On the relation between market access, migration and wages: An empirical analysis *

Rodrigo Paillacar † Laura Hering ‡

February 14, 2008

Abstract

This paper aims at studying the impact of market access on spatial labor adjustment in Brazil, using a New Economic Geography (NEG) framework, enriched with elements of labor economics. In particular, we look at the determinants of bilateral migration between Brazilian states, taking into account sector-region specific market access and the differences in return to education across states. Since regional returns to education risk to be biased due to spatial sorting of skills, we follow the approach of Dahl (2002) and estimate a Roy model that controls for self-selected migration and gives us corrected returns to education. We see that returns to education and access to markets play an important role in the migration decision for all workers. Regions with low market access push their residents to migrate to regions with higher market access, where higher labor demand leads to more jobs and higher wages. Nevertheless, we can see that for manufacturing workers, when the market access per state and industrial sector is used instead of the state's total market access, coefficients and significance increase, indicating that the driving force in migration patterns is the difference in market access in the sector in which the individual is working.

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

JEL classification: F12, F16, R12, R23.

* We thank Pierre-Philippe Combes, José De-Sousa, Thierry Mayer, Vincent Rebeyrol, Thierry Verdier and participants of the RIEF 2008 doctoral meeting and the ETSG 2007 for excellent suggestions and fruitful discussions.

†Centre d'Economie de la Sorbonne (TEAM), 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

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

1 Introduction

In economic literature we find increasing 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; Brakman et al., 2004; Mion, 2004; Hanson, 2005; Head and Mayer, 2006; Paillacar, 2006; Hering and Poncet, 2006) and shown that further trade liberalization can lead to more spatial inequality.

NEG theory explains the emergence of a heterogeneous economic space by appealing to transport costs and increasing returns to scale (Krugman, 1991, and Krugman and Venables, 1995). One of its central tenets is the importance of proximity to consumers, as represented by the region's "market access" which is typically defined as the distance-weighted sum of the market capacity of surrounding locations (Fujita et al., 1999). The NEG "wage equation" then models nominal wages as a function of the region's market access. Wages are predicted to be higher at the economic center, and lower at the periphery. Since locations which are closer to consumer markets enjoy lower transport costs, firms based in these locations can afford to pay higher wages.

A region experiencing a positive shock on its market access is confronted with a higher demand for its goods. To respond to this new demand, labor demand increases in this region. If no additional workers are employed, the production cannot increase sufficiently to respond to the high demand and therefore prices and in consequence wages in these regions will go up, increasing by this the inequality within the country (adjustment by prices). In the case of free labor mobility, migrating workers reduce the upward pressure on wages and limit spatial inequality, but change the production structure (adjustment by quantity).

In most of these recent NEG studies, this indirect impact of trade liberalization on wages within a country via factor mobility is neglected, leading to biased estimations of the impact of market access on inequality.[2]

To contribute to a better understanding of the impact of trade liberalization on spatial inequality, we analyze the relationship between variation in regional market access and bilateral migration within Brazil. High market access being linked to higher wages, regions with high market access should induce migration flows to these regions. Low market access regions should be more susceptible for emigration.

This paper explores implications derived from two economic frameworks, which are rarely explored together on the empirical level: NEG theory, which is especially well suited for analyzing the consequences of trade liberalization on regional wages but which started only recently to consider worker heterogeneity, 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.

Including elements of labor economics into the migration analysis is necessary because in reality, migration cannot readily compensate labor demand shocks or wage pressures. There are numerous determinants of the migration decision, which need to be taken into account and which depend on the characteristics of the home region, of the possible destination and last but not least of the individual.

In our analysis we concentrate on the three following aspects of migration patterns.

In migration literature, gains in expected real wages (e.g. Todaro, 1969; Harris and Todaro,

---

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

[2] Two exceptions are De Sousa and Poncet (2007), who look at the impact of migration and market access on provincial average wage rates in China, and Hering and Poncet (2007) who find that the impact of market access is stronger in Chinese provinces with low immigration rates.

1970), amenities (e.g. Treyz, Rickman, Hunt and Greenwood, 1993) or expected returns to skills (e.g. Borjas et al., 1992) are key variables for the migration decision. 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 state of destination. This is reflected by the observation that immigrants in a specific region often share common characteristics like gender, educational level or other. Since these variables also key wage determinants, the wage level in a region will thus correspond to the average characteristics of the people living there. Numerous studies (e.g. Dahl, 2002; Borjas et al. 1992, Borjas, 1978) show that when workers chose where to live and work based on their comparative advantage, then the estimated returns to education in any given region could be biased upward or downward.

A further issue when searching for determinants of bilateral migration is the possibility that certain variables will influence the migration decision differently depending on individual characteristics. For example, Levy and Wadycki (1974) have shown that educated individuals value amenities much more than uneducated and Schwartz (1973) argues that the negative impact of distance on migration flows diminishes with increasing education of individuals.

The third focus is on a specific characteristic of Brazil: when looking at Brazilian migration data, studies have shown that about 40% of all Brazilians have migrated at least once in their lives (Fiess and Verner, 2003). This high number stands in contrast to the low sector relocation in the country. Few Brazilians change the sector of their main activity. This hints at a spatial adjustment of industry-specific labor demand, where regions with high labor demand in a certain sector will attract workers of this industry from other regions.

In this paper, when looking at the impact of trade on migration, we try to control for these three aspects. First, we control for the fact that the individual migration decision and the destination depend on individual characteristics like age, gender, family status and educational level.

In a second step, we differentiate migrants according to their educational level to see whether high educated people migrate for different reasons or go to different regions than less educated.

In the last section, to test the hypothesis that migration decisions are determined by the sector in which the individual is working, we differentiate migrants according to their industry. We investigate whether the migration decision of people working in sectors which produce tradeable goods is influenced by the market access of their own sector or whether they will migrate in response to total regional market access. In this last case migrants might go to states where total market access is high but market access for their sector low. Brazilian trade data is available at the sectoral level, also for inter-state trade. To our knowledge, we are the first study to calculate a region-sector specific market access variable and analyze its impact in a NEG framework.

Results show that we face indeed a significant selection bias in our data.

We find that migration flows depend negatively on market access of the home region and positively on the one of the destination. Thus, we observe an adjustment to the demand by the quantity. Nonetheless, we further confirm that migration costs and amenities play a significant role in the migration decision, too.

We further state differences in the migration pattern between individuals of primary, secondary and tertiary education. Market access seems to play a more important role for higher educated individuals. These observed differences can lead to a prevailing of spatial inequalities even if there is a high level of internal migration.

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 rest of the paper will proceed as follows: Section 2 presents the theoretical framework, and summarizes some implications for the empirical part; Section 3 indicates the data sources and describes the computation of our market access variables; Section 4 outlines the estimation strategy; empirical results are reported in Section 5 and Section 6 concludes.

2 New Economic Geography theory

We consider a monopolistic competition framework with product differentiation, including firm-level increasing returns to scale and trade costs. The theoretical result shows that, under a zero profit condition, a positive relationship between market access and regional wages could be established, called "NEG wage equation": firms are willing to pay higher wages in regions that are close to large markets, since firms in these regions are able to deliver goods to markets at low transport costs.

Consider a country of R regions and a two-sector economy. The first (A-sector) is characterized by constant returns, perfect competition and no trade costs. This sector offsets all trade imbalances in the other sector, thus permitting spatial specialization. The agglomeration forces take place in the second sector, which we call M-sector hereafter. This sector produces the differentiated good, experiencing trade costs and increasing returns. Preferences are described by a Cobb-Douglas function with a Dixit-Stiglitz sub-utility for the M-good. A proportion μ of the regional income is devoted to consumption of the M-goods.

$$U_i = M_i^{\mu} A_i^{1-\mu} \quad ; \quad 0 < \mu < 1 \tag{1}$$

M_i is a consumption index of the varieties of the M-sector for region i. The varieties are defined as a *continuum* of N goods, where $q_{ji}(v)$ corresponds to the demand of region i for the vth variety coming from region j. As demonstrated by Baldwin et al. (2003), there is one firm per variety, so we can refer indifferently to a variety or a firm, the total number of symmetric firms from a region being n_j. The parameter σ represents the constant elasticity of substitution (CES) between any two varieties.

$$M_i = \left[\sum_{j}^{R}\left(\int_0^{n_j} q_{ji}(v)^{\frac{\sigma-1}{\sigma}} dv\right)\right]^{\frac{\sigma}{\sigma-1}} = \left[\sum_{j}^{R}\left(n_j\, q_{ji}^{\frac{\sigma-1}{\sigma}}\right)\right]^{\frac{\sigma}{\sigma-1}} \quad ; \quad \sigma > 1 \tag{2}$$

As we are interested in the market access of region i, MA_i,we maximize the profit of each firm to obtain region j's demand for a variety coming from region i. This demand $q_{ij}(v)$ is determined by the regional income Y_j, the CIF price p_{ij} and a price index P_j. Trade costs between two regions i and j take the form of iceberg costs. With the FOB price (or mill price) being p_i, products from i are sold in region j for the price $p_{ij} = p_i \tau_{ij}$:

$$q_{ij} = \mu Y_j p_{ij}^{-\sigma} P_j^{\sigma-1} \tag{3}$$

$$P_j = \left[\sum_{i}^{R} n_i \; (p_i \tau_{ij})^{1-\sigma}\right]^{\frac{1}{1-\sigma}} \tag{4}$$

The price index, $P_j^{1-\sigma}$, is defined as the sum over the prices of individual varieties and reflects the potential suppliers of this market, considering trade costs, the elasticity of substitution, and the prices they charge. In this sense, it could be considered as a measure of the market crowding: a well served region is a region where we expect a high competition and therefore lower product prices.

Turning to the supply side of the model, increasing returns in the M-sector are usually modeled by a fixed cost per plant f_i, and a constant marginal cost m_i. Hence, profits of a firm are:

$$\pi = p_i q_i - m_i q_i - f_i \tag{5}$$

Profit maximization results in a constant mark-up:

$$p_i = \frac{m_i \sigma}{\sigma - 1} \tag{6}$$

Using the demand function in (3) and the fact that gross profits are given by $\pi_{ij} = p_{ij} q_{ij} / \sigma$, we can define the profits earned in each market j:

$$\pi_i = \frac{1}{\sigma} \left[p_i^{1-\sigma} \left(\frac{\mu Y_j}{P_j^{1-\sigma}} \right) \phi_{ij} \right] - f_i \tag{7}$$

We adopted the notation of Baldwin et al. (2003) using the term *free-ness* (*phi-ness*) of trade, $\phi_{ij} \equiv \tau_{ij}^{1-\sigma}$, that represents the combined impact of (1) trade costs and (2) the elasticity of substitution on demand. When these variables are too high, trade becomes prohibitive, and only the local demand is relevant ($\phi_{ij} = 0$). A frictionless world is represented by $\phi_{ij} = 1$. To obtain the net profit in each potential location i, we sum the profits from all locations j using equation (6) and (7):

$$\Pi_i = \frac{1}{\sigma} \left[\left(\frac{\sigma}{\sigma - 1} \right)^{1-\sigma} m_i^{1-\sigma} \underbrace{\sum_j^R \left(\frac{\mu Y_j}{P_j^{1-\sigma}} \phi_{ij} \right)}_{MA_i} \right] - f_i \tag{8}$$

The term in the sum is called *market access* or *Real Market Potential*, and is usually abbreviated as MA, where MA_i is defined as the sum of the final demand addressed to region i, weighted by the accessibility from i to these markets j (since it considers ϕ_{ij}) and by the market crowding level of every region j (since it considers the price index $P_j^{1-\sigma}$).

The spatial equilibrium can be achieved under the hypothesis that all firms will earn the same profit. An iso-profit equation that normalizes the profit to zero gives us a relationship between costs and MA:

$$m_i^{\sigma-1} f_i = \frac{1}{\sigma} \left(\frac{\sigma}{\sigma - 1} \right)^{1-\sigma} MA_i \tag{9}$$

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, we 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 firm and employment 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 the next two subsections we will briefly develop these two adjustment versions in order to derive testable equations.

2.1 The price version: Market Access, factor rewards and worker heterogeneity

Tracing a more direct relationship between wages, employment and the MA requires specifying the technology and production factors considered for the M-Sector, as well as assumptions about labor mobility. In our model, labor is the only production factor.[3] We follow Head and Mayer (2006) in introducing worker heterogeneity into the standard Krugman (1980) model, assuming that labor is the only production factor, and positing both a fixed, F , and a variable, a , component of firm-level labor requirements. Apart from notation and the inclusion of individual characteristics, we obtain what Fujita et al. (1999) call the "wage equation", indicating which wages a firm from a given location i can afford to pay to worker k:

$$w_{ki} = \left[\left(\frac{\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma}}{\sigma F \ [a]^{\sigma-1}} \right) MA_i \right]^{\frac{1}{\sigma}} exp(\rho \Lambda_k) \tag{10}$$

Log-linearizing (10) gives us an empirically testable direct relationship between MA and the regional wage.[4]

This equation is estimated using micro data instead of regional data. This has two main advantages: first, including a vector of individual characteristics allows to control for spatial sorting of skills, skills being expected to be highly correlated with MA (Mion and Naticchioni, 2005). Second, individual characteristics will play a critical role in correcting the selection bias in migration. The empirical model is obtained by taking the logs of equation 10:

$$\ln w_{ki} = \zeta + \sigma^{-1} \ln MA_i + \rho \Lambda_k + \nu_{ki} \tag{11}$$

where w_{ki} is the wage from individual k in region i, and Λ_k represents individual characteristics like school years, gender and age etc.

2.2 The quantity version: Migration and the spatial adjustment

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

[3] Redding and Venables (2004) develop a model with labor and vertical linkages in a Cobb-Douglas function. Baldwin et al. (2003) present models with capital and labor.

[4] Forslid and Ottaviano (2003) offer another variant of the model where they maintain the assumption that A-sector demands only unskilled workers, but the M-sector employs both types in a particular way: skilled workers in the fixed cost (F^s) and unskilled workers in the variable cost (a^u). In relation to spatial mobility, unskilled workers can not move while skilled ones are perfectly mobile. These modifications do not alter the linearized version of the NEG wage equation, but facilitate the analytical solving in a general equilibrium.

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):

$$\dot{s}_s = (\omega_i - \omega_j)\, s_s(1 - s_s) \tag{13}$$

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 (Tabushi 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 in the estimation.

3 Data

In this section we will first cite the different data sources and describe briefly our household panel data. 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.

3.1 Data sources

We consider three groups of data. The first data assembly is used for the market access estimation and contains trade data, which is obtained from the Brazilian Foreign Trade Secretariat (SECEX, *Secretaria de Comércio Exterior*) and Vasconcelos (2001)[7]. We further need the latitudes and longitudes of the states (to calculate bilateral distances), contiguity and common language. This data comes from the CEPII (*Centre d'Etudes Prospectives et d'Informations Internationales*) and

[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:

$$L_i^s = n_i l_i = n_i \sigma \beta \frac{1}{exp\,(Z)} \tag{12}$$

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

Table 1: Descriptive statistics for the Hourly wages 1999.

	All workers
Observations	37586
Mean	3.23
Median	1.89
Min	0.10
Max	192.46
SD	4.73
Skewness	8.82
Kurtosis	175.26

Wages expressed in current Reais of 1999 (1 R$ = 0,55 US$).

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*).

Our main data set are the 1999 and 2002 national household surveys (*Pesquisa Nacional por Amostra de Domicílio*, PNAD). [8] The PNAD survey covers individual data on all 27 Brazilian states, considering around 100,000 households. The 1999 data set will be used to estimate the regional wage equations to obtain the corrected returns to education. For these estimations we limit our analysis to non-agricultural workers classified as *private sector workers* and with a formal employment. The first column of Table 1 describes some characteristics of the full sample. We dropped the observations below the 1^{th} and above the 99^{th} percentile of the hourly wage distribution to avoid the effect of outliers. In order to reduce endogeneity issues, migration flows are obtained from the 2002 data set and cover all individuals between 16 and 60 years old.

The last group of variables needed to evaluate the migration equation are the control variables for amenities and stem from IPEADATA.

3.2 Market Access Calculation

Our main variable of interest is the state's market access, as defined in equation 8. To obtain a sound estimate of it, we follow the methodology pioneered by Redding and Venables (2004), who derive market access from a gravity equation. The states total market access can then be calculated as follows:[9]

If we denote X_{ij} the bilateral exports from region i to region j, we can use equation (3) and the iceberg trade costs to show that:

$$X_{ij} = n_i p_{ij} q_{ij} = \underbrace{n_i p_i^{1-\sigma}}_{FX_i} \phi_{ij} \underbrace{\mu Y_j P_j^{\sigma-1}}_{FM_j} \tag{14}$$

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

$$\ln X_{ij} = FX_i + FM_j + \delta \ln d_{ij} + \lambda_1 C_{ij} + \lambda_2 B_{ij} + \lambda_3 (B_{ij} * C_{ij}) + \lambda_4 b_{ij} + u_{ij} \tag{15}$$

[8]This survey is conducted almost every year by the IBGE, and has already been used to study migration issues by Fiess and Verner (2003), among others.

[9] A detailed derivation of the trade gravity equation, the market access construction and additional estimations can be found in Paillacar (2006).

where the phi-ness 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:

$$MA_i^{Total} = MA_i^{Local} + MA_i^{National} + MA_i^{International} \tag{16}$$

with

$$MA_i^{Local} = \exp\left(\widehat{FM}_i\right) d_{ii}^{\hat{\delta}} \quad ; \quad d_{ii} = 2/3\sqrt{area_i/\pi} \tag{17}$$

$$MA_i^{National} = \sum_{j \neq i}^{regions} \exp\left(\widehat{FM}_j\right) d_{ij}^{\hat{\delta}} \exp\left(\widehat{\lambda}_1 + \widehat{\lambda}_2\right) \tag{18}$$

$$MA_i^{International} = \sum_{j}^{countries} \exp\left(\widehat{FM}_j\right) d_{ij}^{\hat{\delta}} \exp\left(\widehat{\lambda}_1 + \widehat{\lambda}_3\right) \exp\left(\widehat{\lambda}_4 C_{ij}\right) \tag{19}$$

The coefficients of the global trade equation are presented in Table 7 in the Appendix.

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.[10] 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 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. We continue to postulate that, by virtue of an advantageous position, firms in a region with high MA can afford to pay higher wages, but we expect workers from low-MA regions to migrate to regions with higher

[10]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.

MA, and by this partially offset the higher wages. The first step of our analysis consists in the estimation of the NEG wage equation to obtain regional returns to education which we can introduce as determinants of migration flows in the second step, the migration equation. Nevertheless, if we estimate the wage equation in (11), we will not be able to make the selection bias correction. In the next two subsections we will therefore first explain how to obtain corrected returns to education and then derive the migration equation which links market access to bilateral migration.

4.1 Selection bias correction

In order to control for the phenomenon of self-selected migration and the resulting spatial sorting of skills and the intrinsic risk of biased regional returns to education, we will estimate region-specific wage equations, one wage equation for each of the 27 Brazilian states, which include a correction function based on a multinomial logit. This method is based on Dahl (2002) who provides a simple way to model and correct for selection bias when there are many choices, based on a Roy model of mobility and earnings. In the following, the theoretical derivation of the estimated specification is given.

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 showed by Dahl (2002), this utility can be decomposed in (1) V_{ij}, a component capturing the mean effect of pecuniary (wages) and non pecuniary considerations (amenities), 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 the wage equation 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.

Traditional selection bias corrections (like the conditional logit model) are not very well suited to cases where individual migration decisions imply numerous potential destinations.[11] Dahl (2002)

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

proposes a non-parametric approach, using the frequency estimator for individuals sharing the same characteristics, namely location, gender, education and age. This means that we are assuming that individuals with the same characteristics are affected in the same way by the determinants of migration. Thereby, by isolating the education variable, the equation 11 enriched with a correction function λ_j^* that uses the first-best migration probability (p_{kij}) and the retention probability (p_{kii}) becomes[12]:

$$\ln w_{kj} = \zeta_j + \rho_j h_k + \psi_j \Lambda_k + \lambda_j^* \left(p_{kij}, p_{kii}\right) + \xi_{kj} \tag{20}$$

where λ_j^* is estimated using a polynomial expansion.

We estimate equation 20 once for each of the 27 Brazilian states to obtain the corrected returns to education, which we will then use in the next step as explaining variables the migration equation. 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 (see also Mion and Naticchioni (2005)).

4.2 Migration equation

The migration equation shows us how the migration probability for each group individual 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.

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

$$V_{kij} = X_{ij}\beta + \xi_{ij} + e_{kij} \tag{21}$$

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 23 leads then to:

$$Pr(V_{kij} > V_{kim}) \forall j \neq m \tag{22}$$

$$Pr(e_{kij} - e_{kim} > X_{im}\beta - X_{ij}\beta + \xi_{im} - \xi_{ij}) \forall j \neq m \tag{23}$$

[12]The wage equations are estimated on the state level, since we cannot localize our individuals on a finer geographical level. These state-specific estimations cannot contain total market access as an explanatory variable, since it is necessarily the same for all observations, but we can nevertheless indirectly control for the impact of agglomeration economies (and among them, for better market access) by including dummy variables for urban vs. rural. For an estimation of the structural version of the NEG wage equation for Brazil see Paillacar (2006). We could include market access at the sectoral level in this step. This issue will hopefully be treated in a future version of the paper

McFadden (1973) shows that by integrating out over the distribution of the logistic distribution, we can obtain the following migration probabilities:

$$Pr(M_{ijk} = 1) = \frac{exp(X_{ij}\beta + \xi_{ij})}{\Sigma_{J=1}^{J} exp(X_{ij}\beta + \xi_{ij})} \tag{24}$$

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

$$s_{ij} = Pr(M_{ijk} = 1) = \frac{exp(X_{ij}\beta + \xi_{ij})}{\Sigma_{J=1}^{J} exp(X_{ij}\beta + \xi_{ij})} \tag{25}$$

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

$$s_{ii} = Pr(M_{ijk} = 1) = \frac{exp(X_{ii}\beta + \xi_{ii})}{\Sigma_{J=1}^{J} exp(X_{ij}\beta + \xi_{ij})} \tag{26}$$

Dividing equation 25 by equation 26 and taking the log yields:

$$ln(\frac{s_{ij}}{s_{ii}}) = ln\left(\frac{exp(X_{ij}\beta + \xi_{ij})}{exp(X_{ii}\beta + \xi_{ii})}\right) = X_{ij}\beta - X_{ii}\beta + \xi_{ij} - \xi_{ii} \tag{27}$$

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

$$\ln\frac{s_{ij}}{s_{ii}} = \beta_1 + \beta_2\rho_j - \beta_3\rho_i + \beta_4 MA_j - \beta_5 MA_i + \beta_4 dist_{ij} + \beta' A_j - \beta' A_i + v_{ij} \tag{28}$$

where A corresponds to a vector of the state's amenities, and $dist$ to the bilateral distance, proxying migration costs.

As we are interested in differentiated effects in migration by educational level, in section 5.2.1we will estimate equation 28 also separately for our four groups of workers - according to their level of education. By this, we stress the fact that the migration decision across these groups may be influenced differently by returns to education and the spatial economy.

Finally, in 5.2.2we will also 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).

5 Results

Before we present estimation results for the wage and migration equation, we want to give 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 chose the answer to the question "where have you lived before?" in the 2002 PNAD data set as the

Table 2: Descriptive statistics for Migrants.

Education	**Total**	**Migrants**	**Percentage**
Tertiary	59650	19263	32.3%
Secondary	27811	8558	30.7%
Primary	52287	17628	33.7%
No School	14810	5128	34.6%
TOTAL	154558	50577	32.7%

region of origin. Since we don't know when the individuals have migrated, this approach will at least assure us that when we consider the migration decision of individuals having lived in several states before, we take into account the two for the migration decision relevant states.

In Table 2 we see that in total 32.7 % of the workers aged between 16 and 60 in our data set have migrated. With this percentage our sample reflects well the estimated percentage of migration of the total Brazilian population, where around 40% are supposed to have migrated at one point in time (Fiess and Verner (2003)). As we see, the percentage of migrants is similar for all four educational levels considered.

Figure 1 in the Appendix reports the percentage of total migration flows per educational category that goes to each of the 27 states.[13] Since our data sample contains more observation for highly populated states and is by this representing the distribution of population within Brazil, it is not surprisingly to see that the state with the highest number of migrants is Sao Paulo. More interesting in this graphic is the difference between the different educational levels, indicating that there are clearly different preferences across these four groups. We can see that migrants with no education tend to move more than the others to states in the north-east (21-29), whereas people with tertiary education go more to the north-west (11 - 17), Brasilia (53), where the capital is located, and to Rio Grande do Sul (43), which is known for its good climate and recent economic development.

Figure 2 shows the part of each migrant group in the actual population of the respective educational group in each state. We see that the states Roraima (11), Amazonas (13) and Brasilia (53) are the one with the highest immigration rates for all four categories. We see that in Sao Paulo uneducated people are to more than 70 % originate from other states, whereas for highly educated people, only 20% come from somewhere else. States with relatively few immigrants are Rio grande do Sul (43), Bahia (29) are Minas Gerais (31).

5.1 Regional returns to schooling

To see whether in Brazil the self-selection of individuals in terms of location is important or not, we first estimate the 27 wage equations without the correction term (equation 11). In a second step, we include the correction function containing migration probabilities that we obtain by a multinomial logit estimation(equation 20). The independent variables of the wage equation are the same in both specifications and refer to the classic Mincer wage equation regressors: age, the square of age, gender and education. We differentiate between four different educational levels: *no education* for those who declare zero school years, *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. The category of reference in the estimations is always

[13] See Table 8 in the Appendix for a list of the 27 states and their codes.

no education. A further independent variable is *urban*, which takes the value 1 if the individual lives in an urban area and 0 in the case of a rural residence.[14]

The selection function contains the following variables: four age group, the four educational level, family status (married, married with child, single mother), gender, ethnic and the five macro-regions of origin.[15]

The estimation of the wage equations gives us for each state the wage differential between primary, secondary or tertiary education and the reference group, the non educated workers. Figures 3 to 5 in the Appendix, plotting the coefficients of corrected versus the non-corrected returns to education, illustrate well the existence of a bias in the estimated returns to education.[16] The bias appears to be less important for tertiary education, but is still significant.

The difference between the two coefficients shows that we have mainly a positive self-selection for tertiary education. Dahl (2002) finds a similar but weaker 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 a better match for their particular skills and talents compared to those with less education. For primary and secondary education, the selection bias takes also often a negative form.

Also, migration choices of highly educated individuals might be more responsive to unobserved earnings because these persons are more likely to move for a fixed moving cost or because variation in unobserved earnings across states is greater for them. This could generate a positive correlation between schooling level and the error term in the wage regressions for the self-selected samples and hence an upward bias in the uncorrected estimates of the return to education.

For the estimation of the migration equation we need to compare the returns of education in two different states. But when looking e.g. at the coefficients of tertiary education, the only information we get when we find a higher coefficient in location i than in location j is that wage differentials between no educated workers and those with tertiary education are higher in i than in j. Returns can only be compared across states if the base wage for non educated workers is the same everywhere. When looking at our data, we find that this is not the case. To be able to compare returns to schooling across regions, we weight our coefficients for tertiary, secondary and primary education with the average base wage of the non educated workers in each state. By definition, the weights for Sao Paulo are one and all other base wages are compared to Sao Paulo base wages.

We then obtain three new variables, ρ^t, ρ^s, ρ^p and ρ^0[17], which represent the comparable state-specific returns to tertiary, secondary, primary and no education, which we will use as regressors in the next step. [18]

[14] For some states from the Northwest, we have only data from urban areas. In these cases *urban* has been dropped.

[15] 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 economic opportunities afforded by living in these regions clearly explain much of their high population density.

[16] We exclude state 21 due to a very low number of observations for this state.

[17] Returns to education for individuals with no schooling are the mean base wage of the state for this category of workers.

[18] We also find any positive 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

5.2 Market Access and migration response

In this section, we finally estimate the migration equation, equation 28. We will look at different aggregation levels of bilateral migration, differentiating between sectoral migration flows and different educational levels. Most regressions include fixed effects for the destination and origin macro-region. Fiess and Verner (2003) compare the migration pattern from and to these macro-regions, showing that these differ significantly. Introducing these dummies has therefore the advantage of capturing structural and cultural differences between the regions for which we have no control variables, but they also change the interpretation of our coefficients. When adding fixed effects, we will explain differences in migration flows within macro-regions, not across macro-regions.

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).

All estimations cluster the error terms at the origin-destination-couple-level.

To give a first impression of the determinants of migration, column 1 of Table 3, reports results for total bilateral migration depending on total market access, bilateral distances and macro-region fixed effects. Next to manufacturing workers, migration flows contain also a high number of workers in service or agricultural sectors. Bilateral migration, 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.

As we see from the reported results, market access plays indeed a significant role in the migration choice of individuals. The signs of the two coefficients correspond to our expectations: the impact of MA_j is clearly positive and significant at the 1% level, indicating that high market access states attract workers. The coefficient of MA_i is negative and significant, confirming that low-MA regions are more likely to see their workers leave.

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

In column 2, we repeat the same regression without regional fixed effects. In this case the destination's market access has a much lower coefficient. This indicates that the destination's market access is a key variable for determining the new location within a given macro-region, though it is also an important attraction factor across macro-regions.

5.2.1 Market access and migration: impact of education

From column 3 of Table 3 on, we differentiate four migration flows according to the four educational levels for each existing couple of states. Bilateral migration m^e_{ij} is defined here as

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).

$migrants^e_{ij}/stayers^e_{ii}$, the number of migrants with educational level e over $stayers^e_i$, the number of workers with educational level e in our sample that stay in region i. Since we do not observe migration flows for all possible combinations of sates and educational level, regressions are on only 1923 positive flows. As additional explanatory variables we can include here the corrected regional returns to education which we have estimated in section 5.1. Since in this table we do not distinguish between sectors, the market access variable is again on the aggregated level, i.e. total market access.

Column 3 has the same specification as column 1 but estimates or obtained at the more disaggregated level. Results are qualitatively the same, with slightly lower coefficients for our market access variables. To show the importance of correcting for self-selection in the estimation of the returns to education, we first use in column 4 the uncorrected returns to education and only from column 5 on the corrected returns to education. While in column 4 it seems that migrants stem mainly from states with high returns to education and the returns to education at the destination do not seem to play a significant role, the corrected returns draw a more realistic picture, indicating that emigration takes place in states with low returns to education and immigration states are those with high returns to education.

In column 6 of Table 3, we add two amenities variables. First, the homicide rate, defined here as the percentage of total population that has died in 1999 because of homicide, which indicates the level of criminality and is therefore used as an index for the security level of a state. We expect that people will leave regions with a high homicide rate and go to areas with a smaller rate. As a second amenity variable, we calculate construct an unemployment rate by taking the difference between the total active population and the employed population per state. Due to lacking data for most years, we use data from 1991 here. Unemployment rate can be correlated to market access, but could also have structural or institutional reasons, independent of the economic development of the state. Naturally, we suppose migrants to come from regions with high unemployment rates and go to states with less unemployment.

The positive and significant sign of the homicide rate in the state of origin is in line with our expectations, as well as the negative and significant unemployment rate of the destination. More difficult to interpret is the negative and significant sign for the unemployment rate of the home state, suggesting that the higher the unemployment rate, the less people leave the state. If this would really be the case, this could either indicate some structural specificities of the state or a poverty trap (people are too poor to bear migration costs and do not migrate even if they would have higher chance to find a job in other region), but this issue needs further investigation before making any conclusions.

In order to know whether determinants vary significantly between migrants with different educational levels, Table 4 displays results for the estimations of equation 28, when run separately for each educational group.

We see, that for all four groups, distance has a similar negative impact on migration, significant at the 1% level, with the exception of less qualified migrants, who seem to be moving more easily to regions which are farer away. The importance of market access though seems to be increasing with education. This can be best seen in the coefficients of MA_j. But also the market access of origin has its strongest impact on high educated migrants and is insignificant for the two less educated groups.

From the two returns to education variables, only ρ^e_i, the respective returns in the state of origin is significant for all groups, though the high positive coefficient for non educated migrants should be considered with caution. We find positive and significant estimates for ρ^e_j for educated workers, except for secondary education, indicating that people choose regions where their educational level is better remunerated.

The homicide rate has more or less the same impact in all four groups. The unemployment rate in the state of destination plays the expected role, having a negative and significant impact, especially for higher qualified individuals. Lower unemployment normally going ahead with higher economic development and higher living standards, it is in the line of our expectancies that educated migrants choose these locations.

5.2.2 Market access and migration: sectoral level

In the first two columns of Table 5, we repeat the same estimations as in the last two column of Table 3, 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. Coefficients for distance and MA_i are smaller. MA_j becomes even non significant in column 1. Also ρ_j loses in magnitude and significance. It seems that, when limiting our analysis to agricultural and manufacturing workers who are actually working in the tradeable sectors, contrary to our expectations, market access is a less clear determinant of bilateral migration. In particular for the choice of destination its explanation power seems limited.[19] This also suggests that the results in the previous regressions could have been driven by migrants in non tradeable service sectors, usually not supposed to be tradeable or produce under increasing returns to scale.

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 more robust relationship between market access and migration on a more disaggregated level.

In the last three columns of Table 5, we therefore differentiate migrants according to the industry in which he works. 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.

At first, in column 2, we keep the total market access as independent variable. Coefficients of MA_i and MA_j become significant again but stay low in comparison to the previous Tables. In Column 3 we finally introduce the region-sector specific market access. We observe much higher and highly significant coefficients for MA_{is} and MA_{js} than for the total market access. In the last column, we replace macro-region fixed effects by state fixed effects. This leads again to a decrease in coefficients, but given that in this regressions, we control for all other origin and destination state specific factors, 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_{is} and MA_{js} suggests that even if a region might have a high 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 them. Note that the coefficient of the bilateral distance, although significant, is strongly reduced: manufacturing workers tend more easily to 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 affect 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.

[19] Results stay similar, when adding additional or different amenities variables or when excluding agricultural workers.

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.

These results 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 more 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 the market access, the benefits will be captured by these workers. In the next section we explore if skill levels are also playing a role in shaping differentiated effects among worker's mobility.

5.2.3 Market access and migration: impact of education and sectors

In this last section, we combine the two dimensions considered above - migration by sector and by educational level. Table 6 shows bilateral migration flows defined as $migrants_{ijs}^{e} = migrants_{ij}^{e}/stayers_{iis}^{e}$, 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. We find that returns to education stay important when controlling for sectoral market access and also when adding state fixed effects. In the last column we add also dummies for the four educational categories. Results stay unchanged. When separating estimations according to education at this fine aggregation level (not reported here), we see that all coefficients are quite similar among all type of workers. There seems not to be any particular differences in the migration determinants according to education once controlled for the sector in which the individuals are working.

Table 3: Bilateral migration flows I

COEFFICIENT	(1) m_{ij}	(2) m_{ij}	(3) m^e_{ij}	(4) m^e_{ij}	(5) m^e_{ij}	(6) m^e_{ij}
$ln\ dist_{ij}$	-0.792***	-0.750***	-0.802***	-0.806***	-0.806***	-0.825***
	(0.071)	(0.075)	(0.071)	(0.071)	(0.072)	(0.073)
$ln\ MA_i$	-0.503***	-0.484***	-0.307***	-0.337***	-0.262***	-0.391***
	(0.090)	(0.045)	(0.099)	(0.10)	(0.10)	(0.11)
$ln\ MA_j$	0.720***	0.377***	0.494***	0.473***	0.403***	0.576***
	(0.086)	(0.039)	(0.094)	(0.095)	(0.096)	(0.11)
ρ_i (uncorr)				0.224*		
				(0.13)		
ρ_j (uncorr)				0.164		
				(0.12)		
ρ_i					-0.252***	-0.324***
					(0.093)	(0.093)
ρ_j					0.516***	0.599***
					(0.091)	(0.090)
$ln\ unemp_i$						-0.429**
						(0.18)
$ln\ unemp_j$						-0.701***
						(0.20)
$ln\ homicide_i$						0.542***
						(0.10)
$ln\ homicide_j$						-0.130
						(0.094)
Fixed effects	macro-r.	no	macro-r.	macro-r.	macro-r.	macro-r.
Constant	-1.089	2.121***	-0.845	-0.585	-0.574	-5.757***
	(1.41)	(0.77)	(1.55)	(1.55)	(1.55)	(2.16)
Observations	632	632	1923	1923	1923	1923
R^2	0.44	0.35	0.25	0.26	0.27	0.30

Robust standard errors in parentheses

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

Table 4: Bilateral migration flows by educational level

COEFFICIENT	(1) $m_{ij}^{noeducation}$	(2) $m_{ij}^{primary}$	(3) $m_{ij}^{secondary}$	(4) $m_{ij}^{tertiary}$
$ln\ dist_{ij}$	-0.692***	-0.910***	-0.895***	-0.809***
	(0.093)	(0.100)	(0.092)	(0.074)
$ln\ MA_i$	0.103	-0.202	-0.502***	-0.657***
	(0.16)	(0.17)	(0.16)	(0.11)
$ln\ MA_j$	0.376**	0.463***	0.557***	0.993***
	(0.15)	(0.15)	(0.14)	(0.12)
ρ_i	1.578***	-0.626**	-0.795***	-0.463***
	(0.35)	(0.27)	(0.19)	(0.14)
ρ_j	-0.439	0.549**	0.0409	0.322**
	(0.35)	(0.24)	(0.24)	(0.15)
$ln\ unemp_i$	-0.692**	-0.748**	0.0798	-0.176
	(0.30)	(0.31)	(0.29)	(0.21)
$ln\ unemp_j$	-0.164	-0.744**	-0.549**	-0.912***
	(0.31)	(0.31)	(0.28)	(0.22)
$ln\ homicide_i$	0.409***	0.607***	0.522***	0.564***
	(0.15)	(0.15)	(0.15)	(0.11)
$ln\ homicide_j$	-0.0852	-0.230*	-0.0909	-0.0808
	(0.14)	(0.14)	(0.13)	(0.10)
Fixed effects	macro-r.	macro-r.	macro-r.	macro-r.
Constant	-8.678***	-7.243**	-1.607	-6.512***
	(3.04)	(3.09)	(2.90)	(2.16)
Observations	383	504	458	578
R^2	0.34	0.27	0.36	0.47

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5: Bilateral migration flows II

COEFFICIENT	(1) $mRest^e_{ij}$	(2) $mRest^e_{ij}$	(3) m_{ijs}	(4) m_{ijs}	(5) m_{ijs}
$ln\ dist_{ij}$	-0.581***	-0.602***	-0.580***	-0.363***	-0.371***
	(0.072)	(0.071)	(0.059)	(0.063)	(0.054)
$ln\ MA_i$	-0.250**	-0.411***	-0.316***		
	(0.10)	(0.11)	(0.077)		
$ln\ MA_j$	0.161	0.241**	0.437***		
	(0.10)	(0.12)	(0.078)		
$ln\ MA_{is}$				-0.583***	-0.243**
				(0.078)	(0.11)
$ln\ MA_{js}$				0.784***	0.454***
				(0.077)	(0.11)
ρ_i	-0.316***	-0.359***			
	(0.094)	(0.091)			
ρ_j	0.176*	0.208**			
	(0.10)	(0.099)			
$ln\ unemp_i$		-0.293			
		(0.21)			
$ln\ unemp_j$		-0.456**			
		(0.23)			
$ln\ homicide_i$		0.486***			
		(0.11)			
$ln\ homicide_j$		-0.0342			
		(0.10)			
Fixed effects	macro-r.	macro-r.	macro-r.	macro-r.	state
Constant	-0.955	-3.826*	-1.133	-5.680***	-3.911***
	(1.54)	(2.20)	(1.24)	(0.72)	(0.68)
Observations	1264	1264	7777	2410	2410
R^2	0.19	0.22	0.22	0.28	0.41

Robust standard errors in parentheses

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

Table 6: Bilateral migration flows III

COEFFICIENT	(1) m^e_{ijs}	(2) m^e_{ijs}	(3) m^e_{ijs}
$ln\ dist_{ij}$	-0.236***	-0.246***	-0.247***
	(0.055)	(0.041)	(0.040)
$ln\ MA_{is}$	-0.510***	-0.233**	-0.201**
	(0.079)	(0.097)	(0.096)
$ln\ MA_{js}$	0.760***	0.494***	0.477***
	(0.080)	(0.10)	(0.098)
ρ_i	0.424***	0.201***	-0.0519
	(0.076)	(0.076)	(0.13)
ρ_j	-0.250***	0.0234	-0.178**
	(0.067)	(0.070)	(0.085)
Fixed effects	macro-region	state	state & education
Constant	-6.883***	-5.200***	-4.896***
	(0.64)	(0.63)	(0.64)
Observations	4072	4072	4072
R^2	0.27	0.38	0.40

Robust standard errors in parentheses

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

6 Conclusion

In this paper, we 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 four educational levels and test if migration patterns vary across these groups. We further look at the impact of returns to education and amenities on migration. To obtain corrected returns to education, we control for the presence of self-selection in the spatial repartition of the individuals. We do this by estimating a Roy model, following the approach of Dahl (2002). In a second step, we look at the determinants of bilateral migration between Brazilian states, taking into account the differences in return to education across states. We see that access to markets plays an important role in the migration decision: regions with low market access push their 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 Tabushi 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. In manufacturing, the worker having obtained a sector-specific human capital maximizes his wage when staying in the same sector. This provides enough incentives to migrate also to region's far away from his original state, if demand for his sector is increasing there. If trade liberalization in Brazil is changing the market access, the benefits will be captured by these workers. Moreover, these implicit benefits measured by worker mobility, are similar for the four schooling levels considered in the study. In sum, although educational level is strongly affecting migration by the channel of selection bias, it is skills that are sector-specific that are more important in determining the dynamic response of workers to market access differentials.

The coefficients obtained in this first exercise for market access and the other migration determinants can be used in simulations for assessing intranational impacts inside Brazil on a series of subjects like a deepening of the integration process in the world economy, infrastructure policies or other measures tending to reduce the strong internal fragmentation in this country, etc.

References

Baldwin, R., R. Forslid, P. Martin, G. Ottaviano, and F. Robert-Nicoud (2003). *Public Policies and Economic Geography*. Princeton University Press.

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.

Fiess, N. and D. Verner (2003). Migration and human capital in Brazil during the 1990s. World Bank Policy Research Working Paper 3093, World Bank.

Forslid, R. and G. Ottaviano (2003). An analytically solvable Core-Periphery model. *Journal of Economic Geography 3*(3), 229–240.

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*. Forthcoming. Also in NBER 12885.

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.

Mion, G. and P. Naticchioni (2005). Urbanization externalities, market potential and spatial sorting of skills and firms. Working Paper 5172, CEPR.

Murata, Y. (2003). Product diversity, taste heterogeneity, and geographic distribution of economic activities: 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.

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 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.

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

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

The Dynamics of U.S. Internal Migration, by George I. Treyz; Dan S. Rickman; Gary L. Hunt; Michael J. Greenwood The Review of Economics and Statistics 1993 The MIT Press

Education and the Decision to Migrate: An Econometric Analysis of Migration in Venezuela, by Mildred B. Levy; Walter J. Wadycki Econometrica 1974 The Econometric Society

Interpreting the Effect of Distance on Migration, by Aba Schwartz The Journal of Political Economy 1973 The University of Chicago Press

Appendix

Here we present the Trade Gravity equation for the Market Access aggregated.

Table 7: Trade Gravity Equations.

	(1)
Log of Distance	-1.02^{a}
	(0.11)
Internal Contiguity	0.73^{a}
	(0.21)
National Contiguity	1.17^{b}
	(0.37)
Internal Border	-2.97^{a}
	(0.64)
National Border	-1.44^{c}
	(0.63)
No. of obs	4309

All specifications include exporting and importing country fixed effects. Robust Standard Errors in parentheses. ^{a}, ^{b} and ^{c} represent respectively statistical significance at the 0.1%, 1% and 5% levels. Dependent variable: Log of Trade flows .

Table 8: 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

Figure 1: Proportion of migrants

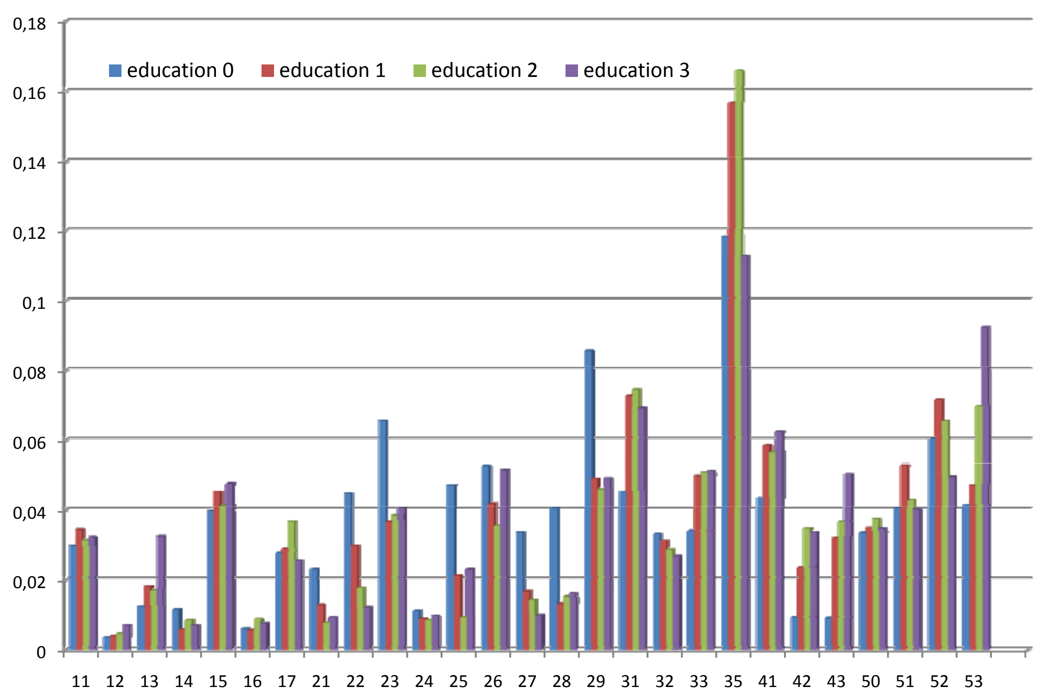

Figure 2: Migrants by respective population group

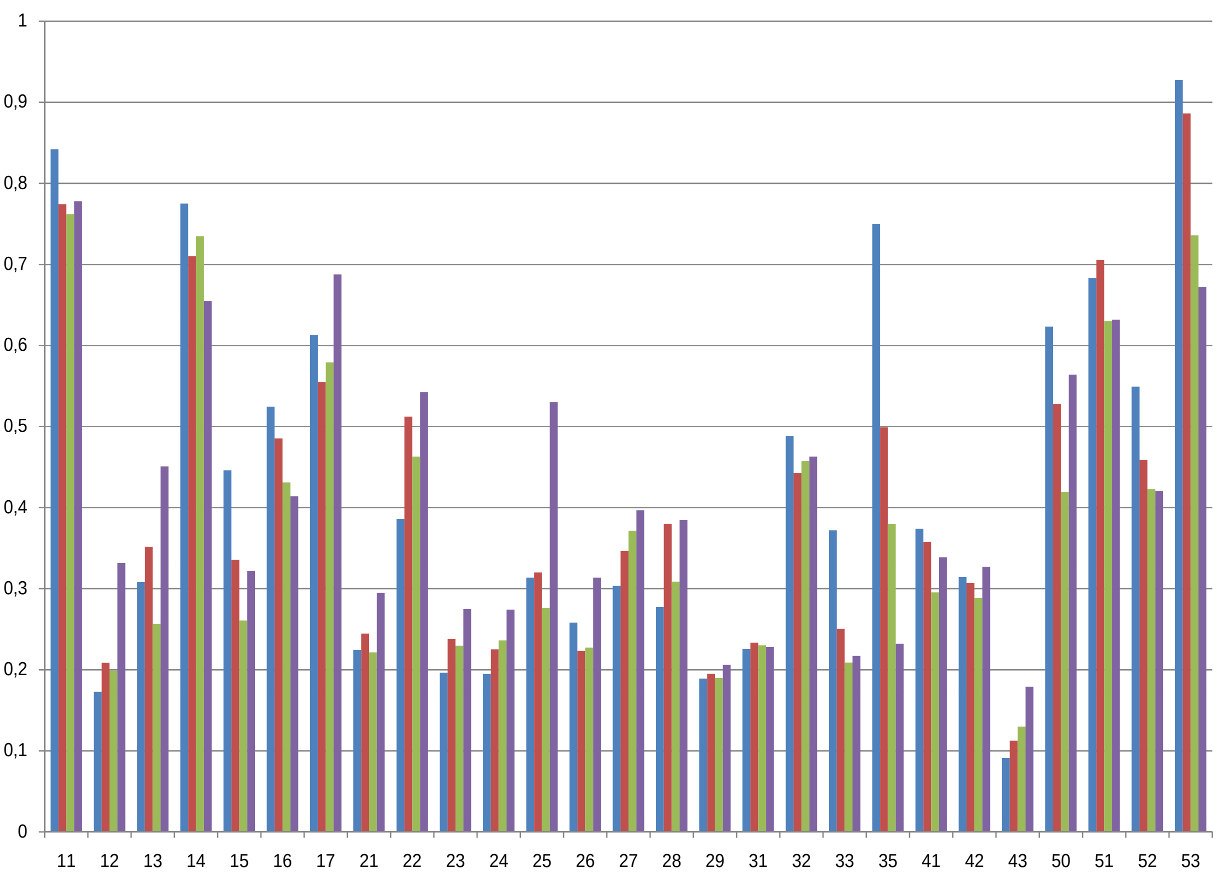

Figure 3: Returns for primary education

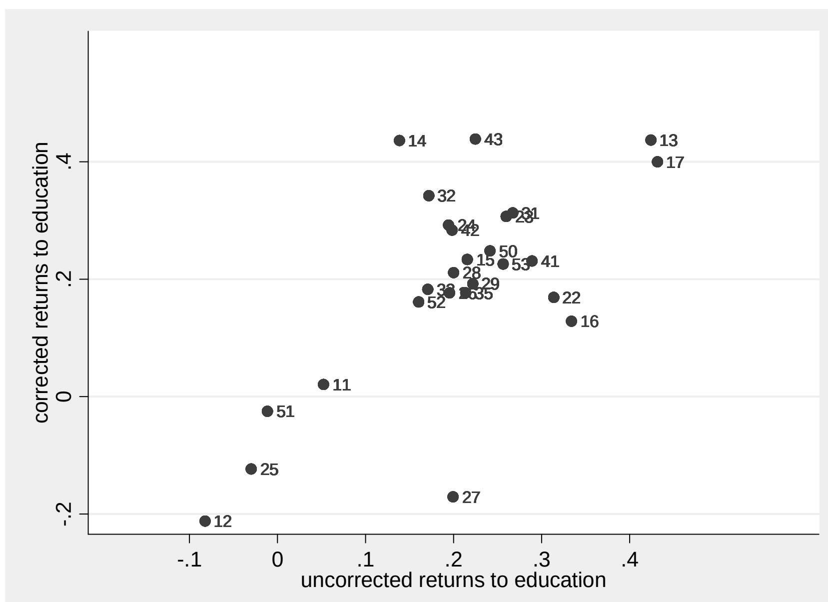

Figure 4: Returns for secondary education

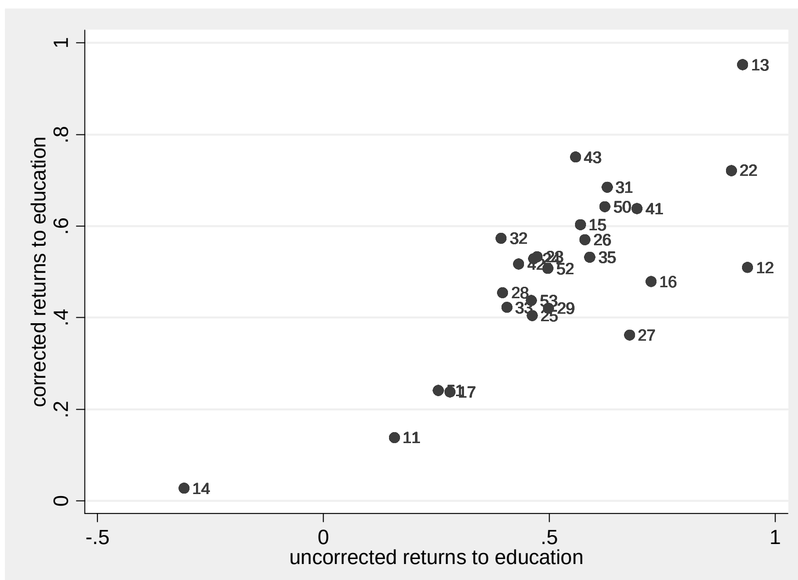

Figure 5: Returns for tertiary education