

How does Market Access shape Internal Migration?*

Laura Hering[†] Rodrigo Paillacar[‡]

September 13, 2011

Abstract

This paper studies the role of international trade on internal migration in an economic geography framework by assessing the impact of regional differences in access to foreign markets on bilateral migration rates between the 27 Brazilian states. We use individual data for 1993 to 2003 to construct yearly migration rates. Bilateral migration rates are also computed separately for each sector and for different educational levels. State and sector-state specific market access indicators are estimated using trade data. We find that regions with low market access see residents migrate to regions with higher market access, where higher labor demand leads to better job opportunities. Low-educated migrants are found to be more sensitive to regional differences in market access than high-educated. This could be explained by the specialization of Brazil in goods that are intensive in unskilled labor. An increase in the demand for Brazilian goods thus leads to a higher demand for low educated labor. All manufacturing workers react strongly to regional differences in market access of their own industry. Simulations of a decrease in international and intranational trade costs lead to an important spatial reorientation of migration flows. Movements are stronger and more homogeneous across sectors in the case of intranational trade liberalization.

Keywords: Economic geography, migration, international trade, wage, selection bias, Brazil.

JEL classification: F12, F16, R12, R23.

* We thank Andrew Bernard, Matthieu Couttenier, Fabian Gouret, Philippe Martin, Sandra Poncet, Vincent Rebeyrol, Jose de Souza, Cristina Terra, Julien Vauday, Vincent Vicard and participants of the ELSNIT conference (Florence), the Migration and Development Conference (Lille), the CREST Macro seminar (Paris) and the EUI Macro seminar (Florence) for excellent suggestions and fruitful discussions.

[†]European University Institute and Paris School of Economics, Université de Paris 1, CNRS. EUI, Via delle Fontanelle 10, 50014 San Domenico, Italy. E-mail: laura.hering@eui.eu

[‡]Paris School of Economics, Université de Paris 1, CNRS. 106-112 Boulevard de l'Hôpital 75647 Paris Cedex 13, France. E-mail: rodrigo.paillacar@univ-paris1.fr

1 Introduction

In economic theory, migration acts as an adjustment mechanism for spatial disequilibria. To understand migration patterns and to design successful regional policies, the identification of the macro-economic determinants of migrants' location choice is essential. The aim of this paper is to emphasize the role of international trade on the domestic migration pattern by studying its impact on inter-state migration within Brazil. Even though the link between migration and trade has been at the center of many studies, these have mainly focused on the reverse causality, the impact of international migration on shaping international trade¹ or the impact of internal migration on internal trade (Combes et al., 2005). Yet, in an open economy intra- and international trade should affect migration patterns.

As for the role of exports, a region which increases its access to foreign consumer markets attracts firms and raises labor demand. This is likely to attract additional workers from less well located regions. The role of imports is more ambiguous. On the one hand, if imports increase, for example as a consequence of a decrease in trade barriers, and the newly imported good has been produced locally before, imports may substitute for local production. The following decrease in labor demand may lead to the emigration of workers. On the other hand, for a good to be imported, it is likely that it is cheaper to buy than to produce it. Increasing imports then result in lower living costs which in turn attract migrants. A region's access to consumer and supplier markets can thus shape migration patterns. With ongoing globalization, this effect of trade on migration is likely to become more and more pronounced. It is thus also of political interest to assess its importance.

The two studies treating the impact of international trade on internal migration we are aware of are Aguayo-Tellez et al. (2008) and Kovak (2008). As this paper, both studies apply to Brazil. Aguayo-Tellez et al. (2008) analyze one particular aspect of exports by looking at the role of multinational enterprises. They find that migrants are attracted by the growth of employment opportunities in states with high concentration of foreign owned establishments. Kovak (2008) highlights the negative effect of imports on labor demand. He shows that decreases in regional wage levels resulting from tariff cuts will trigger migration between the Brazilian states.

This paper concentrates on the export channel of trade by using regional differences in access to consumer markets to explain internal migration patterns. Brazil is a particular interesting country to study this link, since internal migration in Brazil is exceptionally high: in 1999, 40% of all Brazilians lived in a state different from

¹ Studies on the impact of immigration on trade are numerous. See for example Gould (1994), Head and Ries (1998), Rauch and Trindade (2002).

the state they were born in (Fiess and Verner, 2003).² Also the access to consumer markets varies substantially across the 27 states, as has been shown by Fally et al. (2009), and varies sufficiently across years to identify an impact over time.

A theoretical framework that is well suited to study the impact of trade on the spatial structure of a country is the New Economic Geography (NEG) theory. This theory explains the agglomeration of economic activity and how workers will choose their location following changes in trade costs. It underlines the importance of the region's proximity to consumers, which is modeled by the region's access to markets. Regions closer to big consumer markets (i.e. with higher "market access") experience lower transport costs and can sell at lower prices in these markets. Firms located in high market access regions thus face higher demand for their products and consequently higher labor demand. Another important feature of high market access regions is that they are more likely to pay higher wages because of the low transport costs, resulting in higher income levels. This effect is reinforced since the costs of living in these regions are often lower because also imported consumption goods incur lower transport costs (Krugman, 1991).

The fact that previous work on migration determinants has seldom focused on the trade channel, is partly due to the fact that international trade studies concentrate on labor relocation rather than on migration per se. Using such an approach, researchers explore changes in the employment levels, implicitly assuming that factors are mobile. In the context of NEG models, the effect of market access on employment has been estimated by Hanson (1998) for Mexico and Head and Mayer (2006) for the European Union. The first study looking at the impact of market access on migration in a NEG framework is Crozet (2004). For five European countries, he estimates a quasi-structural migration equation based on the model proposed by Helpman (1998) to show that cost advantages due to higher market access shape migration flows within countries. His approach has been adopted by Pons et al. (2007) to explain migration flows within Spain and by Kancs (2005) for migration flows between the Baltic states.³

In our study, we assess empirically whether a change in the region's access to international markets leads indeed to a change in the structure of inter-state migration. We test whether regional differences in access to foreign markets have an impact on bilateral migration rates within Brazil. Our panel data set is constructed

² For a comparison: Dahl (2002) finds that in 1999 in the US, a country known for its high labor mobility, only 30% of all male workers no longer live in the state in which they were born.

³ There is also a long tradition of economic models treating the connection between industrialization processes and rural-urban migration in developing countries, backing at least at Lewis (1954) and Harris and Todaro (1970). Our point is that this literature is not addressing the specific features of a trade channel as NEG models do. In addition, these approaches focus on unilateral rural-urban migration and are thus not suited when migration flows in opposite directions are observed.

using yearly bilateral migration rates between the 27 Brazilian states, obtained from a micro data set for the period 1993 to 2003. We check whether this effect is robust to the introduction of various control variables relating to alternative explanations of migration and control for self-selected migration and the presence of zero-value migration flows.

This analysis allows to understand how the economic geography influences spatial labor movements in the context of a developing country. Contrary to the countries of the European Union, studied in Crozet (2004), Brazil is still highly specialized in goods that are intensive in low-skilled labor. An increase in the demand for these goods is thus likely to affect not the highly-educated but the low-educated workers, whose migration decisions might be different from the in general more mobile educated workers. Our second objective is therefore to investigate whether the sensitivity to market access varies across different educational levels of workers.

Thirdly, we test whether the migration decision is driven by the evolution of market access in the particular industry of the migrant. Thanks to the availability of Brazilian trade data at the sectoral level for the year 1999, we can calculate market access for 16 manufacturing industries. Bilateral migration rates by sector are computed from the 2000 population census. This allows us to assess the impact of market access on a cross-section for the year 2000 for each sector separately and to test for possible heterogeneity across industries. Menezes-Filho and Muendler (2007) find that the high migration rate in Brazil stands in sharp contrast to the low sector relocation in the country. This means that also spatial labor allocation takes place mostly within sectors. As a consequence, workers are likely to choose their destination depending on the spatial distribution of market access in their industry.

Lastly, from simulations with estimated coefficients on sector-specific migration rates, we derive predictions for changes in internal migration flows between the 27 Brazilian states in case of further trade liberalization.

This paper hence contributes to the literature in at least two ways. We are the first to highlight the link between trade and internal migration in an economic geography framework. While Crozet (2004) concentrates on the importance of low living costs generated by a good domestic market access, we concentrate on the role of the access to foreign markets. Using only the access to international markets and not domestic markets has the important advantage that we eliminate an endogeneity problem that arises when immigrants raise local consumption, as highlighted by Head and Mayer (2004).

Since we are interested in the role of trade, our market access is constructed using revealed trade flows, as do Redding and Venables (2004). The use of bilateral trade data between the Brazilian states and the foreign countries reveals both observed

and unobserved trade costs and market characteristics, both for foreign countries and for the Brazilian states. These allow us to obtain a structural measure of market access. For the Brazilian states this is of a particular importance. If international market access is estimated as a simple function of the country's income weighted by the bilateral distance as often done in the literature, it follows that the Northern regions have a higher market access than the Southern regions. This is because they are closer to the US and the European market. Yet, most of the country's export activity is located in the South. Using trade data thus reflects much better the restricted access to foreign markets of the isolated Northern region than the simple income over distance measure.

The second contribution is that we are the first to draw on individual data to demonstrate the link between economic geography and migration. Using bilateral migration rates that are obtained from micro data has several advantages. First, it allows the disaggregation of migration flows and to compare different types of migrants. Previous studies looking at the impact of economic geography on migration have not done this before. Here, we focus on differences in migration behavior across two educational levels and 16 manufacturing industries. Second, it allows us to take into account the migration literature, which highlights the aspect of a possible self-selection of migrants towards certain destinations. In our estimation of regional wage levels, we control for a possible self-selection bias in this variable, as proposed by Dahl (2002). Even though we are looking at bilateral migration rates, thanks to the use of micro data, our empirical strategy is based on the individual migration decision. Our estimated migration equation is derived directly from the utility differential approach that is widely used in the migration literature.

A last novelty of this paper is that it studies the link between migration and economic geography in a developing country. Redding and Schott (2003) show in a theoretical model that if skill-intensive sectors have higher trade costs or stronger increasing returns to scale, highly skilled workers benefit more from a higher market access. In line with this model Breinlich (2006) shows on a panel of European regions, which are specialized in technology and high-skill intensive goods, that higher market access leads to a stronger accumulation of human capital. Since Brazil is mainly producing and exporting goods intensive in unskilled labor in the Brazilian economy, we expect a lower impact of human capital accumulation through a high market access, because a higher market access is more likely to increase demand for low-educated workers.

Our results, obtained from panel data regressions and from the cross-section on manufacturing workers for 2000, indicate that people indeed migrate from states with low to states with high market access. In line with the migration literature, which highlights the importance of pecuniary incentives for migration, especially

spatial wage differentials (Greenwood 1985, 1985), we add also regional wage levels as migration determinants. Findings of a significant impact of market access are robust to the inclusion of regional wages and a number of other controls. We find that the sensitivity to market access is much lower for higher educated workers. We thus find no evidence for the Redding and Schott (2003) prediction. Further, we find that for manufacturing workers the sector-region specific market access plays a more important role than the state's total market access. When looking at each manufacturing sector separately, we see that in nearly all sectors, workers are sensitive to their own sector's market access but that the magnitudes vary across sectors.

In simulations of a decrease in international or intranational trade costs, we predict an overall decrease in migration and an important spatial reorientation or migration flows, whereas movements are stronger and more homogeneous across sectors in the case of intranational trade liberalization.

We finally conclude that altogether, these results suggest that the geography of regional access to international markets can help explain migration patterns within a country. The fact that migration patterns are affected by the location's industrial specialization suggests that implications of NEG theory (i.e. regional advantages generated by the region's position in the spatial economy) are better understood in combination with comparative advantage and sector-specific inputs (e.g. human capital specificity).

The rest of the paper will proceed as follows: Section 2 presents the theoretical framework, and summarizes some implications for the empirical part; Section 3 outlines the estimation strategy; Section 4 indicates the data sources and describes the computation of our market access variables and how we obtain consistent region-specific wage levels; Section 5 gives some descriptive statistics and stylized facts on migration and its link to market access in Brazil; Section 6 gives the empirical results of the migration equation. Simulations are reported in Section 7 and Section 8 concludes.

2 Theoretical background

We consider the standard New Economic Geography (NEG) model (Fujita et al., 1999) that combines monopolistic competition with product differentiation, trade costs and increasing returns to scale at the firm level. The economy is composed of $j = 1, \dots, J$ regions and one agricultural and s manufacturing sectors. The agricultural sector produces a homogeneous agricultural good A , under constant returns and perfect competition, which is freely tradeable. Each manufacturing sector s produces a large variety of differentiated goods, under increasing returns and imperfect

competition, where each firm produces a different variety. In both sectors, labor is the only factor of production, but labor is immobile between sectors.

For simplicity, in this section, we regroup all manufacturing sectors s in one aggregated manufacturing sector M . This allows us to follow the notation of Fujita et al. (1999) in the presentation of the model. We will differentiate between sectors and introduce the superscript s only at the end of this section.

All consumers of region j share the same Cobb-Douglas preferences for the consumption of the goods A and M :

$$U_j = M_j^\mu A_j^{1-\mu}, \quad 0 < \mu < 1, \quad (1)$$

where μ denotes the expenditure share of manufactured goods. M_j is defined by a constant-elasticity-of-substitution (CES) sub-utility function of n_i varieties:

$$M_j = \sum_{i=1}^R \left(n_i q_{ij}^{(\sigma-1)/\sigma} \right)^{\sigma/(\sigma-1)}, \quad \sigma > 1, \quad (2)$$

with q_{ij} representing demand by consumers in region j for a variety produced in region i and σ being the elasticity of substitution. Given the expenditure on manufacturing goods (M) of region j (E_j) and the *c.i.f.* price of a variety produced in i and sold in j (p_{ij}), the standard two-stage budgeting procedure yields the following CES demand q_{ij} :

$$q_{ij} = p_{ij}^{-\sigma} G_j^{\sigma-1} E_j, \quad (3)$$

where G_j is the CES perfect price index for manufactured goods, defined over the *c.i.f.* prices:

$$G_j = \left[\sum_{i=1}^R n_i p_{ij}^{1-\sigma} \right]^{1/1-\sigma}. \quad (4)$$

This price index is a (inverse) measure of competition prevailing on market j . Competition increases with the number of firms active in j (decrease in G_j) and decreases with the price at which they sell (increase in G_j). Each shipment of a manufactured good from i to j involves a variable “iceberg” transport cost T_{ij} . This technology assumes that for every unit of good shipped abroad, only a fraction ($\frac{1}{T_{ij}}$) arrives. This variable is often proxied by the inverse of the bilateral distance between i and j . The consumer price p_{ij} is then proportional to the mill price p_i and T_{ij} . The demand for a variety produced in i and sold in j Equation (3) can thus be written as:

$$q_{ij} = (p_i T_{ij})^{-\sigma} G_j^{\sigma-1} E_j. \quad (5)$$

Total sales of a of the representative firm in region i can be obtained by summing sales across regions. Given that total shipments to one region are T_{ij} times quantities

consumed, this yields:

$$\sum_{j=1}^J p_{ij} q_{ij} = \sum_{j=1}^J (p_i T_{ij})^{1-\sigma} G_j^{\sigma-1} E_j = p_i^{1-\sigma} M A_i, \quad (6)$$

where

$$M A_i = \sum_{j=1}^J T_{ij}^{1-\sigma} G_j^{\sigma-1} E_j = \sum_{j=1}^J \phi_{ij} m_j, \quad (7)$$

represents the market access of each exporting region i (Fujita et al., 1999). This market access indicator is simply the sum of the market capacities of all destinations j , m_j , weighted by the measure of bilateral trade costs, ϕ_{ij} , between i and j .

In order to see the different ways how changes in market access can impact the local economy, we have to model the manufacturing firm's profit function.

We assume that the technology is the same for all varieties of sector M in all locations. Each firm faces a fixed input F and a marginal input requirement c . Given that labor is the only production factor, the quantity of labor l_i required for the production of quantity q_i can be described as

$$l_i = F + c q_i. \quad (8)$$

The firm's profit function can then be written as

$$\pi_i = p_i q_i - w_i (F + c q_i), \quad (9)$$

where

$$q_i = \sum_{j=1}^R (p_i T_{ij})^{-\sigma} G_j^{\sigma-1} E_j T_{ij} = p_i^{-\sigma} M A_i, \quad (10)$$

Maximizing this profit function using Equation (10) yields the usual mark-up pricing rule

$$p_i \left(1 - \frac{1}{\sigma}\right) = c w_i \Leftrightarrow p_i = c w \frac{\sigma}{\sigma - 1}. \quad (11)$$

Given the pricing rule, the profits of a firm in location i are given by

$$\pi = w_i \left[\frac{q_i c}{\sigma - 1} - F \right]. \quad (12)$$

The model supposes free entry and exit of firms in response to profits. According to the zero-profit condition, the equilibrium output q^* is the same for all firms:

$$q^* = F(\sigma - 1)/c. \quad (13)$$

Replacing q^* in the labor demand function Equation (8), yields the equilibrium labor input $l^* = F\sigma$. With L_i the number of workers in sector M in region i , the equilibrium number of manufacturing firms located in i is

$$n_i = L_i/l^* = L_i/F\sigma \quad (14)$$

and total output of region i is $n_i q^*$.

From this definition we can see that when a region i increases total output, in the long run, the number of firms and workers in i will have to increase as well.

Under the condition that the free entry of firms leads to zero-profits in the long run, the equilibrium output q^* produced by each firm has to be equal to the demand function defined in Equation (10).

Each firm maximizes gross profits $\pi_{ij} = p_{ij}q_{ij}/\sigma$ on each market. Together with 10 and 11, this yields

$$\begin{aligned} \pi_i &= \frac{1}{\sigma} \left[\left(\frac{\sigma}{\sigma-1} \right)^{1-\sigma} (w_i c_i)^{1-\sigma} \sum_j^J [\phi_{ij} G_j^{\sigma-1} E_j] \right] - w_i F_i \\ &= \frac{1}{\sigma} \left[\left(\frac{\sigma}{\sigma-1} \right)^{1-\sigma} (w_i c_i)^{1-\sigma} M A_i \right] - w_i F_i \end{aligned} \quad (15)$$

Setting long run profits equal to zero, using Equation (13) and solving Equation (15) for wages, we obtain the so-called wage equation (Fujita et al. 1999):

$$\begin{aligned} w_i &= \frac{\sigma-1}{\sigma c_i} \left[\sum_j^J [\phi_{ij} G_j^{\sigma-1} E_j] \frac{c_i}{F(\sigma-1)} \right]^{1/\sigma} \\ &= \frac{\sigma-1}{\sigma c} \left[\frac{1}{q^*} M A_i \right]^{1/\sigma}. \end{aligned} \quad (16)$$

which shows how wages are influenced by market access. Equation (16) gives the manufacturing wage at which firms in location i break even, when income levels, price indices and transport costs are given. It can be seen from this equation that the wage is higher the higher are the expenditures in the different markets, the better is the firm's access to these markets (the lower the transport costs) and the less competition the firm faces in these markets. Using this equilibrium equation and replacing the price index by Equation (4) and n by Equation (14), we can see how market access, wages and the number of workers in a given region are linked.

$$\begin{aligned}
w_i &= \left(\frac{\sigma-1}{\sigma c}\right) \left[\frac{1}{q^*} \sum_{j=1}^R T_{ij}^{1-\sigma} \left[\sum_{i=1}^R n_i p_{ij}^{1-\sigma} \right]^{\frac{\sigma-1}{1-\sigma}} E_j\right]^{1/\sigma} \\
&= \left(\frac{\sigma-1}{\sigma c}\right) \left[\frac{1}{q^*} \sum_{j=1}^R T_{ij}^{1-\sigma} \left[\sum_{i=1}^R \frac{L_i}{l^*} p_{ij}^{1-\sigma} \right]^{\frac{\sigma-1}{1-\sigma}} E_j\right]^{1/\sigma}
\end{aligned} \tag{17}$$

In the 90's, Brazil experienced with the creation of Mercosur and several unilateral tariff reductions an important decrease of trade costs, T , in a high number of manufacturing sectors. Many Brazilian firms thus enjoyed a gain in market access due to the liberalization of international trade.

According to the theory above, this trade liberalization should result in the creation of new firms and labor reallocation between locations, with workers going to locations that experienced an increase in market access. Given that spatial relocation is very slow, we state that under the period of our study (1993 to 2003), Brazil is still in the phase of adjustment to the lower trade costs.

As can be seen from Equation (17), either wages in region i or the number of firms and workers in region i will have to adjust, for location i to come back to an equilibrium situation.

The adjustment by wages occurs when the production of goods do not increase albeit lower transport costs. Lower transport costs lead to lower consumer prices which in turn results in an increased demand. But since the additional demand cannot be satisfied, prices will raise. Firms transfer this gain to workers by paying higher wages. If the total production of region i stays constant, the increase in the region's market access, will result in a proportional increase in the wage level. An increase in the demand, which is not accompanied by an increase in the production will thus lead to spatial price (wage) disparities. This is what Head and Mayer (2006) call the *price version* of the adjustment to a change in market access.

In the case of the quantity adjustment, the total output of the region increases, resulting in a higher labor demand L_i . New workers are attracted to region i . These newly arriving workers exert a downward pressure on wages, which (at least partially) offsets the upward pressure coming from the higher market access (as described above in the *price version*). This quantity adjustment via an increase in n_i and L_i is reflected in a decrease of the price index of region i , as can be seen from equation (4).⁴ The decrease in T is thus compensated by an increase in L .

Of course, both adjustment mechanisms can also occur simultaneously. When labor is not perfectly elastic, some locations with a higher demand for their manufac-

⁴ If wages stay constant, the decrease in the price index will have to compensate all of the original increase to come back to equilibrium.

tured goods may pay a higher nominal wage even if the number of firms increases. Fujita et al. (1999) highlight that locations with a higher demand for manufactures tend, other things being equal, to offer a higher real wage to manufacturing workers.⁵ This effect creates a strong agglomeration force, resulting in the concentration of economic activity in certain regions. An increase on market access attracts migration thus by two channels. The direct impact comes from the creation of new positions due to higher labor demand or better career opportunities linked to the firm's new export activities. Market access also indirectly influences migration through its impact on wages, which are known to be main drivers of migration.

In this paper, we are interested in the market access indicator of state i , MA_i , and the market access indicator of state i in sector s , MA_i^s . Does an increase in the state's market access MA_i (MA_i^s), resulting in an increase in the number of open positions and the creation of export-specific jobs and (either because firms grow or because new firms open), attract new workers and contribute to an increase in L_i (L_i^s)? In order to control for the indirect impact of market access, we also include the state's wage level.

By taking into account the number of new positions but capturing also better career opportunities, we should capture two different types of migrants. Contracted migration, including migrants that move in response to a concrete job offer, as well as speculative migration, where individuals move to a certain region in the hope of finding a job there, should be responsive to changes in market access.

3 Methodology

The central objective of this paper is to empirically relate bilateral migration rates between the 27 Brazilian states to regional differences in market access. The derivation of the empirical specification of our migration equation used for this purpose is presented in this section and follows Grogger and Hanson (2008) and Sorensen et al. (2007).

We consider a log utility model as in Grogger and Hanson (2008). Every individual k coming from location i maximizes his indirect utility V_{kij} across all possible destinations j . The individual location choice M_{kij} in a general utility differential approach can then be written as:

$$M_{kij} = 1 \quad \text{if and only if} \quad V_{kij} = \max(V_{ki1}, \dots, V_{kiR}),$$

$$= 0 \quad \text{otherwise.}$$

⁵ If expenditure for manufacturing goods is high, real wages might be high both because the nominal wage is high and because the price index is low.

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

$$V_{kij} = X_{ij}\beta + \xi_{ij} + e_{kij} \quad (18)$$

The utility of migrating to region j for an individual k from origin i is determined by X_{ij} , the characteristics of the location j . The product $X_{ij}\beta$ represents the utility the individual receives from these characteristics, where β is a vector of marginal utilities. The subscript i is included because some characteristics of location j can vary across original locations, as for example bilateral distance.

The error term ξ_{ij} represents unobserved location characteristics. $X_{ij}\beta$ and ξ_{ij} assign the same utility level to all individuals coming from i and going to j . In order to still allow individuals from the same region to choose different locations, we include an idiosyncratic error term that varies across both individuals and locations, e_{kij} . We assume that this error term follows an i.i.d. extreme value distribution.

Given that individuals choose the location that maximizes their utility, Equation 18 leads to:

$$\begin{aligned} Pr(V_{kij} > V_{kim}) \forall j \neq m \\ Pr(e_{kij} - e_{kim} > X_{im}\beta - X_{ij}\beta + \xi_{im} - \xi_{ij}) \forall j \neq m \end{aligned} \quad (19)$$

McFadden (1974) shows that thanks to the i.i.d. extreme value distribution of the individual error term, integrating out over the distribution of the logistic distribution yields the following migration probabilities:

$$Pr(M_{kij} = 1) = \frac{\exp(X_{ij}\beta + \xi_{ij})}{\sum_{j=1}^J \exp(X_{ij}\beta + \xi_{ij})} \quad (20)$$

Using the methodology proposed by Berry (1994), this probability of migration from i to j can be interpreted as the share of individuals from i migrating to j . As in Sorensen et al. (2007), we then write the share of migrants from i to j , s_{ij} , as

$$s_{ij} = Pr(M_{kij} = 1) = \frac{\exp(X_{ij}\beta + \xi_{ij})}{\sum_{j=1}^J \exp(X_{ij}\beta + \xi_{ij})} \quad (21)$$

and the share of stayers of region i , s_{ii} , as

$$s_{ii} = Pr(M_{kii} = 1) = \frac{\exp(X_{ii}\beta + \xi_{ii})}{\sum_{j=1}^J \exp(X_{ij}\beta + \xi_{ij})} \quad (22)$$

Dividing Equation 21 by Equation 22 and taking the log yields:

$$\ln\left(\frac{s_{ij}}{s_{ii}}\right) = \ln\left(\frac{\exp(X_{ij}\beta + \xi_{ij})}{\exp(X_{ii}\beta + \xi_{ii})}\right) = X_{ij}\beta - X_{ii}\beta + \xi_{ij} - \xi_{ii} \quad (23)$$

In section 6.1, we estimate this migration equation adding a time dimension t and replacing the vector X by our location variables of interest (expressed in logarithms). This yields:

$$\ln m_{ijt} = \ln \frac{S_{ijt}}{S_{iit}} = \beta_1 + \beta_2 MA_{jt} - \beta_3 MA_{it} + \beta_4 w_{jt} - \beta_5 w_{it} + \varphi_{ij} + \eta_t + v_{ijt} \quad (24)$$

where the dependent variable m_{ijt} is defined as the ratio of the number of migrants in year t going from i to j over the number of stayers. MA and w correspond to the state's international market access to the state's wage level. The construction of these two variables is described in Section 4. η represents yearly fixed effects and v_{ijt} is a i.i.d. bilateral error term. With the bilateral fixed effects φ_{ij} , we attempt to capture time invariant unobserved location characteristics of the destination and the fact that they might vary depending on the state of origin. Due to data limitations, we cannot add variables on the states' amenities even though we suspect them to play a significant role in shaping migration patterns within Brazil. However, if they are constant over time, they are captured by the bilateral fixed effects.

As a consequence of controlling for bilateral and time fixed effects, we exploit the panel dimension of our data. The coefficients of market access indicate how a change in the origin's or destination's market access affects the evolution of migration between a given pair of states.

In Section 6.3, we are interested in the sectoral dimension of market access and migration flows. Introducing the sectoral dimension s , but restricting the analysis to a single year (due to data availability for the year 2000 only), our empirical specification becomes:

$$\ln m_{ijs} = \ln \frac{S_{ijs}}{S_{iis}} = \beta_1 + \beta_2 MA_{js} - \beta_3 MA_{is} + \beta_4 \widehat{w}_{js} - \beta_5 \widehat{w}_{is} + \varphi_{ij} + \theta_s + v_{ij} \quad (25)$$

where θ represents sector fixed effects.

A last important feature of our model is the independence of irrelevant alternatives (IIA). This implies that the probability of choosing one state relative to the probability of choosing another is independent of the characteristics of a third state. IIA arises from the assumption that the error terms are i.i.d. across alternative destinations. IIA may be violated if two or more destinations are perceived as close substitutes by potential migrants. Hausman and McFadden (1984) note that if IIA is satisfied, then the estimated regression coefficients should be stable across all 27 destination sets. To test whether IIA is violated in our sample, we follow Grogger and Hanson (2008) by re-estimating our model 27 times, each time dropping one of the destinations.⁶ Coefficients of our market access and wage variables stay similar

⁶ We undertake this procedure once for the panel data set from Section 6.1 and once for the

across samples, suggesting that the IIA property is not violated here.

4 Data

In this section we will first present the data sources and how we calculate our variables for migration rates, market access and region-specific wages.

4.1 Individual data

In this study, we use individual data for the years 1992 to 2003 (with data missing for 1994) collected and published by the Brazilian Institute of Geography and statistics (IBGE). Data the years 1992-1999 and 2001-2003 come from the yearly household survey *Pesquisa Nacional por Amostra de Domicilios* (PNAD), which covers between 310.000 and 390.000 individuals each year.⁷ In 1994 the PNAD was not conducted because of a strike. 2000 was the year of the population census which is conducted every ten years. We have access to a random sample covering over 20 million individuals which represents nearly 12% of the total Brazilian population. Migration rates and regional wages are obtained from these data sets.

We will use the PNAD data set to assess the importance of international market access for bilateral migration flows. Panel regressions in the first empirical section are based on this data set. The second empirical section concentrates on the sectoral composition of migration flows and the link to sectoral market access. For this, we use the Census data for 2000, given its much higher coverage that allows a disaggregation of migration flows at the sectoral level.

4.1.1 Migration rates

We define 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.

Given our interest in bilateral migration within Brazil, individuals having lived abroad in that period are excluded from our analysis. We further limit our sample to individuals who declare to have a job with a positive wage and who are at least 20 years old at the time of the interview, but not older than 65. In line with migration literature, we restrict our sample for the calculation of migration rates to information of household heads only. This leaves us with 69,000 (1993) to 93,000

cross-section in Section 6.3, where we look at the sectoral composition of migration flows and market access.

⁷ The PNAD does not follow individuals, but interviews a different random sample of residents each year.

(2003) individuals for each round of the PNAD and nearly 4.6 million individuals in the census year 2000.

Bilateral migration rates are defined as the number of migrants from state i to j over the number of workers that stayed in state i . Whereas for 2000, thanks to the access to the census data sample, we observe positive migration flows for 697 out of 702 state combinations, we have only on average 435 flows in each year of the PNAD. For the panel regression analysis we exclude migration rates involving the state Tocantins, leaving us with on average 406 out of 650 migration flows.⁸

Despite this relatively high number of zero-flows, the PNAD is considered as being representative of real migration rates and thus adequate for studying migration patterns within Brazil (Fiess and Verner, 2002). We will nevertheless address the problem of unobserved flows by various robustness checks. In Section 6.5, we will run Poisson-Maximum-Likelihood estimations including the zero-flows. However, when comparing regressions on the aggregated level for 2000 which contains migration rates for all possible flows, and yearly regressions of the PNAD, we find that coefficients do not vary substantially, indicating that a bias resulting from the omission of zero-flows might be small.⁹

To take into account heterogeneity of the impact of market access across different types of individuals, we calculate migration rates also at more disaggregated levels.

On the one hand, we distinguish between low and highly qualified individuals, where a person is defined as qualified when she has more than eight or nine years of schooling.¹⁰ On the other hand, we are interested in bilateral migration rates at the sectoral level. These rates are obtained from the census data for the year 2000. Given that we don't have information about the individual's work in 1995, we make the assumption that individuals already worked in the same sector as they did in 2000. Sectoral market access is computable only for manufacturing goods and agriculture, we hence limit our analysis to individuals who declare working in these industries. This reduces our sample to nearly 465,000 individuals for 2000.

4.1.2 Regional wages

Our migration equation will also contain state-specific wage variables. Regional wage differentials are considered as one of the key determinants of migration patterns. Also it is important to take into account this variable since, as described in Section 2, market access can simultaneously impact wages and migration. The simplest way

⁸ Tocantins has officially separated from Goiás in 1988. Trade data for this state is not available for all years of the PNAD and for the first half of the 90s statistical information on this state is not yet reliable.

⁹ Yearly regressions on the PNAD and aggregated migration flows for the 2000 census are not reported but are available at request.

¹⁰ We use the information on school years and on the highest degree the individual has obtained or is enrolled in to identify the educational level of each individual.

to include a wage variable in our specification would be to take the average wage of each region. There are two reasons, not to do so. First, given that an individual's wage depend crucially on his educational attainment, it would be better to allow the wage variable to vary across the educational levels. Second, as highlighted in the literature, individuals currently living in state j are not a random sample of the population, because similar individuals tend to move to the same locations. This is reflected by the fact that immigrants in a specific region often share common characteristics like gender, ethnic group or educational level. Since these variables are also the main wage determinants, the wage level in a region will correspond to the average characteristics of the people living there. As a consequence, the average wage level is subject to a potential selection bias.

This problem is particularly important in the case of estimating migration as a function of wages. The individual's location choice is not a function of the average wage level in each region, but depends on the wage she will gain in each region. Her personal wage expectations are likely to differ substantially from the mean depending on her personal characteristics. To circumvent this problem, Falaris (1987) proposes to use predicted wages that have been estimated by a Mincerian wage equation using an additional correction term, that accounts for self-selected migration.

We follow closely De Vreyer et al. (2009) in obtaining predicted wages of each individual corrected for self-selection. These authors generalize the approach of self-selection correction developed by Dahl (2002). In his study on inter-state migration in the US, Dahl (2002) includes the individual's migration probabilities into the state-specific wage equation. In a first step, we thus need to determine migration probabilities of all individuals in our data set. For this, we estimate a multinomial logit model based on the individual's location choice.¹¹

This allows us to obtain the probabilities of an individual to migrate to each of the Brazilian states as a function of his personal characteristics. A polynomial of these migration probabilities are then added as additional regressors in the wage equations to get consistent estimates of the coefficients.¹²

The obtained estimates are then used to predict wages for each individual in each of the 27 states.

¹¹ For the sake of brevity, regression results used for the construction of the wage variable are not reported but are available on request. Traditional selection bias corrections (like the conditional logit model) are not very well suited to cases where individual migration decisions imply numerous potential destinations. We thus run a multinomial logit separately for each year, predicting individual migration probabilities for each state based on the individual's educational level, age group, family status, state of origin and a dummy for migrants.

¹² The wage equation includes the individual's education, gender, age, ethnic group and its location (rural vs urban). As in Dahl (2002), we include as correction terms the individual's probability of living in the state of destination and the probability of staying in the original state. We try different specifications of the wage equation including more migration probabilities and different number of polynomials, but find only small differences in predicted wages. Wage equations are run separately for each year.

We are interested in comparing the wage an individual gains in her actual state of residence with the wage she could gain in any other of the 27 Brazilian states.

Since we aggregate migrants for the estimation of bilateral migration rates, we will use also an aggregated wage variable. The aggregation of individual wages to regional average wages is done in the following way: The wage variable for the state of origin i , w_i , will be the average of all workers who actually lived in i five years ago (more precisely, the mean of their predicted wages). In contrast, the wage variable for a given destination j , w_j , varies according to the state of origin of the migrants. For each destination j , we compute the wage variable w_j according to the origin of the residents. The wage variable of São Paulo thus differs between individuals coming from Amazonas and those coming from Acre. For the former we take the mean of the predicted wages for São Paulo of all the workers declaring Amazonas as the origin, and for the later we take the mean of the predicted wages for São Paulo for the individuals having lived in Acre. This method allows us to use the mean of predicted wages of the same individuals for computing the regional wage at the origin and at the destination, what reduces again any potential selection bias.

We compute the wage variable once using the predicted wages corrected by self-selection bias and once using predicted wages that do not take into account the additional correction term. Both variables are highly correlated and in the empirical analysis, results obtained with the corrected and with the uncorrected wages are mostly the same, so in the interest of space, we will mainly report and discuss results with corrected wages.

In section 6.2, we distinguish migration rates and wages according to the educational level of migrants. The wage variables are obtained in the same way as on the aggregated bilateral level, but take different values for high and low educated workers. In section 6.3, we look at migration flows by industry. In these regressions, regional wages are industry-specific.¹³

By introducing the state's wage level in our migration equation, we also control for an impact of market access on wages. The market access impact we observe in our regression is the effect on migration beyond the effect on wages.

4.2 Trade data and Market Access Calculation

Our main variable of interest is the state's market access, as defined in Equation 7 in Section 2. To obtain a sound estimate of market access, we follow the methodology

¹³ Note that we are using nominal wages here, which are deflated for all years using the Brazilian consumer price index. The pertinent variables for migration should be real wages that are deflated by region-specific price indices. Unfortunately, information on state-specific price indices covers only main Brazilian cities in a limited number of states. By including dyadic fixed effects in our final regression, we can control at least for time invariant differences in price indices between two states.

pioneered by Redding and Venables (2004).

We estimate the following gravity trade equation to get estimates for bilateral transport costs and market capacity of each trade partner which can then be used to compute a market access indicator for each Brazilian state.¹⁴

$$EX_{ij} = n_i p_{ij} q_{ij} = \underbrace{n_i p_i^{1-\sigma}}_{FX_i} \phi_{ij} \underbrace{\mu E_j G_j^{\sigma-1}}_{FM_j} \quad (26)$$

where EX_{ij} are bilateral exports from region i to region j . i and j can be either a Brazilian state or one of the 210 countries included in the trade data set. Under the assumption of homogeneous firms, total exports from i to j can be written as a function of the quantity q_{ij} sold at the price p_{ij} on market j by the number of firms in i , n_i .

The first term on the right hand side of the equation, $n_i p_i^{1-\sigma}$, is considered as the supply capacity of region i . It is defined as the number of firms active in i times the *f.o.b.* price. Supply capacity and market capacity are country-specific for country i and country j and can be captured by exporter and importer fixed effects, FX_i and FM_j . The freeness of trade variable, ϕ_{ij} , can be proxied by different variables that enhance or deter trade such as bilateral distance d_{ij} and contiguity C_{ij} .

This methodology is rarely applied in regional studies because of data limitations: bilateral trade flows are rarely available at the intranational level, particularly for developing countries. Brazilian states are a fortunate exception with an interstate trade matrix for 1999, including even information on the sectoral composition of trade flows. Also, international trade flows at the state level are available. For a detailed description of the data and its sources, please refer to Appendix B.

4.2.1 Sectoral market access

For the calculation of a sector-region specific market access variable, we estimate equation 4.2 separately for each of the 20 manufacturing sectors into which we can classify the Brazilian and international trade flows.¹⁵

Taking the logs of equation ??, our estimated specification of the trade equation

¹⁴ Detailed derivation of the trade gravity equation, the market access construction and additional estimations can be found in Paillacar (2006) and Fally et al. (2008), available at <http://team.univ-paris1.fr/teamperso/paillacar/>.

¹⁵ Data is available at ISIC Rev 3 at 2-digits. The classification used in the official statistics is the Brazilian nomenclature CNAE 3.1 which is fully equivalent to ISIC Rev 3 at the level of aggregation that we are considering (See Table A-1 in Appendix A for the full list). Average values of the estimated coefficient across industries are available at Fally et al., 2008, Table 2, second column

becomes

$$\ln EX_{ij} = FX_i + FM_j + \delta \ln d_{ij} + \lambda_1 C_{ij} + \lambda_2 B_{ij} + \lambda_3 (B_{ij} * C_{ij}) + \lambda_4 b_{ij} + u_{ij} \quad (27)$$

The freeness of trade, ϕ_{ij} , is captured by the bilateral distance d_{ij} , the contiguity, C_{ij} , of two trading partners and the crossing of a border. As we include intranational and international data in the sector specific trade regression, we can differentiate between a national border, b_{ij} , and $B_{ij} = 1$ for trade flows between a Brazilian state and a foreign country. We assume that crossing a national border implies several costs and consequently we expect a negative coefficient for this dummy variable. In addition, we introduce the contiguity interacted with national border to consider the possibility of differentiated effects of contiguity depending on whether the neighbor is another Brazilian state or a foreign country. u_{ij} is a random bilateral error term. Note that since we estimate the gravity equation separately for each sector, all coefficients and fixed effects are allowed to vary across sectors. This allows us to build a state-sector specific market access indicator.

Following Head and Mayer (2006) we include internal trade observations X_{ii} , trade flows within the same state or country, calculated as production minus total exports. This allows to identify the internal border b_{ij} between two Brazilian states. To capture the fact that additional costs are implied when a product leaves the region, this dummy is defined as 1 for all trade flows, except the internal flows. We expect its coefficient also to be negative.

According to the differentiation between internal, national and international trade flows, we can thus define a region's market access as composed of three parts reflecting the market access of the own state (trade within the state), of the rest of the country (trade flows with the rest of the country) and market access of the international markets.

$$MA_i^{Total} = MA_i^{Local} + MA_i^{National} + MA_i^{International} \quad (28)$$

with

$$MA_i^{Local} = \exp(\widehat{FM}_i) d_{ii}^{\delta} ; \quad d_{ii} = 2/3\sqrt{area_i/\pi} \quad (29)$$

$$MA_i^{National} = \sum_{j \neq i}^{regions} \exp(\widehat{FM}_j) d_{ij}^{\delta} \exp(\widehat{\lambda}_1 C_{ij} + \widehat{\lambda}_4) \quad (30)$$

$$MA_i^{International} = \sum_j^{countries} \exp(\widehat{FM}_j) d_{ij}^{\delta} \exp(\widehat{\lambda}_1 C_{ij} + \widehat{\lambda}_3 B_{ij} * C_{ij}) \exp(\widehat{\lambda}_2) \quad (31)$$

where \widehat{var} indicates estimates obtained from the trade regressions.

Table 1: International market access by state (1993).

	Code	State	International market access	Rank
North	11	Rondônia	89,36	21
	12	Acre	33,83	25
	13	Amazonas	131,99	12
	14	Roraima	16,45	26
	15	Pará	147,14	7
	16	Amapá	81,71	22
	17	Tocantins		
Northeast	21	Maranhao	110,59	17
	22	Piauí	108,67	19
	23	Ceará	140,06	9
	24	Rio Grande do Norte	110,49	18
	25	Paraíba	112,11	16
	26	Pernambuco	138,49	11
	27	Alagoas	119,01	15
	28	Sergipe	67,12	24
	29	Bahia	144,77	8
Southeast	31	Minas Gerais	153,22	2
	32	Espírito Santo	139,10	10
	33	Rio de Janeiro	150,03	5
	35	São Paulo	156,30	1
South	41	Paraná	149,98	6
	42	Santa Catarina	150,39	4
	43	Rio Grande do Sul	151,17	3
Center-West	50	Mato Grosso do Sul	106,55	20
	51	Mato Grosso	121,12	14
	52	Goiás	124,84	13
	53	Distrito Federal	68,75	23

Source: own calculations.

4.2.2 Market access for panel regressions

Our migration panel data set covers the years 1993 to 2003. We thus calculate market access the years 1992 to 2002, so that we use a lagged market access variable. In order to allow coefficients and fixed effects to vary across years, we run trade regressions on equation 27 separately for each year. Since intranational trade flows are available only for 1999, the market access variable used in the panel data regressions will be computed only with international trade flows in order to be consistent

over years. Although a complete indicator of market access would be interesting to study, as described in the introduction, employing only the international part of market access has the important advantage of reducing endogeneity. Endogeneity between market access and migration can result from the fact that high migration inflows increase local demand for goods and thus local market access. By focusing on the international component of market access, we mitigate the endogeneity concern that has often been addressed to studies using market access including domestic demand. In the regression analysis, we will however also proxy domestic demand by the state’s income per capita and population to control for a possible positive correlation between international and local market access. Summary statistics of this market access indicator are reported in Table 1. We see that international market access varies a lot between states. In 1992, the lowest market access (belonging to Roraima in the North) is only 10% of São Paulo’s, where market access is the highest throughout the sample period. Because of trade data unavailability for Tocantins, we cannot compute the indicator for all years for this state and exclude it from the panel regressions.

5 Pattern of migration flows

In this section, we present some stylized facts and summary statistics about the migration patterns observed in our data and the relation between migration and the state’s access to markets. This section is based on the final data sample we identified in Section 4.1.1. For a detailed description of internal migration in Brazil see Fiess and Verner (2003).

5.1 Descriptive statistics on migration

Column 1 in Table 2 reports the total number of migrants observed for each year. We see that on average 3.31% of all individuals in our final data set have moved within the last five year to another state, but migration has declined over the sample period. Columns 2 and 3 report the numbers and percentages for high and low qualified workers. Qualified individuals are throughout all years more mobile than less qualified workers.¹⁶

When comparing mobility between sectors, we see that it varies substantially across sectors.¹⁷ Whereas in the leather, bags and footwear industry only 2.8% of workers have moved to another state, this percentage is up to 5% in coke, nuclear and petroleum and nearly 4% in food, beverages and tobacco.

¹⁶ Tables and figures in this section would be very similar also if we wouldn’t limit our analysis to household heads only or use a finer classification for education.

¹⁷ In the interest of space, numbers are not reported here.

Table 2: Yearly migration rates

Year	(1)		(2)		(3)	
	All Nb	%	Highly educated Nb	%	Low educated Nb	%
1992	2,635	3.81	782	4.89	1,853	3.49
1993	2,503	3.54	782	4.59	1,721	3.20
1995	2,409	3.22	789	4.31	1,620	2.87
1996	2,278	3.07	776	4.06	1,502	2.73
1997	1,859	3.31	702	4.46	1,157	2.87
1998	2,543	3.21	949	4.31	1,594	2.79
1999	2,685	3.30	1,034	4.48	1,651	2.83
2000*	155,544	3.39	49,704	4.27	105,840	3.09
2001	2,955	3.28	1,180	4.25	1,775	2.85
2002	3,037	3.31	1,264	4.21	1,773	2.87
2003	2,916	3.13	1,239	3.94	1,677	2.72
Mean		3.31		4.31		2.91

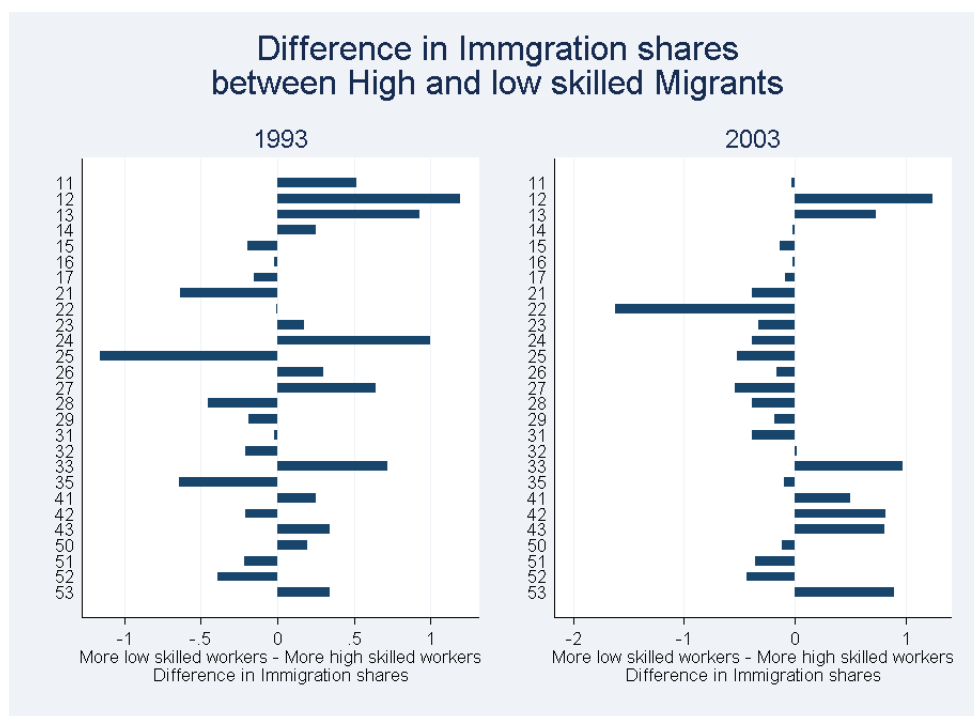
Source: Own calculations. * Data are from the PNAD (1992 - 2003) and the 2000 Population Census (2000). Means are calculated excluding 2000.

To see whether migrants with different educational levels have different location preferences, Figure 1 displays differences in migrant shares between the two educational groups for each state for the years 1993 and 2003. In 1993 migrants with higher education are dominant in at least one state in every Brazilian macro region.¹⁸ Whereas the Distrito Federal (53), where the capital Brasilia is located, Rio de Janeiro (33) and the three states in the South (41 - 43), which are known for their good climate and recent economic development, have stayed the preferred destinations of highly educated workers, the North East and Center region (except the Distrito Federal), have become the main immigration regions 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. They are also an indicator for self-selection or for differences in location choice determinants across different types

¹⁸ The Brazilian states are grouped in five macro-regions (see also Table 1 and the map in Figure A-1 in Appendix A.). This classification is based on the structural and economic development of the different states, regrouping states with similar characteristics. The North is sparsely populated, poor, and largely inaccessible. The Northeast is the poorest macro-region of Brazil with the lowest life expectancy and wages, little access to mineral deposits or navigable rivers, and the highest proportion of low educated persons. The Center-West combines a diverse set of characteristics, mixing poor rural areas, dense forests, and the federal capital city of Brasilia, where income and education levels are high. The Southeast and the South are the most economically developed regions of Brazil. Education levels, income and life expectancy are all high in these regions, and dense highway networks make it easy to get around. These regions offer high economic opportunities and have a high population density.

Figure 1: Differences in migrant shares between high and low qualified workers



Source: Own calculations.

of migrants.

5.2 Correlation between migration and market access

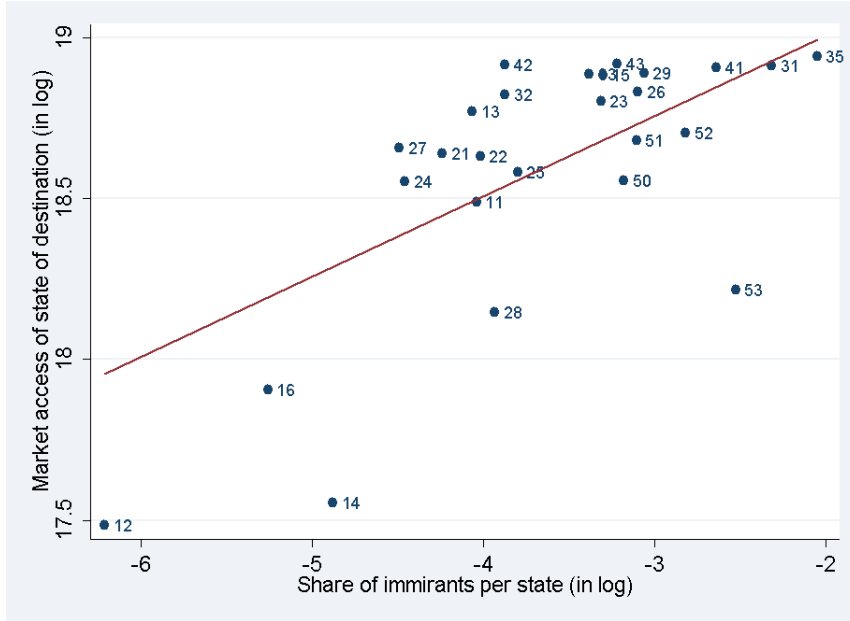
Before we come to the regression analysis, we report some simple statistics and figures to show the link between migration and market access.

Figure 2 plots the share of workers out of all migrants a state attracts against its level of market access. We can see a positive correlation between the two variables, indicating that the higher the state's international market access, the higher its share of migrants.¹⁹

Figure 3 displays the link between bilateral migration and regional differences in market access. In this figure, we use the rank of market access for each region, displayed in Table 1. São Paulo ranks first. The smallest market access, rank number 26, is found in Roraima (14) at the border to Venezuela. We see that when the rank at the state of origin is lower than the rank of the destination, migration is high. Whereas, when the state has a low rank of market access, workers are less likely to go to these states. The bigger the difference between the two states, the higher the migration rate between them.

¹⁹ Graphs and correlations are very similar throughout all years in our sample, we thus report only 1993 here.

Figure 2: Market access and migrant shares



Source: Own calculations.

6 Results - Market Access and migration response

In this section, we estimate the migration equation, equation 25 derived in section 3, using first panel data estimation techniques. We then look at differences in sectoral market access to evaluate the role of sector-specific market access and heterogeneity in the reaction to market access across manufacturing sectors. At last, we will show several robustness test on our market access indicator.

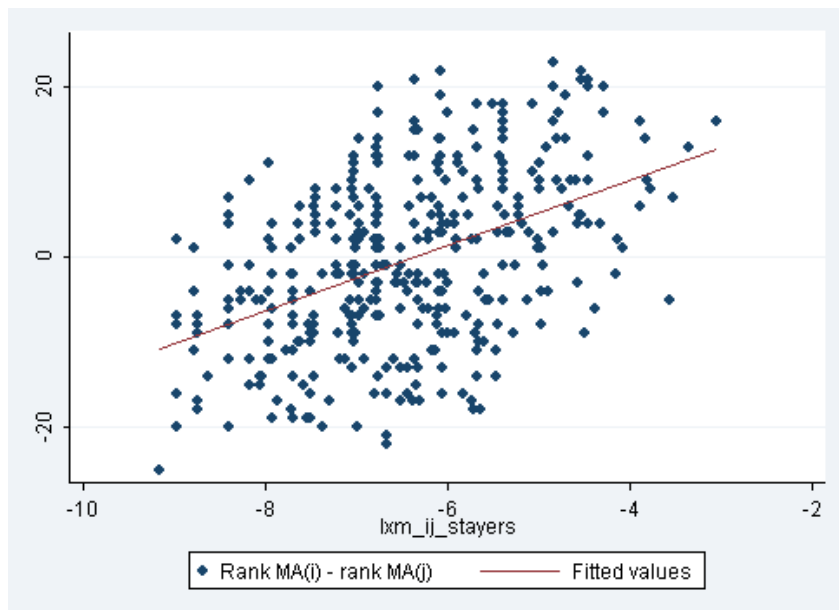
6.1 Panel regressions - aggregated migration rates

In this section, we test empirically whether an increase in a state's market access results in higher immigration and lower emigration rates. Further, we are interested in the role of regional differences in wages, which are considered as important determinants in the migration decision.

We use a panel including seven years between 1993 and 2003.²⁰ Our panel variable is defined as the origin-destination state couple. Accordingly, we include dummy variables for each combination of origin and destination state. The dummy for the couple ij (São Paulo - Amazonas) differs from the dummy for the couple ji (Amazonas - São Paulo). These bilateral fixed effects take into account time-

²⁰ In the panel data set, we use 1993, 1996, 1997, 1998, 1999, 2002 and 2003. Although we have data for 1992, 1995 and 2001, these years are dropped when using lagged wage variables. We thus exclude these years from the beginning to have a consistent number of observations throughout all regressions.

Figure 3: Market access and bilateral migration rates



Source: Own calculations. Migration rates are defined as m_{ijs} , the probability of worker in sector s living in i to migrate to state j .

invariant specificities concerning migration between two particular states. They typically capture differences in climate, price indices or institutions (those which are stable over our sample period).

Since we are looking at bilateral migration, we include each independent variable once for the state of origin and once for the state of destination. Our main interest lies in the state's access to international markets, MA . We expect market access of the origin, MA_i , to have a negative impact on migration: the higher MA_i , the higher the demand for work. Thus, the individual can attain already a high utility in his home region and is not necessarily motivated to look for a job in another state. For the destination's market access (MA_j) we expect a positive coefficient: the higher this indicator, the more the region attracts people in search for a job in export-oriented industries.

A gravity model of migration, as we use, should also include a proxy for moving costs between two states (Greenwood, 1997). Moving costs are expected to have a negative impact on the number of migrants and are increasing in distance since the farther away is the destination, the more expensive is the journey and the less familiar is the new environment (climate, institutions, cultural specificities). Given that distance is constant over time, we will not be able to use this proxy when introducing dyadic fixed effects. A component reducing migration costs is the stock of immigrants in a given region. A region with a higher number of immigrants from origin i disposes of a better social network and thus offers better integration

of an additional arriving migrant from i (Card, 2001). We will thus also include the number of persons who have been born in i but lived already in j at time $t - 5$, when we explain migration in time t between i and j .

Market access and wage variables are all lagged one year in order to reduce endogeneity. Market access is computable only from 1991 on and regional wages from 1992 on. Using a higher number of lags will thus reduce significantly our number of observations. However, we run also regressions with lags of two and three years, which confirm main results. To take into account that market access and regional wages are estimated in a prior regression, displayed standard errors, which are clustered at the origin-destination-couple-level, are bootstrapped.

Before reporting results with dyadic fixed effects, we run regressions with regional fixed effects, where we define one dummy variable for each of the five macro-regions. This allows us to include distance and the stock of migrants in our regression.

In the first column of Table 3, we regress bilateral migration rates only over market access, bilateral distance and year fixed effects and regional dummies to control for time invariant specificities of each region.

The positive impact of MA_j confirms our expectations on the role of market access in migration. States that experience an increase in this variable attract workers. The negative coefficient of MA_i shows that outflows are lower with increasing market access. The negative and significant parameter of the distance is in line with the literature and shows that migrants tend to stay close to their home. In the second column we add the proxy for the stock of migrants. As expected, this variable has a positive impact on migration rates. Market access of the destination becomes non significant.

From Column 3 on, we finally define our panel variable and introduce dyadic fixed effects.

We obtain again significant results for both market access variables, suggesting that this indicator plays indeed a role in the migration choice. From the test on the comparison of the magnitudes of both market access coefficients we can conclude that absolute magnitudes are not statistically different. Adding the stock of migrants when controlling for dyadic fixed effects, leads to a non significant coefficient of the former, probably because the migrant stock does not vary sufficiently in time. Market access of the destination and the origin stay both significant (Column 4).

We consider here that the primary channel by which high market access attracts migrants are the good employment opportunities, including not only the type of job but also the high number of available jobs. This last component is typically also reflected by the local unemployment rate, which is often considered in the estimation of gravity migration. Whereas our market access indicator should be a determinant of contracted as well as speculative migration, the unemployment rate is mostly

Table 3: Bilateral Migration

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Migrants over stayers								
MA_j	0.279 ^a (0.107)	0.040 (0.091)	0.323 ^a (0.087)	0.306 ^a (0.093)	0.327 ^a (0.089)	0.322 ^a (0.089)	0.321 ^a (0.089)	0.312 ^a (0.091)	
MA_i	-0.904 ^a (0.125)	-1.338 ^a (0.126)	-0.360 ^b (0.162)	-0.354 ^b (0.169)	-0.379 ^b (0.167)	-0.360 ^b (0.164)	-0.357 ^b (0.169)	-0.313 ^c (0.163)	
$dist_{ij}$	-0.703 ^a (0.060)	-0.316 ^a (0.062)							
$stock_{ij}$		0.332 ^a (0.028)		0.050 (0.040)					
$unemp_j$					0.213 (0.226)				
$unemp_i$					-0.313 (0.262)				
w_j (unc)						-0.022 (0.174)			
w_i (unc)						-0.009 (0.197)			
w_j (cor)							-0.015 (0.062)	-0.015 (0.064)	
w_i (cor)							0.038 (0.069)	0.044 (0.066)	
pop_j								0.958 ^c (0.513)	
pop_i								-2.895 ^a (0.580)	
$gdp\ cap_j$								0.384 ^c (0.210)	
$gdp\ cap_i$								-0.280 (0.244)	
MA_j/MA_i									0.307 ^a (0.091)
w_j/w_i (cor)									-0.031 (0.053)
pop_j/pop_i									1.791 ^a (0.466)
gdp_j/gdp_i									0.334 ^b
Observations	2473	2473	2473	2473	2473	2473	2473	2473	2473
R^2	0.372	0.495	0.865	0.865	0.865	0.865	0.865	0.868	0.867

^a, ^b and ^c represent respectively statistical significance at the 1%, 5% and 10% levels. All regressions include time fixed effects. Columns 1 and 2 contain region fixed effects. Column 3 to 9 contain dyadic fixed effects. Bootstrapped standard errors in parentheses (1000 replications), clustered at the origin-destination level. All variables are in logs and lagged one year.

important for the speculative type of migration. As data do not provide us with this distinction explicitly, we try to capture this effect through the unemployment rate.

Since official unemployment data at the state level is not available, we include in Column 5 a local unemployment rate computed from the survey data,²¹ None of the two unemployment variables is significant and no changes in our market access indicator can be observed. We do not conclude that unemployment is not a key parameter for migration, but rather prefer the explanation given by Greenwood (1997). He argues that using local unemployment rates in migration equations has many limitations. For instance, a measure of labor turnover could be more appropriate because migrants “are more concerned by rates at which hiring for new jobs is taking place”.

From Column 6 on, we introduce the state’s wage levels. In Column 6 we first use the wage variables which are computed without taking into account a possible selection bias. Wages in Column 7 are corrected for this bias. The introduction of the mean of the predicted wages does not alter the coefficients or significance of the market access indicator, since we observe no significant impact any of the two different wage variables.²² In his survey, Greenwood (1997) underlines that empirical evidence has often failed to confirm the importance of wage differentials on migration. According to this author, the non-significant effect of wages could result from the omission of time variant amenities. While our fixed effects control for location-specific features that are constant over time, it is possible that some amenities evolve in the opposed direction of nominal wages.

A further explanation for the non-significant effect of wage differentials is that they are reflecting spot wage differentials. But migrants might value other job characteristics like stability or career opportunities more than the actual wage. Aguayo-Tellez et al. (2008) have shown at the individual level that the impact of spot wage differentials on the location decision of Brazilian migrants is not robust.

Another explanation could stem simply from the fact that the regional average of the predicted wages is still too general and neglects a strong heterogeneity between educational or levels. In the next subsection, where we distinguish between different educational levels of workers, we will thus use state-education specific average wages.

In the last two columns, we introduce two additional controls, population and GDP per capita.²³ Both variables can be seen as a proxy for local economic activity and are strongly correlated with our market access variable (see Table A-2). It is

²¹ Regional unemployment rates are only available for some of the big cities on the coast.

²² We re-estimated Columns 6 to 9 with different specifications of the wage variables, using different correction terms, different aggregation criteria and just observed wages, but none had a significant sign.

²³ Data on population and GDP for the Brazilian states are from the IPEA (*Instituto de Pesquisa Econômica Aplicada*).

thus possible, that the significant coefficient of our variable of interest captures, at least partially, the impact of the evolution of population or of the per capita income. Indeed, when adding the log of the state's population, the coefficient of market access of the state of origin decreases slightly but stays significant at the 10% level.

The negative sign of the coefficient for population at the state of origin, can also be explained by the fact that a higher population is normally associated with a bigger surface and/or a greater number of cities. It is thus more likely that an individual can find a (better) job within the same state and is less likely to be urged to search for a job farer away (keeping in mind that distance has a strong negative impact on the location choice of migrants).

An increase in GDP per capita has a less important impact than an increase in population and does not alter much the magnitude of market access. Also here, coefficients are according to our expectations. Richer regions are much less likely to send migrants.

In the last column, we replace the origin and destination variables by their respective ratios in order to test directly the importance of the differences in these variables between two different states. The ratios for market access, population and GDP per capita are positive and significant, whereas the ratio of wages is non significant, confirming previous results.

A preliminary conclusion on this first section thus suggests that international market access can be considered as a good indicator for export-oriented economic opportunities and as a determinant of bilateral migration rates.

6.2 Panel regression - high versus low educated migrants

In this section, we split our sample according to the two levels of education to see whether low or highly educated migrants are more sensitive to changes in market access.

Fally et al. (2009) show that for Brazil the impact of international market access varies between workers with high and low education. They find a stronger impact on wages of low educated workers. It is therefore likely that the impact of market access on migration varies also across educational levels.

According to Redding and Schott (2003), who predict a higher wage premium of market access for skilled workers, we could expect that highly educated workers have a stronger incentive to go to regions with good access to foreign markets. In this case, the coefficient of market access of the destination should be higher for highly educated individuals.

But the opposite effect is also possible: highly qualified people might react less to differences in market access because; thanks to their skills, they are more likely

to find a job in the region of their choice. Highly educated individuals are known to be very sensitive to non-pecuniary location characteristics like pollution, crime rate, schools, parks etc. For example, Levy and Wadycki (1974) have shown for Venezuelan data that educated individuals tend to value amenities much more than low qualified and Schwartz (1973) argues that the negative impact of distance on migration flows diminishes with educational attainment.

We first look at low qualified workers with no more than nine years of schooling. Results are reported in Table 4.

Table 4: Low educated migrants

	(1)	(2)	(3)	(4)	(5)
	Migrants over stayers - Low educated only				
MA_j	0.281 ^b (0.143)	0.409 ^a (0.134)	0.402 ^a (0.123)	0.406 ^a (0.128)	
MA_i	-1.170 ^a (0.150)	-0.413 ^c (0.218)	-0.414 ^c (0.227)	-0.363 ^c (0.214)	
$dist_{ij}$	-0.557 ^a (0.068)				
w_j			-0.103 (0.065)	-0.098 (0.068)	
w_i			0.013 (0.070)	0.026 (0.070)	
pop_j				0.153 (0.664)	
pop_i				-1.664 ^b (0.770)	
$gdp\ cap_j$				0.488 ^c (0.250)	
$gdp\ cap_i$				-0.269 (0.290)	
MA_j/MA_i					0.380 ^a (0.116)
w_j/w_i					-0.063 (0.048)
pop_j/pop_i					0.756 (0.551)
$gdp\ pc_j/gdp\ pc_i$					0.390 ^b (0.175)
Observations	1841	1841	1841	1841	1841
R^2	0.339	0.877	0.877	0.879	0.878

^a, ^b and ^c represent respectively statistical significance at the 1%, 5% and 10% levels. All regressions include time fixed effects. Column 1 contains region fixed effects. Column 2 to 5 contain dyadic fixed effects. Bootstrapped standard errors in parentheses (1000 replications), clustered at the origin-destination level. Explanatory variables are in logs and lagged one year.

Table 5: Highly educated migrants

	(1)	(2)	(3)	(4)	(5)
	Migrants over stayers - High-educated only				
MA_j	0.062 (0.103)	0.074 (0.107)	0.083 (0.109)	0.082 (0.107)	
MA_i	-0.969 ^a (0.134)	-0.351 ^c (0.180)	-0.358 ^b (0.177)	-0.320 ^c (0.177)	
$dist_{ij}$	-0.482 ^a (0.053)				
w_j			0.163 ^b (0.083)	0.166 ^b (0.081)	
w_i			-0.052 (0.073)	-0.004 (0.075)	
pop_j				1.026 ^c (0.573)	
pop_i				-3.746 ^a (0.643)	
$gdp\ cap_j$				0.058 (0.225)	
$gdp\ cap_i$				-0.424 ^c (0.246)	
MA_j/MA_i					0.154 (0.095)
w_j/w_i					0.083 (0.053)
pop_j/pop_i					2.171 ^a (0.437)
$gdp\ pc_j/gdp\ pc_i$					0.232 (0.166)
Observations	1884	1884	1884	1884	1884
R^2	0.373	0.838	0.839	0.844	0.842

^a, ^b and ^c represent respectively statistical significance at the 1%, 5% and 10% levels. All regressions include time fixed effects. Column 1 contains region fixed effects. Column 2 to 5 contain dyadic fixed effects. Bootstrapped standard errors in parentheses (1000 replications), clustered at the origin-destination level. Explanatory variables are in logs and lagged one year.

The coefficients for the destination's market access is higher than those found at the aggregated level, but the origin's market access is significant only at the 10% level. The wage variable used in this subsection is the mean of the predicted wages for each state by level of education. It varies between low and highly qualified workers and corresponds therefore much better to the expectations for each individual. Nevertheless, wages are not significant for low qualified workers. In the last column, we display again the ratio between the origin and the destination variables.

When looking at the estimates for qualified workers, presented in Table 5, we

see that market access of the origin has similar coefficients and significance as seen for low qualified migrants. In contrast, in none of the columns, market access of the destination turns out significant. Even in the last column, the ratio of market access of the destination and the origin is non significant.

The most important variable here is the population size, indicating that high educated people search the big agglomerations. But we also observe a positive impact of the destination's wage level. For all educational groups, population size is a significant determinant of bilateral migration rates.

Results at the aggregated level thus seem to be mainly driven by the sensitivity to market access of less educated workers.²⁴ The differences in migration patterns between educational levels can thus be explained in part by their different sensitivity to market access and wages. Economic opportunities associated to international trade are most important for the location choice of low educated individuals. The findings of no impact of changes in market access is in line with the migration literature cited above, highlighting that with a higher educational achievement non-pecuniary benefits are crucial for the location choice. Unfortunately, we do not have data on amenities so that we cannot directly control for their effects.²⁵ Also, since Brazil is exporting mainly goods that are intensive in unskilled-labor, an increase in exports will lead to an increase in demand for low educated workers. Maia (2001) studies the impact of trade and technology on employment in Brazil and finds that both were destroying more unskilled than skilled job. Low-educated workers are thus more likely to be touched by the loss of their previous job and thus are obliged to move to another state to find a new employment.

6.3 Sector-specific market access - pooled regressions

In this section, we will use the 2000 Census data set, restricting our attention to workers in one of the sixteen manufacturing industries in our sample. The migration rates used here are defined as m_{ijs} , the probability of worker in sector s living in i to migrate to state j .

The specification of migration rates at the sectoral level enables us to include market access at the sectoral level as regressor in the migration equation. Since we can calculate sector-specific market access only for a single year, results reported in this section are based on a cross-section for 2000. Magnitudes of estimates based on

²⁴ Note that the numbers of observations are much lower here. Coefficients are therefore likely to be of expected signs but estimated imprecisely. This can lead to non significant parameters.

²⁵ Of course, it is also possible that international market access and amenities evolve simultaneously and into the same direction. In this case, we risk to overestimate the effect of market access. But given that market access is not significant for individuals with high education, who are known to react the most sensitively with respect to amenities, the coefficients on market access are unlikely to be biased by omitted amenities.

the census data can thus not be compared directly with the estimates on the PNAD panel data set. Estimates in this section refer to differences in levels and not, as in the previous sections, to growth rates.

Table 6: Sectoral Market Access

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Migrants over stayers							
MA_j^s	0.139 ^a (0.054)	0.106 ^b (0.050)	0.173 ^a (0.042)			0.166 ^a (0.042)	0.166 ^a (0.040)	
MA_i^s	-0.144 ^a (0.055)	-0.135 ^b (0.052)	-0.245 ^a (0.043)			-0.242 ^a (0.045)	-0.241 ^a (0.042)	
<i>Inter. MA_j^s</i>				0.489 ^b (0.215)				
<i>Inter. MA_i^s</i>				-1.054 ^a (0.266)				
<i>Local MA_j^s</i>					0.145 ^a (0.035)			
<i>Local MA_i^s</i>					-0.226 ^a (0.039)			
$dist_{ij}$	-0.499 ^a (0.076)	-0.336 ^a (0.087)						
$Stock_{ij}$		0.144 ^a (0.034)						
w_j (unc)						0.641 ^a (0.171)		
w_i (unc)						-0.357 ^b (0.149)		
w_j (cor)							0.639 ^a (0.184)	
w_i (cor)							-0.382 ^b (0.153)	
MA_j/MA_i								0.204 ^a (0.033)
w_j/w_i (cor)								0.493 ^a (0.132)
Observations	3201	3201	3201	3201	3201	3201	3201	3201
R^2	0.518	0.531	0.802	0.799	0.802	0.803	0.803	0.803

^a, ^b and ^c represent respectively statistical significance at the 1%, 5% and 10% levels. Columns 1 and 2 contain industry and region fixed effects. Column 3 to 8 contain industry and dyadic fixed effects. Bootstrapped standard errors in parentheses (1000 replications), clustered at the origin-destination level. Explanatory variables are in logs and lagged one year.

As before, we first report in Column 1 and 2 of Table 6 regressions including

only market access, macro-region fixed effects and the distance parameter.

Coefficients of market access for destination and origin in Column 1 have the expected sign and are highly significant. The introduction of the stock of migrants is again highly significant and decreases the parameters of market access but both MA indicators stay significant.

From Column 3 on, in addition to the industry dummies, we use dyadic fixed effects as in the previous section. Note that this doesn't allow us to introduce any other bilateral, state or industry-specific variables. Both market access variables stay highly significant and with the expected signs. Testing for the equality of the magnitude of the two MA coefficients indicates however that the origin's market access has a stronger impact than the destination's market access.

Since we have a sectoral market access including the local, national and international component, we can test for the relative importance of each of the components. Column 4 looks at the impact of international market access only. The market access variable used in Column 5 is limited to the access to the own state. These two regressions show that it is not only the local or the international dimension that drives the impact of market access, but that all of its components play a significant role in the migration decision. The importance of the international and national component confirm that including only GDP per capita or other state characteristics that represent only the local economic activity can not capture migrants that come to a state because of its opportunities related to its relative location in the country or in the world economy.

In the previous section, we concentrated on the international market access, but at the aggregated level. Given that we are here in a cross-section, we cannot compare directly the coefficients on the international part of the sector-specific market access (Column 5) with the parameters of the international market access used in the panel regressions. Since we are however interested in knowing which one of the two is more important in the migration choice, we re-estimate the impact of the sector-specific international market access as in Column 5 but using regional fixed effects plus the bilateral distance instead of dyadic fixed effects. We then re-estimate the same equation but replacing the market access variable with the one employed in the previous section (Section 6.1), the aggregated international market access for all sectors (see Table A-3). Comparing the beta coefficients of these two regressions, we find that beta coefficients of the sector-specific international market access have bigger magnitudes than the aggregated international market access.

These findings, together with the always highly significant impact of MA_i^s and MA_j^s , suggests that even if a region might have a high aggregated market access, this is not necessarily a sufficient argument for a worker to move to that region. The specific sectoral conditions seem to play a stronger role in sending and attracting

migrants.

The next two columns in Table 6, we control for wages (uncorrected and corrected for self-selection). As before, differences between the corrected and uncorrected wages are negligible, and magnitudes of the MA parameters are not affected. But in contrast to the previous section, sector-specific average wages enter the migration equation significantly for the origin and the destination, supporting the explanation that aggregated wages were not accounting for strong regional wage heterogeneity within and across sectors. Estimates in Column 7 represent our benchmark regression, which will be used in Section 7 for the prediction of sector-specific bilateral migration rates and the simulation of the impact on migration following changes in the market access variable.

Also in the last column, when using market access and wage ratios between the two states, wages keep their significant impact.

6.4 Sector-specific market access - the heterogeneous impact across sectors

NEG theory predicts that the location choice of a firm depends stronger on market access when transport costs are high and increasing returns to scale are important. Industries that have higher transport costs and a high degree of increasing returns to scale should thus be more sensitive to market access. Paillacar (2009) shows in a cross-country study that out of 27 manufacturing sectors, only 16 sectors exhibit a robust impact of market access on the wage level in the respective industry. We could thus also expect heterogeneity in the sensitivity of migration rates to market access.

To see whether there are differences in the reaction to market access across sectors, we allow the impact of market access to vary across industries. Having sectoral data only for the year 2000, we cannot run regressions separately by sector while controlling simultaneously for dyadic (or even state) fixed effects. In order to keep these controls, we run one regression including all sectors but interact the market access variable with the respective industry dummy. Coefficients of market access reported in Table 7 are thus sector specific. All but four out of the 16 sectors show a positive and significant sign for the destination's market access and only for wood industry the sign is negative. The sectors Food, beverages and tobacco and other non-metallic mineral products are the only ones with a non significant impact for the origin. Again, for wood manufacturing this coefficient is against expectations, being positive and significant.²⁶

²⁶ Results do not change much when controlling for wages and/or allowing the impact of wages to vary across sectors.

Table 7: Sectoral market access - by sector

Sector	(1)	(2)
	Migrants over Stayers	
	MA_j^s	MA_i^s
Food and beverages & Tobacco products	0.171 ^a (0.060)	0.055 (0.069)
Textiles	0.156 ^a (0.055)	-0.317 ^a (0.068)
Wearing Apparel	0.153 ^a (0.058)	-0.187 ^a (0.071)
Leather, bags and footwear	0.109 (0.104)	-0.464 ^a (0.101)
Wood	-0.075 (0.064)	0.228 ^a (0.067)
Paper and paper products, publishing and printing	0.184 ^a (0.046)	-0.301 ^a (0.052)
Coke, petroleum and nuclear fuel	-0.042 (0.102)	-0.133 (0.092)
Chemicals	0.216 ^a (0.057)	-0.295 ^a (0.055)
Rubber and plastic products	0.273 ^a (0.059)	-0.544 ^a (0.055)
Other non-metallic mineral products	0.118 ^b (0.053)	-0.016 (0.060)
Basic metals	0.028 (0.126)	-0.675 ^a (0.111)
Fabricated metal products	0.224 ^a (0.052)	-0.332 ^a (0.055)
Machinery and equipment	0.206 ^a (0.064)	-0.339 ^a (0.059)
Electrical machinery	0.119 ^c (0.073)	-0.419 ^a (0.066)
Motor vehicles, transport equipment	0.311 ^a (0.067)	-0.416 ^a (0.061)
Furniture	0.193 ^a (0.052)	-0.102 ^b (0.052)
Observations		3201
R^2		0.837

^a, ^b and ^c represent respectively statistical significance at the 1%, 5% and 10% levels. The regression contains industry and dyadic fixed effects. Bootstrapped standard errors in parentheses (1000 replications), clustered at the origin-destination level. Explanatory variables are in logs and lagged one year.

This detailed regression shows us that previous results are not driven by some particular sectors. We see however that there are differences across sectors, in particular in the magnitudes of the parameters of the origin's market access. Nevertheless, market access seems to play an important role in the migration choice for nearly all

manufacturing workers.

These findings contribute to a better understanding of the NEG forces at work in the real economy. Whereas NEG theory knows only one manufacturing sector that produces one differentiated good with labor completely mobile between the different varieties, we know that in Brazil workers do not move freely between industries. Though we observe the agglomeration effect described in NEG (higher demand attracts new workers and leads to bigger agglomerations), this mechanism is not affecting all individuals in the same way. Workers will choose their destination depending on their sector and the spatial structure of their industry's market access. As a consequence, the biggest agglomeration will not necessarily attract all workers. It is more likely that individuals will go to different locations, depending on the region's sectoral composition. Once a region has acquired an advantage in a certain industry (due to comparative advantages or other) and its market access increases for this industry, specialization will be facilitated, because the high sectoral market access will attract corresponding workers. We hence observe an adjustment by the quantity mechanism also at the sectoral level.

Table 8: Sectoral market access - High versus low qualified

	(1)	(2)	(3)	(4)	(5)	(6)
	Migrants over stayers					
	Low educated			Highly educated		
MA_j	0.198 ^a (0.052)	0.174 ^a (0.054)		0.167 ^a (0.049)	0.151 ^a (0.049)	
MA_i	-0.151 ^a (0.052)	-0.254 ^a (0.057)		-0.324 ^a (0.054)	-0.311 ^a (0.055)	
w_j (cor)		0.540 ^a (0.204)			0.909 ^a (0.210)	
w_i (cor)		-0.246 (0.223)			-0.705 ^a (0.205)	
MA_j/MA_i			0.215 ^a (0.042)			0.235 ^a (0.038)
w_j/w_i			0.394 ^b (0.169)			0.799 ^a (0.147)
Observations	2447	2447	2447	1923	1923	1923
R^2	0.716	0.807	0.807	0.832	0.836	0.835

^a, ^b and ^c represent respectively statistical significance at the 1%, 5% and 10% levels. All columns contain industry and dyadic fixed effects. Bootstrapped standard errors in parentheses (1000 replications), clustered at the origin-destination level. Explanatory variables are in logs and lagged one year.

6.4.1 Market access and migration: impact of education and sectors

In Section 6.2 we stated differences in the sensitivity to international market access between highly and low qualified workers. To see whether these two groups also differ in their reaction to sector-specific market access, we again split our sample between low and highly educated workers. Results are reported in Table 8. The wage variable is defined as the mean of the predicted wages (corrected for self-selection) for the individuals from i with education e and working in sector s .

We find that market access has a significant impact both for low and high-educated workers. A difference in the sensitivity to market access can not be clearly seen in these estimations. Results might be driven by restricting the sample only to workers in manufacturing sectors or that we do not look at growth rates. The state's access to markets in the sector the individual is employed seems to be important for all individuals. However, we confirm results on wages from Section 6.2: highly educated workers value differences in wages more than less educated migrants.

6.5 Robustness checks - GMM and ML estimations

In the previous sections, we have seen that the market access variables were robust to the inclusion of additional controls and in the case of sector-specific migration and market access, also to the splitting of the sample between different educational groups. Market access could thus be considered as a good indicator for trade-oriented economic opportunities that attracts migrants. Yet, there are still some concerns about the robustness of our results linked to possible omitted variables and the high number of missing trade flows that could affect our results.

First, there is a concern of reverse causality. A region's market access is likely to increase when new migrants come to the region. Regressing only international market access on internal migration reduces this causality concern. There is however still a risk of some time-varying unobservables like changes in amenities that could be correlated with the error term. To tackle these issues of endogeneity, we perform a two-step GMM estimator applied to first differenced data for the regression reported in Column 3 of Table 3. The endogeneity problem is of course much bigger for the sector-specific market access since we include also the local component, but having data for only one year, we cannot perform any GMM estimations on this data set. In section 6.3 we have presented regressions including only international or local market access to show that results are not driven only by the local component which is the likely to be the most endogenous one.

GMM results are presented in Table 9. Conditions for the validity of GMM are met: the transformed error terms exhibit first-order correlations, and the test rejects second-order correlations. Instruments are not rejected by the Hansen and

Table 9: Non linear estimations - GMM

	(1)	(2)	(3)
	Migrants over stayers		
	lags 3-4	lags 3-5	lags 3-6
$L.mig_{ij}$	0.21 (0.28)	0.02 (0.21)	0.13 (0.15)
MA_j	0.88 ^b (0.45)	0.78 ^c (0.44)	0.61 ^c (0.33)
MA_i	1.09 (0.74)	0.92 (0.65)	0.77 (0.56)
No. of obs	1514	1514	1514
No. of groups	352	352	352
Hansen(p)	0.31	0.70	0.68
Sargan(p)	0.21	0.52	0.56
AR(1)p	0.02	0.02	0.00
AR(2)p	0.59	0.85	0.59
No. inst.	27	37	45

Standard errors in parentheses, with ^a, ^b and ^c denoting significance at the 1, 5 and 10% levels respectively. Standard errors were estimated using the correction proposed by Windmeijer (2005). All regressions include time dummies and dyadic fixed effects. All explanatory variables are lagged one year. AR(1) and AR(2) are tests for Autocorrelation for the first and second lags. Hansen and Sargan tests indicate the validity of instruments.

Sargan tests. In the interest of space, we report only the coefficients for the lagged dependent variable and market access. The three regressions presented are identical except for the number of lags chosen for the instruments. Following the recent literature (Roodman, 2009) we minimize the use of instruments to avoid over-fitting. For instance, the first column shows the regression resulting from using only the third and fourth lags as instruments. Standard errors were estimated by using the correction proposed by Windmeijer (2005).

The first lag for our migration variable is never significant, suggesting the absence of persistence in bilateral migration flows after having controlled for origin-destination-pair fixed effects. Whilst more imprecisely estimated, coefficients for destination market access of the GMM estimations are supportive of the results found in our panel regressions above. Regarding origin market access, we fail to find significant coefficients in the GMM regressions.

As a last test for the robustness of our market access indicator, we address the concern of zero-value flows.

Both the survey data from which we obtain migration rates and the trade data we use to compute our market access variable are characterized by a high number of zero-value flows. In this context, OLS regressions on our gravity equations can be criticized because they lead to biased estimates in the presence of many zero value flows. A high number of zero value trade flows is likely to result in biased estimates of the parameters used to compute market access. To treat zero-flows in a gravity equation researchers often recur to non-linear estimations (e.g. Gamma or Poisson regressions) that allow to deal with some specific problems of trade data.

Table 10: Non-linear estimations - Aggregated Panel data

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Migrants over stayers (aggregated flows) with Zeros							
	Poisson		Gamma		Poisson		Gamma	
MA_j	0.889 ^a (0.249)	0.340 ^b (0.149)	0.762 ^a (0.158)	0.175 (0.191)				
MA_i	0.056 (0.107)	-0.164 (0.198)	-0.008 (0.144)	-0.181 (0.386)				
MA_j (zero)					0.823 ^a (0.267)	0.092 (0.152)	0.863 ^a (0.192)	0.114 (0.218)
MA_i (zero)					0.083 (0.136)	-0.173 (0.152)	0.074 (0.164)	-0.159 (0.292)
$dist_{ij}$	-1.136 ^a (0.073)	-1.213 ^a (0.070)	-1.126 ^a (0.070)	-1.345 ^a (0.064)	-1.143 ^a (0.077)	-1.213 ^a (0.070)	-1.136 ^a (0.071)	-1.345 ^a (0.064)
Fixed effects	region	state	region	state	region	state	region	state
Observations	4550	4550	4550	4550	4550	4550	4550	4550

^a, ^b and ^c represent respectively statistical significance at the 1%, 5% and 10% levels. All columns include time fixed effects. Bootstrapped standard errors in parentheses (1000 replications), clustered at the origin-destination level. Explanatory variables are in logs and lagged one year.

A definitive method on how the problem of zero-flows should be treated has however not yet emerged (Silva and Tenreyro (2006); Martinez-Zarzoso et al. (2007); Martin and Pham (2008); Helpman et al. (2008); Silva and Tenreyro (2009)). The question is even more difficult to answer when non-linear estimations are combined with location fixed effects (Buch et al., 2006). When the proportion of zeros becomes too important, iterative methods for Poisson or Gamma estimations often do not converge. This problem appears to be particularly acute, when we run trade regressions at the sectoral level. The proportion of zeros can easily surpass 80%,

Table 11: Non linear estimations - Sectoral MA

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Migrants over stayers (sector specific) with Zeros							
	Poisson		Gamma		Poisson		Gamma	
MA_j^s	0.427 ^a (0.115)	0.308 ^a (0.092)	0.371 ^a (0.070)	0.205 ^a (0.074)				
MA_i^s	-0.066 (0.123)	-0.118 (0.097)	0.034 (0.069)	-0.067 (0.084)				
MA_j^s (zero)					0.069 (0.050)	-0.018 (0.037)	0.076 ^b (0.035)	0.084 ^b (0.034)
MA_i^s (zero)					-0.049 (0.058)	-0.030 (0.032)	-0.027 (0.034)	-0.070 ^b (0.031)
$dist_{ij}$	-0.992 ^a (0.169)	-1.167 ^a (0.105)	-1.125 ^a (0.098)	-1.395 ^a (0.084)	-1.004 ^a (0.169)	-1.156 ^a (0.107)	-1.083 ^a (0.098)	-1.378 ^a (0.085)
Fixed effects	region	state	region	state	region	state	region	state
Observations	11232	11232	11232	11232	10530	10530	10530	10530

^a, ^b and ^c represent respectively statistical significance at the 1%, 5% and 10% levels. All columns contain industry fixed effects. Bootstrapped standard errors in parentheses (1000 replications), clustered at the origin-destination level. Explanatory variables are in logs and lagged one year.

and iterative methods for Poisson or Gamma estimations often do not converge. In order to explore the sensibility of results to a nonlinear estimation including zeros, we reduce the dimensionality of the zeros, by dropping all countries that trade with less than 20 partners. The proportion of zeros in the more extreme cases then falls to maximal 50%, allowing the convergence of all but one sector. Still, in an important number of the regressions border effects or fixed effects are not identified, reducing dramatically the quality and availability of market access measures. We thus keep as our main variable market access obtained by OLS estimations.

In Table 10, we report Poisson and Gamma regressions on all possible bilateral migration flows, including migration rates that take the value zero in our data set. Two kinds of location fixed effects were considered: macro region fixed effects and origin and destination state fixed effects. Regressions with origin-destination-pair fixed effects as used before do not converge and are thus not displayed here. The absence of dyadic fixed effects allows the inclusion of the distance coefficient.

In the first four columns, we use the same market access variable as in Table 3, but the migration equation is estimated using Poisson or Gamma methods including all zero value migration flows. The count data methods confirm a positive impact

of the destination's market access. However, including the zeros leads to a non significant impact of the origin's market access.

In the second half of the table (columns 5 to 8), we repeat the same regressions, but use a market access variable that is computed including zero trade flows. Hence, estimations in both stages are made using count data methods.

Table 11 shows the corresponding estimations for the sector-specific migration rates and sectoral market access. Market access of the destination is mostly significant and in the last column, also the origin's market access enters significantly the migration equation.²⁷

The non-linear estimations, as well as the use of *MA* measures including zero trade flows confirm main results from the OLS regressions in the previous sections. Results here are also in line with those found in the GMM regressions, which suggest that the destination's market access is robust across specifications.

7 The impact of national and international integration: Simulations

In this final section, we simulate a reduction of impediments to trade in order to better understand the impact of market access and its evolution on the migration pattern within Brazil. The classical question asked in NEG models is the following: How will a reduction in trade costs influence the number of migrants and their destination choices? To approach this question in our context, two issues must be operationalized. First, what kind of aggregation level should be used (regional, sectoral, both)? Second, what kind of trade cost changes should be simulated?

Regarding the first issue, we prefer an approach that takes into account the sectoral as well as the regional dimension, we thus look at bilateral migration rates that are sector-specific as in Section 6.3.

Regarding the choice of trade cost reduction, we have to consider the Brazilian context. In the recent past, Brazil undertook much effort to equilibrate national and federal demands of competitiveness trying to improve internal and international integration. Recent policies aiming at interlinking and developing Brazilian regions are the investment in highways, tax exemptions intended to raise competitiveness for some lagged regions, and, of course, the famous relocation of the capital to the interior of the country in an entirely new city (Brasilia). In parallel, there have been trade liberalization policies aiming at integrating Brazil into the world economy. These reforms took the form of unilateral reduction of tariffs and the creation of

²⁷ We have fewer observations when we use market access which has been calculated using zeros because in this case, trade equations for two sectors do not converge. They are thus excluded from the regression.

a common market within the Mercosur. All these efforts led to higher trade flows and are reflected in an increase of Brazil's market access in a whole. Some states like Amazonas, São Paulo and Santa Catarina, saw their market access increase in particular because of their location close to the new national and international markets.

In this section, we investigate the possible consequences of a further reduction in internal and external trade barriers on the internal migration pattern. Which type of liberalization, the international or the intranational one, changes more the spatial distribution of the work force in Brazil? Which sectors appear more concerned?

To answer these questions, we perform a simulation exercise of migration flows similar to Hunt and Mueller (2004), who study intra- and international migration between Canada and the US, focussing on returns to skill and on the international border effect as main migration determinants. These authors predict changes in migration rates based on estimates of their migration equation. In order to evaluate the consequences of a decrease in the international border effect, they reduce gradually the value of the coefficient of the border effect obtained in the estimation of their original migration equation. Changes in migration flows are calculated as the difference between migration flows predicted with the observed border effect (the benchmark) and the migration flows predicted with the reduced border effect.

Instead of just mechanically reducing the whole component of trade costs, we prefer to study the impact of the specific part of market access that could be eventually modified through policy makers by tariff reductions, infrastructure policies or others: the average border effect estimated in the gravity trade equation (Equation 27).

We distinguish between two border effects. The international border for trade between Brazil and another country, and the intranational border effect for trade between two Brazilian states. For international (intranational) trade liberalization we reduce the value of the estimated coefficient of the international (intranational) border, which are obtained separately for each sector from the gravity equation. In the following sections, we comment results from a 50% decrease in the border effects.²⁸ The coefficient of the international border effect is not changed and vice versa. It is important to note that the applied reduction is on the sector-specific coefficient on the border dummy. Since border effects vary across sectors, a 50% reduction can represent a much larger reduction for one sector than for another. Because of regional specialization, the impact of this reduction is not balanced among Brazilian states.

²⁸ Results do not change qualitatively between different levels of border effect reduction. Only magnitudes are affected, since effects are linear. We concentrate on the 50% reduction which is already a very strong decrease - a complete annulation of any border effect being even less realistic.

Computing market access with the reduced border effects shows that the overall increase in market access when reducing the international border effect is sizeable, but not as big as when reducing the intranational border effect.

The counterfactual sector-specific bilateral migration rates are obtained using the new values of sectoral market access in the specification from Column 7 of Table 6. In order to be able to make predictions in absolute numbers, we will employ as our dependent variable migrants over the whole population instead of migrants over stayers.²⁹ As shown in Table A-4 in Appendix A, coefficients do not change significantly.

Finally we want to stress that our simulations capture only the direct effect of a reduction in the border effect on the migration rates. Since we cannot predict neither the indirect effect of market access on migration via its impact on wages nor do we control for physical endowment and comparative advantages of industries, we cannot draw conclusions on the final impact of trade liberalization on regional specialization or on the spatial distribution of the work force. Our simulations mainly highlight that a reduction in trade costs can have an economically significant impact on migration rates and that effects are likely to vary substantially across states and sectors.

7.1 Changes in sector-specific migration rates

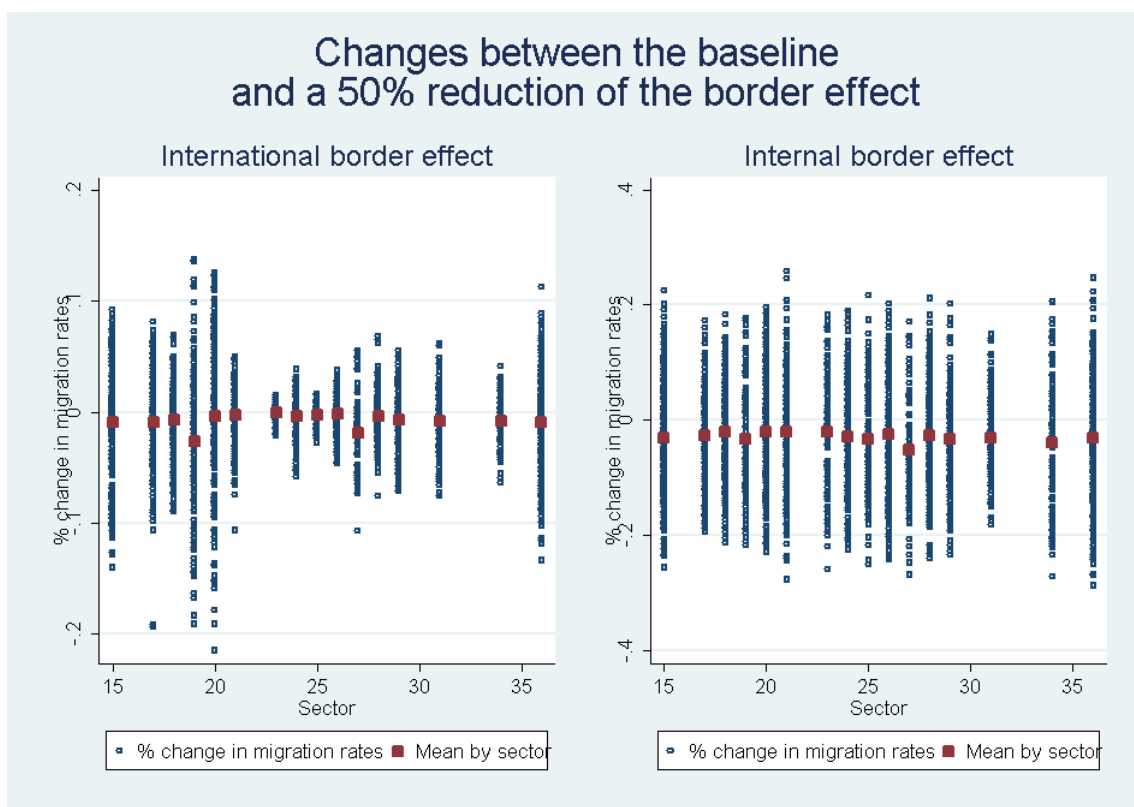
The source of regional differentials in trade performances can result in different consequences for the migration patterns. In traditional trade models, workers relocate between sectors. However, empirical evidence for Brazil points out that sector relocation is weak (Menezes-Filho and Muendler, 2007). Hence, spatial wage differentials should reflect differences in regional specializations. States specialized in sectors in which Brazil holds a comparative advantage should have the higher wages.

Under these conditions, migrants are likely to go to regions that host firms in the sector they are specialized in. When in addition increasing returns come into play, migration rates will also reflect advantages linked to the relative location of the regions (irrespective of comparative advantage). Since most manufacturing industries are likely to exhibit some increasing returns to scale, increasing returns could thus reinforce comparative advantages.

In the case of trade liberalization we thus expect that regions that have a traditional comparative advantage in one of Brazilian's main exporting industries, will exhibit more intense net immigration and that the net inflows are higher the more important the industry's returns to scale. Estimating potential changes in migration

²⁹ When obtaining an increase of i.e. 5% of the migration rates when using migrants over stayers, this results could be due to an increase in migrants or a decrease in stayers. Using total population which is very big alleviates this problem.

Figure 4: Simulation: changes in migration rates by sector



Source: Own calculations. Codes for sectors s are reported on the x-axis (see Table A-1 for the corresponding names of each sector). Migration rates are defined as m_{ijs} , the probability of worker in sector s living in i to migrate to state j .

rates at the sectoral level appears thus to be the appropriate aggregation level.

Figures and tables in this and the following section concentrate on the computation of migration rates obtained with a reduction of 50% of the original border effect (either of the international or the intranational border). Figure 4 plots the change in percentages in each sector-specific bilateral migration rate, once for international and once for intranational trade liberalization. On the x-axis, we report the sector of origin of the migrants.

For both types of liberalization each sector will experience an increase in some migration rates and a decrease in others. Changes are more important for a decrease in the intranational border effect, but are still very low on average. The averages below zero indicate that we see a decrease in the total number of migrants. Most importantly, the figure reveals important variations in the impact across sectors, mainly in the case of international trade liberalization. Coke and petroleum (23), for example, shows only little reaction to a decrease in the international border effect. In Table 7, we couldn't find any significant impact of market access in this sector. Variations are more stronger for fabricated metal products (28) and

Machinery and equipment (29), where we also observed an important impact of market access. Variations for Leather (19), and especially Wood production (20), should be interpreted more cautiously since the destination's market access has not been found significant, when allowing the market access coefficient to vary across industries.

Muendler (2007) calculates a comparative-advantage indices measure, proposed by Balassa (1965), for Brazilian industries for the period 1986 to 2001. The manufacturing sectors with the highest international comparative advantages³⁰ expose important changes in migration rates when international trade is facilitated. This suggests that already a relatively small change in the spatial distribution of market access in these sectors is sufficient to initiate migration. This confirms the idea stated above that increasing returns might reinforce the comparative advantage of a region.

These findings highlight the fact that NEG implications for wage inequality are highly sector-specific: we can expect that wage differentials in certain industries remain higher, when migrants are less mobile. In other sectors, a human capital specificity may allow workers to reap important gains, which provides enough incentives to migrate even to places far away. If trade liberalization in Brazil is changing market access, the benefits are likely to be captured mainly by these manufacturing workers.

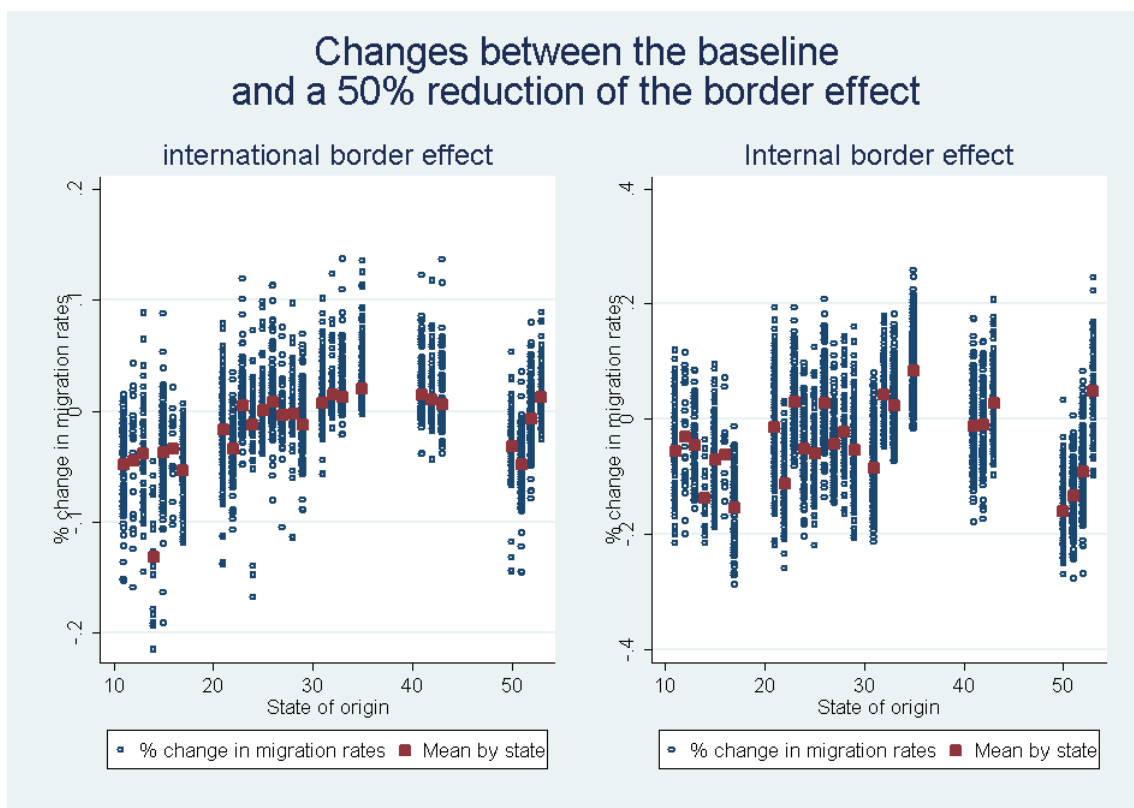
Since the location choice depends strongly on the sector in which the individual is working, spatial wage inequality can persist since regions specialize in different industries and workers are not sufficiently mobile between these industries to induce a factor price equalization.

7.2 Changes in state-specific migration rates

Having uncovered the heterogeneity across industries, we want to see how the spatial migration pattern changes. Are certain industries agglomerating in certain states and are there regions that will see an increase or a decrease in net immigration? Figure 5 plots again the change in percentages for each type of trade liberalization for each sector-specific bilateral migration rate. Here, we report on the x-axis the state of origin i of the migrant working in sector s . That means, a negative change corresponds to the result that less people working in a specific sector s are leaving i to go to j . We find that each state sees certain migration rates decrease and others increase, but the average is for some positive and for some negative.

³⁰ Muendler (2007) finds that Brazil has the highest comparative advantages in food and beverages and tobacco (15); leather, bags and footwear (19); wood (20); paper and paper products (21); rubber and plastic products (25); other non-metallic mineral products (26); basic metals(27) and fabricated metal products (28).

Figure 5: Simulation: changes in migration rates by state of origin



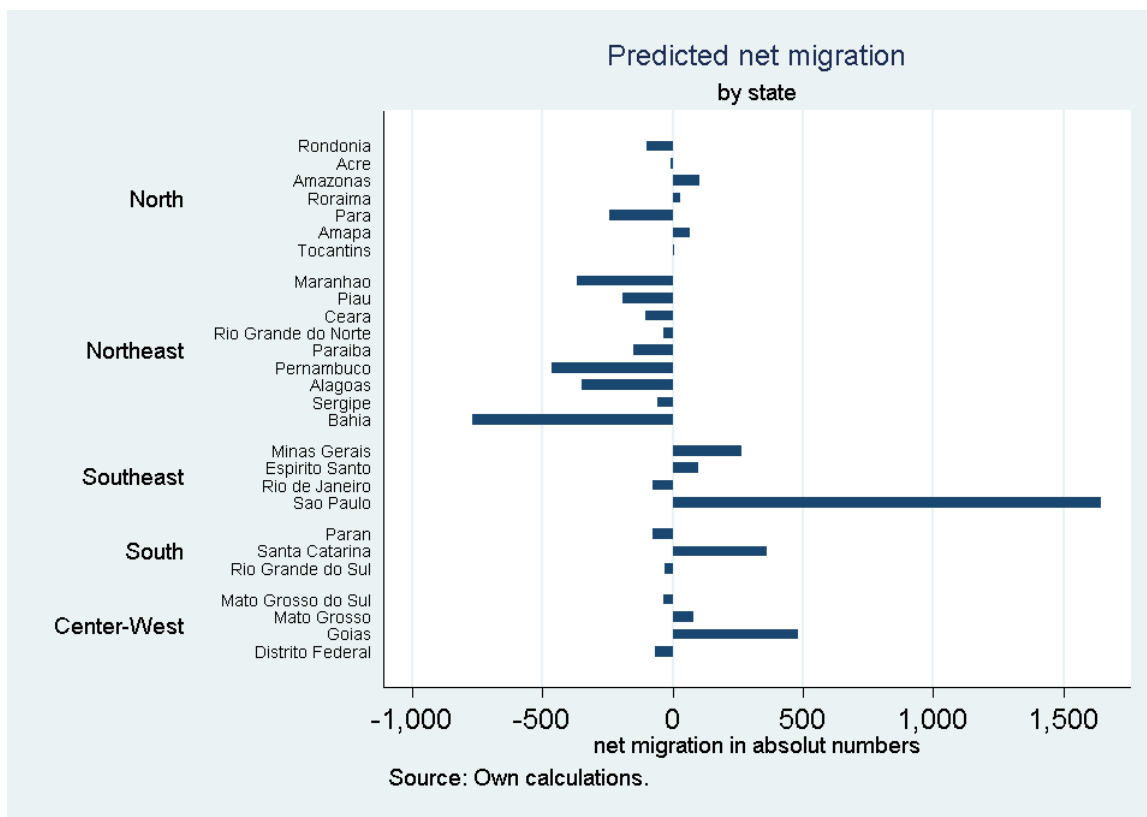
Source: Own calculations. Codes for states of origin i are reported on the x-axis (see Table 1 for the corresponding names of each state). Predicted migration rates are defined as the probability of worker in sector s living in i to migrate to state j .

In São Paulo (35) and the other states of the South East (30 - 34) international trade liberalization leads to an increase of most of the migration rates, meaning that they see more people leave these states than before. On the contrary, emigration of the Northern states (11 - 17) is less pronounced than before. Many of the predicted emigration rates from this region are much lower after a reduction in the international border effect. Hence, following an international trade liberalization, the North becomes relatively more and the South East relatively less attractive for workers. A similar pattern is observed as a consequence of intranational liberalization.

We then calculate the absolute number of migrants and find that overall migration has decreased after trade liberalization. In our data set we observe 16,298 migrating manufacturing workers in 2000 (household heads only). Our benchmark regression from Column 7 in Table 6 underestimates slightly the total number of migrants, predicting only 14,301 migrants.³¹ Reducing the international border effect by half will result in a prediction of only 14,158 migrants and no border effect at

³¹ Absolute numbers are obtained using Column 2 or Table A-4.

Figure 6: Predicted migration rates



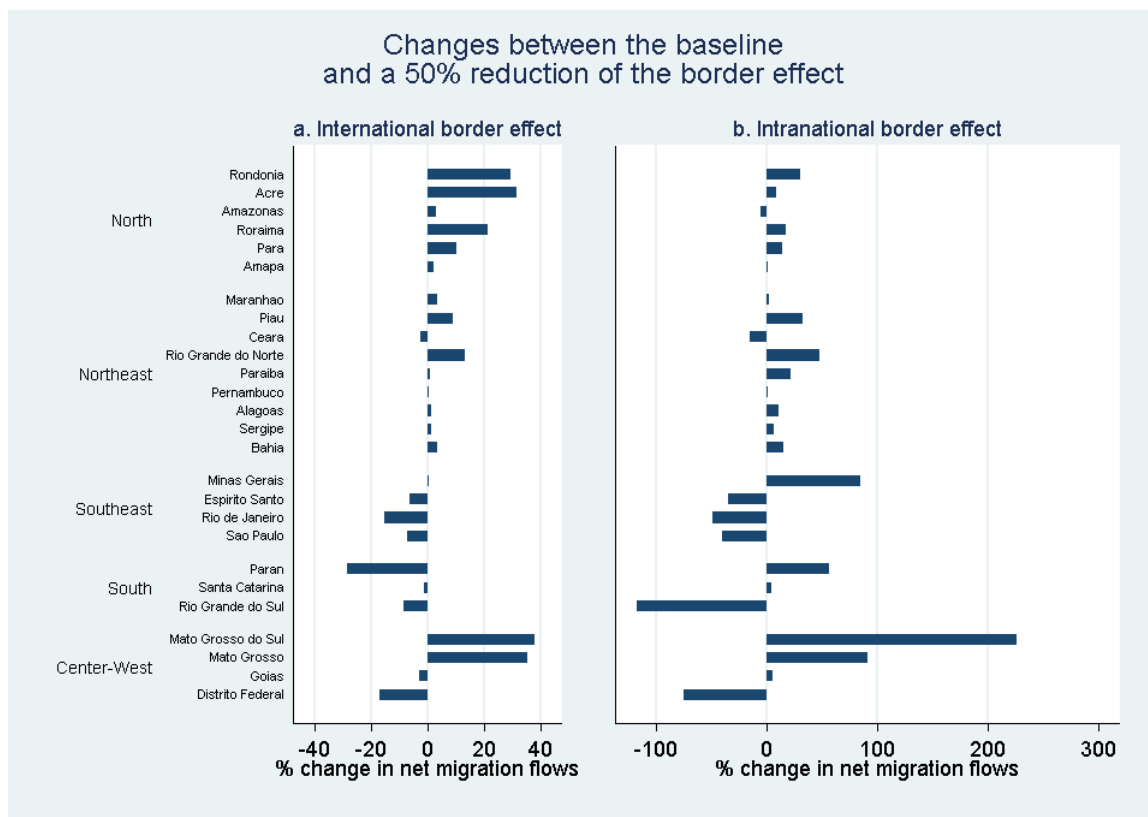
Source: Own calculations. Net migration rates are calculated using sector-specific bilateral migration rates predicted based on estimates of Table A-4, Column 2.

all in 13,651 migrants. Nevertheless, there are several states that will receive more migrants than before. In order to see which states will benefit in terms of additional migrants from the decrease in the border effect, we calculate net migration rates before (obtained from the benchmark regression) and after its reduction.

Figure 6 shows the resulting net migration for each state calculated with predicted sector-specific bilateral migration rates from our benchmark regression. As can be seen, São Paulo (35), Santa Catarina (42) in the South and Goiás (52) in the Center-West exhibit a positive net migration, while the North and Northeast (with the important exception of Amazonas) tend to exhibit negative net migration. The dramatic emigration from the poorest regions in the Northeast is a very well known fact of Brazilian migration (Fiess and Verner, 2003). This picture is simply the actual pattern of migration in Brazil.

The impact of our exercise on net migration flows for each state can be seen in Figure 7. The percentage changes for each flow after a hypothetical reduction in the international border effect are reported in Panel (a). The simulations predict important movements towards two states of the Center-west, Mato Grosso do Sul

Figure 7: Simulation of changes in migration rates



Source: Own calculations.

(50) and Mato Grosso (51), the neighboring Northern state Rondônia (11) and Acre (12). We confirm the decrease in net immigration to the states of the South-East that we have stated already in Figure 5. Hence, trade liberalization favors the “penetration to the interior” of Brazilian territory, the intended strategy of the country since its Independence.

Patterns are very similar in the case of intranational migration (7, panel (b)), but magnitudes are stronger, exceeding changes of 100 % for several states. In particular, Mato Grosso (50) and Mato Grosso do Sul (51) in the Center-west and the neighboring state Minas Gerais (31) see their net inflows increase substantially. The Distrito Federal (53) and Rio Grande do Sul (43), experience the most important losses. Interestingly, for Southern state Paraná (41), we observe a sizeable reversal of the net migration flow between intranational and international trade liberalization. This suggests that this small state could benefit from its central position within Brazil, next to São Paulo and its equidistant proximity to the two rich Southern states and to the emerging Center-West. In the case of a better integration of the Brazilian states, its location close to the economic centers of Brazil becomes even more advantageous.

Also the state of Amazonas exhibit opposite outcomes in net migration after the simulation of international versus internal trade liberalization, also if to a much lesser extend then Paraná. This can be explained by the fact that, on the one hand, Amazonas holds an Export Processing Zone which allows preferential access to foreign countries. Also Amazonas is engaged in a very active trade with the neighboring countries Columbia and Venezuela, as stated by Fally et al. (2009). On the other hand, its relatively high access to foreign markets stands in opposition to a very weak national market access (Fally et al, 2009). Under the international border effect reduction, Amazonas neighbors in the North (Rondonia, Acre, Roraima) also benefit. After the reduction of the intranational border effect, Amazonas exhibit a negative net migration, and also the net gains of its neighbors are slightly lower.

When looking at net migration into each state for particular sectors, we can see that all of them reduce their net inflows to São Paulo (35). There is however not one state that attracts positive net inflows all sectors. The distribution of net migrants for each state varies across sectors, suggesting that different regions concentrate in different sectors.

8 Conclusion

This study analyzes the impact of trade on internal migration decisions in a New Economic Geography framework by looking at the relationship between market access and bilateral migration rates. We regroup migrants into 16 sectors and two educational levels and show how migration patterns and reaction to changes in market access vary across different types of workers.

We see that access to markets plays an important role in the migration decision beyond the effect that market access has on wages which are themselves important determinants of migration flows. States with low market access push residents to migrate to states with higher market access, where higher labor demand offers more jobs and higher wages. Low educated workers seem to be more sensible to changes in market access than highly educated workers.

For manufacturing workers, we see that sector-state specific market access plays a more important role in their migration decision than the state's total market access and that sector-specific market access is significant for all high and low educated migrants. Market access is significant for nearly all manufacturing sectors. Our robustness checks, including GMM and count data methods confirm a positive and significant impact of the destination's market access on bilateral migration rates.

The fact that migration patterns are apparently also driven by industrial specialization suggests that implications of NEG theory (i.e. regional advantages generated by the region's position in the spatial economy) are better understood in combina-

tion with comparative advantage and sector-specific inputs (e.g. human capital specificity).

The coefficients obtained in this exercise for market access and the other migration determinants have been used in simulations for gaining some insight on possible consequences of a international or intranational trade liberalization on migration patterns within Brazil. We find that a deepening of the integration process within Brazil leads to a slight decrease in the absolute number of migrants and has a very heterogeneous impact across sectors and across states. Changes in trade pattern are thus likely to have a significant impact on the redirection of migration flows.

References

- Aguayo-Tellez, E., M.-A. Muendler, and J. P. Poole (2008). Globalization and formal sector migration in Brazil. WIDER Research Paper 22, UNU-WIDER.
- Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *RAND Journal of Economics* 25(2), 242–262.
- Breinlich, H. (2006). The spatial income structure in the european union - what role for economic geography? *Journal of Economic Geography* 6:5, 593–617.
- Buch, C. M., J. Kleinert, and F. Toubal (2006). Where enterprises lead, people follow? Links between migration and FDI in Germany. *European Economic Review* 50(8), 2017–2036.
- Card, D. (2001). Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. *Journal of Labor Economics* 19(1), 22–64.
- Combes, P.-P., G. Duranton, and H. Overman (2005). Agglomeration and the adjustment of the spatial economy. *Papers in Regional Science* 84(3), 311–349.
- Crozet, M. (2004). Do migrants follow market potentials? An estimation of a New Economic Geography model. *Journal of Economic Geography* 4(4), 439–458.
- Dahl, G. (2002). Mobility and the return to education: Testing a Roy model with multiple markets. *Econometrica* 70(6), 2367–2420.
- De Vreyer, P., F. Gubert, and F. Roubaud (2009). Migration, Self-selection and Returns to Education in the WAEMU. *Journal of African Economies Forthcoming*.
- Falaris, E. M. (1987). A nested logit migration model with selectivity. *International Economic Review* 28(2), 429–43.
- Fally, T., R. Paillacar, and C. Terra (2008). Economic geography and wages in Brazil: Evidence from micro-data. Technical report. Available at <http://team.univ-paris1.fr/teamerso/paillacar/>.
- Fally, T., R. Paillacar, and C. Terra (2009). Economic geography and wages in Brazil: Evidence from micro-data. *Journal of Development Economics*. Forthcoming. Available at <http://team.univ-paris1.fr/teamerso/paillacar/>.
- Fiess, N. and D. Verner (2003). Migration and human capital in Brazil during the 1990s. World Bank Policy Research Working Paper 3093, World Bank.

- Gould, D. M. (1994). Immigrant links to the home country: Empirical implications for U.S. bilateral trade flows. *The Review of Economics and Statistics* 76(2), 302–16.
- Greenwood, M. (1997). Internal migration in developed countries. In M. Rosenzweig and O. Stark (Eds.), *Handbook of Population and Family Economics*, Volume 1, Chapter 12, pp. 647–720. Amsterdam: Elsevier.
- Grogger, J. and G. H. Hanson (2008). Income maximization and the selection and sorting of international migrants. NBER Working Paper 13821.
- Hanson, G. (1998). Regional adjustment to trade liberalization. *Regional Science and Urban Economics* 28(4), 419–444.
- Harris, J. and M. Todaro (1970). Migration, unemployment and development: A two-sector analysis. *American Economic Review* 60(1), 126–142.
- Hausman, J. and D. McFadden (1984). Specification tests for the multinomial logit model. *Econometrica* 52(5), 1219–1240.
- Head, K. and T. Mayer (2004). The Empirics of Agglomeration and Trade. In V. Henderson and J. F. Thisse (Eds.), *Handbook of Regional and Urban Economics*, Volume 4, pp. 2609–2669. Amsterdam: Elsevier.
- Head, K. and T. Mayer (2006). Regional wage and employment responses to market potential in the EU. *Regional Science and Urban Economics* 36(5), 573–594.
- Head, K. and J. Ries (1998). Immigration and trade creation: Econometric evidence from Canada. *Canadian Journal of Economics* 31(1), 47–62.
- Helpman, E. (1998). the size of regions. In S. E. Z. I. Pines, D. (Ed.), *Topics in Public Economics*, pp. 33–54. Cambridge University Press, Cambridge.
- Helpman, E., M. Melitz, and Y. Rubinstein (2008). Estimating Trade Flows: Trading Partners and Trading Volumes. *Quarterly Journal of Economics* 123(2), 441–487.
- Hunt, G. L. and R. E. Mueller (2004). North american migration: Returns to skill, border effects, and mobility costs. *The Review of Economics and Statistics* 86(4), 988–1007.
- Kancs, D. (2005). Can we use NEG models to predict migration flows? An example of CEE accession countries. *Migration Letters* 2(1), 32–63.
- Kovak, B. (2008). Regional labor market effects of trade policy: Evidence from Brazilian liberalization. Mimeo.

- Krugman, P. (1991). Increasing returns and Economic Geography. *Journal of Political Economy* 99(3), 483–499.
- 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.
- Lewis, W. A. (1954). Economic development with unlimited supplies of labour. *The Manchester School of Economic and Social Studies* 22, 139–191.
- Martin, W. and C. S. Pham (2008). Estimating the Gravity Equation when Zero Trade Flows are Frequent. MPRA Paper 9453, University Library of Munich, Germany.
- Martinez-Zarzoso, I., F. Nowak-Lehmann, and S. Vollmer (2007). The Log of Gravity Revisited. Center for European, Governance and Economic Development Research (cege) Discussion Papers 64, University of Goettingen (Germany).
- McFadden, D. (1974). *Frontiers of Econometrics*, Chapter Conditional Logit Analysis of Qualitative Choice Behavior, pp. 105–142. Academic Press.
- Menezes-Filho, N. A. and M.-A. Muendler (2007). Labor reallocation in response to trade reform. Technical report.
- Paillacar, R. (2006). Market potential and worker heterogeneity as determinants of Brazilian wages. Working paper, University of Paris 1. Available at <http://team.univ-paris1.fr/teamerso/paillacar/>.
- Paillacar, R. (2009). An empirical study of the world economic geography of manufacturing industries (1980-2003). Mimeo. university of paris 1.
- Pons, J., E. Paluzie, J. Silvestre, and D. Tirado (2007). Testing the New Economic Geography: Migrations and industrial agglomerations in Spain. *Journal of Regional Science* 47(2), 289–313.
- Rauch, J. E. and V. Trindade (2002). Ethnic Chinese networks in international trade. *The Review of Economics and Statistics* 84(1), 116–130.
- Redding, S. and P. K. Schott (2003). Distance, skill deepening and development: Will peripheral countries ever get rich? *Journal of Development Economics* 72(2), 515–541.
- Redding, S. and A. Venables (2004). Economic geography and international inequality. *Journal of International Economics* 62(1), 53–82.
- Roodman, D. (2009). How to Do xtabond2: An Introduction to Difference and System GMM in Stata. *Stata Journal* 9(1), 86–136.

- Schwartz, A. (1973). Interpreting the effect of distance on migration. *Journal of Political Economy* 81(5), 1153–69.
- Silva, J. M. C. S. and S. Tenreyro (2006). The Log of Gravity. *Review of Economics and Statistics* 88(4), 641–658.
- Silva, J. M. C. S. and S. Tenreyro (2009). Comments on "The Log of Gravity Revisited". Manuscript, London School of Economics.
- Sorensen, T., P. Fishback, S. Allen, and S. Kantor (2007). Migration creation, diversion, and retention: New Deal grants and migration (1935-1940). IZA Discussion Papers 3060, Institute for the Study of Labor (IZA).
- Vasconcelos, J. and M. Oliveira (2006). Analise da Matriz por Atividade Economica do Comercio Interestadual no Brasil. Working paper 1159, IPEA. Brazil.
- Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics* 126(1), 25–51.

Appendix A - Additional tables and figures

Figure A-1: Map of Brazil



Source: <http://volunteerbrazil.com/photos%202008/map-of-brazil.gif>

Table A-1: Manufacturing industries

industry code	industry
15	Food and beverages & Tobacco products
17	Textiles
18	Wearing Apparel
19	Leather, bags and footwear
20	Wood
21	Paper and paper products, publishing and printing
23	Coke, petroleum and nuclear fuel
24	Chemicals
25	Rubber and plastic products
26	Other non-metallic mineral products
27	Basic metals
28	Fabricated metal products
29	Machinery and equipment
31	Electrical machinery
34	Motor vehicles, trailers & other transport equipment
36	Furniture

Table A-2: Correlation table

	m_{ij}/s_i	MA_j	MA_i	w_j (cor)	w_i (cor)	pop_j	pop_i	$gdp pc_j$	$gdp pc_i$
$m_{ij}/stayers_i$	1.0000								
MA_j	0.0885	1.0000							
MA_i	-0.3167	-0.0266	1.0000						
w_j (cor)	0.0119	-0.1604	-0.0184	1.0000					
w_i (cor)	-0.1423	-0.0060	-0.0370	0.4666	1.0000				
pop_j	0.1993	0.7499	-0.0532	-0.0147	0.0059	1.0000			
pop_i	-0.4046	-0.0204	0.7224	0.0765	0.1478	-0.0518	1.0000		
$gdp pc_j$	0.1837	0.0198	-0.0301	0.5499	0.0995	0.2682	-0.0207	1.0000	
$gdp pc_i$	-0.1856	-0.0157	0.0487	0.4622	0.7940	0.0209	0.2901	0.2106	1.0000

All variables are in logs. Dependent variables MA_j to $gdp pc_i$ are lagged one year.

Table A-3: Aggregated versus sectoral market access

	(1)	(2)	(3)	(4)
	OLS	beta	OLS	beta
$Inter MA_j^s$	0.495 ^a (0.172)	0.310 ^a		
$Inter MA_i^s$	-0.726 ^a (0.179)	-0.453 ^a		
$Inter MA_j$ (agg)			0.491 ^a (0.059)	0.116 ^a
$Inter MA_i$ (agg)			-1.257 ^a (0.062)	-0.279 ^a
$dist_{ij}$	-0.500 ^a (0.030)	-0.217 ^a	-0.485 ^a (0.028)	-0.210 ^a
Observations	3201	3201	3201	3201
R^2	0.510	0.510	0.576	0.576

^a, ^b and ^c represent respectively statistical significance at the 1%, 5% and 10% levels. All columns contain industry fixed effects. Standard errors in parentheses, clustered at the origin-destination level. Explanatory variables are in logs and lagged one year.

Table A-4: Coefficients for simulation

	(1)	(2)
	$m_{ij}^s/stayers_i$	m_{ij}^s/pop_i
MA_j	0.166 ^a (0.042)	0.165 ^a (0.040)
MA_i	-0.242 ^a (0.045)	-0.232 ^a (0.042)
w_j	0.639 ^a (0.171)	0.654 ^a (0.172)
w_i	-0.357 ^a (0.149)	-0.305 ^b (0.132)
Observations	3201	3201
R^2	0.803	0.0805

^a, ^b and ^c represent respectively statistical significance at the 1%, 5% and 10% levels. Both columns contain industry fixed effects. Bootstrapped standard errors in parentheses (1000 replications), clustered at the origin-destination level. Explanatory variables are in logs and lagged one year.

Appendix B - Data sources for market access calculations

In order to estimate sector-specific market access we need three sets of trade data: (1) trade data between Brazilian states, which is drawn from Vasconcelos and Oliveira (2006) who processed the information of the value-added tax provided by the National Council of Financial Policy (CONFAZ, Conselho Nacional de Política Fazendária) from the Ministry of Finance (Ministerio da Fazenda); (2) trade data between Brazilian states and foreign countries (Secretaria de Comércio Exterior, Ministry of Trade); and (3) between foreign economies (BACI: Base pour l'Analyse du Commerce International, CEPII). Moreover, using total sales across regions and industry from the PIA database allows computing internal flows within regions by subtracting intra and international exports. These sets of data provide a complete and consistent picture of all trade flows, defined at the 2 digit ISIC Revision 3 level (corresponding to the Brazilian CNAE 2 digit industry classification).

Trade flows for constructing market access for the panel data set, rely on yearly Brazilian trade data for the years 1991 to 2002 coming from the Brazilian Foreign Trade Secretariat (SECEX, *Secretaria de Comércio Exterior*).

We also need a variety of data on geography, infrastructures and regulations. Distances, colonial links, languages, coordinates, GDP, areas and demographic densities are provided by CEPII (Centre d'Etudes Prospectives et d'Informations Internationales) and IBGE. The distance between states in the geodesic distance between their respective capitals (computed in km using the coordinates).