UNITED NATIONS
UNIVERSITY
UNU-WIDER

WIDER Working Paper 2019/113

Migrants leaving mega-cities

Where they move and why prices matter

Eva-Maria Egger*

December 2019

Abstract: Traditional economic models predict rural to urban migration during the structural transformation of an economy. In middle-income countries, it is less clear which direction of migration to expect. In this paper I show that in Brazil as many people move out of as into metropolitan cities, and they mostly move to mid-sized towns. I estimate the determinants of out-migrants' destination choice, accounting for differences in earnings, living costs, and amenities, and I test whether the migrants gain economically by accepting lower wages but enjoying lower living costs. The findings suggest that the destination choice of out-migrants minimizes the costs of moving. On average, city-leavers realize higher real wages, including low-skilled migrants who would lose out in nominal terms. The paper thus provides evidence on economic incentives to leave big cities in a middle-income country.

Keywords: Brazil, internal migration, prices, secondary towns

JEL classification: J61, R23, C35

Acknowledgements: I am indebted to Andy McKay and Amalavoyal Chari for their constant support throughout this work. My thanks also go to Corrado Giulietti, Panu Pelkonen, L. Alan Winters, Forhad Shilpi, John Gibson, Joachim De Weerdt, Luc Christiaensen, and Ravi Kanbur, as well as participants at the Annual Meeting of the Latin American and Caribbean Economics Association 2014 and the XII ENABER and Second Ibero-American Meeting on Regional Development 2014 for helpful comments. I am grateful for support in the form of a doctoral grant by the Economic and Social Research Council (ESRC) and the research programme consortium Migrating out of Poverty (UK Department for International Development). Any errors are mine alone.

---

* UNU-WIDER, Maputo, Mozambique; corresponding author: egger@wider.unu.edu

This study has been prepared within the UNU-WIDER project on Social mobility in the global south – concepts, measures, and determinants.

Information and requests: publications@wider.unu.edu

ISSN 1798-7237 ISBN 978-92-9256-749-1

https://doi.org/10.35188/UNU-WIDER/2019/749-1

Typescript prepared by Gary Smith.

The United Nations University World Institute for Development Economics Research provides economic analysis and policy advice with the aim of promoting sustainable and equitable development. The Institute began operations in 1985 in Helsinki, Finland, as the first research and training centre of the United Nations University. Today it is a unique blend of think tank, research institute, and UN agency—providing a range of services from policy advice to governments as well as freely available original research.

The Institute is funded through income from an endowment fund with additional contributions to its work programme from Finland, Sweden, and the United Kingdom as well as earmarked contributions for specific projects from a variety of donors.

Katajanokanlaituri 6 B, 00160 Helsinki, Finland

1 Introduction

Migration of workers from rural to urban areas forms an integral part of the structural change of a country's economy (Harris and Todaro 1970). With higher income levels, this migration flow is expected to slow down or even reverse, as seen in the 1970s and 1980s in the USA and Europe as the so-called 'population turnaround' (Cochrane and Vining 1988). However, there is little evidence on this reversal of migration patterns in general and the determinants of out-migration from cities in countries that are transitioning from low to middle and higher income. At the individual level, the migration into urban areas is driven by earnings differentials, and cities offer higher returns than rural areas (Sjaastad 1962). Contrary to these traditional models of rural–urban migration, this paper documents that as many migrants move out of cities as into them in Brazil. Their destinations are medium-sized towns. I investigate which local characteristics are preferred by the metropolitan out-migrants and whether their destination choice can still be explained by a gain in earnings.

Census survey data from Brazil show that around 20 per cent of first-time internal migrants[1] moved out of metropolitan cities between 2009 and 2010, which equals the share of migrants moving into the metropolises in the same period. The majority of out-migrants (around 78 per cent) move to live and work in medium-sized destinations, [2] not small and rural locations. It appears that high- and low-educated out-migrants are equally likely to move, which gives rise to the questions: What drives these workers out of the cities? And where do they choose to live?

Urban areas attract workers due to job opportunities, high wages, and better services. Yet, with waves of urbanization in developing countries, large cities face many problems associated with overcrowding, such as informal housing, congested infrastructure, and unemployment. City growth increases demand for housing and amenities, whose supply is rather inelastic. Higher living costs put pressure on workers whose budget share for these goods is relatively high (Giannetti 2003). These factors could give rise to migration out of cities. While wages and amenities are expected to be higher in cities than in smaller towns, living costs are also higher. The migration choice therefore becomes a balancing act between these factors across a large set of possible destinations.

I estimate how these factors affect the individual destination choice conditional on migration with a conditional logit model as in Fafchamps and Shilpi (2013). The focus lies on the established determinants of migration: expected wages based on individual characteristics, moving and living costs, local factors such as crime rates, and the quality of public education and health service provision. Coarsened exact matching (CEM) is applied to control for selection bias in the prediction of expected wages. Other papers investigate the sorting decision from an individual choice perspective. These studies found that workers sort themselves to destinations by balancing the highest return to their skills and the chance to find employment against local living costs and the presence of local amenities according to their individual preferences (Borjas 1987; Borjas et al. 1992; Dahl 2002; Grogger and Hanson 2011; Ham et al. 2011; Moretti 2011), but only a few studies use data from transitioning countries (Aguayo-Téllez et al. 2010; Aroca Gonzalez and Maloney 2005; Fafchamps and Shilpi 2013; Lokshin et al. 2007).

Following this analysis, I further investigate how the realized instead of expected wages reflect a gain or loss resulting from moving out of metropolitan cities. I use counterfactual wages of migrants to compute the return to out-migration in nominal and real terms. A few studies analyse the counterfactual situation of households had their member not migrated, but not of the migrants themselves (Adams

[1] These migrants leave their place of birth and are not return migrants.

[2] The median population size of the administrative unit, a *microregião*, is 173,453 inhabitants. I classify a medium-sized *microregião* as one that has between 170,000 and one million inhabitants, and above one million as metropolitan city.

2006; Adams and Cuecuecha 2013; Barham and Boucher 1998; Brown and Jimenez 2008; Lokshin et al. 2007; Rodriguez 1998; Tunalı 2000).

I find that the metropolitan out-migrants prefer smaller cities where costs of living are lower. The counterfactual analysis confirms that the return to metropolitan out-migration is positive in real wages. The difference in living costs between metropolitan origins and non-metropolitan destinations appears to exceed migration costs. This result is strongest for low-skilled workers, who would normally experience a decline in nominal wages from leaving cities. Furthermore, out-migrants prefer towns closer to their origin and within their own state of birth, which reduces the economic and social costs of moving. They seem to be willing to accept lower-quality health service provision, but prefer destinations where education quality is relatively better. These results suggest that high prices are pushing workers out of metropolitan cities.

With this paper I contribute to the related literature in several ways: To the best of my knowledge, it is the first work to empirically document the economic determinants of out-migration from metropolitan cities at the individual level in a middle-income country.[3] I exploit the detailed information on migration from a unique census survey. These data allow measurement of migration between local labour markets so that I capture the largest share of labour mobility within Brazil, improving on studies that investigate internal migration only at the regional level (Aguayo-Téllez et al. 2010; dos Santos Júnior et al. 2005; Fally et al. 2010; Lall et al. 2009; Santos and Ferreira 2007; Yap 1976). This approach also allows testing the importance of secondary towns as potential destinations for migrants not only from rural areas, adding to the recent literature (Christiaensen et al. 2017). Furthermore, by combining the destination choice and counterfactual analysis, I assess the capacity of migrants to evaluate expected earnings across a large set of possible destinations. These insights into the location choices of workers are relevant for regional planning (Moretti 2011).

The paper is structured as follows. The data used for the empirical analysis are described in Section 2. Descriptive maps, graphs, and tables explore the nature of migration from metropolitan to non-metropolitan cities in detail in Section 3. Thereafter, the conceptual framework of the destination choice model is discussed and results are presented in Section 4. The results of the counterfactual exercise are presented in Section 5. Section 6 then concludes.

2 Data

2.1 Data source

Every 10 years, the Brazilian National Institute for Geography and Statistics (IBGE) conducts a 10 per cent nationally representative household survey, the Census survey (*Censo Demográfico* 2010; IBGE 2012). The 2010 survey comprises around 20 million individual observations in all municipalities of Brazil. It contains information on household composition, living conditions, labour market, education, geographic location, and migration.

2.2 Definition of migration

The 2010 Census allows identification of migrants in the sample using the questions ‘Were you born in this municipality?’ to establish whether people are living in their birthplace, ‘When did you move to this

[3] One exception is McCormick and Wahba (2005), who analyse migration in and out of big cities in Egypt. However, their sample of migrants moving out of the big cities includes only 82 observations, and their hypothesis focuses on the movement into compared to out of large cities and the concentration of specific skill and age groups in large cities.

municipality?' to provide the year of migration, and 'In which municipality (in which state) did you live before you moved to this municipality that you are currently living in?' to provide the exact origin of migrants. It further asks for the municipality of the respondent's current job as well as of their previous job. Migrants are individuals who used to live and work in a different location than the one they were living in at the time of the survey.

2.3 Sample

The sample comprises working-age migrants and non-migrant residents. The legal working age in Brazil is 16 years, and the retirement age for men is 65 years. The age group for the sample has been restricted to ages 25–65 years. This way it can be assumed that students are excluded. All individuals in the sample are currently not in school and are participating in the labour market, which means they are either employed or unemployed but looking for work. I restrict the sample of migrants to those who moved within the past year, between 2009 and 2010, in order to minimize recall bias.

2.4 Definition of origins and destinations

Migration is measured as a change in living and working location at the level of a *microregião*. *Microregiões* are geographic and administrative agglomerations of municipalities sharing a labour market and economic activities, a bit larger than counties in the USA. I define 22 of these *microregiões* as metropolitan based on their population size of one million and above.[4] There are 551 *microregiões*, 22 of which are metropolitan, while 529 are non-metropolitan. Information on the local characteristics is aggregated to the *microregião* level using individual-level data from the Census survey. I use survey weights to obtain local estimates of wages and housing prices measured as the amount of rent per room. At the level of the federal state, 43 per cent of the metropolitan out-migrants leave their state of birth; the other 57 per cent stay within the same state when they move, and even more stay within their region. These observations confirm that the level of analysis at the *microregião* level also captures intra-regional population dynamics, the largest movements in the country.

2.5 Other variables

Other information on local characteristics is obtained from *Ipeadata*. This is an online data pool provided by the Instituto de Pesquisa Econômica Aplicada (Ipea), a Brazilian public research institute that collects data from several ministries and other public sources. It contains information at the *microregião* level on gross domestic product (GDP), quality of education and health provision, and homicide rates as a measure of crime.

The quality of education and health are measured using an index constructed and annually updated by the Industrial Federation of the federal state of Rio de Janeiro (FIRJAN). The index for education provision combines information about the subscription rate of pre-school children, dropout rate, rate of teachers with higher education, average daily teaching hours, as well as the results of a national education development score. The health provision quality index comprises the number of prenatal consultations, deaths due to poorly defined causes, and child deaths due to preventable causes.

All variables used and their source are specified in the Appendix in Table A10.

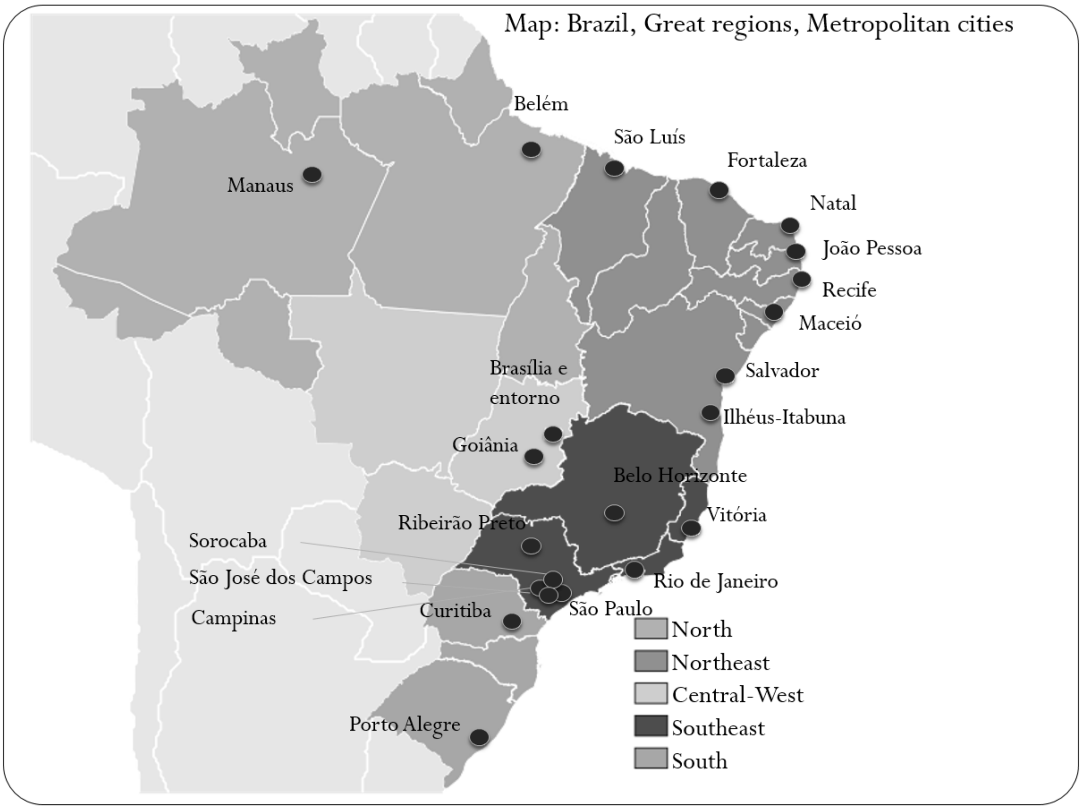
[4] This definition follows that of the United Nations' World Urbanization Prospects (UNWUP) (Christiaensen et al. 2013).

3 Descriptive statistics

3.1 Patterns of internal migration in Brazil

Figure 1 shows a map of Brazil, highlighting its five greater regions and the 22 metropolitan cities that are the focus of this analysis. The metropolitan cities are located mainly along the coast, with the exception of the state capitals in the South-eastern region, Goiânia in the Central-West, Manaus in the Amazon, and the national capital Brasília.

Figure 1: Map of greater regions and metropolitan cities of Brazil



Source: author's elaboration.

Labour migration within Brazil is historically very common and mainly attributed to socio-economic differences between regions and between the underdeveloped rural areas and several large urban centres (Yap 1976). In recent years, migration patterns in Brazil have been changing. Of all Brazilian internal migrants in the year before the 2010 Census, 47 per cent moved between non-metropolitan areas (Table 1). The second largest movement is into and out of metropolitan cities from and to non-metropolitan *microregiões*, comprising around 20 per cent each of all recent migrants, a substantial share of migration in the country. The remaining 12 per cent of migrants move between the metropolises.

The rate at which people leave big cities has been increasing in the past decade, as illustrated in figure 2. The graph plots the out-migration rate from cities with over one million inhabitants in Brazil from 2004 to 2009.

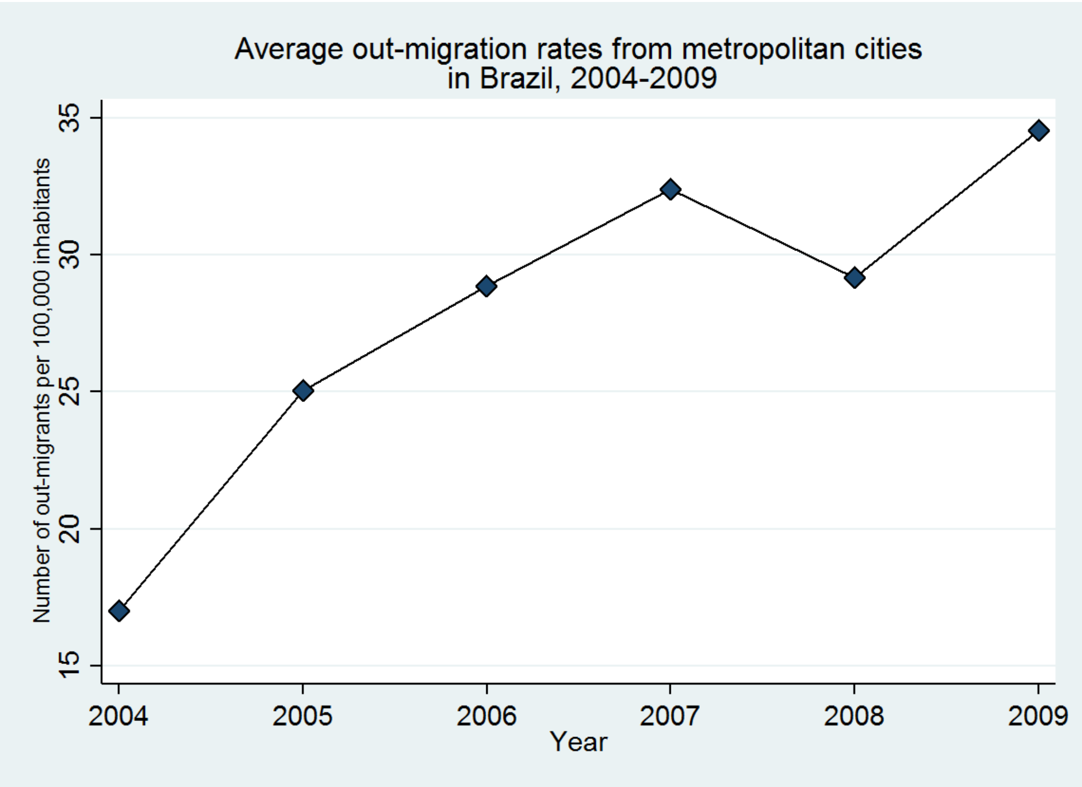
Table 1: Migration between metropolitan and non-metropolitan *microregiões* between 2009 and 2010.

	Destination			
	Non-metropolitan		Metropolitan	
Origin	*N*	%	*N*	%
Non-metropolitan	380,627	46.9	167,781	20.7
Metropolitan	162,647	20.1	99,143	12.2

Notes: Total N = 810,196, using survey weights.

Source: author's compilation based on data.

Figure 2: Out-migration rate from metropolitan cities 2004-2009



Source: author's elaboration.

3.2 Comparing origins and destinations

Table 2 compares metropolitan and non-metropolitan *microregiões* in terms of socio-economic characteristics. In the second and fourth column, I also include the coefficient of variation for the metropolitan and non-metropolitan *microregiões* to illustrate how diverse non-metropolitan areas are.

Metropolitan cities have on average around 10 times more inhabitants than non-metropolitan *microregiões*. In terms of prices, metropolitan residents face room rents that are more than 50 per cent higher than in non-metropolitan areas. At the same time they earn similarly higher wages. As expected, high-skilled occupations are concentrated in the metropolitan areas and labour markets are much more formalized in these big cities. The employment share of various sectors is highest for services, with 53 per cent in the metropolitan areas and 35 per cent in non-metropolitan *microregiões*. Yet, agriculture in the non-metropolitan areas employs around 30 per cent compared to only 9 per cent in metropolitan cities. While GDP is higher in metropolitan regions, it is growing faster in the non-metropolitan regions. In terms of living standards, almost 60 per cent of metropolitan residents live with adequate sewerage, water, and electricity provision, compared to only 36 per cent outside of these cities.[5] This illustrates the stark spatial inequality not only in economic but also in social aspects. Similarly lower are the indices for the quality of health and education provision in non-metropolitan areas in contrast to higher standards in the big cities. In contrast, crime is concentrated in big cities, with a homicide rate of 38 homicides per 100,000 inhabitants compared to around 19 in non-metropolitan areas.

Table 2: Characteristics of metropolitan and non-metropolitan *microregiões* in 2010

	Metropolitan		Non-metropolitan	
	Mean	Coefficient of Variation	Mean	Coefficient of Variation
Population	2,679,687	1.11	213,680	0.79
Room rent (R$, median)	72.47	0.23	45.22	0.42
Hourly wage (R$)	12.11	0.22	7.23	0.29
Share of				
Unskilled workers	0.37	0.09	0.37	0.14
Skilled workers	0.31	0.11	0.40	0.14
High-skilled workers	0.24	0.17	0.16	0.23
Formally employed	0.58	0.11	0.40	0.36
Unemployed	0.06	0.29	0.05	0.41
Share of workers in				
Agriculture	0.09	0.36	0.30	0.39
Industry	0.21	0.23	0.18	0.37
Services	0.53	0.08	0.35	0.23
Public services	0.11	0.25	0.12	0.24
People living in				
Adequate living conditions	0.57	0.28	0.36	0.67
Other measures				
GDP growth 2005-2010	0.16	0.31	0.18	0.79
Health quality index (0 to 1)	0.82	0.09	0.79	0.11
Education quality index (0 to 1)	0.77	0.14	0.73	0.14
Homicide rate (per 100,000)	38.00	0.54	18.58	0.77

Notes: skill level of workers is based on the occupation classification by the International Labor Organization (ILO). Industries include extractive industry, processing industry, electricity/gas, sanitation/sewerage, construction. Services include commerce, transport, housing/food, information/communication, financial services, real estate, professional consulting, science and technology, administrative services, arts/culture/sports, domestic services, and other services. Public services include public administration, security, education, health and social services, international organizations/foreign institutions. Six of the *microregiões* had missing values for homicide rates. I replaced them with 0 due to the way homicides are reported.

Source: author's compilation based on data.

The variation in these characteristics among non-metropolitan *microregiões* is large. The second and fourth columns of Table 2 show the coefficients of variation for metropolitan and non-metropolitan *microregiões* respectively. It is relatively larger for non-metropolitan areas in almost all categories but population, public service worker share, and education quality. This motivates the analysis of the metropolitan out-migrants' destination choice. Labour mobility is expected to respond to this spatial variation of real wages and other socio-economic characteristics.

3.3 Comparing migrants and residents

Metropolitan out-migrants are unlikely to be representative of the population of metropolitan cities. In Tables 3 and 4 I compare the characteristics of non-metropolitan residents, metropolitan out-migrants,

[5] The definition of which type of sewerage, water, and electricity provision is adequate comes from the report on subnormal agglomerations in Brazil (IBGE 2010).

and metropolitan residents. The comparison of these residents with metropolitan out-migrants sheds light on the differences between migrants and residents at the origin and destination.[6]

Table 3: Characteristics of migrants and non-migrants, 2010

	Non-metropolitan residents	Metropolitan out-migrants	Metropolitan residents
Number of observations	4,184,904	19,318	1,598,869
Population size	31,054,034	155,288	24,508,716
Age	40.25	36.85	40.22
Female	0.41	0.37	0.45
White	0.51	0.51	0.51
Education level			
None, primary incomplete	0.47	0.29	0.29
Primary, secondary incomplete	0.16	0.16	0.17
Secondary, higher incomplete	0.26	0.33	0.36
Higher complete	0.11	0.21	0.19

Notes: proportions and means computed using survey weights.

Source: author's compilation based on data.

Table 4: Labour market characteristics of migrants and non-migrants, 2010

	Non-metropolitan residents	Metropolitan out-migrants	Metropolitan residents
Unemployed	0.05	0.12	0.06
Log(monthly wages)	6.59	6.95	6.98
Sector			
Formal private	0.40	0.43	0.56
Formal public	0.06	0.08	0.06
Informal	0.26	0.23	0.21
Self-employed	0.02	0.02	0.01
Small business	0.26	0.24	0.15
Industry, ISIC			
Agriculture	0.26	0.14	0.09
Industry	0.22	0.27	0.21
Services	0.38	0.44	0.54
Public services	0.15	0.17	0.17

Notes: proportions and means computed using survey weights. Industries include extractive industry, processing industry, electricity/gas, sanitation/sewerage, construction. Services include commerce, transport, housing/food, information/communication, financial services, real estate, professional consulting, science and technology, administrative services, arts/culture/sports, domestic services, and other services. Public services include public administration, security, education, health and social services, international organizations/foreign institutions.

Source: author's compilation based on data.

Migrants are on average slightly younger than residents and relatively more of them are male. Overall, they are much better educated than the average resident at the non-metropolitan destination, and their education is very similar to that of metropolitan residents. A slightly larger share of migrants has a tertiary education compared to metropolitan residents. From this comparison, it does not seem that low- or high-educated workers are more likely to leave metropolitan cities than the respective other group.

Table 4 documents the labour market characteristics of out-migrants and residents. Around 12 per cent of workers who left the metropolitan cities for non-metropolitan destinations are unemployed, in contrast

[6] For simplification, in the descriptive analysis this comparison does not account for the fact that migrants could all be concentrated in a specific subset of non-metropolitan *microregiões* so that residents in all non-metropolitan *microregiões* might not represent the exact comparison group of that specific subset.

to an unemployment rate of only 5 per cent among non-metropolitan residents. This indicates that metropolitan out-migrants are a heterogeneous group and some clearly lose out at their new destination. However, the high unemployment rate might just capture a period of adjustment for very recent migrants who have not found a job yet at their new destination.

In terms of wages, migrants earn on average more than their non-migrant counterparts at non-metropolitan destinations, and they earn almost as much as residents in metropolitan areas. This might only reflect differences in the productivity of locations where migrants live as well as different observable and unobservable characteristics of migrants in contrast to residents. The regression analysis in this paper aims to disentangle these factors.

More than 60 per cent of non-migrants in the metropolitan cities are employed either in the public formal or private formal sectors, whereas only around 46 per cent of non-migrants in non-metropolitan towns work in the formal sector. Migrants appear to find relatively more formal employment at the destinations outside of the big cities compared to the residents there.

Most migrants work in service sectors. Few work in agriculture at the non-metropolitan destinations even though this is the main sector of activity there. In the metropolitan cities, services are the main employment sectors. This suggests that most migrants are unlikely to change their sector of activity when they move out of metropolitan areas.

These observations highlight three findings. First, there is a significant difference in economic and social characteristics between metropolitan and non-metropolitan *microregiões* that are likely to determine migration between these. Prices of housing—non-tradable living costs—are much higher in the metropolitan cities, and the non-metropolitan areas are catching up economically. Second, there is a large spatial variation in the characteristics of non-metropolitan *microregiões* across the country. Hence, metropolitan out-migrants are unlikely to be indifferent between destinations in their choice of where to move. Third, migrants are not a random draw of the population, and they are a heterogeneous group. It is important to account for migrants' characteristics and underlying selection in the econometric analysis of this paper.

4 Destination choices of migrants

4.1 Empirical methodology

The empirical analysis now focuses on the estimation of the effect of various local attributes on the destination choices of migrants. The analysis is based on a multiple-choice setting, such as presented by McFadden (1974). The empirical application is restricted to those who migrated.[7] As in Fafchamps and Shilpi (2013), I model destination choice conditional on the individual being a migrant.

Migrants are assumed to choose their location in order to maximize their utility. Motivated by a random-utility model, a migrant i residing in current location o chooses among all possible destinations J. Let $\mathbf{z}_{ij}$ be a vector of destination attributes that vary across alternatives and can vary by migrant i, and let c_j be the cost of moving to destination j from the current location o. Therefore, I define $c_j = 0$ if $j = o$.

[7] The model allows inclusion of residents in the analysis and assumes that they chose not to move. In the empirical application this would result in a sample so large that it is not feasible to handle. The decision to migrate itself yields a selection bias distinct from the location choice. Costs of moving are heterogeneous for workers, so some of those who did not move might have done so due to high costs or risks, which gives rise to a selection bias in the decision to migrate. By excluding the choice to stay at one's origin, and estimating the destination choice model with migrants only, this specific selection bias does not arise.

The utility of moving to destination j is assumed to have the following form:

$$U_{ij} = \beta\text{'}z_{ij} - c_j + \varepsilon_{ij} \tag{1}$$

The utility of migrant i from moving to destination j depends on the destination attributes, moving costs, and an idiosyncratic random component ε_{ij}. The observed choice by the migrant is assumed to reflect the maximum utility of all J utilities. The probability that migrant i chooses destination j is therefore

$$\text{Prob}(U_{ij} > U_{ik}) \qquad \text{for all other} \qquad k \neq j \tag{2}$$

It is assumed that the error terms are distributed independently and identically with Weibull distribution, as in McFadden (1974):

$$F(\varepsilon_{ij}) = \exp(-e^{-\varepsilon_{ij}}) \tag{3}$$

The probability of moving to destination j is now modelled conditional on migration (i.e. leaving location o). If Y_i represents a random variable indicating the destination choice of migrant i, the probability that this choice is destination j conditional on migration can then be expressed as:

$$\text{Prob}(Y_i = j | Y_i \neq o) = \frac{e^{\beta' z_{ij} - c_j}}{[\,\sum_{j=1}^{J} e^{\beta' z_{ij} - c_j}] - e^{\beta' z_{io} - c_o}} \tag{4}$$

This is equal to:

$$\text{Prob}(Y_i = j | Y_i \neq o) = \frac{e^{\beta' z_{ij} - c_j}}{\sum_{j \neq o}^{J} e^{\beta' z_{ij} - c_j}} \tag{5}$$

Equation 5 represents a conditional logit model. The vector z_{ij} may comprise individual-specific but destination-invariant characteristics w_i and the attributes of each destination x_{ij}, which vary across destinations and can also vary across individuals

$$z_{ij} = g(w_i, x_{ij}) \tag{6}$$

In this analysis, the interest lies in the attributes of destinations and not on migrants' characteristics. Greene (2000) shows how w_i drops out of the probability in Equation 5 so that this model automatically controls for any individual-specific factors in the destination choice.[8] However, this also implies that I cannot estimate the effect of such factors as age of the migrant etc. Hence, the alternative specific conditional logit model takes the following form:

$$\text{Prob}(Y_i = j | Y_i \neq o) = \frac{e^{\beta' x_{ij} - c_j}}{\sum_{j \neq o}^{J} e^{\beta' x_{ij} - c_j}} \tag{7}$$

This model can be estimated by the method of maximum likelihood. Let $d_{ij} = 1$ if $Y_i = j$, and 0 otherwise. Then the log-likelihood function is:

$$\log\, L = \sum_{i=1}^{N} \sum_{j=1}^{J} d_{ij} \log \text{Prob}(Y_i = j | Y_i \neq o) \tag{8}$$

For the estimation, there are N observations and regressors for each of the 17,501 metropolitan out-migrants.[9] They choose from $J = 529$ possible non-metropolitan *microregiões* as destinations. Only one of the destinations will have a positive outcome as the chosen destination—that is, the one observed in the data. This results in 9,258,029 individual–destination observations for the multivariate analysis. I

[8] In some applications, such as that of Fafchamps and Shilpi (2013), this is called individual fixed effect alternative specific conditional logit. It is, however, not to be confused with the inclusion of fixed effects as in a panel model.

[9] The sample of metropolitan out-migrants is slightly reduced as I only include those in the analysis who are matched so that the results are comparable. Those dropped were not matched.

cluster standard errors at the metropolitan *microregião* of origin o because it is likely that migrants from the same origin share a pattern in their destination choice.

Based on the human capital migration model (Sjaastad 1962), the destination attributes of interest in this analysis are wages and prices. One measure for prices would be a consumer price index, which is not available at the level of *microregiões*. Deaton and Dupriez (2011) showed that food trade is strongly integrated across Brazil, so that prices for these consumption goods do not vary much across space. Price differences in non-tradable goods, such as housing, therefore appear more suitable to capture differences in cost of living between metropolitan origins and non-metropolitan destinations. Living costs are proxied with the amount of rent per room reported in the survey. If individuals do not report rent, the average rent in the *microregião* of residency is imputed (see Morten and Oliveira (2016), who use the same approach for Brazil). This is the most disaggregated available measure of prices for non-tradable goods in Brazil. In order to measure the impact of local development, I additionally include population and GDP, as well as homicide rates, as destination attributes. I also look at public service provision, which is an index for the quality of education and health provision at the local level.

Migration costs are measured by the Euclidean distance between origin and destination in kilometres, and additionally by a dummy for whether a migrant moved out of her state of birth to a different state. This captures social proximity of a destination to the migrant's origin, as in Brazil people have a strong identity with their birth state. Both of these variables imply also the social cost of being farther away from one's family and friends. I further include an interaction term of distance and the 'other-state' dummy in order to disentangle the impact of these two factors.

Average wages in a location need not reflect the wages a migrant can expect to earn. I therefore predict expected wages for migrants based on their characteristics and the coefficients from a wage estimation of residents at each location.

First, I estimate a wage regression separately for all 6.9 million resident observations in each *microregião*. The wage regression takes the following form:

$$W_i^j = \alpha_j(a_i^j - \bar{a}_j) + \beta_j(E_i^j - \bar{E}_j) + \gamma_j(S_i^j - \bar{S}_j) + \chi H_i^j + \delta_j + v_i^j \tag{9}$$

Log hourly wages of individual i in location j are determined by the individual characteristics a_i^j, E_i^j, S_i^j, household characteristics H_i^j, a dummy for the *microregião* δ_j, and an idiosyncratic error term v_i^j. The variable a_i^j summarizes age and age-squared, E_i^j is the education level, and S_i^j measures gender and race (white vs non-white). Each of these variables is demeaned at the level of the *microregião*, so that the coefficients α_j, β_j, and γ_j capture the return to these characteristics specific to each location. Additionally, this implies that δ_j measures the unconditional *microregião*-specific average wages. Household characteristics, H_i^j, include the proportion of children and a dummy for whether the partner works, as these might vary by region—for example, in more rural areas households tend to be larger and female labour force participation lower, so that wages would be overestimated in these areas if this was not controlled for. I use the survey weights in these regressions to make the estimates representative of the population.

For each *microregião*, the coefficients $\widehat{\alpha}_j$, $\widehat{\beta}_j$, $\widehat{\gamma}_j$ from this regression are then used to predict a measure of expected wages for each migrant. This predicted wage reflects what each migrant can expect to earn in each *microregião*, conditional on her characteristics a_i, E_i, and S_i, and the unconditional local wage level $\widehat{\delta}_j$:

$$\widehat{E[W_i^j|X_i]} = \widehat{\delta}_j + \widehat{\alpha}_j(a_i - \bar{a}_j) + \widehat{\beta}_j(E_i - \bar{E}_j) + \widehat{\gamma}_j(S_i - \bar{S}_j) \tag{10}$$

The coefficients of the wage predictions corresponding to Equation 10 (see Appendix Table A8) confirm the relationships documented in the literature: age has a positive but diminishing effect on expected wages; women earn less than men; white Brazilians earn more than non-whites; and wages increase with the level of education.

This approach assumes that migrants are a random draw from the resident population, so that the returns to individual characteristics should be the same for migrants and residents. In the descriptive statistics I showed that migrants differ from the resident population in a number of observable characteristics. This implies that the expected wage measures used in the analysis so far could be biased by unobservable characteristics. I thus estimate another measure for expected wages that should reduce the selection bias. I predict expected wages in non-metropolitan destinations from a sample of previous migrants from the same origin as the migrant. These migrants moved to the destinations more than a year ago. They are assumed to be more comparable to migrants than residents in terms of unobservable characteristics specific to migrants, such as risk-taking preferences.

For the metropolitan origins, I predict expected wages based on a matched sample of residents at origin. I apply CEM to use only those residents that look most similar to the migrants. CEM bounds the degree of model dependence in the main analysis and the data are automatically restricted to common support. The large dataset of the Census at hand is very suitable for this matching method, without facing the trade-off of conventional matching methods between bias and variance. Migrants and non-migrants are matched on sex, age (for migrants age at the year of migration), race, education level, marital status, sector of activity, and city of origin. Balance statistics are presented in the Tables A1–A3.

4.2 Results

In the specific application of this paper, metropolitan out-migrants choose their destination not only based on destination attributes, but on these attributes relative to the attributes of migrants' metropolitan origins. For each location attribute, I thus compute the difference between the destination and origin—for example, the cost of living in destination j minus the cost of living in origin o. Table 5 gives an overview of the differences between destinations and origins of all variables of interest and how these differences vary between the destinations that migrants chose to those that they did not choose.

Table 5: Difference between non-metropolitan destination and metropolitan origin, comparing chosen destination to alternative destinations

	Chosen destination	Alternative destinations	t-statistic, difference in mean
Expected hourly wages (log)	–0.52	–0.58	–21.16
Rent per room (log)	–0.43	–0.63	–57.24
Other state than origin	0.57	0.92	174.73
Distance to origin (km)	1,020.31	1,123.94	28.83
Population (log)	–2.32	–2.81	–58.14
GDP (log)	–2.82	–3.59	–68.47
Homicide rate	–9.52	–12.52	–16.77
Education provision quality index (0 to 1)	–0.03	–0.07	–33.04
Health provision quality index (0 to 1)	–0.02	–0.05	–32.03

Source: author's compilation based on data.

Table 5 already indicates some patterns of destination choice. Similar to what we observed earlier in the descriptive part, nominal wages are on average always lower in non-metropolitan areas. Migrants tend to choose locations where this gap is relatively smaller, -0.52 compared to -0.58. Similarly, prices for housing, measured in rent per room, are on average higher in the big cities. Migrants settle in locations where this price gap is not as big as in other possible destinations. This could indicate a trade-off between higher wages and lower prices at destinations. The difference of prices between chosen destination and alternative locations is larger than that of wages. This could indicate that prices matter more than wages for the location decision.

A clearer pattern is revealed with regards to the geographic and social distance of chosen destinations: 57 per cent of chosen destinations are in a different state than the origin (which is also the state of birth of migrants), contrasting to 92 per cent of the destination alternatives, and reflected also in a lower average distance of chosen destinations to the migrants' origin.

The difference in GDP is also statistically significant. Migrants choose locations with a higher GDP compared to alternative destinations, but GDP is on average always lower in non-metropolitan destinations than in metropolitan origins. The same is the case for population size. Chosen destinations are on average larger than their alternatives. Based on their average size, they are, however, not the smallest locations, but still medium-sized *microregiões*. In terms of amenities, chosen destinations have relatively higher levels of homicide rates in contrast to alternative destinations. Education and health service provision is on average higher in the chosen *microregião* than in alternative destinations.

These averages are all statistically different between chosen and alternative destinations. Many of these factors are highly correlated with each other, which makes it necessary to apply a multivariate analysis to disentangle their influence on the metropolitan out-migrants' destination choices.

Table 6 reports the results of the alternative specific conditional logit model that estimates the probability for destination choice conditional on migration as specified in Equation 8. The interpretation of coefficients in the alternative specific conditional logit model is not straightforward. It is not possible to compare the coefficient size directly, but only in relative terms, which I will do in Section 4.4.

Table 6 reports the results. The specification in column 1 is that of only wage and price differences. The first coefficient in column 1 is that of wage differences, with a value of -0.817. Wage differences are on average negative, as documented in Table 5, implying that an increase in wage differences means that wages get closer to those in the metropolitan origin of migrants. Metropolitan out-migrants hence prefer to move to non-metropolitan *microregiões* that have a relatively smaller wage difference to the origin than alternative destinations.

For living costs, the opposite is the case. The second coefficient in column 1 is positive. Due to much higher prices in metropolitan cities, the difference in the cost of living is on average negative, but migrants prefer to keep this difference as large as possible. Metropolitan out-migrants prefer destinations that are cheaper than their origin and relatively cheaper than alternative destinations. Both coefficients are strongly significant.

In all models, I control for the costs of migration, the distance to move, and the social distance. The latter is the dummy for whether a destination is in a different state than a migrant's origin. The distance between destination and origin appears insignificant. Migrants prefer to stay within their state of birth.

Wages and prices remain with their expected signs and are both significant predictors of destination choice once I add other control variables in columns 2 and 3, but their coefficients become smaller. In column 2, I add population, GDP, homicide rates, and local amenities. By construction, metropolitan out-migrants move to locations with fewer inhabitants than their metropolitan origin. They also seem to prefer relatively smaller locations among their destination choices. The level of economic activity measured in the log of GDP does not appear to significantly predict the location choice. Metropolitan out-migrants significantly prefer destinations where the homicide rate is relatively lower than in their metropolitan origins. In terms of public service provision, it appears that migrants show preferences for locations with relatively better education quality, but they accept relatively lower health provision quality.

Table 6: Destination choice conditional on migration, alternative specific logit

	Likelihood to select chosen destination		
	(1)	(2)	(3)
Difference in:			
Expected wages (log)	–0.817***	–0.464**	–0.272
	(0.205)	(0.224)	(0.183)
Rent per room (log)	1.184***	0.961***	0.854***
	(0.163)	(0.282)	(0.295)
Population (log)		0.689***	0.684***
		(0.175)	(0.179)
GDP (log)		–0.077	–0.048
		(0.123)	(0.122)
Homicide rate		0.010**	0.009*
		(0.005)	(0.005)
Education quality index		–3.972**	–4.281***
		(1.650)	(1.637)
Health provision quality index		4.167**	4.171**
		(1.774)	(1.828)
Destination specific:			
Distance to origin (log)	–0.046	–0.170	–0.195
	(0.357)	(0.435)	(0.426)
Other state	–2.197***	–2.318***	–2.302***
	(0.365)	(0.301)	(0.306)
Observations	9,258,029	9,258,029	9,258,029
χ^2	166	1,356	1,558
Number of cases	17,501	17,501	17,501
Number of alternatives	529	529	529

Notes: significance levels * 10, ** 5, and *** 1 percent. Standard errors are clustered at the metropolitan *microregião* of origin. The estimator is alternative specific conditional logit. In each column, the first set of regressors are the difference between destination and origin for each destination alternative. The second set, indicated as *Destination specific*, are measured at the destination relative to the origin. Prices are measured using rent per room. Column 3 uses expected wage differences based on past migrants at the destination and matched residents at the origin, as explained in Section 4.1.

Source: author's compilation based on data.

In column 4 I use expected wages that are corrected for selection as explained in the previous section. The selection-corrected wage differences between non-metropolitan destinations and metropolitan origins are thus based on wage predictions using the previous migrants at destinations and matched residents at origins. The coefficient of expected wage differences becomes smaller and insignificant now that I control for selection. It is plausible that wage differences are less stark once observable and some unobserved characteristics are accounted for. This indicates that such characteristics of migrants led to an upward bias in expected wage differences. The coefficients of other variables remain the same. The coefficient of housing prices becomes a bit smaller, but it remains positive and strongly significant.

In summary, the results of the multivariate analysis confirm that prices matter significantly for the destination choice of metropolitan out-migrants. Wages are weakly significant and become insignificant when I account for potential selection bias. Furthermore, physical and social moving costs appear to matter. Lastly, I document that also amenities in terms of public service provision are significantly correlated with the destination choice.

4.3 Robustness: hedonic housing prices

The second main variables of interest in the destination choice analysis are the prices or living costs, proxied so far by the average rent per room in a *microregião*. In the Census data, households are asked to state the monthly rent they pay if they live in a rented apartment or house and the number of rooms in the unit. I used this data to aggregate the average room rent at the *microregião* level. This measure ignores the possibility that the price differences might just reflect differences in housing quality. Similar to Li and Gibson (2014), I construct a hedonic housing index that measures the differences in housing costs based on location-specific amenities rather than housing-specific characteristics.

Households are asked to provide information on the quality of walls, floors, and the presence of toilets. Additional questions inform about the quality of sewerage, waste water, and electricity access. With these variables I can estimate a hedonic housing price for each location. I regress the rent per room on these characteristics weighted by the household survey weight, and I include a dummy for each *microregião*. These regression results are presented in Table A7. The coefficients of the *microregião* dummies capture any location-specific amenities that contribute to spatial price differences. I extract these estimates to construct a location-specific hedonic living cost measure. This variable is independent of differences in housing quality.

Using this measure provides no change in the signs of the estimates of the influence of price differences on destination choice compared to the initial results (see Table 7). The coefficients of the hedonic price index are positive and strongly significant. Other coefficients of the analysis do not change significantly.

Table 7: Destination choice of metropolitan out-migrants: hedonic prices—inclusion of unemployment rate

	Likelihood to select chosen destination	
	(1)	(2)
Difference in:		
Expected wages (log)	–0.286*	–0.292*
	(0.174)	(0.156)
Rent per room (log)	0.774***	0.748***
	(0.276)	(0.252)
Population (log)	0.685***	0.672***
	(0.144)	(0.145)
GDP (log)	–0.006	0.008
	(0.084)	(0.091)
Homicide rate	0.009*	0.009**
	(0.005)	(0.004)
Education quality index	–4.071**	–4.107**
	(1.723)	(1.814)
Health provision quality index	4.225**	4.101**
	(1.929)	(1.617)
Unemployment rate		–1.196
		(4.969)
Destination specific:		
Distance to origin (log)	–0.230	–0.233
	(0.424)	(0.423)
Other state	–2.312***	–2.321***
	(0.304)	(0.279)
Observations	9,258,029	9,258,029
χ^2	1,371	1,426
Number of cases	17,501	17,501
Number of alternatives	529	529

Notes: significance levels * 10, ** 5, and *** 1 per cent. Standard errors are clustered at the metropolitan *microregião* of origin. The estimator is alternative specific conditional logit. In each column, the first set of regressors are the difference between destination and origin for each destination alternative. Wage differences are based on predictions using the sample of matched residents at the origin and previous migrants at the destination, and price differences are computed using a hedonic price index. In column 2, unemployment is added as regressor.

Source: author's compilation based on data.

In column 2 of Table 7 I add the difference in unemployment rates to the regression. In the migration literature, it has been suggested that not only expected wages, but also the likelihood to earn these, matter for migration decisions. In the case of Brazilian workers moving out of metropolitan cities, differences in unemployment rates are not significantly related to their destination choice. This could be due to the fact that the unemployment rates in non-metropolitan *microregiões* are on average only 1 percentage point lower than in metropolises.

4.4 Relative effect size

Marginal effects of the alternative specific logit model can be computed for each possible location choice, but this is computationally burdensome and ineffective in presenting the results. To illustrate and compare the effect sizes, we can look at one destination alternative: the one with a price difference very close to the average price difference to metropolitan origins. I take the significant regressors from the full specification with selectivity robust price and wage measures as in column 1 of Table 7 and compute their elasticities for this example location.

Let the probability of choosing destination j be P_j, then the elasticity of P_j with respect to an attribute x_{ij} evaluated at the mean $\bar{x_{ij}}$ can be written as:

$$\frac{\delta \log(P_j)}{\delta \log x_{ij}} = \bar{x_{ij}}(1-P_j)\beta_x \tag{11}$$

where β_x is the coefficient of the destination attribute from the conditional logit estimation (Greene 2000). The means and elasticities for significant covariates are presented in Table 8.

Table 8: Elasticities of independent variables for example location

	Mean	Elasticity
Other state	0.923	–2.132
Population (log)	–2.812	–1.923
Expected wages	–2.764	0.790
Living costs (hedonic prices)	–0.492	–0.381
Education quality index	–0.072	0.293
Homicide rate	–12.519	–0.113
Health provision quality index	–0.048	–0.202

Source: author's compilation based on data.

The elasticities reveal that by far the largest effect on migrants' destination choice is that of a migrant leaving her or his state of birth. This captures the physical and social costs of moving and confirms that migration costs in Brazil are still high and a significant factor in labour mobility (Morten and Oliveira 2016). The effect of population size is equally large, which reflects the large gap in terms of city sizes between the metropolis and small towns. Despite weak significance, the elasticity of destination choice to wage differences is relatively larger than that to prices. A 1 per cent increase in the price difference between the metropolitan origin of a migrant and this specific location make it on average 0.4 per cent more likely to be chosen as the destination, compared to 0.8 per cent as a response to a 1 per cent increase in wage gaps.

Local amenities concerning public service provision are significant predictors for the destination choice, but their effect size is smaller than that of prices, with education appearing slightly more important than health, followed by the homicide rates.

5 Counterfactual earnings of metropolitan out-migrants

The previous results showed that prices play a sizeable and significant role in the destination choice of metropolitan out-migrants, whereas expected wages do not appear significant once I control for self-selection of migrants. It is possible that this is due to incorrect expectations of migrants about their earnings. Thus, this section focuses on the actual observed earnings of migrants at their destination in contrast to expected wages. The actual earnings are compared to a prediction of what a migrant would have earned had she not moved out of the metropolitan city—her counterfactual wage. With this comparison, this section aims to see whether and how metropolitan out-migration is associated with a wage loss or gain and what role living costs play in this question.

5.1 Empirical methodology

The wage return to migration is defined as the difference between income at the destination, y_d, and at the origin, y_o:

$$r = y_d - y_o \tag{12}$$

In the empirical application, income is proxied by the log of hourly wages, W.[10] The comparison of migrant wages between origin and destination can be interpreted as an evaluation problem. Let migration be the treatment with $M_i = 1$ if the individual moved, $M_i = 0$ if not. For each individual, two outcomes in terms of wage differences can be defined as

$$Y_i^0 = \log(w_{i,0}) - \log(w_{i,0}) \qquad \text{if} \qquad M_i = 0 \tag{13}$$

$$Y_i^1 = \log(w_{i,1}) - \log(w_{i,0}) \qquad \text{if} \qquad M_i = 1 \tag{14}$$

Thus, the wage difference due to migration can be identified for migrants as average treatment effect on the treated (ATT):

$$\text{ATT} = E(Y^1 - Y^0 \mid M = 1) = E(Y^1 \mid M = 1) - E(Y^0 \mid M = 1) \tag{15}$$

The first term on the right-hand side is observable in the data at hand—the wages of migrants at their destination. The second term represents the counterfactual outcome, what migrants would have earned had they not migrated, which cannot be observed. I only observe wages for migrants at their destination and for non-migrants at the origin. If wages were estimated using ordinary least squares (OLS) and then compared, the wage differences would be biased due to selection into migration arising from individual-specific unobservable characteristics.

It is necessary to account for this potential bias in the empirical estimation of migrants' counterfactual wages. This is especially important in a context in which spatial wage differences have been found to reflect variation in labour force composition and industry concentration. In Brazil, the labour force is distributed unequally across space, concentrating better-educated workers in metropolitan areas and economically stronger regions. Thus, the returns to education based on observable characteristics explain around half of the spatial wage differences (Almeida dos Reis and Paes de Barros 1991; Ferreira et al. 2006; Foguel et al. 2015). Furthermore, Brazilian workers have shown little mobility across industries, so it seems reasonable to focus on the self-selection by location and not by sector (Hering and Paillacar 2015; Menezes-Filho and Muendler 2011).

I therefore use the predicted wages from the matched sample of residents in metropolitan origins of migrants, as described in Section 4.1. The difference between the actual observed wages at the destination and the predicted counterfactual wages at the origin is the return to migration out of metropolitan

[10] I choose to look at hourly wages earned in the main job instead of total income as hourly wages in the main job should mostly reflect the return to individual characteristics based on location, whereas total income also depends on household composition and other factors.

cities. Real wages are computed using the local average rent per room as the denominator of actual and predicted nominal wages.

5.2 Results

This section presents the results of the counterfactual analysis. Table 9 presents the average return to migration as the difference between average actual and counterfactual wages for migrants moving out of metropolitan areas. These migrants earn significantly lower wages at their non-metropolitan destinations. Once I account for the local living costs by using real wages, the difference becomes positive. This indicates that metropolitan out-migrants lose in nominal terms, but gain in real wages due to lower living costs in non-metropolitan destinations.

The results without matching for nominal wage differences are around 0.1 log points larger than when matching is applied (see Table A4). This indicates an overestimation of wages at the origin when not accounting for selection, and suggests that out-migrants are negatively selected from the metropolitan working population.

Table 9: Differences in actual and predicted wages for metropolitan out-migrants, after matching

Log (nominal hourly wages)	*N*	Mean
Observed	15,424	1.816
Predicted	15,424	2.069
Difference		–0.253***
Log (real hourly wages)	***N***	**Mean**
Observed	15,424	–2.237
Predicted	15,424	–2.396
Difference		0.159***

Notes: significance levels * 10, ** 5, and *** 1 per cent for t-test of the difference in means between observed and predicted wages. Predicted wages are based on a matched sample of metropolitan residents.

Source: author's compilation based on data.

Table 10 documents heterogeneity in wage returns along the education level of migrants. I define high-educated workers as those who completed high school or any higher level of education. Low-educated workers are those who did not complete high school or any lower level of education. The results show that real wages are higher at the destination than the origin for both groups. For high-educated individuals leaving the big cities the real wage gains are larger, because their loss in nominal wages is relatively small. In contrast, low-educated workers see a large loss in nominal terms, and a relatively smaller gain in real wages. For both groups the nominal and real wage differences are statistically significant.

Table 10: Differences in actual and predicted wages for metropolitan out-migrants, by education level

	High-educated	
Log (nominal hourly wages)	*N*	Mean
Observed	3,107	2.846
Predicted	3,107	2.930
Difference		–0.084***
Log (real hourly wages)	*N*	Mean
Observed	3,107	–1.270
Predicted	3,107	–1.544
Difference		0.274***
	Low-educated	
Log (nominal hourly wages)	*N*	Mean
Observed	12,317	1.556
Predicted	12,317	1.851
Difference		–0.295***
Log (real hourly wages)	*N*	Mean
Observed	12,317	–2.481
Predicted	12,317	–2.611
Difference		0.130***

Notes: Significance levels * 10, ** 5, and *** 1 per cent for t-test of the difference in means between observed and predicted wages. Predicted wages are based on a matched sample of metropolitan residents.

Source: author's compilation based on data.

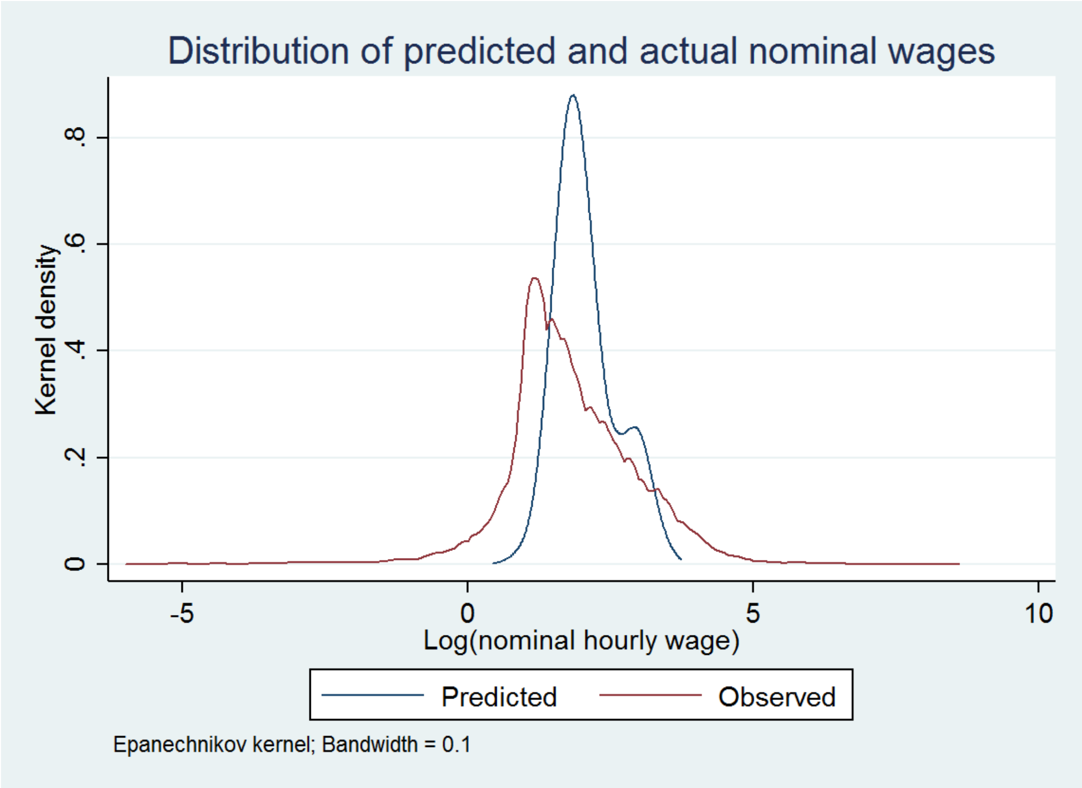
So far, the counterfactual wage comparison has focused on the average wage return. However, the distributional graphs of actual and counterfactual wages document the return to metropolitan out-migration along the wage distribution. Figures 3 and 4 show the wage distributions of workers who have moved out of a metropolitan area. They compare the observed wages of migrants at their destination and the predicted counterfactual wages at the origin. As suggested from the results in Table 9, for nominal wages the distribution of observed wages lies left of the predicted earnings in metropolitan origins. Wages are generally higher in origins, and out-migration implies a loss in nominal terms.

For real wages (see Figure 4) the distribution of observed wages lies now a bit to the right of counterfactual wages in metropolitan origins, reflecting the positive return in real terms to leaving expensive cities.[11]

In the analysis of the destination choice of migrants we learned that metropolitan out-migrants face a trade-off in choosing between destinations where their loss in nominal wages is smallest but their gain in lower living costs largest. This can explain why some individuals do not experience a positive return to migrating out of metropolitan areas across the income distribution. Some might fail to successfully evaluate their destination alternatives, some might lack the information about wages and prices at all destinations, and others might just not be successful in acquiring the wage employment they had expected or they migrate for other reasons, such as family. In this way, some metropolitan out-migrants lose out, while on average they gain in real wage returns.

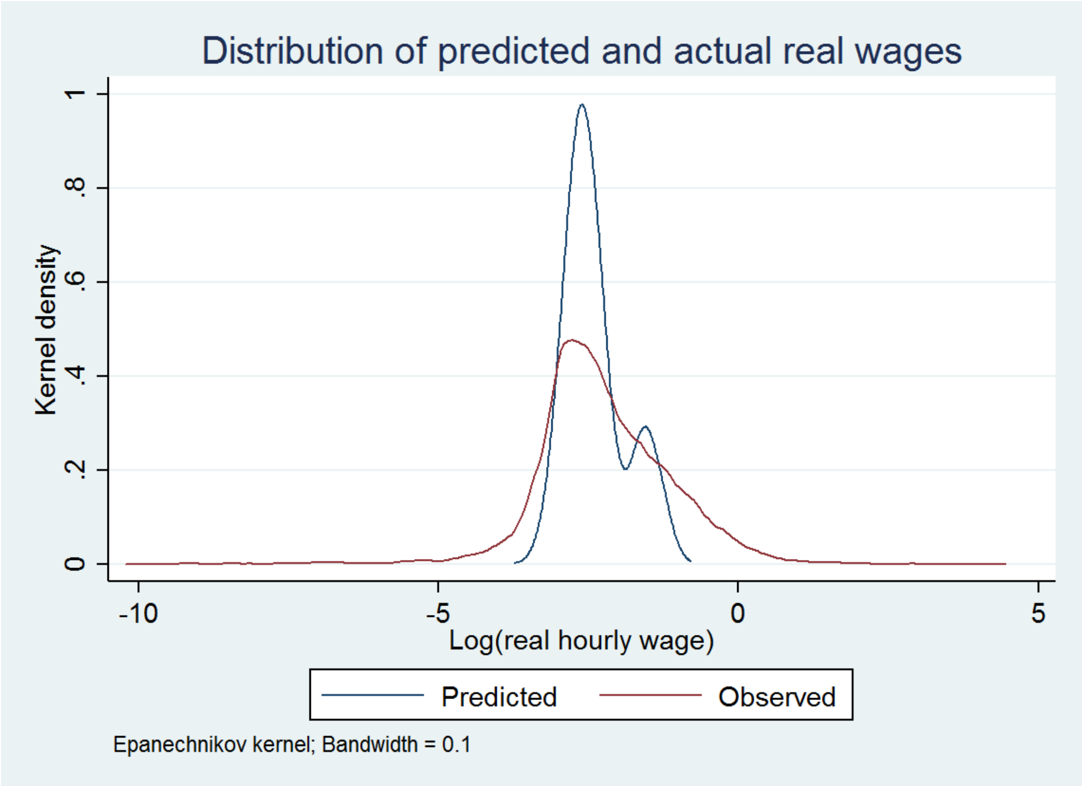
[11] The distributions are tested to be significantly different with a Kolmogorov–Smirnov test for equality of distributions. Both the nominal and real wage distributions are significantly different.

Figure 3: Kernel density plot of actual and predicted nominal wages of metropolitan out-migrants with matching



Source: author's elaboration.

Figure 4: Kernel density plot of actual and predicted real wages of metropolitan out-migrants with matching



Source: author's elaboration.

To understand better how migrants evaluate possible destinations, I contrast the previous results with those of people moving into metropolitan cities applying the same methodology. Table A9 in the Appendix confirms that in nominal terms metropolitan cities present a huge income gain, but especially low-skilled migrants experience significant and large losses in real terms. This points to the importance of experience and information. Metropolitan residents face high prices, and this variable thus dominates their evaluation of possible destinations. Workers in non-metropolitan areas might not have considered living costs as a significant factor, so they fail to account for them in their destination choice.

5.3 Robustness: price measures

In Section 4.3, I computed a hedonic price measure to control for location-specific differences in housing quality. The results did not change qualitatively, but the coefficients became slightly smaller, thus indicating that this measure better captures unconditional housing prices. I therefore also compute the real wages using the hedonic price as the denominator. The results are presented in Table 11. The wage difference remains positive and statistically significant, but it is smaller by around one-third than in the initial results.

Table 11: Differences in actual and predicted real wages for metropolitan out-migrants using hedonic prices as the denominator, after matching

Log (real hourly wages)	*N*	Mean
Observed	15,424	–2.105
Predicted	15,424	–2.155
Difference		0.050***

Notes: significance levels * 10, ** 5, and *** 1 per cent for t-test of the difference in means between observed and predicted wages. Predicted wages are based on a matched sample of metropolitan residents.

Source: author's compilation based on data.

Further concerns regarding the measure of living costs could arise from the fact that low- and high-educated workers probably face different housing markets with different average prices. I therefore conduct the counterfactual analysis using education group-specific living costs as the denominator when computing real wages (see Tables A5 and A6). I also use the median prices in the *microregião* instead of the mean. The signs and significance of wage differences remain the same as in the initial results, but their size changes. For high-educated workers the estimates yield larger positive wage differences when I use education group-specific living costs. The differences are 0.37 and 0.3 log points for average and median rents specific to high-educated workers. For low-educated workers the education-specific wage differences are smaller than unadjusted ones. Applying the median prices faced by low-educated workers yields the smallest difference of 0.07 log points compared to 0.13 in the initial results. Rents are on average lower for low-educated workers, so that their gain from leaving metropolitan cities becomes smaller than when I did not account for the education group-specific rents. The opposite applies for high-educated workers.

6 Conclusion

The economic literature on migration in developing countries focuses on rural to urban movements because this was the dominant observation in most countries, among them Brazil during its transition from a low- to middle-income country. In the decade of the 2000s, the movements of workers across Brazil have shown to lead equally out of metropolitan cities as into them. This paper uses the Brazilian Census data from 2010 to study this movement. High- and low-skilled workers are equally likely to leave the big cities and out-migrants move to smaller towns, not to remote rural areas. These secondary cities have been rapidly growing economically, also due to targeted government investment in previously lagging areas of the country (do Planejamento 2010; Lall et al. 2009; Mata et al. 2005).

City-leavers are on average comparable to metropolitan residents in terms of education, but they differ in their age and sex composition. In their medium-sized destination towns, the out-migrants tend to work for slightly lower wages in 'urban' sectors, such as services and manufacturing, and less in agriculture. The descriptive part of the paper documents that non-metropolitan areas in Brazil are significantly different from the big cities, but also exposes a large variation among them. Wages are lower than in cities, but so are the costs of amenities, resulting in lower living costs. The out-migrants face the balancing act to reconcile lower earnings with lower living costs and worse amenities.

I therefore estimate the importance of real wages in the destination choice of metropolitan out-migrants and find that migrants maximize their utility by moving into smaller towns not far from their metropolitan origins. In these destinations they face lower nominal wages, but also lower prices. The counterfactual analysis reveals that on average the migrants achieve a positive return in real wages by leaving the city. This finding is especially strong for low-educated workers who would lose out from leaving the big cities if only nominal wages were considered. Non-metropolitan areas have on average worse-quality public service provision. The metropolitan out-migrants reveal a preference for education provision over health services, emphasizing that preferences vary between amenities.

The findings are in line with the literature on wage returns to migration. It is confirmed that the comparison of wages conditional on individual skills is important for the destination choice, but migrants seem to consider them only jointly with living costs (Dahl 2002; Kennan and Walker 2011; Moretti 2011; Tunalı 2000). Furthermore, selection-corrected expected wages entered the model of destination choice with only weak significance, which could indicate that workers have incorrect expectations about their wages and they could do even better in their destination choice. Rather, metropolitan out-migrants choose destinations that reduce their costs of moving as well as their living costs, which is why I find a positive return in real wages. These results suggest that high prices are pushing workers out of metropolitan cities.

References

Adams, R.H. (2006). ‘Remittances and Poverty in Ghana’. Policy Research Working Paper 3838. Washington, DC: World Bank.

Adams, R.H., and A. Cuecuecha (2013). ‘The Impact of Remittances on Investment and Poverty in Ghana’. *World Development*, 50: 24–40.

Aguayo-Téllez, E., M.-A. Muendler, and J.P. Poole (2010). ‘Globalization and Formal Sector Migration in Brazil’. *World Development*, 38(6): 840–56.

Almeida dos Reis, J.G., and R. Paes de Barros (1991). ‘Wage Inequality and the Distribution of Education’. *Journal of Development Economics*, 36(1): 117–43.

Aroca Gonzalez, P., and W.F. Maloney (2005). ‘Migration, Trade, and Foreign Direct Investment in Mexico’. Policy Research Working Paper 3601. Washington, DC: World Bank.

Barham, B., and S. Boucher (1998). ‘Migration, Remittances, and Inequality: Estimating the Net Effects of Migration on Income Distribution’. *Journal of Development Economics*, 55(2): 307–31.

Borjas, G.J. (1987). ‘Self-Selection and the Earnings of Immigrants’. *The American Economic Review*, 77(4): 531–53.

Borjas, G.J., S.G. Bronars, and S.J. Trejo (1992). ‘Self-Selection and Internal Migration in the United States’. *Journal of Urban Economics*, 32(2): 159–85.

Brown, R.P., and E. Jimenez (2008). ‘Estimating the Net Effects of Migration and Remittances on Poverty and Inequality: Comparison of Fiji and Tonga’. *Journal of International Development*, 20: 547–71.

Christiaensen, L., J. De Weerdt, and R. Kanbur (2017). ‘Cities, Towns, and Poverty: Migration Equilibrium and Income Distribution in a Todaro-Type Model with Multiple Destinations’. Discussion Paper 10692. Bonn: Institute for the Future of Labor.

Christiaensen, L., J. De Weerdt, and Y. Todo (2013). ‘Urbanization and Poverty Reduction: The Role of Rural Diversification and Secondary Towns’. *Agricultural Economics*, 44(4–5): 435–47.

Cochrane, S.G., and D.R. Vining (1988). ‘Recent Trends in Migration between Core and Peripheral Regions in Developed and Advanced Developing Countries’. *International Regional Science Review*, 11(3): 215–43.

Dahl, G.B. (2002). ‘Mobility and the Return to Education: Testing a Roy Model with Multiple Markets’. *Econometrica*, 70(6): 2367–420.

Deaton, A., and O. Dupriez (2011). ‘Spatial Price Differences within Large Countries’. Working Paper 1321. Princeton, NJ: Princeton University, Woodrow Wilson School of Public and International Affairs.

do Planejamento, M. (2010). ‘Balanço 4 Anos 2007–2010. PAC Programa de Aceleração do Crescimento’. Report. Brasília: Ministério do Planejamento.

dos Santos Júnior, E.d.R., N. Menezes-Filho, and P.C. Ferreira (2005). ‘Migração, Seleção e Diferenças Regionais de Renda no Brasil’. *Pesquisa e Planejamento Econômico*, 35(3): 299–331.

Fafchamps, M., and F. Shilpi (2013). ‘Determinants of the Choice of Migration Destination’. *Oxford Bulletin of Economics and Statistics*, 75(3): 388–409.

Fally, T., R. Paillacar, and C. Terra (2010). 'Economic Geography and Wages in Brazil: Evidence from Micro-Data'. *Journal of Development Economics*, 91(1): 155–68.

Ferreira, F.H.G., P.G. Leite, and J.A. Litchfield (2006). 'The Rise and Fall of Brazilian Inequality: 1981–2004'. Policy Research Working Paper 3867. Washington, DC: World Bank.

Foguel, M.N., I. Gill, R. Mendonça, and R. Paes De Barros (2015). 'The Public–Private Wage Gap in Brazil'. Ipea Discussion Paper 95. Brasília: Ipea.

Giannetti, M. (2003). 'On the Mechanics of Migration Decisions: Skill Complementarities and Endogenous Price Differentials'. *Journal of Development Economics*, 71: 329–49.

Greene, W. (2000). *Econometric Analysis*, 4th edition. London: Prentice-Hall.

Grogger, J., and G.H. Hanson (2011). 'Income Maximization and the Selection and Sorting of International Migrants'. *Journal of Development Economics*, 95(1): 42–57.

Ham, J.C., X. Li, and P.B. Reagan (2011). 'Matching and Semi-Parametric IV Estimation, a Distance-Based Measure of Migration, and the Wages of Young Men'. *Journal of Econometrics*, 161(2): 208–27.

Harris, J.R., and M.P. Todaro (1970). 'Migration, Unemployment and Development: A Two-Sector Analysis'. *American Economic Review*, 60: 126–42.

Hering, L., and R. Paillacar (2015). 'Does Access to Foreign Markets Shape Internal Migration? Evidence from Brazil'. *World Bank Economic Review*, 30(1): 78–103.

IBGE (2010). 'Censo Demográfico 2010: Características urbanísticas do entorno dos domicílios'. Technical Report. Rio de Janeiro: IBGE.

IBGE (2012). 'Censo Demográfico 2010: Microdados'. Available at: www.ibge.gov.br.

Kennan, J., and J.R. Walker (2011). 'The Effect of Expected Income on Individual Migration Decisions'. *Econometrica*, 79(1): 211–51.

Lall, S.V., C. Timmins, and S. Yu (2009). 'Connecting Lagging and Leading Regions: The Role of Labor Mobility'. Policy Research Working Paper 4843. Washington, DC: World Bank.

Li, C., and J. Gibson (2014). 'Spatial Price Differences and Inequality in the People's Republic of China: Housing Market Evidence'. *Asian Development Review*, 31(1): 92–120.

Lokshin, M., M. Bontch-Osmolovski, and E. Glinskaya (2007). 'Work-Related Migration and Poverty Reduction in Nepal'. Policy Research Working Paper 4231. Washington, DC: World Bank.

Mata, D.D., U. Deichmann, J.V. Henderson, S.V. Lall, and H.G. Wang (2005). 'Examining the Growth Patterns of Brazilian Cities'. Technical Report 3724. Washington, DC: World Bank.

McCormick, B., and J. Wahba (2005). 'Why Do the Young and Educated in LDCs Concentrate in Large Cities? Evidence from Migration Data'. *Economica*, 72(285): 39–67.

McFadden, D. (1974). 'The Measurement of Urban Travel Demand'. *Journal of Public Economics*, 3(4): 303–28.

Menezes-Filho, N.A., and M.-A. Muendler (2011). 'Labor Reallocation in Response to Trade Reform'. NBER Working Paper 17372. Cambridge, MA: NBER.

Moretti, E. (2011). 'Local Labor Markets'. In O. Ashenfelter and D. Card (eds), *Handbook of Labor Economics*, volume 4b. Amsterdam: Elsevier.

Morten, M., and J. Oliveira (2016). ‘Paving the Way to Development: Costly Migration and Labor Market Integration’. Working Paper 568. Stanford, CA: Stanford Center for International Development.

Rodriguez, E.R. (1998). ‘International Migration and Income Distribution in the Philippines’. *Economic Development and Cultural Change*, 46(2): 329–50.

Santos, C., and P.C. Ferreira (2007). ‘Migração e distribuição de renda no Brasil’. *Pesquisa e Planejamento Econômico*, 37(3): 405–26.

Sjaastad, L.A. (1962). ‘The Costs and Returns of Human Migration’. *Journal of Political Economy*, 70(2): 80–93.

Tunalı, I. (2000). ‘Rationality of Migration’. *International Economic Review*, 41(4): 893–920.

Yap, L. (1976). ‘Internal Migration and Economic Development in Brazil’. *Quarterly Journal of Economics*, 90(1): 119–37.

Appendix

Table A1: Balancing statistics before matching

Multivariate L1 distance:	0.83541258						
Univariate imbalance:							
	L1	Mean	Min	25%	50%	75%	max
Age at migration	0.10634	–2.2895	–1	–1	–3	–3	–1
Sex	0.0979	–0.0979	0	0	0	0	0
Education level	0.07352	–0.12922	0	–1	0	0	0
Race	0.00576	0.00576	0	0	0	0	0
City of origin	0.14503	0.72656	0	0	4	0	0
Marital status	0.01909	–0.0098	0	0	0	0	0
Sector of activity	0.15388	–0.96689	0	–3	–1	0	–1

Source: author's compilation based on data.

Table A2: Balancing statistics after matching

Multivariate L1 distance:	0.78689126						
Univariate imbalance:							
	L1	Mean	Min	25%	50%	75%	max
Age at migration	0.04019	–0.15014	0	0	0	–1	0
Sex	0.06926	–0.06926	0	0	0	0	0
Education level	0.03117	–0.03117	0	0	0	0	0
Race	0.00467	–0.00467	0	0	0	0	0
City of origin	0.00479	0.00479	0	0	0	0	0
Marital status	0.02571	–0.02571	0	0	0	0	0
Sector of activity	0.00355	0.00355	0	0	0	0	0

Source: author's compilation based on data.

Table A3: Matching summary

Number of strata		9,796
Number of matched strata		3,785
	Non-migrants	Migrants
All	683,517	16,172
Matched	587,346	15,401
Unmatched	96,171	771

Source: author's compilation based on data.

Table A4: Differences of actual and predicted wages for metropolitan out-migrants, before matching

Log (nominal hourly wages)	*N*	Mean
Observed	14,810	1.767
Predicted	14,810	1.874
Difference	–0.107***	
Log (real hourly wages)	*N*	Mean
Observed	14,810	–2.466
Predicted	14,810	0.303
Difference	0.303***	
Log (real hourly wages) *Hedonic price as denominator*	*N*	Mean
Observed	14,810	–2.353
Predicted	14,810	0.204
Difference	0.204***	

Notes: significance levels * 10, ** 5, and *** 1 per cent for t-test the of difference in means between observed and predicted wages.

Source: author's compilation based on data.

Table A5: Observed and predicted real wage differences using different measures of living costs

Log (real hourly wages)		
High-skilled *Skill-specific mean rents*	*N*	Mean
Observed	2,702	–1.587
Predicted	2,702	–2.098
Difference		0.510***
Skill-specific median rents	*N*	Mean
Observed	2,702	–1.508
Predicted	2,702	–1.948
Difference		0.439***
Median hedonic prices	*N*	Mean
Observed	2,702	–1.161
Predicted	2,702	–1.469
Difference		0.308***

Notes: significance levels * 10, ** 5, and *** 1 per cent for t-test the of difference in means between observed and predicted wages. Predicted wages are based on the matched sample. Rent for room is aggregated at the micro-region level. Skill-specific rents are the rent per room aggregated only for the high- or low-skilled observations respectively in a micro-region applying population survey weights. Once the mean is aggregated, in another case the median. Lastly, I also use the median to aggregate the hedonic housing price measure. These different price measures are used as the denominator to compute real hourly wages.

Source: author's compilation based on data.

Table A6: Observed and predicted real wage differences using different measures of living costs

Log (real hourly wages)		
Low-skilled		
Skill-specific mean rents	*N*	Mean
Observed	11,393	–2.456
Predicted	11,393	–2.691
Difference		0.235***
Skill-specific median rents	*N*	Mean
Observed	11,393	–2.365
Predicted	11,393	–2.589
Difference		0.224***
Median hedonic prices	*N*	Mean
Observed	11,393	–2.374
Predicted	11,393	–2.555
Difference		0.181***

Notes: significance levels * 10, ** 5, and *** 1 per cent for t-test of the difference in means between observed and predicted wages. Predicted wages are based on the matched sample. Rent for room is aggregated at the micro-region level. Skill-specific rents are the rent per room aggregated only for the high- or low-skilled observations respectively in a micro-region applying population survey weights. Once the mean is aggregated, in another case the median. Lastly, I also use the median to aggregate the hedonic housing price measure. These different price measures are used as the denominator to compute real hourly wages.

Source: author's compilation based on data.

Table A7: Regression of housing prices on housing characteristics, OLS estimates

	log(rent per room)		
Urban area	0.256***		
	(0.005)		
Type of dwelling (base = house)		*Waste water (base = network)*	
Townhouse/condominium	0.146***	Well on property	0.007**
	(0.003)		(0.003)
Flat	0.396***	Well outside property	–0.088***
	(0.002)		(0.005)
Hut	0.196***	Carro-pipa	–0.072***
	(0.006)		(0.014)
Wall material (base = bricks coated)		Rainwater cistern	–0.074***
Bricks not coated	–0.160***		(0.028)
	(0.002)	Rainwater other	–0.097
Wood	–0.265***		(0.068)
	(0.003)	Rivers, lakes, etc.	–0.081***
Plaster coated	–0.461***		(0.023)
	(0.015)	Other	–0.155***
Plaster not coated	–0.521***		(0.010)
	(0.020)	Well in village	0.165**
Wood unprepared	–0.344***		(0.066)
	(0.010)	*Canalization (base = inside)*	
Straw	–0.073	On the property	–0.052***
	(0.155)		(0.004)
Others	–0.146***	No	–0.148***
	(0.015)		(0.006)
Bathroom (base = none)		*Garbage (base = collected directly)*	
1	–0.213***	Collected in collective	–0.054***
	(0.006)		(0.002)
2	–0.095***	Burnt	–0.229***
	(0.006)		(0.008)
3	0.047***	Buried	–0.017
	(0.007)		(0.043)
4	0.220***	Tossed in public area	–0.229***
	(0.012)		(0.008)
5	0.355***	Tossed in river, lake, or sea	–0.195***
	(0.027)		(0.036)
6	0.517***	Other	0.005
	(0.054)		(0.027)
7	0.430***	*Electricity (base = company)*	
	(0.119)	Yes, other	–0.094***
8	1.046***		(0.010)
	(0.237)	No electricity	–0.238***
9 or more	0.356***		(0.021)
	(0.083)	Constant	4.235***
Sanitation (base = network)			(0.016)
Septic sump	–0.089***		
	(0.002)		
Rudimentary sump	–0.200***		
	(0.002)		
Ditch	–0.225***		
	(0.005)		
River, lake or, sea	–0.152***		
	(0.004)		
Other	–0.212***		
	(0.009)		
Micro-region dummies	Yes		
Observations	927,192		
R^2	0.539		

Notes: Significance levels * 10, ** 5, and *** 1 per cent. Standard errors are robust. Observations are households.
Source: author's compilation based on data.

Table A8: Coefficients and t-statistics of prediction of wages for migrants based on past migrants at destination, OLS

	Log(hourly wage)	
	Coefficient	t-statistic
Age	0.048	18.250
Age-squared	–0.045	–13.939
Female	–0.368	–61.567
White	0.123	19.385
Education (base = none)		
Primary, secondary incomplete	0.240	28.118
Secondary, higher incomplete	0.530	71.485
Higher complete	1.441	154.831
Mean (Log(hourly wage))	–0.754	

Notes: OLS estimates weighted with population weights. The sample were all migrants who moved more than one year ago to the destinations.

Source: author's compilation based on data.

Table A9: Observed and predicted real wage differences using different measures of living costs, metropolitan in-migrants

Log (real hourly wages)		
High-skilled		
Skill-specific mean rents	*N*	Mean
Observed	1,068	–1.931
Predicted	1,068	–1.974
Difference		0.043*
Skill-specific median rents	*N*	Mean
Observed	1,068	–1.795
Predicted	1,068	–1.894
Difference		0.099***
Low-skilled		
Skill-specific mean rents	*N*	Mean
Observed	7,357	–2.775
Predicted	7,357	–2.501
Difference		–0.274***
Skill-specific median rents	*N*	Mean
Observed	7,357	–2.680
Predicted	7,357	–2.394
Difference		–0.286***

Notes: significance levels * 10, ** 5, and *** 1 per cent for t-test of the difference in means between observed and predicted wages. Predicted wages are based on the matched sample. Rent for room is aggregated at the micro-region level. Skill-specific rents are the rent per room aggregated only for the high- or low-skilled observations respectively in a micro-region applying population survey weights. Once the mean is aggregated, in another case the median. Lastly, I also use the median to aggregate the hedonic housing price measure. These different price measures are used as the denominator to compute real hourly wages.

Source: author's compilation based on data.

Table A10: Variables and data sources

Variable	Description	Source
Variables for descriptive statistics and destination choice model on micro-region level		
Housing prices	Average rent on micro-region level	Census, IBGE
Education provision quality index	Index from 0 to 1, computed based on: subscription rate of pre-school children, dropout rate, rate of teachers with higher education, average daily teaching hours, results of the IDEB (indicator of development of education in Brazil)	FIRJAN*
Health provision quality index	Index from 0 to 1, computed based on: number of prenatal consultations, deaths due to badly defined causes, child deaths due to preventable causes	FIRJAN*
Homicide rate	per 100,000 inhabitants in 2008	Ipeadata
Distance to state capital	Indicator for market access (Fally et al. 2010)	Ipeadata
GDP	Log of GDP in 2009	Ipeadata
Distance between origin and destination	geodesic distance as indicator for fixed moving costs, author's calculation from coordinates	Census, IBGE
Additional variables for wage regression, on individual level		
Partner participation	Dummy whether partner is working	Census, IBGE
Proportion of children in household		Census, IBGE
Marital status	Separated/divorced/widowed, single, married	Census, IBGE
Sector	Public, private, informal, self-employed	Census, IBGE
Industry	21 industries according to International Standard Industrial Classification of all Economic Activities (ISIC)	Census, IBGE
Federal state	27 states	Census, IBGE
Variables for matching, on individual level		
Age	At time of migration, i.e. one year ago	Census, IBGE
Race	White and non-white	Census, IBGE
Education level	Primary, middle, high school, college	Census, IBGE
Micro-region of origin/residency	City of origin for migrants and city of residency for comparison group of non-migrants	Census, IBGE

Source: author's compilation based on data.