustrates well how much the spatial structure of the domestic economy is influenced by what happens abroad.

In the last scenario (column 5), we consider a decrease in the internal distance to the next harbor by 10 percent, i.e., a reduction in bilateral trade costs (ϕ_{ij}). This improves relatively more foreign market access of inland states.

[33] However, the additional effects mentioned here should play only a minor role. Notably, Morten and Oliveira (2014) show that congestion costs associated with housing would be negligible in the case of Brazil.

[34] This positive demand shock is modeled as an increase by 3 percent of the market capacity (the estimated importer fixed effect FM_{jst} in equation 12) of these countries. An increase of 3 percent corresponds to an increase by one standard deviation of the estimated market capacities.

This last finding also has implications for domestic policies: the reduction in internal distance can also be interpreted as an improved domestic infrastructure that facilitates the access to the sea for the inland states. However, as our results point out, policy makers aiming at regional development need to be aware that due to the country's integration into the world economy the effects of their measures can be reinforced or opposed by events happening outside of the country.

CONCLUSION

This paper shows that workers move away from states with low market access and prefer states with higher market access. By controlling for region and sector-specific wages, we can identify the direct impact that market access has on the migration decision beyond the wage channel. We further find differences in the sensitivity of migration rates to changes in foreign demand across sectors and educational levels. This heterogeneity can reinforce the industrial specialization of regions and explain differences in migration patterns between groups of workers.

Our findings highlight the importance of interactions with foreign countries in shaping the internal spatial distribution of the labor force. This aspect is generally excluded from regional migration studies, which rely only on purely domestic migration determinants. This paper employs household survey data, which has the advantage of considering the informal sector that represents over one third of the Brazilian workforce. However, our data doesn't allow us to identify the main driving force behind the observed direct effect of foreign market access. Linked employer-employee data could be used to study the evolution of the wage profile after migration, a potential improvement of matching between firms and workers, or to assess nonpecuniary aspects of the jobs (e.g., job tenure) and how they are linked to the export activity of a region.

REFERENCES

Adamson, D., D. Clark, and M. Partridge. 2004. "Do Urban Agglomeration Effects and Household Amenities Have a Skill Bias?" *Journal of Regional Science* 44 (2): 201–23.

Aguayo-Tellez, E., M.-A. Muendler, and J. P. Poole. 2010. "Globalization and

Formal Sector Migration in Brazil.” *World Development* 38 (6): 840–56.
Anderson, J. E., and E. van Wincoop. 2003. “Gravity with Gravitas: A Solution to the Border Puzzle.” *American Economic Review* 93 (1): 170–92.
Autor, D. H., D. Dorn, and G. H. Hanson. 2013. “The China syndrome: Local Labor Market Effects of Import Competition in the United States.” *American Economic Review* 103 (6): 2121–68.
Berry, S. T. 1994. “Estimating Discrete-Choice Models of Product Differentiation.” *RAND Journal of Economics* 25 (2): 242–62.
Bertoli, S., and J. Fernández-Huertas Moraga. 2013. “Multilateral Resistance to Migration.” *Journal of Development Economics* 102 (C): 79–100.
Bertoli, S., J. Fernández-Huertas Moraga, and F. Ortega. 2013. “Crossing the Border: Self-Selection, Earnings and Individual Migration Decisions.” *Journal of Development Economics* 101 (C): 75–91.
Chaney, T. 2008. “Distorted Gravity: The Intensive and Extensive Margins of International Trade.” *American Economic Review* 98 (4): 1707–21.
Corseuil, C. H., M. Foguel, G. Gonzaga, and E. P. Ribeiro. 2013. “Youth Labor Market in Brazil Through the Lens of the Flow Approach.” Proceedings of the 41st Brazilian Economics Meeting, ANPEC.
Cunha, J.-M. P. d. 2002. “O uso das PNADs na análise do fenômeno migratório: Possibilidades, lacunase desafios metodológicos [Using PNAD Surveys to Study Migrations: Potentialities, Shortcomings and Methodological Challenges].” IPEA Working Paper 875.
Dahl, G. 2002. “Mobility and the Return to Education: Testing a Roy Model with Multiple Markets.” *Econometrica* 70 (6): 2367–420.
Dustmann, C., T. Frattini, and I. Preston. 2013. “The Effect of Immigration along the Distribution of Wages.” *Review of Economic Studies* 91 (1): 154–63.
Eaton, J., and S. Kortum. 2002. “Technology, Geography, and Trade.” *Econometrica* 70 (5), 1741–79.
Fally, T., R. Paillacar, and C. Terra. 2010. “Economic Geography and Wages in Brazil: Evidence from Micro-Data.” *Journal of Development Economics* 91 (1):

155–68.
Fiess, N., and D. Verner. 2003. “Migration and Human Capital in Brazil during the 1990s.” Policy Research Working Paper 3093, World Bank, Policy Research Department, Washington, DC.
Fujita, M., P. Krugman, and A. Venables. 1999. *The Spatial Economy: Cities, Regions, and International Trade*. Cambridge: MIT Press.
Grogger, J., and G. H. Hanson. 2011. “Income Maximization and the Selection and Sorting of International Migrants.” *Journal of Development Economics* 95 (1): 42–57.
Hanson, G. 2005. “Market Potential, Increasing Returns, and Geographic Concentration.” *Journal of International Economics* 67 (1): 1–24.
Head, K., and T. Mayer. 2006. “Regional Wage and Employment Responses to Market Potential in the EU.” *Regional Science and Urban Economics* 36 (5): 573–94.
———. 2011. “Gravity, Market Potential and Economic Development.” *Journal of Economic Geography* 11 (2): 281–94.
———. 2013. “Gravity Equations: Workhorse, Toolkit, and Cookbook.” CEPR Discussion Papers 9322, C.E.P.R. Discussion Papers.
Helpman, E., O. Itskhoki, and S. Redding. 2010. “Inequality and Unemployment in a Global Economy.” *Econometrica* 78 (4): 1239–83.
Henley, A., G. R. Arabsheibani, and F. G. Carneiro. 2009. “On Defining and Measuring the Informal Sector: Evidence from Brazil.” *World Development* 37 (5): 992–1003.
Hering, L., and S. Poncet. 2010. “Market Access Impact on Individual Wages: Evidence from China.” *Review of Economics and Statistics.* 92 (1): 145–59.
Kovak, B. 2011. “Local Labor Market Effects of Trade Policy: Evidence from Brazilian Liberalization.” mimeo.
———. 2013. “Regional efects of trade reform: What is the correct measure of liberalization?” *American Economic Review* 103 (5): 1960–76.
Letouzé, E., M. Purser, F. Rodríguez, and M. Cummins. 2009. “Revisiting the

Migration-Development Nexus: A Gravity Model Approach." MPRA Paper 19227, University Library of Munich, Germany.

Levy, M. B., and W. J. Wadycki. 1974. Education and the Decision to Migrate: An Econometric Analysis of Migration in Venezuela." *Econometrica* 42 (2): 377–88.

McFadden, D. 1974. "Conditional Logit Analysis of Qualitative Choice Behavior." In P. Zarembka, ed., *Frontiers of Econometrics*. New York, NY: Academic Press.

Menezes-Filho, N. A., and M.-A. Muendler. 2011. "Labor Reallocation in Response to Trade Reform." NBER Working Papers 17372, National Bureau of Economic Research, Inc.

Moretti, E. 2011. "Local Labor Markets." In O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics*. North Holland: Elsevier.

Morten, M., and J. Oliveira. 2014. "Migration, Roads and Labor Market Integration: Evidence from a Planned Capital City." mimeo.

Muendler, M.-A. 2007. "Balassa (1965) Comparative Advantage by Sector of Industry, Brazil 1986–2001." mimeo.

———. 2008. "Trade and Workforce Changeover in Brazil." In Bender, Lane, Shaw, Andersson, and von Wachter, eds., *The Analysis of Firms and Employees: Quantitative and Qualitative Approaches*. Chicago and London: The University of Chicago Press.

Muriel, B., and C. Terra. 2009. "Sources of Comparative Advantage in Brazil." *Review of Development Economics* 13 (1): 15–27.

Ortega, F., and G. Peri. 2013. "The Effect of Income and Immigration Policies on International Migration." *Migration Studies* 1 (1): 47–74.

Redding, S., and P. K. Schott. 2003. "Distance, Skill Deepening and Development: Will Peripheral Countries Ever Get Rich?" *Journal of Development Economics* 72 (2): 515–41.

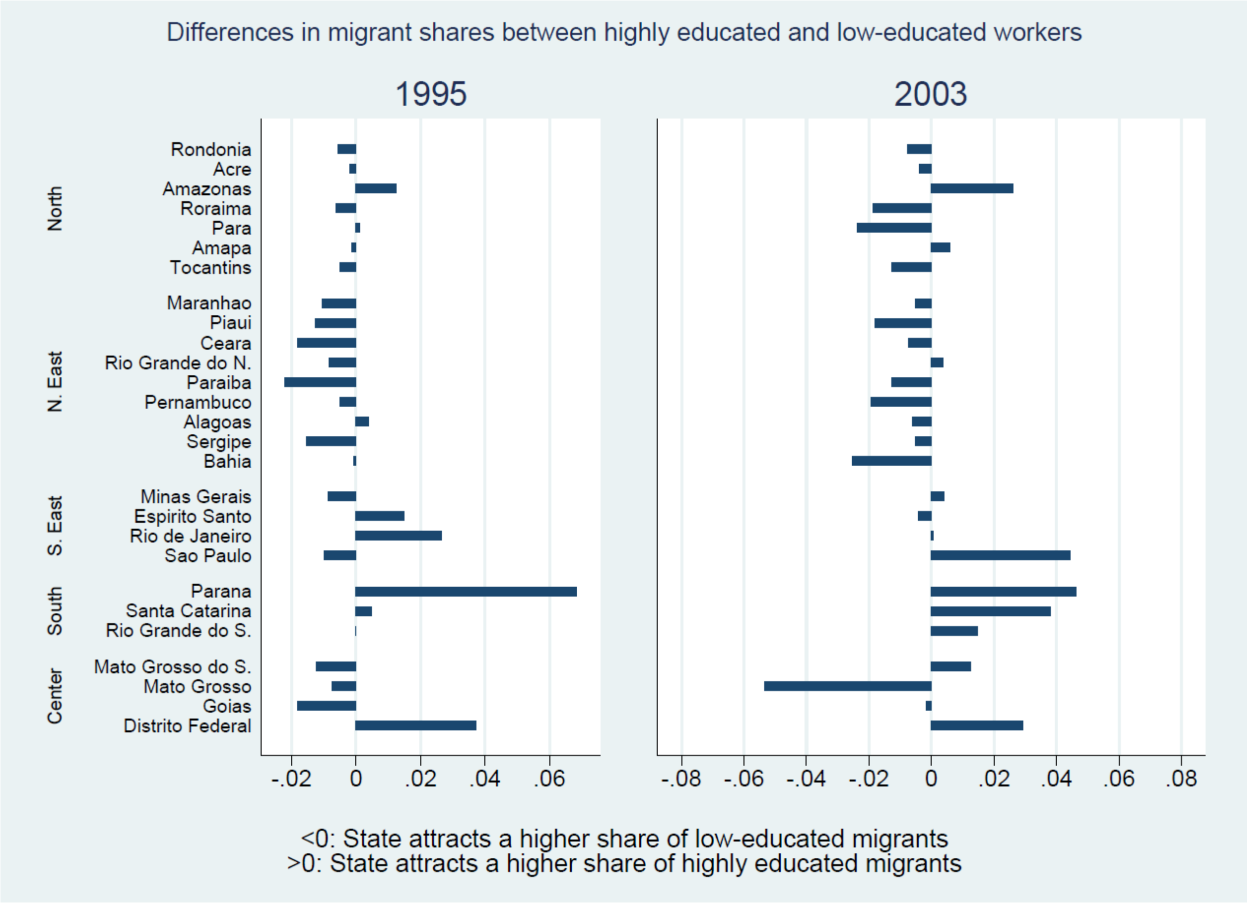
Redding, S., and A. Venables. 2004. "Economic Geography and International Inequality." *Journal of International Economics* 62 (1): 53–82.

Wagner, J. 2012. "International Trade and Firm Performance: A Survey of

Empirical Studies since 2006." *Review of World Economics* 148 (2): 235–67.

APPENDIX

Figure A-1: Differences in migration patterns across educational groups



Source: Own calculations. Migrant shares for each state are calculated as migrants from *i* to *j* with educational level *e* over the total number of migrants in the respective educational group.

Table A-1: Estimation Results of the Trade Equation (Equation 11): Averages of Coefficients by Industry

Industry	Distance	Border	RTA	WTO	Market capacity
Agriculture	-1.134	1.058	0.549	0.417	26.44
	[0.0455]	[0.245]	[0.127]	[0.248]	[1.777]
Mining	-1.457	1.040	0.317	0.230	27.19
	[0.0393]	[0.0885]	[0.0972]	[0.146]	[1.944]
Food	-0.739	0.726	0.289	0.183	18.66
	[0.0899]	[0.171]	[0.199]	[0.373]	[2.017]

Textiles	-1.334	1.078	0.392	0.417	28.82
	[0.0155]	[0.229]	[0.119]	[0.211]	[1.760]
Wood	-1.401	0.584	0.778	0.312	28.97
	[0.0315]	[0.187]	[0.167]	[0.178]	[2.021]
Paper &	-1.689	0.735	0.727	0.503	30.74
Printing	[0.0631]	[0.151]	[0.0801]	[0.205]	[1.866]
Chemical &	-1.520	0.577	0.608	0.346	30.78
Pharmaceuticals	[0.0191]	[0.149]	[0.0944]	[0.119]	[1.861]
Plastic & non-	-1.616	0.768	0.619	0.560	30.75
metallic	[0.0309]	[0.144]	[0.171]	[0.180]	[1.687]
Basic metals	-1.572	0.589	0.558	0.375	31.04
	[0.0330]	[0.187]	[0.142]	[0.166]	[2.028]
Electrical &	-1.398	0.596	0.580	0.734	30.14
Electronics	[0.0287]	[0.200]	[0.186]	[0.212]	[1.914]
Machinery	-1.430	0.799	0.546	0.660	31.25
	[0.0319]	[0.124]	[0.157]	[0.201]	[1.740]

Equation 11 is run separately for every industry-year combination. This corresponds to 12 regressions for each industry. This table shows averages of coefficients by industry. Standard deviations of the coefficients are indicated in parentheses. Source: Authors' analysis based on data described in the text.

Table A-2: Migration Rates by Sectors

Industry	Migration rates	Nb of individuals
	Averages over all years	
Agriculture	2.66	16026.50
Mining	3.37	414.00
Food	3.54	3402.88
Textiles	2.80	5376.13
Wood	4.14	1043.00
Paper & Printing	2.78	966.13
Chemical & Pharmaceuticals	3.54	1075.25

Plastic & non-metallic	3.13	1621.63
Basic metals	2.87	3414.13
Electrical & Electronics	3.04	583.50
Machinery	2.89	1567.50

Source: Own calculations. Data are from the PNAD (1995–2003).

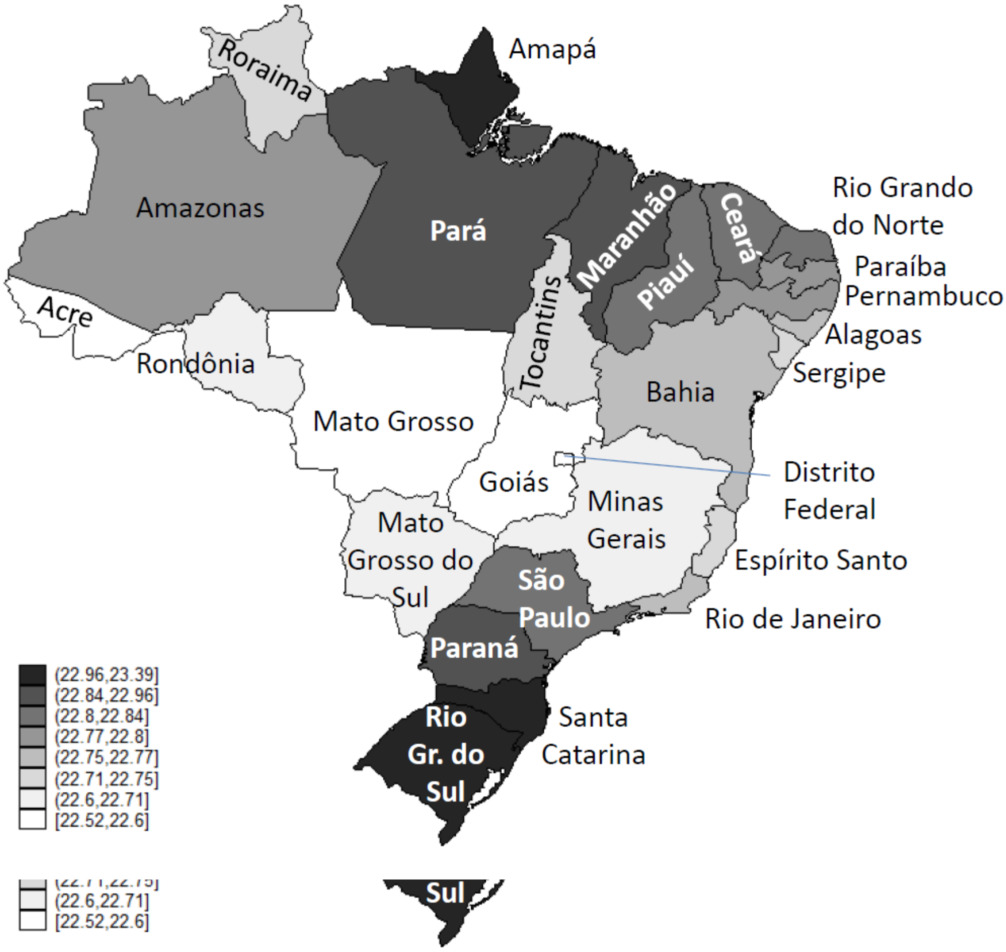
Appendix S1 - Descriptive statistics and robustness of market access

S1.1 Results of the trade equation and market access description

Figure S1.1 shows the spatial distribution of market access across states (reporting aver- ages over years and sectors). Rio Grande do Sul in the South, bordering Uruguay and Argentina, has the highest average market access over our sample period, followed by its neighbor Santa Catarina, which is also sharing a border with Argentina. Amapá on the Northern coast comes third. The state with the lowest market access average is Mato Grosso in the Center-West, which has no direct access to the sea. To identify our market access coefficient in our regression analysis, we exploit however the variation of market access within the same pair of states over time and across industries. Therefore, the absolut ranking of market access across states is of no importance for our analysis.

Tables S1.2 and S1.3 report summary statistics and the correlation between our main variables.

Figure S1.1: Average market access by state (over time and sectors)



Source: own calculations. The figure shows the spatial distribution of the average market access over our sample period. States are grouped into eight quantiles.

Table S1.1: Summary Statistics

	Obs.	Mean	Std. dev.	Min.
m_{ijst}	4183	0.0277	0.054	0.0004
$ln(m_{ijst})$	4183	-4.587	1.401	-7.873
Δpop_{ijt}	4183	-0.186	1.394	-4.802
$\Delta death_{ijt}$	4183	0.0284	0.949	-2.407
Δu_{ijt}	4183	-0.0435	0.574	-1.919
Δw_{ijt}	4183	-0.0050	0.809	-4.656
$\Delta supply_{ijt}$	3798	-0.0409	3.162	-9.950
ΔMA_{ijst}	4183	-0.0075	0.232	-1.572
ΔMA_1	4183	0.0177	0.193	-1.299
ΔMA_2	4183	-0.0108	0.230	-1.525
ΔMA_H	4183	-0.0121	0.280	-2.025
ΔMA_T	4183	-0.0302	0.916	-3.169

ΔMA_R	4183	-0.0175	0.183	-1.001

Table S1.2: Correlation table

	$ln(m_{ijst})$	Δw_{ijt}	ΔMA_{ijst}	ΔMA_1	ΔMA_2	ΔMA_H	ΔMA_T	ΔMA_R
$ln(m_{ijst})$	1.000							
Δw_{ijt}	0.225[a]	1.000						
ΔMA_{ijst}	0.079[a]	-0.003	1.000					
ΔMA_1	0.037[b]	-0.058[a]		1.000				
ΔMA_2	0.039[b]	-0.075[a]	0.988[a]	0.753[a]	1.000			
ΔMA_H	0.055[a]	-0.037[b]	0.992[a]	0.783[a]	0.989[a]	1.000		
ΔMA_T	0.334[a]	0.555[a]	0.231[a]	0.052[a]	0.141[a]	0.203[a]	1.000	
ΔMA_R	-0.046[a]	-0.198[a]	0.822[a]	0.519[a]	0.886[a]	0.857[a]	-0.021	1.000
Obs.	4183							

[a], [b] and [c] indicate significance at the 1%, 5% and 10% confidence levels. For the definition and construction of MA_1, MA_2, MA_H, MA_T and MA_R please see Appendix S1.2.

S1.2 Alternative measures for market access

We construct several market access measures to verify the robustness of our results. First, we estimate Eq. 11 using different specifications for φ_{ij} . For the construction of our preferred market access measure, the distance between a Brazilian state and a partner country is defined as the sum of two components: the internal distance between the state and the closest sea harbor and the external distance between the harbor and the foreign country. In contrast, MA_1 uses the geodesic distance between the Brazilian states and the foreign trading partners, which gives less weight to the internal distance. MA_2 uses the same definition of the bilateral distance as our preferred specification, but estimates the trade equation using only common border and distance as variables for bilateral trade costs.

Second, MA_H addresses the concern of zero-value trade flows. We follow the methodology proposed by Helpman et al. (2008) and control for zeros by adopting a Heckman two-step estimator, using a bilateral "doing business" measure as selection variable in the first step (see Appendix S4.2).

Third, since not only foreign, but also domestic demand should matter for the migration decision, we construct total market access (MA_T) that includes also a domestic dimension. As a proxy for the spatial distribution of domestic demand, we add here the trade cost weighted importer fixed effects of the Brazilian states.

Table S1.3: Alternative market access measures

	Dep. variable: ln(migrants$_{ijst}$/stayers$_{iist}$)				
	(1)	(2)	(3)	(4)	(5)
	ΔMA_1	ΔMA_2	ΔMA_H	ΔMA_T	ΔMA_R
ΔMA_{ijst}	0.633[a]	0.560[a]	0.411[a]	0.254[a]	0.500[a]
	(0.100)	(0.102)	(0.080)	(0.031)	(0.180)
Δw_{ijt}	0.320[a]	0.315[a]	0.316[a]	0.285[a]	0.325[a]
	(0.048)	(0.048)	(0.048)	(0.043)	(0.050)
FE$_{ij}$	yes	yes	yes	yes	yes
FE$_{st}$	yes	yes	yes	yes	yes
FE$_{it}$ & FE$_{jt}$	yes	yes	yes	yes	yes
Obs.	4183	4183	4183	4183	4183

Heteroskedasticity-robust standard errors clustered at the state of origin-year level appear in parentheses. [a], [b] and [c] indicate significance at the 1%, 5% and 10% confidence levels.

.

Finally, we provide a measure based on a reduced set of countries, MA_R. Even though our market access variable can be in general considered exogenous to domestic migration, there might arise an endogeneity concern if ever some of Brazil's industries are large enough to affect global prices.[35] In that case, immigration in state i may increase i's exports capacity in industry s in such a significant way that it reduces market shares by other exporters. This would then be reflected by a reduced export capacity of these countries.

Even though export capacities are not used for the construction of market access, in this case endogeneity might still enter our market access variable through an indirect channel: if losses in market shares of partner country **j** in sector s (induced by internal migration within Brazil) entail a reduction in country **j**'s trade revenues, this can lead to a decrease of **j**'s market capacity in all sectors.

However, no Brazilian state represents more than 1.5% of world exports in any of the eleven industries. Thus it is unlikely that global prices are affected by migration within Brazil. Nevertheless, the risk of endogeneity could remain in a few specific subsectors for some specific destinations.

Over our period of analysis, we observe sensible increases in Brazil's market shares for a few products such as sugar, poultry or coffee. Nevertheless, the impact of domestic migration within Brazil on country **j**'s market capacity depends on the importance of these sectors in **j**'s national economy. This restricts this problem to a limited number of countries, which we can eliminate from our market access calculation.

In our last robustness check, we therefore build the measure MA_R with a reduced set of countries, whose trade balance is unlikely to be affected by Brazilian exports (see Appendix S4.2.2 for details on the construction).

This measure serves also as a robustness check for the importance of long distance trade relationships in shaping internal migration. Since we exclude here also many South American and Central American countries (including Argentina and Mexico), we verify that the effect we identify is not only driven by short-distance shipments to Brazil's neighbors.

The correlation of our preferred measure ΔMA_{ijt} with the alternative market access variables is overall very high (see Table S1.2). For all these variables, we also expect a positive impact, which is confirmed by the estimates reported in Table S1.3.[36]

[35] We thank an anonymous referee for bringing this point to our attention.

[36] Results by industry and education also hold when employing these different market access measures (results available upon request).

Appendix S2 - Occupations

In the tables presented in the main text, we make the implicit assumption that migration takes place only within sectors. In this robustness check, we provide some additional evidence that our results also hold when we restrict our analysis to a sample of individuals that work in industry-specific occupations. This makes it less likely that the individuals were previously working in a different sector and therefore biasing our results.

The PNAD contains ten wide occupational groupings (e.g. administrative, scientific & high tech, retail, manufacturing), which can be broken down into in total 381 occupations. Our subsample includes only workers in occupations which can be mapped easily into our ISIC industries. This leaves us with 156 occupations for which we can safely argue that they are industry-specific enough to assume that workers remain in the same industrial sector also when changing employment (e.g. goldsmiths or clock and watch repairs/makers).[37] Table S2.1 presents results on this subsample. Even though the number of observations is reduced substantially, results hold.

Table S2.1: Main results for sector-specific occupations only

	$\ln(\text{migrants}_{ijst}/\text{stayers}_{iist})$				$\ln(\text{migrants}^{e}_{ijst}/\text{stayers}^{e}_{iist})$
	(1)	(2)	(3)	(4)	(5)
	Benchmark I	PPML	ΔMA_H	ΔMA_T	Benchmark II
ΔMA_{ijst}	0.443[b]	0.636[c]	0.327[b]	0.259[a]	
	(0.189)	(0.374)	(0.154)	(0.038)	
x High edu					-2.608
					(2.128)
x Low edu					0.519[a]
					(0.195)
Δw_{ijt}	0.361[a]	0.184[c]	0.367[a]	0.322[a]	
	(0.105)	(0.100)	(0.104)	(0.098)	
x High edu					0.365
					(0.313)
x Low edu					0.428[a]
					(0.132)
Observations	2181	6500	2181	2181	2392

[37] Since the occupational classifications used in the PNAD changes in 2002, this sample is limited to the years 1995 to 2001.

Heteroskedasticity-robust standard errors clustered at the state of origin-year level appear in parentheses. [a], [b] and [c] indicate significance at the 1%, 5% and 10% confidence levels. Columns 1 to 4 include the fixed effects FE_{it} & FE_{jt}, FE_{st} and FE_{ij}. Column 5 includes fixed effects for FE^{e}_{it} & FE^{e}_{jt}, FE^{e}_{st} and FE^{e}_{ij}.

Appendix S3 - The independence of irrelevant alternatives

Our empirical specification of the migration equation is a grouped logit model, which relies on the underlying assumption of the independence of irrelevant alternatives (IIA). The probability of choosing one state relative to the probability of choosing another may be dependent on the characteristics of a third state. If two or more destinations are perceived as close substitutes by potential migrants, this leads to the violation of the assumption of the IIA. We control for the potential correlation across alternative destinations through the inclusion of origin-time fixed effects as proposed by Bertoli and Fernández-Huertas Moraga (2013). The additional bilateral and destination-time fixed effects further reduce the risk of obtaining biased coefficients due to a violation of the IIA.[38]

Appendix S4 - Data sources and classifications

S4.1 Industry classification

The industrial classification used in the PNAD is PNAD/CD91 for the years 1992 to 2001 and CNAE 1.0 for 2002 and 2003. Using the correspondence tables proposed by the Brazilian National Commission CONCLA (IBGE), we can map these industry codes into the ISIC Rev 3 classification at the two digit level. However, to assure that all categories are well-matched over the entire period, we further need to aggregate up to 12 sectors. We discard Petroleum and Gas because of a very low number of workers in this sector. Our final sample consists of 11 tradeable industries (see Table A-2 for the complete list).

S4.2 Trade flows and trade costs variables

In order to estimate sector-specific market access we need two sets of trade data for the years 1991 to 2002: (1) trade data between Brazilian states and foreign countries (Secretaria de Comercio Exterior, Ministry of Trade); and (2) between foreign economies (BACI: Base pour l'analyse du Commerce International, CEPII). The original industry classification in these datasets corresponds to the HS-2 level (Brazilian exports and imports) and the HS-6 level (BACI). These are matched into the eleven industrial sectors defined in the migration data.
The data for RTA, WTO and common border are provided by the CEPII. For the construction of MA_H we estimate the trade equation using a two-stage Heckman estimation procedure. As

[38] See Bertoli and Fernandez-Huertas Moraga (2013) for a more detailed discussion on the possible methodologies to deal with the IIA assumption. Notably, these authors caution that problems may remain also in our case and suggest to test for the presence of cross-sectional dependence in the residuals. Unfortunately, our panel is too short to allow for such tests. However, Hausman and McFadden (1984) note that if the IIA is satisfied, then the estimated regression coefficients should be stable across all 27 destination sets. We thus re-estimate our model 27 times, each time dropping one of the destinations (following Grogger and Hanson, 2011). Coefficients of our market access and wage variables stay similar across samples, suggesting that the IIA property is not violated here.

selection variable, we use a Doing Business variable, which estimates the costs of starting a business (measured as % of GDP). For the foreign countries, we use the information on this variable from the World Bank. For the Brazilian states, the data comes from the Federaçao das Indústrias do Estado do Rio de Janeiro FIRJAN (2010). By interacting the values of the two trading partners, we obtain a bilateral measure for every exporter-importer combination.

S4.2.1 Bilateral Distances

Distances between two foreign countries are defined as the geodesic distance between the two largest cities (provided by the CEPII). Bilateral distance between Brazilian states and foreign countries are the sum of an internal component and an external component. The internal component is the average geodesic distance of the state's cities with more than 100,000 inhabitants to their nearest harbor.[39] The external component is the mean of the geodesic distances between these nearest harbors and the main city of the partner country.

We choose to incorporate the internal distance to the nearest harbor since about 95% of the volume of Brazilian exports leave the country via the sea (Ministry of External Relations, Brazil, 2008). Transport modes via land, air and rail being much more expensive, a location far away from a harbor increases trade costs substantially, which is captured by the higher internal distance. The alternative market access measure MA_1 uses the great circle distance between the Brazilian states and the foreign country.

The construction of the total market access variable (MA_T) requires also the bilateral distances between the Brazilian states. These are calculated as geodesic distances between their respective capitals.

S4.2.2 Sensitive products

For the construction of MA_R, we identify the products for which Brazil has a high market share and the exporting countries that could be affected by a surge in Brazilian exports.

BACI-COMTRADE gives us the agricultural products at the HS6-level for which Brazil is a major exporter: 020714 (frozen chicken, cut), 090111 (coffee, not roasted),
120100 (soybeans), 150710 (soybean oil crude), 230400 (soy oil-cake) 170111 (raw sugar),

200911 (orange juice) and 240120 (tobacco). We then look at the broader categories with a clearer economic meaning: poultry, coffee, tobacco, orange juice, soy products and sugar. To obtain the list of potential competitors in these industries, we use rankings re- ported in various datasets and reports focussing on these specific agricultural products (FAOSTAT and USDA trade datasets; reports from the European Commission and World Bank). These reports provide also information on the countries whose trade balance is likely to be affected by the evolution of exports in the industries listed above. Based on this, we construct MA_R

[39] For states with numerous cities with more than 100,000 inhabitants, we take the largest five cities. The harbors used to calculate the shortest internal distance are: Maceio (AL), Macapa (AP), Ilheus (BA), Aratu (BA), Pecem (CE), Fortaleza (CE), Vitoria (ES), Itaqui (MA), Antonina (PA), Belem (PA), Cabedelo (PB), Suape (PE), Paranagua (PR), Angra dos Reis (RJ), Niteroi (RJ), Natal (RN), Porto Alegre (RS), Rio Grande (RS), Imbituba (SC), Itajai (SC), San Sebastiao (SP) and Santos (SP). Abbrevations in parentheses correspond to the official acronyms of the Brazilian states. The coordinates of the ports are taken from http://www.worldportsource.com/ports/BRA.php.

excluding the market capacity of 36 of the originally 100 destinations included in our preferred market access measure.[40]

S4.3 Control variables in the migration equation

Data on homicide rates and population of the Brazilian states come from the IPEA (Instituto de Pesquisa Econômica Aplicada). Since official unemployment data at the state level is not available, we compute yearly unemployment rates at the state level with the information of the employment status of the total working age population in the PNAD. In the section Empirical Results by Sector and Education, we use unemployment rates that are constructed separately for each educational group.

S4.4 Comparative advantage of industries

To determine whether an industry enjoys a comparative advantage or disadvantage on the international market, we employ the measure of comparative advantage for Brazilian industries proposed by Muendler (2007). He computes the Balassa indicator of revealed comparative advantage for every industry at the ISIC 2 level. The Balassa indicator is defined as the ratio of the export share of a specific industry in Brazil over the average export share of the same industry at world level. A sector with a ratio higher than 1 is said to have a Revealed Comparative Advantage (RCA). As this differentiation may be considered too strict, it is safer to use the distribution of the RCA indicator. Muendler reports the distribution in quintiles and comments that industries are stable across quintiles during the period 1990-97.
We classify the different industries according to their comparative advantage into three groups: High, Medium and Low. The Balassa indicator of the industries in the comparative advantage group (High) lie in the fifth or fourth quintile of the distribution, while the one of the comparative disadvantage group (Low) lie in the lowest two quintiles. The remaining industries are considered as not having any particular comparative advantage (Medium).

Additional References

Bertoli, S. and J. Fernandez-Huertas Moraga (2013). Multilateral resistance to migration. Journal of Development Economics 102 (C), 79–100.

Federaçao das Indústrias do Estado do Rio de Janeiro FIRJAN (2010). Quanto custa abrir uma empresa no Brasil? Estudos para o desenvolvimento do Estado do Rio de Janeiro, no. 6, Rio de Janeiro.

Hausman, J. and D. McFadden (1984). Specification tests for the multinomial logit model. Econometrica 52 (5), 1219–1240.

[40] The excluded countries are: South America: Argentina, Bolivia, Colombia, Ecuador, Paraguay, Peru. Rest of Americas: Belize, Costa Rica, Dominican Republic, El Salvador, Guatemala, Honduras, Jamaica, Mexico. Europe: Belgium, Denmark, Greece, France, Hungary, Italy, Netherlands, Poland, Spain. Africa: Ivory Coast, Kenya, Malawi, South Africa. Asia: Australia, Fiji, India, Indonesia, Israel, Malaysia, Philippines, Thailand, Turkey

Helpman, E., M. Melitz, and Y. Rubinstein (2008). Estimating Trade Flows: Trading Partners and Trading Volumes. Quarterly Journal of Economics 123(2), 441–487.

Ministry of External Relations, Brazil (2008). Brazilian ports: a safe haven for international investment. Technical report, Trade Promotion department.

Muendler, M.-A. (2007). Balassa (1965) comparative advantage by sector of industry, brazil 1986-2001. mimeo.