SELF-SELECTION OF RETURN MIGRANTS IN BRAZIL AND INEQUALITY

Silvio H. T. Tai *

Ana Carolina B. M. Ribeiro †

Resumo

Esse artigo estuda a autosseleção dos emigrantes retornados no Brasil, de 2001 a 2010. Verificou-se empiricamente o impacto da desigualdade de renda na autosseleção através de dois canais opostos. Em primeiro lugar, os custos de migração impedem indivíduos com baixa qualificação de emigrar, o que contribui para a seleção positiva. Essa seleção é acentuada em locais com muita desigualdade. Em segundo lugar, indivíduos com baixa qualificação têm incentivos salariais para emigrar de estados mais desiguais, o que contribui para a seleção negativa. Considerando os dois canais, os efeitos dos custos de emigração são mais importantes, com predominância da seleção positiva.

Palavras-chave: Desigualdade, Emigrantes retornados, Autosseleção, Brasil.

Abstract

This paper investigates self-selection of returnees migrants in Brazil, from 2001 to 2010. We find evidence of the impact of inequality on self-selection through two opposing channels. First, migrations costs prevent low-skilled individuals from emigrating, which support positive selection. This selection is accentuated in locations with high inequality. Secondly, we find that low-skilled individuals have wage incentives to emigrate from states with relative high inequality, implying a negative impact on selection. All in all, we find more important effects of migration costs, with the predominance of positive selection.

Keywords: Inequality, Return Migration, Self-selection, Brazil.

* RITM - University Paris XI Sud, France and Pontifical Catholic University of Rio Grande do Sul, Brazil. E-mail: silviotai@gmail.com.

† Ph.D. candidate at the Pontifical Catholic University of Rio Grande do Sul, Brazil. E-mail: carolbmribeiro@hotmail.com.

1   Introduction

Several studies indicate that migration from developing to developed countries tend to be relatively qualified, in relation to the individuals in the origin country. This phenomena may be harmful and prevent developing countries from keeping human capital, a fundamental input of economic development (Mankiw et al. 1992) and an essential factor to devise and develop new technologies (Romer 1990).[1]

The recent crisis in developed countries, combined with stabilization and growth in Brazil, has progressively attracted not only foreign migrants, but also Brazilian returnees.[2] The flows of Brazilian returnees had a three-fold increase from 2005 to 2010.[3] One could expect the return of human capital that "drained" in the past, as well as the return of labor. Considering immigration as an inflow of human capital, returnees have some advantages compared to foreign immigrants. Firstly, they do not face bureaucratic, cultural, or linguistic barriers in migrating back to Brazil. Secondly, their return migration does not trigger, at first, negative attitudes in compatriots that remained in Brazil.[4]

Returnees are considered in this paper because they are paid at the same prices of other Brazilians, allowing a direct comparison. We find evidence that these migrants have more education and earn higher wages than Brazilians who remained in the country. Since Brazil is one of the most unequal countries worldwide, one would expect the opposite, that highly educated Brazilians would remain in the country. Considering that they earn relative high salaries, why do these individuals tend to migrate to other countries (and return some time later)?

This article quantifies the self-selection of return migrants in Brazil from 2001 to 2010, and estimates the impact of the underlying theoretical mechanisms of selection through inequality. Using microdata from the 2010 census, this paper compares wages of native-returnees to those of non-migrant natives in each Brazilian state, providing a direct wage comparison. The role of inequality on the correlation between education and migration costs is a key issue in our study.

We found evidence of the impact of inequality on self-selection through two opposing channels. On one hand, migration costs prevent low-skilled individuals from migrating, which supports positive selection. This selection is more elevated in Brazilian states with high inequality, because they present high return to skills. On the other hand, low-skilled individuals have incentives for migrating to countries with low inequality where, their wage is higher, which supports negative selection. Both channels are empirically verified, with a predominance of the former.

Borjas (1987) shows that migration from countries with relatively high inequality and high rates of return to schooling is negatively selected. This theoretical implication is based on a simplified assumption: migration costs are a constant proportion of individuals' income, regardless of their educa-

---

[1]Nonetheless, recent studies point to positive network externalities. See Docquier & Rapoport (2012) for a survey.

[2]Between 2001 and 2010, the inflow of returnees was 2.5 higher than the inflow of foreigners.

[3]See Figure 2.

[4]A stream of the literature (Facchini & Mayda 2009, Müller & Tai 2016) indicates that natives can have negative opinions regarding immigration, either because they can represent a competition in the labor market or because foreigners impose a net cost on the redistribution system.

tion level or wage. Without this hypothesis, migration costs would represent larger proportions of low-skilled individuals' wage, reducing their motivations to migrate (Chiswick 1999, Assunção & Carvalho 2013).

Chiquiar & Hanson (2005) consider variable time-equivalent costs correlated with skills to be what theoretically underpins negative, intermediate and positive selection, depending on the amount of costs and their correlation with skills. Empirically, they find intermediate selection as they compare wage densities of Mexican residents to counterfactual constructed wage densities that would be obtained were Mexican immigrants in the US paid accordingly to Mexican prices. This paper presents the advantage of direct wage comparisons (without price heterogeneity) and also the possibility of econometric examination of different theoretical arguments determining selection, notably the implication of variable proportional costs on the liquidity constraint. Although Chiquiar & Hanson (2005) justify intermediate selection via variable time-equivalent costs, they did not empirically test this hypothesis.

Belot & Hatton (2012) estimate the self-selection determinants for OECD's immigrants. They focus on the poverty constraint and propose that less educated individuals have higher time-equivalent migration costs if their source country is poor. This approach is close to that applied to the second-step regressions in this paper. Although our empirical results are similar, there is a fundamental difference: we use inequality instead of poverty. The theoretical model indicates that inequality determines selection, not poverty.

For instance, one can consider a very poor country where everybody is poor. In this case, the poorest individuals cannot migrate, but selection is not observed because richer individuals cannot migrate either. A country may be poor or rich, but if the correlation between time-equivalent costs and skills is low (otherwise stated, if the inequality is low), cost-based positive selection is low.

In addition, differently from Belot & Hatton (2012), this paper makes use of individual data, which allows us to investigate self-selection on observed and unobserved characteristics by education level.

Borjas & Bratsberg (1996) extend Borjas (1987) return migration model, arguing that there is a negative selection in the migration and a positive selection in the return migration.[5] It is important to note that the selection in (the first) migration compares migrants to native non-migrants whereas the selection in the return (second) migration compares migrants that are coming back to those that remained in the destination country.

This paper compares returnee migrants to natives that have never migrated. In the context of Borjas and Bratsberg's model (with constant time-equivalent costs), only the "first" selection counts, in this case returnees migrants would be negatively selected. Once again, variable time-equivalent costs would change these implications.[6]

---

[5]As migration to countries with relatively low return to skills is negatively selected, the first individuals to migrate are those with the lowest skills (alternatively, individuals with the lowest skill are those with the highest incentives to migrate). Eventually, a certain stock of immigrants lives in the foreign country. When these immigrants decide to go back to their source country, the move starts with the most skilled within the settled immigrants' stock, or the latest to arrive (alternatively, those persons that migrate from source to destination country, but with the lowest incentives.)

[6]Education obtained in a foreigner country can not be controlled and may bias the results.

The main contribution of this paper is to clarify migration self-selection mechanisms based on migration costs and inequality. It is evident that the brain drain caused by positive selection in unequal countries like Brazil, frequently found in the literature, is largely explained by migration costs and high levels of inequality.

The remainder of this paper is divided into five sections. Section 2 sketches Borjas (1987) model and includes variable time-equivalent migration costs. Section 3 presents the econometric specification. Section 4 details individual data of return migration in Brazil from 2001 to 2010. Section 5 analyzes the results of the empirical analysis, and finally section 6 presents our conclusions.

2   Theory

2.1   Self-selection

Borjas (1987), based on Roy's 1951 model, shows that migrants from countries with relatively high returns to skills and also high earning inequality are negatively selected. A strong hypothesis for this result is that time-equivalent migration costs are constants across all individuals. Therefore, low-skilled individuals living in countries with high returns to skills tend to migrate to more equalitarian countries where low-skilled individuals are not so penalized with relative low wages.

Nevertheless, if time-equivalent migration costs are higher for low-skilled individuals, those costs represent an obstacle for them. Chiquiar & Hanson (2005) consider a negative correlation between time-equivalent costs and skills,[7] which underpins the intermediate selection for migrants from Mexico to the US that they found.

Self-selection occurs twice in return migration; when the individual migrates from home country to foreign country, and when she or he migrates back. Borjas & Bratsberg (1996) indicates that if there is positive selection in the first migration, there will be negative selection in the second migration. Nevertheless, if one compares returned migrants with native non-immigrant individuals, what determines their differences is the first selection. This article compares wages of Brazilian returned migrants with those native non-immigrant individuals. This comparison is driven by the first selection, despite the second selection.[8]

Figure 1 illustrates the self-selection on return migration. Considering the "first" migration positively selected, the "second", or return migration, would be negatively selected. However, the latter, those individuals that "migrate temporarily" are on average more educated than those individuals that "stay in the source country".

Following Borjas (1987), Borjas & Bronars (1991), Borjas & Bratsberg (1996), wages w in Brazilian states i are distributed according to the mean log of

---

[7] See Chiquiar & Hanson (2005, pp.243) for examples justifying negative correlation between time-equivalent costs and skills.

[8] For instance, if the first selection is positive, individuals in the foreign country are more educated than those natives that remained in the home country. Immigrant that return, are in average more educated than non-migrant natives, even if they among the least educated among those migrants in the foreign country (even if the second selection is negative.
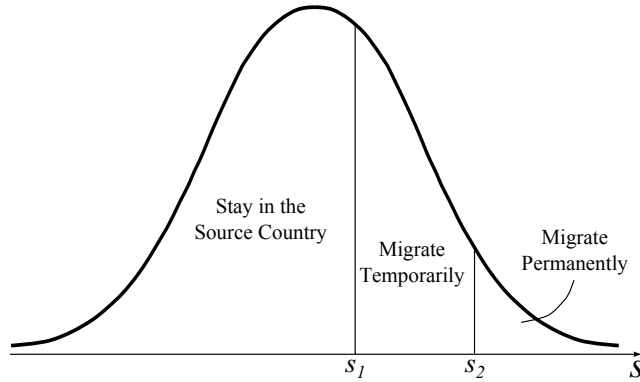
Figure 1: Self-Selection of Return Migration

wages in each state μ_i, the return to skills σ_i and according to the individual skill S:

$$\ln(w_i) = \mu_i + \sigma_i S \tag{1}$$

If the population of the source location i (henceforth Brazilian state, for the sake of clarity) migrated to destination country j, they would face the following distribution of wages:

$$\ln(w_j) = \mu_j + \sigma_j S \tag{2}$$

Considering time-equivalent migrations costs C_{ij},[9] an individual migrates from state i to country j if migration index I is positive:

$$I = \mu_j + \sigma_j S - \mu_i - \sigma_i S - C_{ij} \tag{3}$$

Self-selection is then determined by its partial derivative on skills:

$$\frac{\partial I_{ij}}{\partial S} = \sigma_j - \sigma_i - \frac{\partial C_{ij}}{\partial S} \tag{4}$$

Borjas (1987) assumes that time-equivalent migration costs are constant for all individuals, which eliminates the last term of equation (4). Chiquiar & Hanson (2005) defend a more realistic view where these costs are constant in "financial terms", but time-equivalent costs vary according to the individual income.

This hypothesis, on one hand, accentuates positive selection since migration costs represent a higher proportion of income (thus, a higher migration' barrier) for least-skilled individuals. Even when the differential of return to skill imposes a negative selection, migration costs are more severe for the least-skilled people, preventing them from migrating.

On the other hand, if one considers variable time-equivalent costs, heterogeneity in returns to skill has to be taken into account. The higher the return to education, the higher the negative correlation between migration costs and

---

[9]Considering F_{ij} as total migration costs, we considered that $\ln(w_i + F_{ij}) \approx \ln(w_i) + C_{ij}$, where C_{ij} are time-equivalent costs.

skills, thus, leading to even more positive selection. Time-equivalent costs of migration is a function of:[10]

$$C_{ij} = f(F_{ij}, S, \sigma_i) \tag{5}$$

Instead of assuming or imposing very specific functional forms for time-equivalent migration costs,[11] we would rather have skills S interact with financial costs F_{ij}. Considering equation (6), if β_3 is negative, migration costs are the highest for those individuals with no education at all, and they progressively decrease according to the skill level.[12]

$$C_{ij} = \beta_0 + \beta_1 S + \beta_2 F_{ij} + \beta_3 F_{ij} S \tag{6}$$

This specification assumes that time-equivalent migration costs decrease with skills in the same way everywhere. However, these costs may decrease differently according to the return to skills within each location. Specification (7) takes into account that migration costs represent different proportions of a given individual's income, according to the location's return to skills (σ_i).[13]

$$C_{ij} = \beta_0 + \beta_1 S + \beta_2 F_{ij} + \beta_3 F_{ij} S + \beta_4 \sigma_i F_{ij} S \tag{7}$$

The partial derivative of equation (7) on skills leads to:

$$\frac{\partial C_{ij}}{\partial S} = \beta_1 + \beta_3 F_{ij} + \beta_4 \sigma_i F_{ij} \tag{8}$$

We expect that time-equivalent costs decreases with skills, therefore: $\beta_1 + \beta_3 F_{ij} + \beta_4 \sigma_i F_{ij} < 0$ and that they decrease even more if the return to skills is higher, then: $\beta_4 < 0$.

The coefficient β_1 captures the effect of migration costs that are constant for any pair of state-country, we expect $\beta_1 < 0$.

Replacing equation (8) in equation (4) leads to:

$$\frac{\partial I_{ij}}{\partial S} = \sigma_j - \sigma_i - \beta_1 - \beta_3 F_{ij} - \beta_4 \sigma_i F_{ij} \tag{9}$$

Borjas (1987) considers only the $\sigma_j - \sigma_i$ difference to explain the self-selection of migrants. In the case of an unequal country such as Brazil, this difference is predominantly negative and implies negative selection of Brazilian emigrants. Where time-equivalent costs of migration is a function of skills and return to skills, the selection equation would include other terms ($-\beta_1 - \beta_3 F_{ij} - \beta_4 \sigma_i F_{ij}$). One can notice that even if the $\sigma_j - \sigma_i$ difference has a negative impact on selection, the other terms have a positive effect on self-selection. The econometric analysis shows that the latter prevails.

---

[10] Rate of return to skill in source locations is considered, as migration costs are incurred before the migration itself.

[11] E.g. Chiquiar & Hanson (2005, pp.243) consider the following form for time-equivalent migration costs (noted as π): $\pi = exp(\mu_p - \delta_p S)$, where μ_p and δ_p are constants.

[12] Interaction terms with higher order for skills "S" were tested and are readily available under request. Nevertheless, they do not present significant effects.

[13] The rate of return to skills is interacted only with the interaction between Financial Costs and Skills (coefficient β_4). Other terms do not depend on this rate, for instance β_3 quantifies the effect of Financial Costs on Time-Equivalent costs for null skill level. In this case, it is meaningless to consider the rate of return to skills, since skills are equal to zero.

The inequality of origin location σ_i has two opposite impacts on self-selection. Firstly, in line with Borjas, the impact is negative since it is equivalent to a high return to education rate in the origin place (preventing highly educated people from migrating). Secondly, it impacts self-selection positively because high levels of inequality raise the correlation between time-equivalent costs of migration and education. In the extreme case with very high levels of inequality, time-equivalent costs are huge for low educated people (preventing them from migrating) and tiny for highly educated people. The two opposing impacts of the inequality of origin location on selection is given by equation (10), as $\beta_4 < 0$.

$$\frac{\partial^2 I_{ij}}{\partial S \partial \sigma_i} = -1 - \beta_4 F_{ij} \tag{10}$$

Therefore, Borjas' specification would be affected by a missing variable bias. In particular, if the magnitude of β_4 is large enough, the biased coefficient of the $\sigma_j - \sigma_i$ difference would be even positive. This bias is indeed verified in the econometric analysis.

3   Econometric Specification

3.1   Estimations of Self-Selection - First-Step Regressions

We followed Borjas & Bratsberg (1996) methodology. The term of selection $\frac{\partial I_{ij}}{\partial S}$ is defined for each Brazilian state i as the average log wage difference between migrants from country j (Brazilian states' natives that have already migrated to country j and then returned to state i) and non-migrants (natives in state i that have never migrated). Thus, a first-step regression is thus estimated for each Brazilian state i:[14]

$$ln(w_i^k) = \sum_{j}^{n} \delta_{ij} D_{ij}^k + \zeta_i \mathbf{X}_i^k \tag{11}$$

where D_{ij} is a dummy indicating if individual k is a migrant returned from country j to Brazilian state i. The coefficient δ_{ij} gives, therefore, the migrant's log wage relative to natives. A vector of demographic controls $\mathbf{X}_i^{k}$[15] is included for the regressions of adjusted wages.

These first-step regressions are estimated for two cohorts: the first one for those migrants that returned between 2001 and 2005, and the second one for those migrants that returned between 2006 and 2010. This procedure controls for wages' differences related to the time of migration. Returnees that migrated before 2005 had more time to reintegrate to their homeland than those that migrated after 2005. Moreover, Brazil's economical situation changed over this period,[16] and so did migration flows.

---

[14]Index i is not meaningful in regressions represented by equation (11), as they are run for each state i; however, they are used because coefficients δ_{ij} are going to be pooled from each regression, in the second step.

[15]The controls are age, age squared, education level, marital status, health and metropolitan residence.

[16]In average, the Brazilian GDP grew from 2.80% in the first period to 4.49% in the second period. Source: World Bank Development Indicators.

Figure 2 shows the evolution of these flows between 2001 and 2010. The average over the first period from 2001 to 2005 is about 7000 returnees while the average over the second period is about 24000 returnees. According to Borjas (1996, pp.173), larger amounts of migrants lead to a "more diluted quality of the typical immigrant in a flow that is positively selected".[17] Therefore, selection is sharper in the first period than that in the second period, increasing wage heterogeneity.
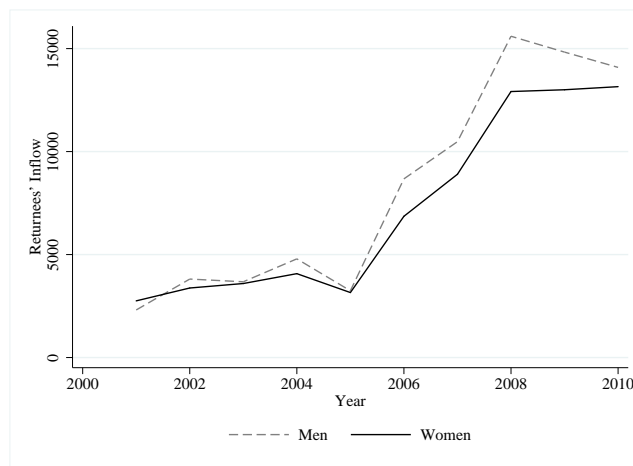
This figure considers 24-64 year-old individuals that were born is the state where they answered the 2010 Brazilian census. The total is 153298 individuals.

Figure 2: Inflows of Returnees by Year

3.2   Proxies of Migration Costs and Return to Schooling

Migration costs are represented by three different variables: the distance between the Brazilian state and the country of migration (for diminishing marginal effects, distance is also considered squared), common language, and contiguity, which are dummy variables that equal to one if the Portuguese is spoken in the country of migration, and if this country shares a border with the Brazilian state, respectively.

$$F_{ij} = \lambda_1 Distance_{ij} + \lambda_2 Distance_{ij}^2 + \lambda_3 Language + \lambda_4 Contiguity \qquad (12)$$

We expect that these costs are an increase function of distance: $\lambda_1 > 0$, $\lambda_2 < 0$, and that it decreases with common language and contiguity: $\lambda_3 < 0$ and $\lambda_4 < 0$

We followed Borjas & Bratsberg (1996) and proxy for the rates of return to skills using a measure of income inequality, the (quintile) ratio, henceforth q:[18]

---

[17]Recalling that Borjas (1996) does not consider variable time-equivalent costs and then positive selection is due only to differences in the rates of return to skill.

[18]We use the ratio of the average income of the 20% richest individuals over the average income of the 20% poorest individual. Doing so, this ratio is consistent with data used for the inequality of migration countries, from UNDP (2011).

$$q_i = \theta \sigma_i \qquad (13)$$

3.3 Determinants of Self-Selection - Second Step Estimable Equations

The estimable equation of (9) is:

$$\delta_{ij} = \alpha_0 + \alpha_1(q_j - q_i) + \alpha_2 Dist_{ij} + \alpha_3 Dist_{ij}^2 + \alpha_4 Lang_{ij} + \alpha_5 Cont_{ij}$$
$$+ \alpha_6 Dist_{ij}\sigma_i + \alpha_7 Dist_{ij}^2\sigma_i + \alpha_8 Lang_{ij}\sigma_i + \alpha_9 Cont_{ij}\sigma_i \quad (14)$$

Where: $\alpha_0 = -\beta_1$

$$\alpha_0 = -\beta_1$$

The model predicts $\alpha_0 > 0$

$$\alpha_1 = \theta$$

$$\alpha_1 > 0$$

$$\alpha_2 = -\beta_3\lambda_1 \text{ and } \alpha_6 = -\beta_4\lambda_1 >$$

The model predicts $\alpha_2 Dist_{ij} + \alpha_6 Dist_{ij}\sigma_i > 0$ and $\alpha_6 > 0$

$$\alpha_3 = -\beta_3\lambda_2 \text{ and } \alpha_7 = -\beta_4\lambda_2$$

The model predicts $\alpha_3 Dist_{ij}^2 + \alpha_7 Dist_{ij}^2\sigma_i < 0$ and $\alpha_7 < 0$

$$\alpha_4 = -\beta_3\lambda_3 \text{ and } \alpha_8 = -\beta_4\lambda_3$$

The model predicts $\alpha_4 Lang_{ij} + \alpha_8 Lang_{ij}\sigma_i < 0$ and $\alpha_8 < 0$

$$\alpha_5 = -\beta_3\lambda_4 \text{ and } \alpha_9 = -\beta_4\lambda_4$$

The model predicts $\alpha_5 Cont_{ij} + \alpha_9 Cont_{ij}\sigma_i < 0$ and $\alpha_9 < 0$

The two opposing effects of the inequality of origin location on selection given by equation 10 are estimated by equation (15) below. Besides the role of inequality *per se*, given by $-\alpha_1$, one can notice that its interaction with factors that reduce migration costs have a positive impact on self-selection, as α_8 and α_9 are expected to be negative. Conversely, the interaction of inequality with the factor that raises migration costs has a positive impact on self-selection, as $\alpha_6 Dist_{ij} + \alpha_7 Dist_{ij}^2$ is expected to be positive

$$-\alpha_1 + \alpha_6 Dist_{ij} + \alpha_7 Dist_{ij}^2 + \alpha_8 Lang_{ij} + \alpha_9 Cont_{ij} \qquad (15)$$

4   Data

The Brazilian 2010 census counted 204,374 returnees that came back to Brazil from 2001 to 2010. In this study, we considered only individuals that returned to the state where they were born,[19] doing so prevents counting people that are not returning, but migrating again. For example, we do not count as a returnee the individual that migrated from Bahia to the US and then migrated again to São Paulo. The self-selection of this individual is related to individuals in Bahia, not in São Paulo. As regressions are estimated by Brazilian states, we compared only individuals born within each state. Table 1 details these individuals by country of migration.

Table 1: Migration Country of Returnee Migrants

Country	Returnees	Perc.	Country	Returnees	Perc.
Africa	1870	1.22%	Asia	29369	19.16%
Angola	1077	0.70%	Japan	27179	17.73%
Mozambique	167	0.11%	China	570	0.37%
Namibia	102	0.07%	Un. Arab Emirates	280	0.18%
Guinea-Bissau	92	0.06%	Lebanon	271	0.18%
Equatorial Guinea	79	0.05%	Israel	233	0.15%
Others	353	0.23%	Others	836	0.55%
America	58739	38.32%	Europe	58922	38.44%
United States	36058	23.52%	Portugal	14617	9.54%
Paraguay	7561	4.93%	United Kingdom	11579	7.55%
Bolivia	3095	2.02%	Spain	9817	6.40%
Argentina	2947	1.92%	Italy	8056	5.26%
Canada	2378	1.55%	France	4278	2.79%
Uruguay	1273	0.83%	Germany	3434	2.24%
Chile	1034	0.67%	Switzerland	1912	1.25%
Mexico	882	0.58%	Ireland	1641	1.07%
Peru	689	0.45%	Netherlands	937	0.61%
French Guiana	525	0.34%	Belgium and Lux.	842	0.55%
Others	2297	1.50%	Austria	501	0.33%
			Others	1308	0.85%
Pacific	2907	1.90%			
Australia	2069	1.35%			
New Zealand	838	0.55%			

Notes: 1491 (1%) individuals out of this table answered "unknown" for migration country.This table considers 24-64 year-old individuals that were born in the state where they answered the 2010 Brazilian census. The total is 153298 individuals.

Rich countries are the top destination places for returnee migrants. The U.S., Japan, Portugal, the U.K., Spain and Italy had been the choice for 70% of returnees.[20]

Table 2 shows the distribution of return migrants in Brazilian states.[21] Returnees are over-represented in the most developed Brazilian regions: the Center-West has 9.54% of migrants, while the population's share is 7.37%. The South has 24.89% of the returnees and 14.36% of the population. The Southeast has 53.81% of the returnees and 42.13% of the population.

---

[19]Which reduced the volume to 153298 individuals.

[20]For comparison, foreign migrants have more dispersed origins, but proximity seems to count more. Bolivia, Paraguay, Argentina, Peru and Uruguay were the origin of 39% of foreign immigrants in Brazil, from 2001 to 2010.

[21]See figure 5 in section Appendix A to locate each Brazilian region in a map.

Table 2: State of Origin of Returnee Migrants

State	Returnees	Perc.	State	Returnees	Perc.
North	4591	2,99%	Northeast 13440	8,77%	
Pará	2008	1,31%	Bahia	5119	3,34%
Rondônia	890	0,58%	Pernambuco	2983	1,95%
Amazonas	760	0,50%	Ceará	2224	1,45%
Amapá	341	0,22%	Paraíba	961	0,63%
Tocantins	311	0,20%	Maranhão	883	0,58%
Acre	141	0,09%	Rio Grande do Norte	470	0,31%
Roraima	140	0,09%	Alagoas	372	0,24%
Piauí	295	0,19%			
Center-West	14626	9,54%	Sergipe	133	0,09%
Goiás	8543	5,57%			
Mato Grosso do Sul	3356	2,19%	Southeast	82491	53,81%
Distrito Federal	1631	1,06%	São Paulo - Metro	23142	15,10%
Mato Grosso	1096	0,71%	Minas Gerais	22290	14,54%
São Paulo - Rest	21913	14,29%			
South	38150	24,89%	Rio de Janeiro	11600	7,57%
Paraná	22963	14,98%	Espírito Santo	3546	2,31%
Rio Grande do Sul	8837	5,76%			
Santa Catarina	6350	4,14%			

Notes: This table considers 24-64 year-old individuals that were born in the state where they answered the 2010 Brazilian census. The total is 153298 individuals.

Returnees are under-represented in the least developed regions: while the North has 2.99% of returnees, it accounts for 8.33% of the Brazilian population. The Northeast has 8.77% of returnees and 27.82% of the population.

Table 3 reports data on education level and gender of non-migrant natives and returnees. Individuals that migrated are over-represented in categories of higher education, as Tertiary Education and Secondary Education. Conversely, they are under-represented in the other lower education categories: Primary Education and No Education. The difference in the No Education category is striking: while 43.3% of natives are in this category, this is the case for only 16% of returnees. All in all, statistics on individuals' education suggest positive selection. In regards to gender, male returnees are over-represented compared to those non-migrant male individuals. The percentage of men is 48% in the native (non-immigrant) population, while this percentage is 53% in the returnee sample.

The measure of inequality used in the econometric analysis is based on the quintile ratio.[22] Data on countries' inequality is obtained from the United Nations Development Programme - UNDP (2011).[23] Brazilian states' quintile ratio is calculated using census 2010 data. The average differences of countries' quintile ratio and Brazilian states' quintile ratio are mostly negative. In 90%[24] of the cases that difference is negative, the average is −9.60.

Figure 3 shows the k-density of the wages of non-migrants and returnees. Some locations[25] were chosen due to their importance or specificity. Figure

---

[22]The quintile ratio is the proportion between the average income of the 20% richest individuals over the average income of the 20% poorest individuals

[23]Data is available at `http://www.undp.org/content/undp/en/home/librarypage/hdr/human_developmentreport2011.html`

[24]Considering 638 pairs of Country-Brazilian State in the largest samples of estimations, those in table 4.

[25]See figure 5 in section Appendix A to locate each Brazilian state in a map.

Table 3: Education Level and Gender

| | Returnees | | | | | |
| | Men | | Women | | Total | |
	Individuals	Perc.	Individuals	Perc.	Individuals	Perc.
No Education	14445	17,70%	10066	14,00%	24511	16,00%
Primary	11523	14,10%	8155	11,40%	19678	12,80%
Secondary	30079	36,90%	25624	35,70%	55703	36,30%
Tertiary	25276	31,00%	27743	38,60%	53019	34,60%
Unknown	190	0,20%	197	0,30%	387	0,30%
	Natives (non-migrants)					
	Men		Women		Total	
	Ind. (000's)	Perc.	Ind. (000's)	Perc.	Ind. (000's)	Perc.
No Education	16900	45,70%	16300	41,10%	33200	43,30%
Primary	5995	16,20%	6138	15,50%	12133	15,80%
Secondary	10200	27,60%	11500	29,00%	21700	28,30%
Tertiary	3793	10,30%	5548	14,00%	9341	12,20%
Unknown	112	0,30%	135	0,30%	247	0,30%

Notes: This table considers 24-64 year-old individuals that were born in the state
where they answered the 2010 Brazilian census. The total is 153298 individuals.

3 (a) refers to the metropolitan area of São Paulo[26], the biggest state in the
country, and figure 3 (b) refers to Rio de Janeiro, the second most important
economy in Brazil. Figure 3 (c) shows the wage distribution for Distrito Fed-
eral, the location with the highest average wage and the most unequal loca-
tion, while figure 3 (d) refers to Piauí, the state with the lowest average wage.
Figure 3 (e) shows data for Santa Catarina, the state with the highest equal-
ity, and figure 3 (f) shows the distribution of wages in Pernambuco, the state
presenting the highest average relative wages in the econometric analysis.[27]

One can notice in these figures that returnees are distributed at higher
wages than non-migrants natives, which suggests positive selection. This is
more evident for relatively poor states like Piauí and Pernambuco, and un-
equal locations like Distrito Federal than for large areas like São Paulo and
Rio de Janeiro[28]. Remarkably, the most egalitarian state in Brazil, Santa Cata-
rina, presents similar distribution of wages, when returnees and non-migrants
are compared.

5 Results

5.1 First-Step Estimations

Equation (11) is regressed twice. First without demographic controls (vector
$\mathbf{X}_i^k$) estimating unadjusted relative wages, and then with demographic con-
trols estimating adjusted relative wages. While the former indicates selection
on observed and unobserved characteristics of migrants, the latter indicates
selection only on unobserved characteristics. We considered hourly wages for

---

[26]The state of São Paulo is broken down into two regions: the metropolitan area and the rest.

[27]The econometric section shows that returnee migrants residing in Pernambuco have Brazil's
highest average wage, relative to natives. Figure 6 illustrates this finding.

[28]These regions are precisely the São Paulo metropolitan area and the state of Rio de Janeiro.
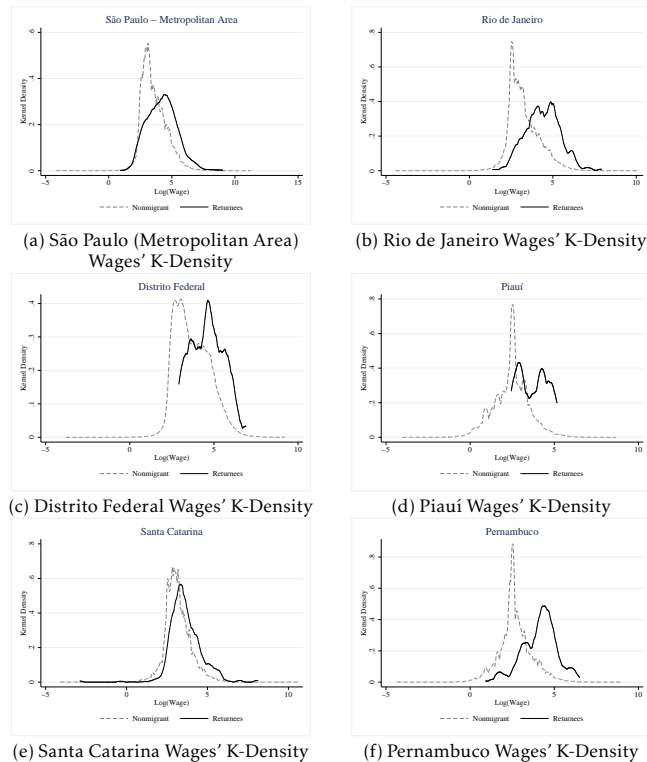
(a) São Paulo (Metropolitan Area) Wages' K-Density

(b) Rio de Janeiro Wages' K-Density

(c) Distrito Federal Wages' K-Density

(d) Piauí Wages' K-Density

(e) Santa Catarina Wages' K-Density

(f) Pernambuco Wages' K-Density

Figure 3: Wages K-Density

all estimations.[29]

Two sets of first-step regressions are estimated. The first set regresses equation (11) 28 times for *unadjusted* wages: for each Brazilian state and Distrito Federal.[30] Each regression estimates the relative unadjusted wage (δ_{ij}) for an average of 11.39 countries, which comprises 638 observations (11.39 countries × 28 states × 2 cohorts = 638) for the second step. Figure 4 shows the distribution of these relative wages, and table 4 reports the results from second-step estimations using them as a dependent variable.

The second set of regressions estimates equation (11) 28 times for *adjusted* wages, resulting in 638 values of adjusted δ_{ij}. Figure 5 shows the distribution of these adjusted relative wages, and table 5 reports the results from second-step estimations where they are the dependent variable.

We restricted the sample to male workers between 24 and 64 years who were born in the state they reside (where they answered the census). The gender restriction prevents any issue related to discrimination and mitigates bias due to tied migration.[31] Estimations are made twice, for two cohorts: one for migrants that came back to Brazil before 2005 and another for those who arrived after 2006.

---

[29]Hourly wages are obtained dividing the individual monthly salary earned in the principal job by the quantity of hours worked.

[30]Brazil has 26 states, but the very large state of São Paulo is broken down into two regions: the metropolitan area and the rest.

[31]Tied migration occurs when the decision of migrating considers the wage difference of the family, and not the individual wage difference (Borjas and Bronars, 1991).

We chose to regress at state level for two reasons. Firstly, we eliminated geographical heterogeneities that could impact wages or migration choice. In the sample studied, the average monthly wage in Distrito Federal is BRL2,545 (the highest in the country), while the average wage in Piauí is BRL545.00 (the lowest in the country). Amenities also differ between states. For instance, average annual temperatures vary from 18.1° (in Santa Catarina) to 27.4° (in Roraima).[32]

Secondly, we considered a returnee an individual that was born in a given Brazilian state and returned to the same state, even if their current city of residence is not their birthplace. By doing so, we did not discard from our analysis the individual that came back to Brazil and does not live exactly in the same city of birth, but on another city in the same state. In addition, we did not consider the sequential migrant.

Estimations of unadjusted δ_{ij} indicate that selection on observed and unobserved skills is primarily positive. Indeed, return migrants earn higher wages than nonmigrants in 80% of the observations.[33] Figure 4 shows the distribution of log-wage differences, the average is 0.79.[34]
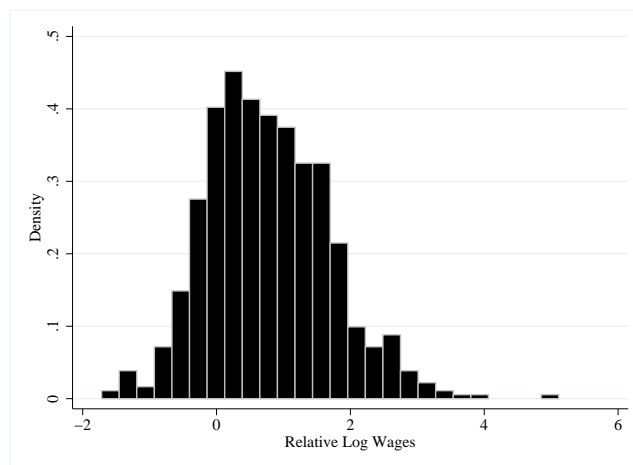

Figure 4: Histogram of Relative Log Wages

Estimations of adjusted relative wages indicate that selection on unobserved is also primarily positive. Return migrants earn higher adjusted wages than non-migrants in 68% of the observations. Figure 5 shows the distribution of wage differences. Relative adjusted wages present smaller values than those of unadjusted wages. As observable characteristics such as education and age are controlled, differences in migrants' wages relative to natives' are reduced. The average wage is 0.33.

---

[32]Source: http://www.inmet.gov.br/portal/index.php?r=clima/normaisClimatologicas

[33]Based on the table 4 sample with 638 observations, where each observation is a triplet state-country-cohort.

[34]This finding does not verify the predictions of Assunção & Carvalho (2013), who expect emigration of middle-class individuals from a country like Brazil, resulting in more equality on their return to the country. Instead we find that returnees are positively selected. The Gini index for the entire population is 0.52520, while the Gini index for non-migrant natives alone is
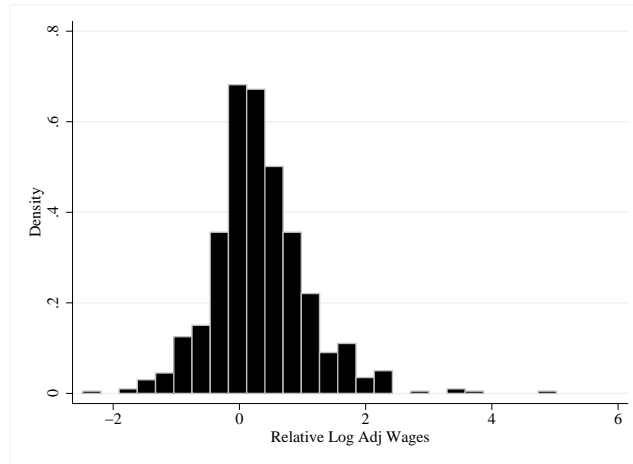
Figure 5: Histogram of Relative Log Adjusted Wages

Figure 6 shows the unadjusted relative average wage by Brazilian state[35]. The Northeast presents the highest wages, where returnees earn from 54% (in Sergipe) to 174% (in Alagoas) higher wages than natives. This region seems to receive the biggest benefits from return migration. Nevertheless, Table 2 indicates that returnees are under-represented in this region. One possible explanation is that positive self-selection is stricter in this region and only the very high-skilled individuals succeed in migrating. It is worth to note that these values are much larger than those found in Borjas & Bronars (1991), which may reflect the higher inequality in Brazil compared to the US.
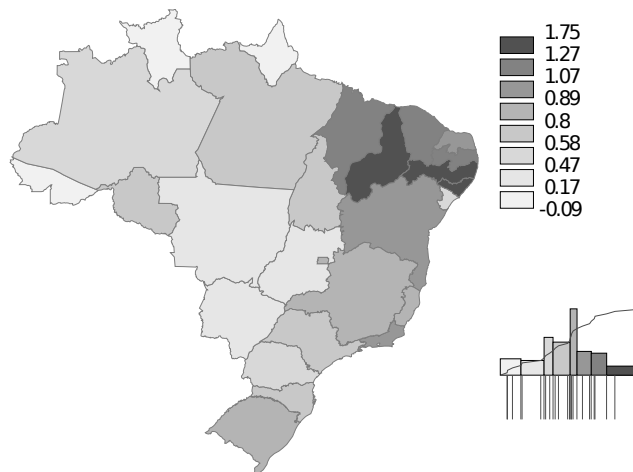

Figure 6: Average (by State) Immigrants' Relative Log Wages

Other Brazilian regions also receive a human capital increase, but to a

---

0.52455. The impact of return migration on inequality, if any, is positive.

[35]See figure 5 in section Appendix A to locate each state Brazilian in a map.

lesser degree. On average, return migrants earn from −9% (in Roraima) to 104% (in Rio de Janeiro) across states of these regions.

5.2 Second-Step Estimations

As the dependent variable is a coefficient estimated in the first step, residuals are heteroscedas-tic.In order to correct this, second-step estimations are obtained through a generalized least squares estimator.[36]

Unadjusted Wages

Table 4 shows the results for the estimable second-step equation (14) for unadjusted wages. Regression (1) estimates the returnee's relative wage on the inequalities' difference $q_j - q_i$ that indicates the difference of return to skills between country of migration and Brazilian state. This regression would test Borjas' specification (1987, 1991, 1996), where time-equivalent costs do not depend on skills. However, contrarily to the model's prediction, the coefficient of $q_j - q_i$ is not positive and highly significant. As expected, the omission of migration time-equivalent costs may lead to a missing variable bias.

We estimated different specifications by taking into account migrations costs. Firstly, we considered in columns (2) and (3) that time-equivalent costs decrease with skills uniformly, in spite of differences in rates of return to skills across Brazilian states. Secondly, we considered that differences in rates of return to skills may imply differences to the extent to which time-equivalent costs decrease with skills: column (4) reports results for the complete equation (14).

Regression (2) estimates a positive effect of distance on migration selection, as predicted by the theoretical model.[37] Brazilians that migrated to far-off countries had to pay a high cost of migration. Low-skilled individuals, regardless of the occasional possibility of higher wages, were less likely to be able to afford such costs as these amounts represented large proportions of these individuals' income.

Nonetheless, it is interesting to note that, although this result confirms the theory, the effect of distance on selection has an unexpected inverted U-shape. The maximum positive effect of distance is 5438 km (which roughly corresponds to Europe). The effects of distance decrease until the maximum value reached by distance in the sample is 11,702 km.

Column (3) introduces other variables that are correlated to migration costs. Common language is a dummy variable that equals to one if Portuguese is one of the languages spoken in the migration country. Contiguity is a dummy variable that equals to one if the Brazilian state and the migration country have a common border. Both variables are negatively correlated with migration costs, which reduces the amount of expenses an immigrant has to pay. Therefore, these variables indicate a reduction in migration costs, allowing low-skilled individuals to migrate and contributes to a negative selection. Coefficients of both variables have, indeed, a negative sign, presenting a significant coefficient. The coefficient of $q_j - q_i$ still remains negative in regressions (2) and (3).

---

[36]See Borjas (1987) for details.

[37]The quadratic function of distance has roots 0 and 10.88 (thousands km). The maximum value reached by distance is 11,702 thousands km.

Table 4: Results: Unadjusted wages

Dependent Variable: Δ of Ln	(1)	(2)	(3)	(4)
$q_j - q_i$	−0.017*** (−0.004)	−0.014*** (−0.005)	−0.012** (−0.007)	0.013* (−0.007)
Distance meters $*10^3$	α_2	0.174*** (−0,045)	0.128** (−0,051)	−0.165** (−0,083)
Distance^2 meters $*10^2 *10^6$	α_3	−0.016*** (−0.004)	−0.013** (−0.004)	0.021** (−0.01)
Common Language	α_4		−0.237** (−0.12)	−0.969** (−0.44)
Contiguity	α_5		−0.557*** (−0.209)	−0.542 (−0.542)
Distance $*q_i$	α_6			0.022** (−0.005)
Distance$^2 *q_i$	α_7			−0.002*** (−0,001)
Common Language $*q_i$	α_8			0.03 (−0.022)
Contiguity $*q_i$	α_9			0.004 (−0.027)
Constant	0.668*** (−0,071)	0.332*** (−0,127)	0.543*** (−0,147)	0.439*** (−0,146)
Observations	638	638	638	638
R-squared	0	0.02	0.05	0.11

In column (4) we allow for differences in the return to skill and establish differences in the correlation between time-equivalent costs and skills. We took all costs variables considered in column (3). The difference in return to skills between the Brazilian state and the country, q_j-q_i, presents the expected positive sign predicted by the theory (which confirms Roy's model) and it is significant at a 10% significance level.

Interestingly, one can note that the shape of the distance's effect on skill changes according to the location's inequality, but it is mostly strictly increasing as the range of q_i varies from 11 to 39.[38] This result helps to elucidate the apparent U-shape impact of distance on selection found in the previous regression. We can see that, once the heterogeneity of inequality across locations is considered, the impact of distance on migration selection is predominantly strictly increasing.

The theoretical predictions regarding migration costs given by equation 10 and estimated according to equation 15, are completely found only for distance in regression (4). The positive impact of distance on self-selection is accentuated by the inequality level of the origin place. Regarding other variables, only language presents a significant and negative coefficient in this specification, although contiguity presents the expected negative sign.

Adjusted Wages

Table 5 reports results for adjusted wages in equation (14). A vector of demographic characteristics is included in the regressions of the first step (equation

---

[38]The variable q_i varies from 11 to 39, with a mean equals to 18. At the lowest limit of q_i, the impact of distance on selection is described by $0.077Dist-0.001Dist^2$, which is strictly increasing in the interval. At the mean (18), the impact is $0.23Dist - 0.015Dist^2$, which is increasing until distance$= 7.67$ (thousands km), and includes 89.34% of the observations. Finally, at the highest limit of q_i, i.e. 39, we have 73.67% of the observations included in the interval where the impact of distance on migration selection is increasing.

Table 5: Results: Adjusted wages

Dependent Variable: Δ of Ln	(1)	(2)	(3)	(4)
$q_j - q_i$	−0.002	0	0.001	0.012**
	−0.004	−0.004	−0.004	−0.006
Distance	α_2	0.097***	0.081*	−0.023
meters $*10^3$		−0.037	−0.042	−0.07
Distance^2	α_3	−0.009***	−0.008**	0.004
meters $*10^2*10^6$		−0.003	−0.003	−0.008
Common Language	α_4		−0.082	−0.746**
			−0.1	−0.373
Contiguity	α_5		−0.197	−0.332
			−0.175	−0.46
Distance $*q_i$	α_6			0.008**
				−0.004
Distance$^2 *q_i$	α_7			−0.001*
				0
Common Language $*q_i$	α_8			0.031*
				−0.019
Contiguity $*q_i$	α_9			0.009
				−0.023
Constant	0.317***	0.118	0.193	0.149
	−0.059	−0.105	−0.123	−0.124
Observations	638	638	638	638
R-squared	0	0.01	0.02	0.03

Notes: Standard errors in brackets *** p<0.01, ** p<0.05, * p<0.1. All regressions include a dummy for the 2001-2005 cohort. The regressions use a generalized least squares estimator to correct the heteroscedasticity in the dependent variable.

11). In line with Borjas & Bratsberg (1996), these characteristics are: age, age squared, education level, marital status, health and metropolitan residence.

Regression (1) estimates the returnee's relative adjusted wage on the inequalities' difference $q_j - q_i$ that indicates the difference of return to skills between country of migration and Brazilian state. Unlike the results for unadjusted wages, $q_j - q_i$ does not present a negative and significant coefficient in any specification.

Columns (2) and (3) introduce variables that capture migration costs, namely, distance, common language and contiguity. Only distance presents a positive and significant impact on selection.[39]

Similarly to estimations of unadjusted wages, regression (4) also confirms theoretical predictions for non-observable characteristics. Inequality differences have a positive and significant coefficient, indicating that individuals migrate to the place where their (non-observable) characteristics are valued the most. Concerning time-equivalent costs, they are significant only when interacting with the inequality ratio q_i.

Theoretical predictions given by equation 10 and estimated according to equation 15 are completely found only for distance, in column (4). It is surprising that Common Language presents a positive and significant coefficient, because the theoretical prediction is a positive coefficient.

Self-selection is, therefore, driven by migration costs and return to skills, both observable and non-observable. More productive Brazilians migrate to the countries that value workers' high productivity the most.

---

[39]The quadratic function of distance has roots 0 and 10.13 (thousands km)in regression (3). The maximum value reached by distance is 11.702 thousands km.

6 Conclusion

This paper analyzes the self-selection of returnee migrants in Brazil from 2001 to 2010. As we compare the wages of returnees to those of non-migrant natives, the observed selection occurred when these individuals migrated from Brazil to a foreign country. Primarily positive selection is estimated using individual data from the 2010 census.

This selection is explained by time-equivalent migration costs and inequality. Time-equivalent costs are higher for low-skilled individuals, as they spend a longer period of time than high skilled individuals to amass the same income. This period is still longer if the inequality is high. Alternatively, one can think that low-skilled individuals face more difficulties with bureaucratic requirements than high-skilled individuals. Yet, credit constraints may raise migration costs for low-income individuals.

We find empiric evidence supporting the role of the correlation between migration costs and skills determining positive selection. This selection is accentuated in Brazilian locations with high inequality and high rates of return to skill, as the correlation is higher.

The empiric analysis also confirms the predictions of Borjas (1987). Low-skilled individuals have incentives to emigrate from Brazilian states with relative (to destination country) high inequality and high returns to skills, implying a negative impact on selection. This impact is only verified once the interaction between migration costs and inequality is controlled. In spite of this partial negative effect on selection, migrations costs mechanisms prevail, resulting in positive selection.

In Brazil, locations with high levels of inequality are the ones that suffer the most with the brain drain. Consequently, they are the most benefited with the increase of human capital when those emigrants returns.

Acknowledgements

Bibliography

Assunção, J. & Carvalho, F. (2013), 'Financial constraints, endogenous educational choices and self-selection of migrants', *Brazilian Review of Econometrics* **33**(2), 99–122.

Belot, M. V. K. & Hatton, T. J. (2012), 'Immigrant selection in the oecd', *The Scandinavian Journal of Economics* **114**(4), 1105–1128.

Borjas, G. J. (1987), 'Self-selection and the earnings of immigrants', *The American Economic Review* **77**(4), 531–553.

Borjas, G. J. & Bratsberg, B. (1996), 'Who leaves? the outmigration of the foreign-born', *The Review of Economics and Statistics* **78**(1), 165–176.

Borjas, G. J. & Bronars, S. G. (1991), 'Immigration and the Family', *Journal of Labor Economics* **9**(2), 123–48.

Chiquiar, D. & Hanson, G. H. (2005), 'International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States', *Journal of Political Economy* **113**(2), 239–281.

Chiswick, B. R. (1999), 'Are immigrants favorably self-selected?', *The American Economic Review* **89**(2), 181–185.

Docquier, F. & Rapoport, H. (2012), 'Globalization, brain drain, and development', *Journal of Economic Literature* **3**(50), 681–730.

Facchini, G. & Mayda, A. M. (2009), 'Does the welfare state affect individual attitudes toward immigrants? evidence across countries', *Review of Economics and Statistics* **2**(91), 295–314.

Mankiw, N. G., Romer, D. & Weil, D. N. (1992), 'A contribution to the empirics of economic growth', *The Quarterly Journal of Economics* **107**(2), 407–437.

Müller, T. & Tai, S. H. T. (2016), Individual attitudes towards migration: reconciling opposing views, Working paper series WPS 16-02-1, University of Geneva.
URL: *Available at: http://www.unige.ch/ses/dsec/repec/files/16021.pdf*

Romer, P. M. (1990), 'Endogenous Technological Change', *Journal of Political Economy* **98**(5), S71–102.

Roy, A. D. (1951), 'Some thoughts on the distribution of earnings', *Oxford Economic Papers* **3**(2), 135–146.

UNDP (2011), Human develpment report 2011, susteinability and equity: a better future for all, Technical report, United Nations Development Programme, New York.

Appendix A

Figure A.1: Brazilian States and Regions