How does market access shapes internal migration? The Brazilian case*

Laura Hering † Rodrigo Paillacar‡

October 8, 2008

Abstract

This paper studies the effect of economic geography and in particular of market access on the migration pattern inside Brazil using Brazilian census data for the year 2000. After having controlled for self-selected migration, we estimate the impact of market access and of regional returns to education on bilateral migration flows between the 27 Brazilian states. Our results show that regions with low market access push residents to migrate to regions with higher market access, where higher labor demand leads to more jobs and higher wages. We further show that migrants react in particular to the market access of their industry. This increases the specialization of regions in different industries and leads to the persistence of spatial income inequalities within the country. Our results highlight that a further integration of Brazil into the world economy could reinforce the uneven spatial distribution of production and income within the country.

Keywords: New Economic Geography, Migration, Wage, Selection Bias, Spatial Adjustment, Brazil.

JEL classification: F12, F16, R12, R23.

* We thank Maria Cristina Terra for access to the Population Census data and Christopher Timmins for data on amenities. Further, we thank Pierre-Philippe Combes, Matthieu Crozet, José De-Sousa, Thierry Mayer, Vincent Rebeyrol, Thierry Verdier and participants of the RIEF 2008 doctoral meeting (Barcelona) and the ETSG 2007 (Athens) for excellent suggestions and fruitful discussions.

†CREST and Centre d'Economie de la Sorbonne, Université de Paris 1, CNRS. 106-112 Boulevard de l'Hôpital 75647 Paris Cedex 13, France. E-mail: laura.hering@gmail.com

‡Centre d'Economie de la Sorbonne, Université de Paris 1, CNRS. 106-112 Boulevard de l'Hôpital 75647 Paris Cedex 13, France. E-mail: rodrigo.paillacar@gmail.com or rodrigo.paillacar@univ-paris1.fr

1 Introduction

In economic literature we find increasing empirical evidence that trade liberalization can have a negative impact on income inequality.[1] Studies based on New Economic Geography (NEG) theory have underlined the positive relation between trade and income (Redding and Venables, 2004; Mion, 2004; Hanson, 2005; Head and Mayer, 2006) and can explain growing spatial inequality also within countries due to trade liberalization, as has been shown on the example of Brazil by Paillacar (2006).[2]

One classic adjustment mechanism, economists would expect to come into effect in this case to offset spatial inequality within a country is internal migration. As soon as income levels diverge substantially, people should migrate from low to high income regions until income levels are equalized across all regions. However, the prevailing of spatial inequalities suggests that internal migration isn't responding in the expected way.

Few research has actually been conducted so far on how internal migration flows react to spatial income inequalities created by trade and trade liberalization. One theoretical model, explaining how workers will choose their locations following changes in trade is proposed by NEG theory.

This theory underlines the importance of a region's proximity to consumers in the determination of the local wage rate and its size in terms of workers. This proximity to consumers is modeled by the region's "market access", which is defined as the distance-weighted sum of the market capacity of surrounding locations (Fujita et al., 1999). Regions closer to consumer markets (i.e. with a higher "market access") experience lower transport costs and enjoy therefore higher income levels (Fujita et al., 1999).

But NEG theory predicts also that wages increase less in reaction to an increase in the demand for local production following trade liberalization, when immigration is high because new workers can ensure the additional demand and wages. In regions, where immigration is low, additional demand can only be partially satisfied, prices and wages would go up to compensate.

According to this model, workers migrate to regions that are closest to the big markets, since regions with high market access can afford to pay higher wages. Workers from low market access regions are thus expected to migrate to regions with higher market access, and by this offset the higher wages.

The first to analyze the impact of market access on migration flows in a NEG framework has been Crozet (2004). Following the model of Tabuchi and Thisse (2002), he derives a quasi-structural migration equation from a NEG model to test how cost advantages due to higher market access shape international migration flows within the European Union.

A potential drawback in the NEG models is that labor migration is considered to continue until real wages are equal across locations.

However, the persistence of spatial inequalities in many countries shows clearly that migration cannot compensate the whole spatial income gaps. In Brazil, where

[1] See Goldberg and Pavcnik (2007) for a recent revision.

[2] Hering and Poncet (2008) find similar evidence for China.

about 40% of all residents have migrated at least once in their lives (Fiess and Verner, 2003), spatial income inequality has stagnated over the last years.

This phenomenon is explained in the migration literature, which shows that migration can under certain circumstances even reinforce spatial inequalities. In migration models, individuals choose to migrate, when they can realize gains in expected real wages (e.g. Todaro, 1969; Harris and Todaro, 1970), amenities (e.g. Treyz et al., 1993) or expected returns to skills (e.g. Borjas et al., 1992), which are key variables for the migration decision. But a well-known problem in comparing expected wages or returns to education when studying migration is the fact that we are facing a selection bias induced by certain unobservable individual characteristics that allow migrants to better match their skills with the location of destination. [3]

This is then reflected by the observation that immigrants in a specific region often share common characteristics like gender, educational level or other. 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 and spatial wage inequality will thus persist.

A related issue is the possibility that certain variables will influence the location decision differently depending on individual characteristics. For example, Levy and Wadycki (1974) have shown that educated individuals tend to value amenities much more than low qualified individuals and Schwartz (1973) argues that the negative impact of distance on migration flows diminishes with educational attainment. These differences can also lead to spatial income inequalities.

A further reason why spatial inequalities prevail, specific to the Brazilian case, is that the high migration rate stands in sharp contrast to the low sector relocation in the country found by Menezes-Filho and Muendler (2007).[4] This suggests that spatial labor adjustment takes place mainly within industries. People are more likely to migrate to regions with high labor demand in their respective sector. Given that different industries are clustered in different regions, income heterogeneity across sectors can translate into spatial wage inequality.

Using individual data of the 2000 population census, we want to assess the impact of market access on the migration patterns within Brazil while following the migration literature and control for self-selection of migrants. This helps to better understand determinants of internal migration and why spatial inequality still persists in Brazil. By analyzing the role of the location of consumer markets in the migration decision, we can derive predictions for internal migration flows between the 27 Brazilian states in case of further trade liberalization (see the simulations in Section 5.1.3). To our knowledge, we are the first to look at internal spatial labor adjustment in reaction to trade liberalization.

[3] Numerous studies (e.g. Dahl, 2002; Borjas et al., 1992) show that when workers chose where to live and work based on their comparative advantage (Roy model), then the estimated returns to education in any given region could be biased upward or downward. This combination of selection and sorting effects has received recent attention in the empirical studies on international migration. See for example, Grogger and Hanson (2008) and Belot and Hatton (2008).

[4] Also Pavcnik et al. (2004) found that industry premiums play an important role in wage determination.

One main contribution of this study is that it explores implications derived from two economic frameworks, which are rarely explored together at the empirical level: NEG theory, which is especially well suited to analyze spatial inequalities, and the labor economics tradition, which stresses the role of individual characteristics as determinants of spatial inequality and a possible selection bias in individual migration decisions.

Our empirical strategy is divided into two parts. First, we control for self-selection in the individual migration decision and second, we derive a migration equation, which assess the role of the different determinants of bilateral migration flows between the 27 Brazilian states.

The chosen methodology for the control of self-selection follows Dahl (2002), who develops a Roy model of mobility and earnings where workers choose in which of the 50 US states to live and work. He proposes a semiparametric correction of the selection bias that is easy to implement. Like Dahl (2002), we estimate a Roy model to obtain returns to education that are corrected for self-selection. These corrected returns will later be included next to market access as determinants for bilateral migration flows in the migration equation.

Thanks to the availability of Brazilian trade data at the sectoral level, we can address the fact that migration takes place within industries by calculating a market access also at the sectoral level. The census data allows also to obtain bilateral migration flows between Brazilian states for each industry so that we can estimate the impact of market access for each sector separately.

As expected, we find that migration flows depend negatively on market access of the home region and positively on the one of the destination. An increase in demand addressed to a region is adjusted by the arrival of additional workers. Nonetheless, migration costs and amenities play a significant role in the migration decision, too.

We see that the sector-region specific market access plays a much more important role than the state's total market access. 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 combination with comparative advantage and sector-specific inputs (e.g. human capital specificity). Since the location choice depends strongly on the sector in which the individual is working, spatial wage inequality can persist given that regions specialize in different industries and workers are not sufficiently mobile between these industries to induce a factor prize equalization. This evolution can be even amplified through trade liberalization if its benefits are not allocated homogeneously to all regions.

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 returns to education; in Section 5 empirical results of the migration equation and simulations are reported and Section 6 concludes.

2 New Economic Geography theory: derivation of market access

We consider a monopolistic competition framework with product differentiation, including firm-level increasing returns to scale and trade costs.

2.1 Demand side

The agricultural sector produces a homogeneous agricultural good, under constant returns and perfect competition. The manufacturing sector produces a large variety of differentiated goods, under increasing returns and imperfect competition. All consumers of region j share the same Cobb-Douglas preferences for the consumption of both types of goods (A and M):

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

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:

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

where q_{ij} represents demand by consumers in region j for a variety produced in region i and σ is the elasticity of substitution. Given the expenditure 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}:

$$q_{ij} = \mu \; p_{ij}^{-\sigma} \; G_j^{\sigma-1} \; E_j, \tag{3}$$

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

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

2.2 Supply side

Transporting manufactured products from one region to another is costly. The iceberg transport technology assumes that p_{ij} is proportional to the mill price p_i and shipping costs T_{ij}, so that for every unit of good shipped abroad, only a fraction $(\frac{1}{T_{ij}})$ arrives. Thus, the demand for a variety produced in i and sold in j eq. (3) can be written as:

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

To determine the total sales, q_i, of a representative firm in region i we sum sales across regions, given that total shipments to one region are T_{ij} times quantities consumed:

$$q_i = \mu \sum_{j=1}^{R} (p_i T_{ij})^{-\sigma} G_j^{\sigma-1} E_j T_{ij} = \mu p_i^{-\sigma} MA_i, \tag{6}$$

where

$$MA_i = \sum_{j=1}^{R} T_{ij}^{1-\sigma} G_j^{\sigma-1} E_j, \tag{7}$$

represents the market access of each exporting region i (Fujita et al., 1999).

2.3 Market access and spatial adjustment

Since NEG models assume that in the short run, profits are entirely transferred to the production factors, a higher demand, that is not accompanied by an increase in the production, will lead to spatial price disparities (which Head and Mayer (2006) call *price version*).

In the long run, this differential is eliminated by a perfect spatial adjustment of firms and workers resulting in (urban) agglomerations (the *quantity version*). NEG theory attempted originally to explain a "quantity effect", that is, the uneven distribution of the economic activity. Consequently, Krugman (1991) proposes perfect migration as the mechanism allowing to put all the effect in the agglomeration.

In general equilibrium, migration can (at least partially) eliminate the effects of market access on wages. Suppose that trade liberalization (a fall in ϕ_{ij}) affects unequally the regions inside a country. This will generate differences in market access across regions. Restoring the equilibrium demands a spatial equalization of profits. Head and Mayer (2006) explore this question by using the two extreme cases of no migration at all and completely free migration: in the first case, no migration is possible, allowing hence an increase in wages in higher market access regions due to higher product prices. In the other extreme, with workers migrating to high market access regions, factor price equalization holds. As a consequence, the number of firms in the region will increase in response to the decline of trade costs. This agglomeration of firms will rise the price index $P_j^{1-\sigma}$, which in turn will lower the MA in that region.[5] Head and Mayer (2006) consider the employment level as an indicator of this quantity version[6] and exploit regional variations of wages and employment levels in time to test which version potentially prevails. A problem with this approach is that it prevents us from integrating the migration, and the reasons behind an imperfect response to market access differentials. Fujita et al. (1999) solve the general equilibrium by postulating a very simple migration dynamic (here reduced to the canonical case of two regions):

[5] Depending on the specific formulation of the model, the agglomeration process can be catastrophic, and often multiple equilibria are possible. Determining empirically the existence of these features have proven a difficult task, and we are not exploring them in this paper.

[6] The employment level depends on the number of firms in each region and on the individual skill level (which we note as Z). Regional labor force is not proportional to the number of enterprises, because regions with higher skilled workers should have a smaller number of employment per firm:

$$L_i^s = n_i l_i = n_i \sigma \beta \frac{1}{exp\,(Z)} \tag{8}$$

$$\dot{s}_s = (\omega_i - \omega_j)\, s_s(1 - s_s) \tag{9}$$

where s_s represents the share of skilled workers in region i, and ω represents real wages (that is, w_i deflated by the price index P_j). This equation has been modified in subsequent theoretical works to accommodate other components that provide more realism, and to explain why the factor price equalization would not be attained. Some of these elements are: heterogeneous preferences about amenities (Tabuchi and Thisse, 2002; Murata, 2003), migration costs (Crozet, 2004; Kim, 2006), labor frictions (Epifani and Gancia, 2005; Kim, 2006), trade costs in the agricultural sector (Picard and Zeng, 2005), and commuting costs (Murata and Thisse, 2005). Some studies have looked for empirical verifications, especially in the European context (Crozet, 2004; Pons et al., 2007; Kancs, 2005). All this studies have in common, that they introduce the intuitive concept of utility differentials as the determinant of migration. We will follow this strand of literature in the formulation of our empirical framework, adding some important improvements to the estimation.

3 Methodology

The central objective of this paper is to empirically relate the bilateral migration flows between Brazilian states to market access while controlling for spatial sorting of skills.

The first step of our analysis consists in assessing the selection bias by estimating a correction term that allows us to obtain corrected regional returns to education (Subsection 3.1). In a second step, we derive a migration equation which links market access and returns to education to bilateral migration (Subsection 3.2).

3.1 Selection bias correction

In this subsection, we present briefly the theoretical derivation of the selection bias in a general utility differential approach as used here and how we obtain corrected returns to education. Region-specific returns to education are normally obtained by estimating simple earning equations for each region, where the returns to education are the estimated coefficient of the education variable. In the following, we will show that estimating simple OLS earning equations will yield biased returns to education and propose a correction function of this bias.

The individual location choice M_{kij} in a general utility differential approach is considered as:

$$\begin{aligned} M_{kij} &= 1 \quad if\ and\ only\ if \quad V_{kij} = max\ (V_{ki1}, \ldots, V_{kiR}), \\ &= 0 \quad otherwise. \end{aligned}$$

Every individual k coming from location i maximizes his indirect utility V_{kij} across all possible destinations j. As shown by Dahl (2002), this utility can be decomposed in (1) V_{ij}, a component capturing the mean effect of pecuniary (wages) and non pecuniary (amenities) considerations, and (2) ϵ_{kij}, an idiosyncratic individual error term. Since earnings of each individual can be observed in only one region at a time, we cannot be sure that she wouldn't have earned more in another place. We assume therefore that observed earnings correspond to the individual's utility maximizing choice ($M_{kij} = 1$). Since individuals currently living in state j are not a random sample of the population we find in general

$$
\begin{aligned}
E\left[\nu_{ki} \mid w_{kj}\right] &= E\left[\nu_{ki} \mid M_{kij} = 1\right] \\
&= E\left[\nu_{ki} \mid \epsilon_{kim} - \epsilon_{kij} \leq V_{ij} - V_{im}, \forall m\right] \\
&\neq 0.
\end{aligned}
$$

where $E\left[\nu_{ki} \mid M_{kij} = 1\right]$ is the selectivity bias for individual k. In the case of a correlation between the conditional expectation and one of the independent variables of the wage equation, as education or age, OLS regressions of earning equations will lead to biased estimates. The direction and size of the bias for each individual depends on the joint distribution of ν_{ki} and the error terms from the R migration equations. In order to control for the phenomenon of self-selected migration and the intrinsic risk of biased regional returns to education due to the resulting spatial sorting of skills, a correction term has to be introduced in the earnings equations.

Traditional selection bias corrections (like the conditional logit model) are not very well suited to cases where individual migration decisions imply numerous potential destinations.[7] Dahl (2002) estimates the selection probabilities semiparametrically by grouping individuals with the same discrete characteristics together and taking cell means for the different migration paths.[8] This means that individuals with the same characteristics, namely location, gender, education and age, are assumed to be affected in the same way by the determinants of migration.

A polynomial of these migration probabilities is then used as correction functions in the earnings equation to get consistent estimates of the return to education.

3.2 Migration equation

The migration equation shows us how the migration probability for each group is affected by different variables. Here we concentrate on state-specific returns to skills, market access and amenities as key factors of the individual migration decision.

[7] The main problem consists in the fact that they impose the independence of irrelevant alternatives and joint normality distribution of the errors.

[8] The same migration probabilities can be obtained by estimating a multinomial logit model, which is what we will do in section 4.4.

By the derivation of our migration equation we follow Sorensen et al. (2007). The utility function described above, $V_{kij} = V_{ij} + e_{kij}$, can also be decomposed as

$$V_{kij} = X_{ij}\beta + \xi_{ij} + e_{kij} \tag{10}$$

The utility of migrating to region j for an individual k from origin i is determined by X_{ij}, the characteristics of the locations i and 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 distance. By introducing the error term ξ_{ij}, we assume that there are location characteristics that cannot be observed by the econometrician. Whereas $X_{ij}\beta$ and ξ_{ij} assigns the same utility level to all individuals coming from i and going to j, we still allow individuals from the same regions to chose different locations, by including a random error term that varies across both individuals and locations, e_{kij}

As described in the section before, individuals go to the location that maximizes their utility. Equation 12 leads then to:

$$Pr(V_{kij} > V_{kim}) \forall j \neq m \tag{11}$$

$$Pr(e_{kij} - e_{kim} > X_{im}\beta - X_{ij}\beta + \xi_{im} - \xi_{ij}) \forall j \neq m \tag{12}$$

McFadden (1974) shows that by integrating out over the distribution of the logistic distribution, we can obtain the following migration probabilities:

$$Pr(M_{ijk} = 1) = \frac{exp(X_{ij}\beta + \xi_{ij})}{\Sigma_{J=1}^{J} exp(X_{ij}\beta + \xi_{ij})} \tag{13}$$

Following the approach of Berry (1994), used also by Sorensen et al. (2007), we then derive our migration equation from the obtained migration probabilities.

Since the probability of an individual from i moving to j can also be interpreted as the share of individuals from i moving to j, we can write the share of migrants from i to j, s_{ij}, as

$$s_{ij} = Pr(M_{ijk} = 1) = \frac{exp(X_{ij}\beta + \xi_{ij})}{\Sigma_{J=1}^{J} exp(X_{ij}\beta + \xi_{ij})} \tag{14}$$

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

$$s_{ii} = Pr(M_{ijk} = 1) = \frac{exp(X_{ii}\beta + \xi_{ii})}{\Sigma_{J=1}^{J} exp(X_{ij}\beta + \xi_{ij})} \tag{15}$$

Dividing equation 14 by equation 15 and taking the log yields:

$$ln(\frac{s_{ij}}{s_{ii}}) = ln\left(\frac{exp(X_{ij}\beta + \xi_{ij})}{exp(X_{ii}\beta + \xi_{ii})}\right) = X_{ij}\beta - X_{ii}\beta + \xi_{ij} - \xi_{ii} \tag{16}$$

Replacing the vector X by our location variables of interest, we obtain our migration equation:

$$\ln \frac{s_{ij}}{s_{ii}} = \beta_1 + \beta_2 \rho_j - \beta_3 \rho_i + \beta_4 \ln MA_j - \beta_5 \ln MA_i + \beta_4 \ln dist_{ij} + \beta^{'} A_j - \beta^{'} A_i + v_{ij} \quad (17)$$

where A corresponds to a vector of the state's amenities, *dist* to the bilateral distance, proxying migration costs, and v_{ij} is the error term. [9]

In 5.1.1 we will estimate the impact of the region-sector specific market access, which allows us to explore also an specialization effect. We want to know whether an individual migrates to a region from which he knows that it is facing a high demand in his industry. If we find a positive (and stronger) impact of the sectoral MA, this would indicate that the migrant is informed ex-ante about his or her job opportunities in the destination region. The argument is reinforced if we also find a reduced importance of migration costs. If the total market access plays a less important role, this is an indicator of some human capital specificity shaping industrial specialization (the standard NEG framework indicates that everyone will migrate to regions with high market access, independent of the industry).

4 Data

In this section we will first present the data sources. In the second subsection, we present very shortly the computation of our main variables of interest: the regional market access and the region-sector specific market access (Subsection 4.3) and the estimation of consistent region-specific returns to education (Subsection 4.4).

4.1 Data sources

Our main data set is a sampling of the 2000 Brazilian population census data, covering over 22 million individuals in all of the 27 states.[10] We limit our sample to workers between 16 and 60 who declare a salary. We use these data to estimate the Roy model to obtain the corrected returns to education and also to construct bilateral migration flows, used in the estimation of the migration equation.

For the market access calculation we use Brazilian trade data from the Brazilian Foreign Trade Secretariat (SECEX, *Secretaria de Comércio Exterior*) and from Vasconcelos (2001)[11]. Data of latitudes and longitudes of the states (to calculate bilateral distances), contiguity and common language comes from the CEPII (*Centre*

[9]Note that the Equation 17 is considering components of the *nominal* wages, while the pertinent variables should be the *real* wages. Unfortunately, in Brazil information on price indexes consider only main cities, and for some of the States. In the case of the sector-specific Market Access, we will introduce Origin and Destination State Fixed Effects that may allow for a reasonable control of state-level price indexes.

[10] Table 1 displays the list of the 27 states and the codes and abbreviations used in this paper.

[11] Detailed description of these sources, advantages and limitations can be found in Paillacar (2006).

d'Etudes Prospectives et d'Informations Internationales) and the Brazilian Institute of Geography and statistics (IBGE). Data on population and GDP for the Brazilian states are from the IPEA (*Instituto de Pesquisa Econômica Aplicada*).

The control variables for amenities needed to evaluate the migration equation stem from IPEADATA. In the interest of space, in this version of the paper, results including amenities are not reported.

4.2 Pattern of migration flows

Here, we display some summary statistics about the migration patterns observed in our data set. In contrast to other studies which normally use the birth state/region as the place from where the people migrate, we consider migration rates based on the individual's state of residence in 1995 and in 2000. This means, that the migration decision has been taken within these five years. For assessing the role of the different determinants of migration, this short time span allows a much more precise analysis.

Table 2 reports migration rates by population. We see that in total 2.9 % of the workers aged between 16 and 60 in our data set have migrated between 1995 and 2000. Workers with primary education migrated less then those with secondary or tertiary education.[12]

Figure 1 displaying immigration rates for each state for each of the three educational levels, confirms this picture.[13] In most states, immigration rates for tertiary education are highest, underlining the high mobility of these workers in comparison to less educated workers. The part of each migrant group in the actual population of the respective educational group is the highest for all three groups in the states Roraima (14), the least populated state of Brazil, Brasilia (53) and Amazonas (13). We see that in Sao Paulo, in 2000 nearly 48% of workers with primary education have arrived in the state within the last 5 years, whereas for highly educated people, it were little more than 2%. The state with the less immigrants is Rio grande do Sul (43), but also Rio de Janeiro (33) had an overall low immigration rate for the period 1995 to 2000.

Figure 2 indicates that there are different location preferences across these four groups. Migrants with less education tend to move relatively more often to states in the north (11-29), whereas people with higher education prefer the south (41 - 50), which is known for its good climate and recent economic development, and to Brasilia (53), where Brazil's capital is located. These differences in the location choice signals that self-selection of migrants is probably an issue in Brazil and should be controlled for when looking at determinants of migration.

[12] We differentiate between three different educational levels: *primary*, which is between 1 and 7 years of schooling, corresponding to primary school; *secondary*, 8 to 10 years of schooling, corresponding to high school and *tertiary*, corresponding to more than 10 school years. Individuals with 0 years of education or no information on school years have been dropped.

[13] Immigration rates are calculated as migrants with educational level e over all population with educational level e in state j.

4.3 Market Access Calculation

Our main variable of interest is the state's market access, as defined in equation 7. To obtain a sound estimate of market access, we follow the methodology pioneered by Redding and Venables (2004).[14]

If we denote X_{ij} the bilateral exports from region i to region j for all firms n_i, and defining $\phi_{ij} = T_{ij}^{1-\sigma}$ as the "freeness" of trade (Baldwin et al., 2003), we can use equation (5) to show that:

$$X_{ij} = n_i p_{ij} q_{ij} = \underbrace{n_i p_i^{1-\sigma}}_{FX_i} \phi_{ij} \underbrace{\mu E_j G_j^{\sigma-1}}_{FM_j} \tag{18}$$

The region-specific variables can be captured by exporter and importer fixed effects FX_i and FM_j, respectively. Taking the logs, our estimated specification of the trade equation is

$$\ln X_{ij} = FX_i + FM_j + \delta \ln d_{ij} + \lambda_1 C_{ij} + \lambda_2 B_{ij} + \lambda_3 (B_{ij} * C_{ij}) + \lambda_4 b_{ij} + u_{ij} \tag{19}$$

where the freeness of trade ϕ_{ij} is defined based on variables that enhance or deter trade such as bilateral distance d_{ij} and contiguity C_{ij}. As we include intranational and international data in our regression, we can identify a national border $B_{ij} = 1$ if the trade flow is between a Brazilian state and a foreign country. We hypothesize that crossing a national border implies several costs and consequently we expect a negative coefficient for this dummy variable. We introduce the contiguity interacted with national border to consider the possibility of differentiated effects at national and international level.

Following Head and Mayer (2006) we include internal trade observations X_{ii} calculated as production minus total exports what allows to identify an 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.

Hence, a region's MA is composed of three parts reflecting the market access to the local level (inside the region), to the national level (inside the country) and to the international markets:

$$MA_i^{Total} = MA_i^{Local} + MA_i^{National} + MA_i^{International} \tag{20}$$

with

$$MA_i^{Local} = \exp\left(\widehat{FM}_i\right) d_{ii}^{\hat{\delta}} \quad ; \quad d_{ii} = 2/3\sqrt{area_i/\pi} \tag{21}$$

$$MA_i^{National} = \sum_{j \neq i}^{regions} \exp\left(\widehat{FM}_j\right) d_{ij}^{\hat{\delta}} \exp\left(\widehat{\lambda}_1 + \widehat{\lambda}_2\right) \tag{22}$$

[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/* .

$$MA_i^{International} = \sum_j^{countries} \exp\left(\widehat{FM}_j\right) d_{ij}^{\hat{\delta}} \exp\left(\widehat{\lambda}_1 + \widehat{\lambda}_3\right) \exp\left(\widehat{\lambda}_4 C_{ij}\right) \quad (23)$$

The coefficients of the global trade equation are available in (Fally et al., 2008, Table 2, first column).

For the calculation of a sector-region specific market access variable, we run the trade equation separately for each of the 20 manufacturing sectors into which we can classify the Brazilian and international trade flows. Data is available at ISIC Rev 3 at 2-digits.[15] Consequently, we obtain sector-specific coefficients for all the market access components (including the importer fixed effects), what allows to build a regional-sector market access.

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. International trade flows disaggregated at the state level are also available, allowing us to construct a very complete indicator of market access.

4.4 Regional returns to schooling

In this section, we will determine the selection bias in migration flows within Brazil and then estimate consistent returns to education using individual migration probabilities.

The individual migration probabilities for each of the 27 Brazilian states can be easily obtained by estimating a multinomial logit model on the individual's location choice. In order to be able to correct for individuals self-selection, it is important to include in the model specification one or more variables that explain location choice but do not influence earnings. We employ the following specification:

$$\begin{aligned} M_{kij} = \beta_j^0 + \beta_j^2 age_k + \beta_j^3 woman_k + \beta_j^4 family_k \\ + \beta_j^5 secondary_k + \beta_j^6 tertiary_k + \beta_j^7 migrant_k + FE_i + \varepsilon_{kj} \end{aligned} \quad (24)$$

where we suppose that the location choice depends on the bilateral distance between the two potential locations, as well as the age and gender of the individual. Further we control whether the individual has children or not and what is her highest educational level (primary, secondary or tertiary education).[16] We add also a

[15] Actually, the classification used 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 3 for the full list). Average values of the estimated coefficient across industries are available at Fally et al., 2008, Table 2, second column

[16] Primary is the reference group here.

dummy for being a migrant and origin and destination fixed effects.[17] The individual's migration probabilities for each state are then predicted based on the estimated coefficients.

A polynomial of these choice probabilities has then been added as a correction term to the set of explanatory variables in the earnings equation. Several specifications of the correction term have been tried. Theoretically, all but one destination probabilities could enter the correction functions. However, this leads to multicollinearity problems in the earnings equations so we retain a more restricted set of choice probabilities. We follow De Vreyer et al. (2007) and use: the first best choice probability, that is the probability to reside in the actual residence state (p_{kij}); the retention probability, that is the probability to remain in the state of origin (p_{kii}) and finally the highest predicted probability, excluding the retention probability (p_{kim}). As De Vreyer et al. (2007), we complete this set of selection correction terms by adding the interactions between these probabilities as explanatory variables, so the final correction term is given by $\lambda_{kj}\left(p_{kij}, p_{kii}, p_{kim}\right)$.

The resulting earnings equations can then be easily estimated by OLS:

$$\begin{aligned}\ln w_{kj} = \zeta_j^0 + \rho_j^2 secondary_{kj} + \rho_j^3 tertiary_{kj} + \zeta_j^1 age_{kj} + \zeta_j^2 woman_{kj} \\ + \beta_j \lambda_{kj}\left(p_{kij}, p_{kii}, p_{kim}\right) + \xi_{kj}\end{aligned} \tag{25}$$

where β_j is a vector of estimated parameters for the different migration probabilities and their interaction terms. We estimate equation 25 once for each of the 27 Brazilian states to obtain the corrected returns to education. We then obtain three new variables, ρ^t, ρ^s and ρ^p[18], which represent the comparable state-specific returns to tertiary, secondary and primary education, which we will use as regressors in the next step.[19] which we will then use in the next step as explaining variables in the migration equation.

Figures 3 and 4 in the Appendix plots the coefficients of corrected versus the non-corrected returns to education. The latter are obtained from the estimation of the 27 earnings equations (equation (25)) without the correction term. These figures illustrate well the existence of a bias in the estimated returns to education for secondary and primary education.

The difference between the two coefficients shows that we have mainly a positive self-selection for tertiary education. Dahl (2002) finds a similar selection bias for the United States and explains the upward bias in the OLS estimates by the fact that individuals with tertiary education are more likely to sort into states that provide

[17] For the interest of space, the results of the multinomial logit are not displayed here, but can be obtained on request from the authors.

[18] Returns to education for individuals with primary education are the mean base wage of the state for this category of workers.

[19] We also find a positive but weak relation between the market access and the corrected returns to schooling, as suggested by Redding and Schott (2004). These two authors develop a NEG model where a higher market access can increase the skill premium. However, in their model, workers are immobile. In a migration context, workers can partially arbitrate these spatial wage disparities for similar occupations, leaving disparities due to skill sorting and imperfections in the labor market (discrimination for example).

a better match for their particular skills and talents compared to those with less education.

With the correction of the selection bias, this work is - to our knowledge - one of the first to reconcile empirically NEG models with economic literature on labor migration.

5 Results

5.1 Market Access and migration response

In this section, we finally estimate the migration equation, equation 17. We will look at different aggregation levels of bilateral migration, differentiating between sectoral migration flows and different educational levels.

Since we are looking at bilateral flows, we include each independent variable once for the state of origin and once for the state of destination. Our exogenous variables of main interest are MA_i, total market access of the state of origin i and MA_j, total market access of the state of destination. We expect MA_i to have a negative impact on migration: the higher MA_i, the higher the demand for work and the higher the wages. Thus, the individual can attain already a high utility in his home region and is not necessarily motivated to look for a better job in another state. The same logic applies to MA_j: the higher this indicator, the more the region attracts people in search for a job or a better paid one.

Moving costs are proxied by the inverse of the bilateral distance between the origin and the destination state and should have a negative impact on the number of migrants: the farer away the destination, the more expensive the journey and the less familiar the new environment (climate, institutions, cultural specificities).

In all estimations, standard errors are clustered at the origin-destination-couple-level.

To give a first impression of the determinants of migration, column 1 of Table 4, reports results for total bilateral migration, where the dependent variable m_{ij} is defined as $migrants_{ij}/stayers_{ii}$, the number of persons that migrate from state i to j ($i \neq j$) over $stayers_i$, the total number of stayers in our sample that are originate from region i and stay in i. Besides manufacturing workers, migration flows in this regression contain also a high number of workers in service or primary sectors.

Independent variables are total market access, bilateral distances and macro-region fixed effects for the destination and origin macro-region.[20]Fiess and Verner

[20]The Brazilian states are regrouped in five macro-regions. 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. The eco-

(2003) compare the migration pattern from and to these macro-regions, showing that these differ significantly. Introducing macro-region dummies has therefore the advantage of capturing structural and cultural differences between the regions, but they also change the interpretation of our coefficients. When adding region fixed effects, we will explain differences in migration flows within macro-regions, not across macro-regions.

As we see from the reported results, market access plays indeed a significant role in the migration choice of individuals. The impact of MA_j is positive and significant, indicating that high market access states attract workers. The coefficient of MA_i is not significant here.

The highly negative and significant parameter of the distance shows that migrants prefer to go to states in the neighborhood.

In the following regressions, we differentiate migrants according to industry. Here, bilateral migration, m_{ijs} is defined as $migrants_{ijs}/stayers_{is}$, the number of persons that migrate from state i to j ($i \neq j$) over $stayers_{is}$, the number of stayers in our sample working in sector s, originating from region i.

In column 2, we keep all sectors. When looking at the migration flows at this disaggregated level, next to the positive and significant impact of MA_j, we find a negative and significant coefficient, confirming that low-MA regions are more likely to see their workers leave.

In the third columns of Table 4, we repeat the same estimation as in column 1 and 2, but we reduce the sample to individuals working in agriculture or one of the 20 manufacturing sectors that figure in the trade data used to calculate the market access to allow us better compare with the following column 4.

As we are able to obtain a region-sector specific market access and region-sector specific migration, we will exploit this additional dimension to see whether we can find a robust relationship between market access and migration on a more disaggregated level.

We observe much higher and highly significant coefficients for MA_{is} and MA_{js} than for the total market access. The impact of distance is lower here.[21]

In column 5, we replace macro-region fixed effects by state fixed effects (for the origin and the destination state). These fixed effects capture all types of state specific advantage or disadvantage of the state, as unemployment, infrastructure, amenities etc. Even though their introduction leads to a decrease in absolute values of the coefficients, the impact of sectoral market access is still very important.

These results indicate a reaction of individuals to demand in their relevant sector. The highly significant impact of MA_i^s and MA_j^s suggests that even if a region might have a high total 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 much stronger role in attracting migrants. Note that the coefficient of the bilateral distance, although significant, is strongly reduced: manufacturing workers tend more easily to

nomic opportunities afforded by living in these regions clearly explain much of their high population density.

[21] Results stay similar, when adding additional or different amenities variables (e.g. Government spending in sports or culture; road accident death rate, among others) or when excluding agricultural workers. State fixed effects in the sector-specific regressions should capture amenities.

migrate to regions that are farer away than workers in the service sector.

These findings are interesting for the 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 see here that workers do not move freely between industries. Though we observe the agglomeration effect described in the NEG (higher demand attracts new workers and leads to bigger agglomerations), this mechanism is not affecting all individuals in the same way, attracting all workers to the biggest agglomeration.

Instead, we observe agglomeration at the sectoral level. Once a region has acquired an advantage in a certain industry and its market access increases for this industry, specialization will be facilitated, because the high sectoral market access will attract corresponding workers. We thus observe an adjustment by the quantity mechanism.

These findings highlight the fact that NEG implications for wage inequality are highly sector-specific: we can expect that wage differentials in services remain higher, because migrants are limited in their spatial movements. In manufacturing, a human capital specificity allows 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 prize equalization.

In the next section we explore if skill levels are also playing a role in shaping differentiated effects among worker's mobility.

5.1.1 Market access and migration: impact of education and sectors

In this section, we disaggregate migration flows even further by taking into account the three different educational level e, primary, scondary and tertiary education. In Table 5 bilateral migration flows are defined as $migrants_{ij}^{es} = migrants_{ij}^{es}/stayers_{ii}^{es}$, the number of migrants in sector s with educational level e over the stayers of manufacturing workers in sector s with educational level e from region i.

This disaggregation allows us to finally include our calculated consistent returns to education, ρ, as additional independent variables. To show the importance of correcting for self-selection in the estimation of the returns to education, we first use in column 2 the uncorrected returns to education, ρ_i^u and ρ_j^u, and only from column 3 on the corrected returns to education, ρ_i^c and ρ_j^e.

We find that market access stays important also when controlling for returns to education and state fixed effects. The negative and significant impact of ρ_i^e indicates further that emigration takes place in states with low returns to education and the positive and significant coefficient of ρ_j^e signifies that immigration states are those with high returns to education.

In the last column, we add also dummies for the industries and the three educational categories. Here, the coefficients of the destination variables decrease significantly for market access and returns to education, turning the last one non significant. The variables of the state of origin gain in force, indicating that people are pushed out of the states where returns to their educational level and their sectoral market access is low.

5.1.2 Market access and migration: winners and losers

In this section, we want to test whether the impact of market access and returns to education on the migration flows might differ across industries. Industries enjoying a comparative advantage in the world economy are expected to have grown in the last years with the opening of the country. Trade liberalization should have increased the international demand for the products of these industries. In order to respond to the increasing demand, more workers are needed. The link between market access and migration inflows should be very strong. On the other hand, workers in industries that have been the "losers" in terms of international comparative advantage, should not respond much to international market access, the relation between market access and migration should be weaker.

In order to test for this, we regroup industries into three groups, based on a comparative-advantage indices measure, proposed by Balassa (1965). Muendler (2007) calculates this measure for Brazilian data for the period 1986 to 2001 and ranks sectors by their comparative advantage. We follow this ranking by defining our groups of "winners" and "losers" as well as the "rest" as described in Table 3.

Regressions for each of the three groups are reported in Columns 1 to 3 in Table 6. As expected, we find a very strong and significant impact of market access for the group of "winners". In comparison, for the group of losers, only the coefficient of the sectoral market access of the state of origin is significant, indicating that for these workers, low sectoral market access is also a reason to leave, but they are not attracted by big centers, where their industry is still growing, given that they are working in an industries that has difficulties to be competitive with foreign countries.

For the last group of industries, that do neither particularly good nor bad in international comparison, we still find a positive impact of the destination's market access. However, once we take out agriculture, we don't find any significant influence of market access for this group anymore. Migration is much more driven by the returns to education: both coefficients are highly significant.

These results indicate, that even though we find an overall positive impact of market access for the choice of destination and a negative influence of market access on the choice of leaving, we see that these impacts vary strongly across industries. For returns to education, we can conclude, that they stay similarly important for all type of industries.

5.1.3 The impact of national integration: Simulations

As a last exercise, we explore the impact of a change in a parameter of interest on migration outcomes. Ideally, we would like to present the impact of a reduction in internal and external trade barriers on internal migration flows. Here, we will present the consequences of a more integrated Brazil, which yield the more important and interesting results.

In the past, Brazil made great efforts to equilibrate national and federal demands of competitiveness. Examples are investment in highways, reduction of tariffs with Mercosur members, offering tax exemptions intended to raise competitiveness for some lagged regions, and of course, the famous relocation of the capital to the interior in an entirely new city (Brasilia). In our model, all these efforts are summarized in the market access of every Brazilian state. While this variable includes internal demand and international demand, we think the most interesting variable to start is the national demand itself. The high level of trade costs still existing among regions reduce the market size they can encompass, affecting competitiveness of the whole nation.

There are two interesting papers on simulation of migration flows. One is Sorensen et al. (2007), who are studying the impact of the grant programs developed by the US during the Great Depression. Another one, is Hunt and Mueller (2004) who perform an exercise closer to ours, by considering intra- and international migration between Canada and the US. They also introduce returns to skill and an international border effect. The simulation they propose consists in reducing the value of the coefficient of the border effect they found in their migration equation. The impact is measured in terms of migration rate. It goes from a baseline of 0.36% to 4.06% in the total absence of a border effect among both countries.

Instead of just mechanically reducing the whole component of trade costs among Brazilian states ($\phi_i j$), we prefer to study the impact of the specific variable that could be eventually modified by policymakers: the average border effect among regions. This reflects the reduction of trade flows only for national trade. It is important to recall that this is not a direct effect on migration, like in the case of Hunt and Mueller (2004), but an indirect effect through the economic attractiveness of Brazilian states, captured by the trade gravity equation.

We reduce the value of the estimated coefficients of the national border in the gravity equation (equation 19). These new values result in an increase in the freeness of trade among regions. While the applied reduction is on an "average" border effect, its impact on the market access does not have to be balanced among Brazilian states, because of the regional specialization (border effects differs among sectors, so a 20% reduction can be very different for one sector or another).

Finally, we generate counterfactual predictions of the odds ratios of bilateral migration. We chose the specification from Column 3 of Table 5), where migration flows are disaggregated at the sectoral and educational level. Consequently, individuals continue to evaluate each Brazilian state considering the same variables (returns to skill, market access, bilateral distance, as well as unobserved state effects), but values of sectoral market access have changed due to changes in the sector-specific internal border effect. Therefore, the prediction of the migration odd ratios is education and industry-specific. In the following discussion, we concentrate on the consequences for those sectors which have a high revealed comparative advantage as defined by the Balassa indicator in Section 5.1.2.

Table **??** shows average results for each of the three educational levels and the 8 "winning" industries.[22] The ratios, read from left to right, reflect the counterfactual values resulting from subsequent reductions of the national border effect. We see that the average impact is very low, reflecting compensating movements across regions, inside of industries.

Mean values hide much more interesting variations. When we look at specific cases of state-to-state migration for specific industries we see quite a lot of heterogeneity. If we take the difference between the baseline and the counterfactual of zero effect, we get the distributions of changes plotted in graph 5. In some cases, like manufacturing of rubber and plastic products (industry 25), most of the values are around zero, suggesting that a better integration of internal trade for this products will not entail big migration flows. By contrast, other sectors like manufacture of food products (15), leather and footwear (19), wood Products (20) or metal products (28) can have important migration responses.

Table 8 reports the potential changes in migration flows for specific pairs of Brazilian states due to the reduction of the average national border effect. For the sake of brevity, we present only the information for tertiary education. We can see that the four sectors mentioned above as having a high adjustment rate, are leading the spatial relocation.

Regarding Brazilian states, the simulations predict important movements towards Center-west (Mato Grosso do Sul 50, Mato Gross 51), the neighbour Northern states (Para 15, Rondonia 11) and to some extent to Bahia (29) also a neighbor state to the richest regions.

[22] Differences between coefficients for the same industry across educational levels come from the introduction of the variables of returns to schooling.

6 Conclusion

This paper analyzed the impact of trade liberalization on inequality in a New Economic Geography framework by looking at the relationship between market access and migration. We regroup migrants into 20 sectors and three educational levels and test to analyze determinants of migration patterns. We further look at the impact of returns to education on migration. To obtain unbiased returns to education, we control for the presence of self-selection in the spatial repartition of the individuals, following the approach of Dahl (2002).

We see that access to markets plays an important role in the migration decision: regions with low market access push residents to migrate to regions with higher market access, where higher labor demand offers more jobs and higher wages. We see that high wages are not the only reason to migrate. Individuals also value good living conditions, accepting even a smaller revenue in compensation to security or a good health system etc. This is in line with recent theoretical developments in NEG as stated by Tabuchi and Thisse (2002) and Murata (2003).

Our results provide important details about the NEG implications for wage inequality. The first is that impacts are highly sector-specific: we can expect that wage differentials in services remain higher, because migrants seem to be more limited in their spatial movements. For manufacturing workers, we see that the sector-region specific market access plays a much more important role than the state's total market access. 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 combination with comparative advantage and sector-specific inputs (e.g. human capital specificity).

The choice of migrants to go to regions with high market access in the respective sector will lead in the long run to a specialization of each state in one or few specific sectors. The industrial specialization described by classic trade theories like Hecksher-Ohlin or Ricardo will take place inside the country. Since the gains of trade liberalization in terms of market access increases are unevenly distributed across the country, a further integration of Brazil into the world economy would probably reinforce the differences in market access and by this initiate even more workers to migrate and regions to specialize.

This is also confirmed by dividing the 20 sectors into three different groups according to their comparative advantage in the world economy. We find, that workers in industries with a high international comparative advantage are much more sensible to changes in market access than the other sectors.

The coefficients obtained in this exercise for market access and the other migration determinants have been used in simulations for assessing intranational impacts inside Brazil. We find that a deepening of the integration process within Brazil, could a have significant impact on the creation and redirection of migration flows.

References

Belot, M. and T. J. Hatton (2008). Immigrant selection in the OECD. CEPR Working paper 6675.

Berry, S. T. (1994, Summer). Estimating discrete-choice models of product differentiation. *RAND Journal of Economics 25*(2), 242–262.

Borjas, G., S. Bronars, and S. Trejo (1992). Assimilation and the earnings of young internal migrants. *Review of Economics and Statistics 74*(1), 170–175.

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.

Epifani, P. and G. Gancia (2005). Trade, migration and regional unemployment. *Regional Science and Urban Economics 35*(6), 625–644.

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/teamperso/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.

Fujita, M., P. Krugman, and A. Venables (1999). *The Spatial Economy: Cities, Regions, and International Trade*. Cambridge: MIT Press.

Goldberg, P. and N. Pavcnik (2007). Distributional effects of globalization in developing countries. *Journal of Economic Literature 45*(1), 39–82. Also in NBER 12885.

Grogger, J. and G. H. Hanson (2008). Income maximization and the selection and sorting of international migrants. NBER Working Paper 13821.

Hanson, G. (2005). Market potential, increasing returns, and geographic concentration. *Journal of International Economics 67*(1), 1–24.

Harris, J. and M. Todaro (1970). Migration, unemployment and development: A two-sector analysis. *American Economic Review 60*(1), 126–142.

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.

Kancs, D. (2005). Can we use NEG models to predict migration flows? An example of CEE accession countries. *Migration Letters 2*(1), 32–63.

Kim, M. (2006). Economic geography, migration and search. Working paper.

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.

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.

Mion, G. (2004). Spatial externalities and empirical analysis: The case of Italy. *Journal of Urban Economics 56*(1), 97–118.

Murata, Y. (2003). Product diversity, taste heterogeneity, and geographic distribution of economicactivities: Market vs. non-market interactions. *Journal of Urban Economics 53*(1), 126–144.

Murata, Y. and J.-F. Thisse (2005). A simple model of economic geography à la Helpman-Tabuchi. *Journal of Urban Economics 58*(1), 137–155.

Paillacar, R. (2006). Market potential and worker heterogeneity as determinants of Brazilian wages. Working paper, University of Paris 1.

Pavcnik, N., A. Blom, P. Goldberg, and N. Schady (2004). Trade policy and industry wage structure: Evidence from Brazil. *World Bank Economic Review 18*(3), 319–344.

Picard, P. and D.-Z. Zeng (2005). Agricultural sector and industrial agglomeration. *Journal of Development Economics 77*(1), 75–106.

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.

Redding, S. and P. Schott (2004). Distance, skill deepening and development: Will peripheral countries everget 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.

Schwartz, A. (1973, Sept.-Oct). Interpreting the effect of distance on migration. *Journal of Political Economy 81*(5), 1153–69.

Sorensen, T., P. Fishback, S. Allen, and S. Kantor (2007, September). Migration creation, diversion, and retention: New deal grants and migration: 1935-1940. IZA Discussion Papers 3060, Institute for the Study of Labor (IZA).

Tabuchi, T. and J.-F. Thisse (2002). Taste heterogeneity, labour mobility and economic geography. *Journal of Development Economics 69*(1), 155–177.

Todaro, M. (1969). A model of labor migration and urban unemployment in less developed countries. *American Economic Review 59*(1), 138–148.

Treyz, G. I., D. S. Rickman, G. L. Hunt, and M. J. Greenwood (1993). The dynamics of U.S. internal migration. *The Review of Economics and Statistics 75*(2), 209–14.

Vasconcelos, J. (2001). Matriz do fluxo de comércio interestadual de bens e serviços no Brasil- 1999. Working paper 817, IPEA. Brazil.

Table 1: List of states.

Rondônia	11
Acre	12
Amazonas	13
Roraima	14
Pará	15
Amapá	16
Tocantins	17
Maranhao	21
Piau	22
Ceará	23
Rio Grande do Norte	24
Paraíba	25
Pernambuco	26
Alagoas	27
Sergipe	28
Bahia	29
Minas Gerais	31
Espírito Santo	32
Rio de Janeiro	33
Sao Paulo	35
Paran	41
Santa Catarina	42
Rio Grande do Sul	43
Mato Grosso do Sul	50
Mato Grosso	51
Goiás	52
Distrito Federal	53

Table 2: Descriptive statistics for Migrants between 1995-2000

Education	**Total**	**Migrants**	**Percentage**
Tertiary	2882548	106389	3,7%
Secondary	2352212	83567	3,5%
Primary	1.48e+07	402241	2.7%
TOTAL	2.01e+07	592197	2.9%

Figure 1: Migrants by respective population group

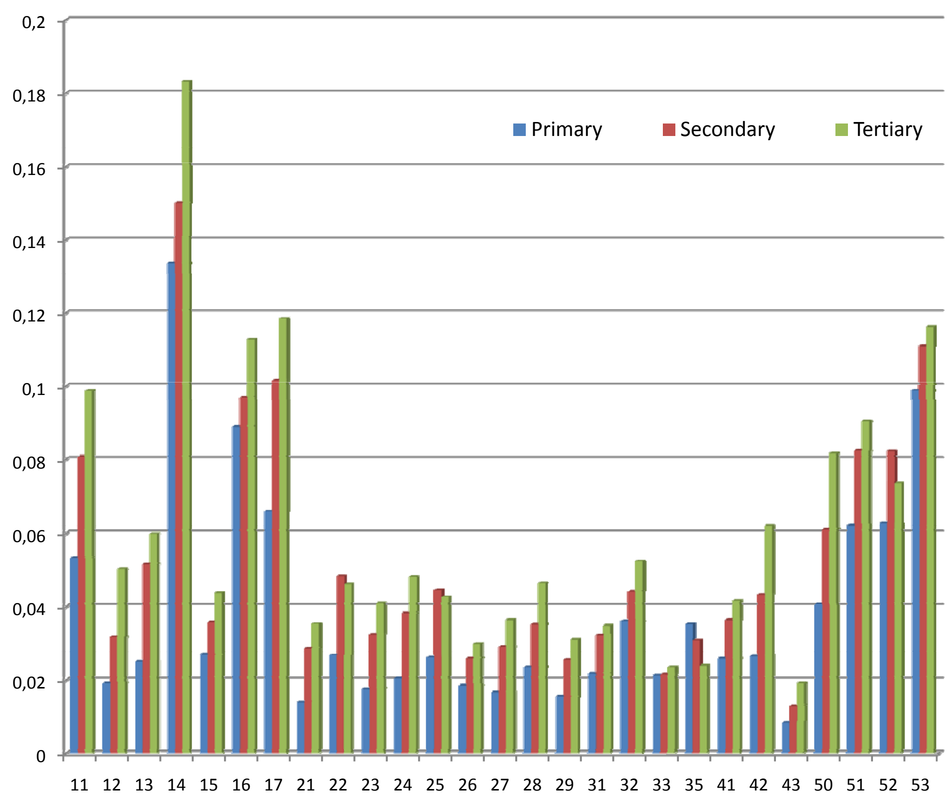

Figure 2: Proportion of migrants

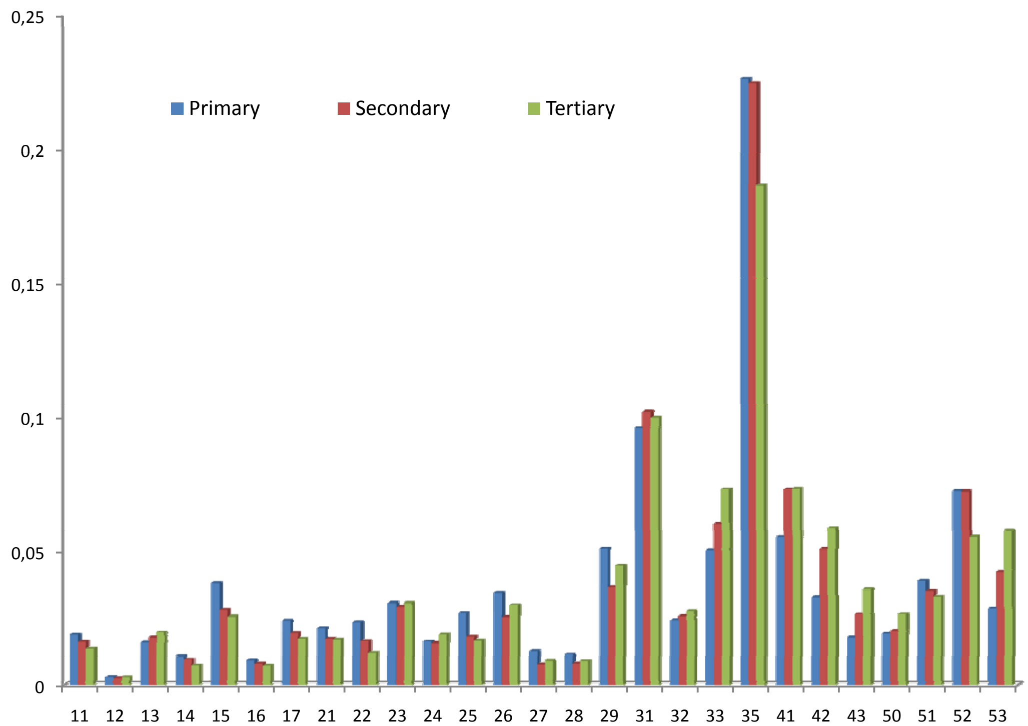

Figure 3: Returns for secondary education

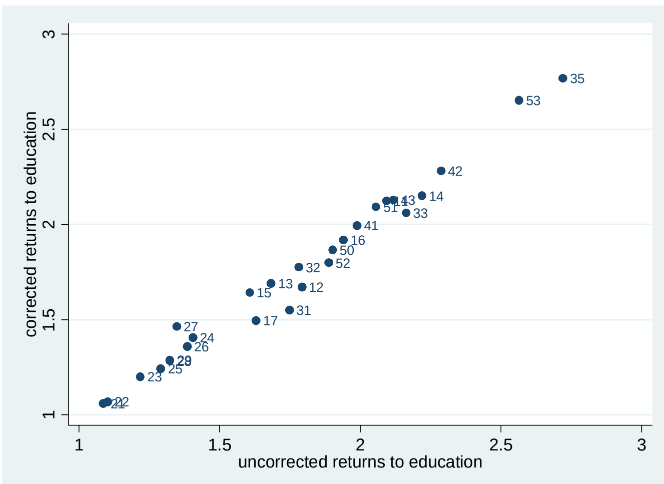

Figure 4: Returns for tertiary education

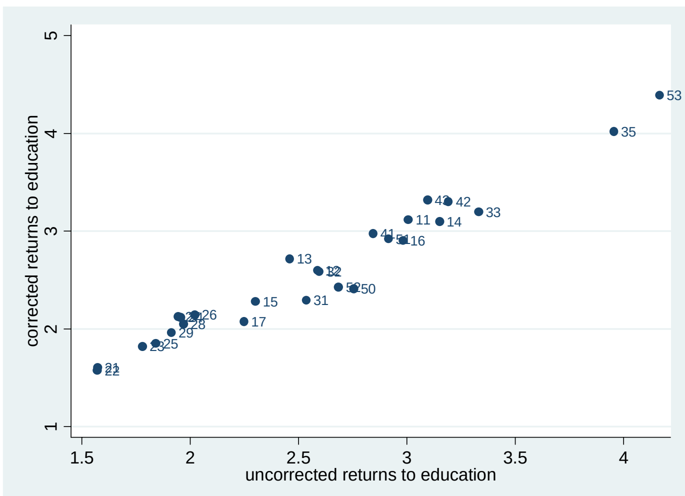

Table 3: Bilateral migration flows at the sectoral level

industry code	industry	group
01	Agriculture, Forestry & Fishing	R
10	Mining and quarrying	R
15	Manuf. Food and beverages & Tobacco products	W
17	Manuf. Textiles	R
18	Manuf. Wearing Apparel	L
19	Manuf. Leather, bags and footwear	W
20	Manuf. Wood	W
21	Manuf. Paper and paper products	W
22	Publishing, printing	L
23	Manuf. Coke, petroleum and nuclear fuel	R
24	Manuf. Chemicals	R
25	Manuf. Rubber and plastic products	W
26	Manuf. Other non-metallic mineral products	W
27	Manuf. Basic metals	W
28	Manuf. Fabricated metal products	W
29	Manuf. Machinery and equipment	R
30	Manuf. Office and computing machinery	L
31	Manuf. Electrical machinery	R
32	Radio, television and communication equipment	L
33	Manuf. Medical, precision and optical instruments	R
34	Manuf. Motor vehicles, trailers & other transport equipment	R
36	Manuf. of furniture	R

W=Winners, L=Losers, R=Rest

Classified by Balassa (1965)comparative advantage indices measure following Muendler (2007)

Table 4: Bilateral migration flows at the sectoral level

	m_{ij}	m_{ij}^s	m_{ij}^s (w.o. services)			
	(1)	(2)	(3)	(4)	(5)	(6)
$ln\ dist_{ij}$	-1.260***	-0.815***	-0.906***	-0.796***	-0.859***	-0.736***
	(0.13)	(0.059)	(0.073)	(0.070)	(0.052)	(0.049)
$ln\ MA_i$	0.0505	-0.233***	-0.249***			
	(0.075)	(0.075)	(0.088)			
$ln\ MA_j$	0.567**	0.447***	0.544***			
	(0.23)	(0.076)	(0.094)			
$ln\ MA_i^s$				-0.502***	-0.212***	-0.333***
				(0.063)	(0.072)	(0.070)
$ln\ MA_j^s$				0.770***	0.486***	0.372***
				(0.064)	(0.071)	(0.074)
Macroregion FE	yes	yes	yes	yes		
State FE					yes	yes
Sector FE						yes
Constant	-4.635	-3.599***	-3.212**	-7.355***	-4.669***	-2.117
	(2.84)	(1.24)	(1.45)	(0.78)	(0.63)	(2.03)
Observations	702	14914	4994	4994	4994	4994
R^2	0.49	0.38	0.35	0.39	0.52	0.65

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 5: Bilateral migration flows at the sectoral and educational level

	m_{ij}^{se}			
	(1)	(2)	(3)	(4)
$ln\ dist_{ij}$	-0.649***	-0.650***	-0.651***	-0.508***
	(0.050)	(0.050)	(0.050)	(0.043)
$ln\ MA_i^s$	-0.232***	-0.235***	-0.235***	-0.432***
	(0.075)	(0.075)	(0.075)	(0.067)
$ln\ MA_j^s$	0.523***	0.519***	0.519***	0.151**
	(0.077)	(0.077)	(0.077)	(0.074)
ρ_i^u		-0.156		
		(0.096)		
ρ_j^u		0.489***		
		(0.096)		
ρ_i^c			-0.151*	-0.620***
			(0.087)	(0.081)
ρ_j^c			0.467***	-0.126
			(0.082)	(0.079)
State FE	yes	yes	yes	yes
Sector FE				yes
Education FE				yes
Constant	-6.092***	-6.603***	-6.555***	5.773***
	(0.64)	(0.64)	(0.65)	(1.90)
Observations	9432	9432	9432	9432
R^2	0.45	0.47	0.47	0.68

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 6: Regressions by sectors

	winners	losers	rest	rest (w.o agri)
	(1)	(2)	(3)	(4)
$ln\ dist_{ij}$	-0.523***	-0.699***	-0.745***	-0.354***
	(0.054)	(0.063)	(0.054)	(0.048)
$ln\ MA_i^s$	-0.722***	-0.514***	-0.107	-0.106
	(0.11)	(0.11)	(0.12)	(0.090)
$ln\ MA_j^s$	0.744***	-0.0702	0.444***	0.119
	(0.11)	(0.091)	(0.12)	(0.090)
ρ_i	-0.0380	0.0188	-0.431***	-0.229**
	(0.091)	(0.12)	(0.11)	(0.096)
ρ_j	0.304***	0.220**	0.820***	0.346***
	(0.084)	(0.11)	(0.12)	(0.097)
State FE	yes	yes	yes	yes
Constant	-1.631*	12.21***	-7.327***	-1.682***
	(0.94)	(2.84)	(0.68)	(0.59)
Observations	3900	827	4052	2468
R^2	0.50	0.68	0.47	0.60

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 7: Simulation of a decrease in the internal border effect

Industry	Education	100%(Baseline)	80%	60%	40%	20%	0%
15	primary	.0435163	.0432126	.0427088	.0420471	.0414945	.0415595
19	primary	.0232151	.0229278	.0224427	.0219053	.0217913	.0227861
20	primary	.0215867	.0214301	.0211836	.0209091	.0208072	.021205
21	primary	.0585228	.0582888	.0578763	.0571999	.0562009	.0549117
25	primary	.1563752	.1560721	.1552986	.1534176	.1492997	.1418178
26	primary	.0373711	.0373062	.0371772	.0369383	.0365496	.0360427
27	primary	.0729977	.0723645	.0713291	.0697798	.0677178	.0653447
28	primary	.0695924	.069345	.0687473	.0675179	.065606	.0637523
15	secondary	.0540684	.0537943	.0533282	.0526912	.0521235	.0521454
19	secondary	.0284316	.0279971	.0272408	.02633	.0258836	.0267719
20	secondary	.0325696	.0322793	.0317761	.0310843	.0304509	.0303615
21	secondary	.092464	.0919962	.0911685	.0897987	.0877357	.0849762
25	secondary	.1784977	.1782593	.1776466	.1761316	.1726939	.1660552
26	secondary	.0584745	.0583549	.0581137	.057654	.0568601	.0556869
27	secondary	.0903645	.0892977	.0875502	.084924	.081392	.0772204
28	secondary	.0991116	.0988285	.0981096	.096475	.0934868	.0896621
15	tertiary	.0855305	.0847086	.0833085	.0813043	.0790629	.077399
19	tertiary	.0336886	.0331724	.0322536	.0310932	.0303971	.0312098
20	tertiary	.0431506	.0427464	.0420252	.0409825	.0398605	.0392043
21	tertiary	.1047992	.1042418	.1032749	.1017183	.0994507	.0965224
25	tertiary	.3009763	.3005582	.299485	.2968367	.2908542	.279362
26	tertiary	.0751903	.074988	.0745818	.0738129	.0724956	.0705511
27	tertiary	.1317894	.1301036	.1273696	.1232967	.1178608	.1114977
28	tertiary	.154368	.153893	.152678	.1498718	.1445663	.1372421

Figure 5: Distribution of changes in migration flows due to a decrease of the internal border effect

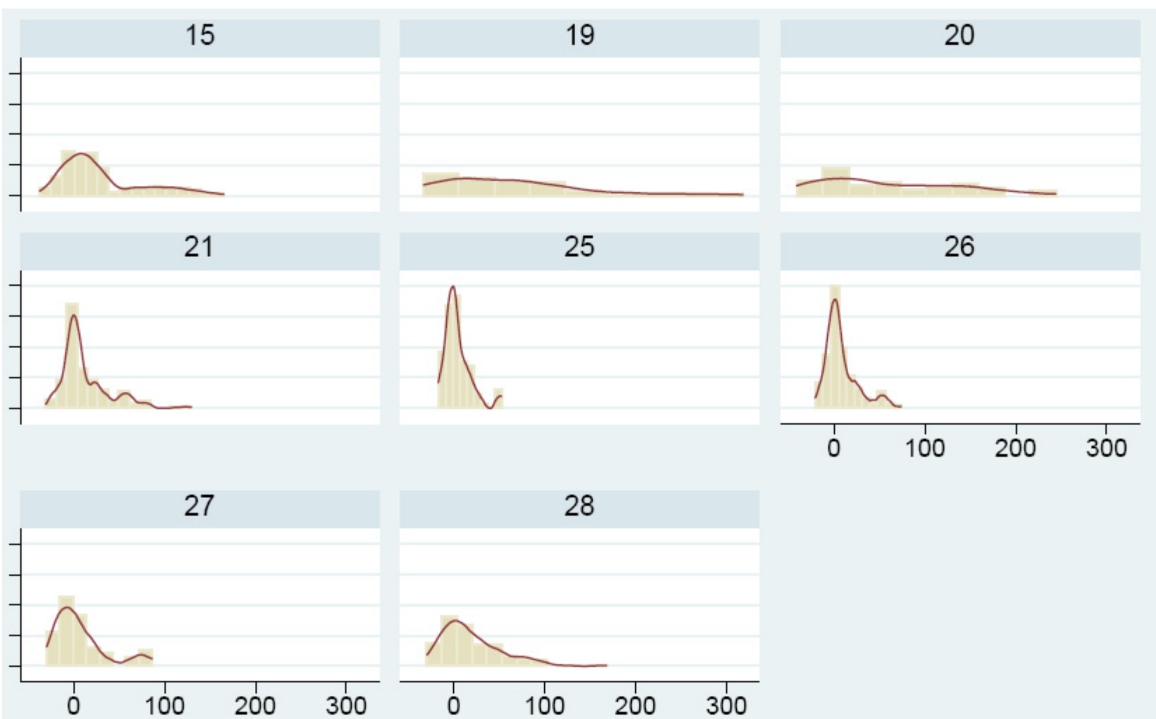

Table 8: Maximal increase in bilateral migration flows for tertiary education

Ranking	industry	state of origin	destination	baseline	0%	Increase in %
1	19	15	35	.0003939	.0016464	318.0027
2	19	24	43	.0002247	.0008571	281.3976
3	19	24	35	.0002327	.0008794	277.9902
4	20	11	35	.0003262	.0011221	243.9341
5	19	51	43	.0008666	.0029712	242.8458
6	19	51	35	.000849	.0028849	239.7829
7	20	11	33	.0006247	.002119	239.1866
8	20	11	41	.0008141	.002734	235.8239
9	20	11	42	.0006175	.0020057	224.814
10	20	14	13	.0205751	.0666459	223.9147
11	19	24	25	.0116374	.0363501	212.3551
12	19	50	33	.0011539	.0034496	198.9612
13	19	52	35	.0012774	.0036215	183.5101
14	20	21	31	.0006408	.0017731	176.6887
15	19	50	41	.0019774	.0054485	175.5356
16	20	15	35	.0004344	.0011907	174.1236
17	20	13	35	.0004148	.0011321	172.9196
18	20	15	33	.0008864	.0023962	170.3397
19	20	15	32	.001705	.0046047	170.0771
20	20	13	33	.0008121	.0021858	169.1523
21	28	14	26	.0088389	.0236867	167.9813
22	20	15	41	.0010134	.0027125	167.6595
23	15	16	35	.000389	.0010308	165.0084
24	15	13	33	.0013874	.0035949	159.1168
25	20	15	42	.0007728	.0020006	158.8844
26	20	13	42	.0007691	.0019823	157.7473
27	19	21	33	.000999	.0025711	157.3616
28	19	50	31	.0010085	.0025891	156.7307
29	19	29	43	.0006403	.0016274	154.1574
30	15	16	53	.0082913	.0209668	152.8784
31	19	29	35	.0007296	.0018379	151.8868
32	15	11	35	.000795	.0019857	149.7717
33	15	13	41	.0020849	.0052062	149.7094
34	20	17	31	.000946	.0023599	149.4499
35	20	22	35	.0002605	.0006494	149.3172
36	20	51	35	.0008378	.0020525	144.9793
37	20	21	29	.0031571	.0076999	143.8912
38	20	12	51	.005363	.0130605	143.5295
39	19	14	15	.018907	.0459764	143.1712
40	20	51	32	.0026283	.0063437	141.363