

Egger, Eva-Maria

Article

Migrating out of mega-cities: Evidence from Brazil

IZA Journal of Development and Migration

Provided in Cooperation with:

IZA – Institute of Labor Economics

Suggested Citation: Egger, Eva-Maria (2021) : Migrating out of mega-cities: Evidence from Brazil, IZA Journal of Development and Migration, ISSN 2520-1786, Sciendo, Warsaw, Vol. 12, Iss. 1, pp. 1-35, <https://doi.org/10.2478/izajodm-2021-0003>

This Version is available at:

<https://hdl.handle.net/10419/298696>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>

Eva-Maria Egger*

Migrating out of mega-cities: Evidence from Brazil

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 article, the author shows that in Brazil as many people move out as into metropolitan cities and they mostly move to mid-sized towns. The author estimates the determinants of out-migrants' destination choice accounting for differences in earnings, living costs, and amenities and tested whether the migrants gain economically by accepting lower wages but enjoying lower living costs. The findings suggest that in their destination choice, out-migrants aim to minimize costs of moving. On average, city-leavers realize higher real wages, including low-skilled migrants who would lose in nominal terms. The article thus provides new evidence on economic incentives to leave big cities in a middle-income country.

Current version: May 17, 2020
Keywords: internal migration, secondary towns, living costs, Brazil
JEL Codes: J61, R23, C35
Corresponding author: Eva-Maria Egger
egger@wider.unu.edu

United Nations University World Institute for Development Economics Research (UNU-WIDER), Rua da Argélia 116,
Maputo, Mozambique

1 Introduction

The 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 70s and 80s in the U.S. and Europe as the so-called "population turnaround" (Cochrane and Vining, 1988). Recently, the growing literature on migration and labor misallocation in low- and middle-income countries finds evidence for migration patterns that seem somewhat surprising to standard theories of rural–urban migration (Morten and Oliveira, 2018; Bryan and Morten, 2019; Munshi and Rosenzweig, 2016; Hicks et al., 2017). These studies emphasize selectivity and migration costs as major drivers of such unexpected patterns and document implied overall productivity or welfare costs. One driver of such costs could be migrant sorting that is contrary to what classic models of rural–urban migration predict. At the individual level, the migration into urban areas is driven by earning differentials and by the fact that cities offer higher returns than rural areas (Sjaastad, 1962). In this article, the author thus investigates human capital-specific sorting of migrants in one of these unexpected migration patterns the move out of metropolitan cities to mid-sized towns.

Census survey data from Brazil show that around 20% of first-time internal migrants¹ 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%) move to live and work in medium-sized destinations,² not in 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 to move out of the cities, where do they choose to live, and whether their destination choice can still be explained with a gain in earnings.

Urban areas attract workers with job opportunities, high wages, and better services. Yet, with the urbanization waves in developing countries, large cities face many problems associated with overcrowding, such as informal housing, congested infrastructure, and unemployment. City growth increases the 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 higher too, the migration choice, therefore, becomes a balancing act between these factors across a large set of possible destinations.

The author estimates how these factors affect the individual destination choice conditional on migration with a conditional logit model as in Fafchamps and Shilpi (2013) or Chernina (2020). The focus lies on the established determinants of migration: Expected wages based on individual characteristics, moving and living costs, as well as local amenities such as crime rates, and the quality of public education and health service provision. Coarsened exact matching (CEM) is applied to control the selection bias in the prediction of expected wages. Wages in neighboring locations are used as an instrument for living costs. Other papers investigate the sorting decision from an individual perspective. These studies found that workers sort

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

2 The median population size of the administrative unit, a *microregião*, is 173,453 inhabitants. The author classifies a medium-sized *microregião* as one that has between 170,000 and 1 million inhabitants and above 1 million as a metropolitan city.

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; Moretti, 2011; Ham et al., 2011; Grogger and Hanson, 2011), but only a few studies use data from transitioning countries (Aroca Gonzalez and Maloney, 2005; Lokshin et al., 2007; Aguayo-Tellez et al., 2010; Fafchamps and Shilpi, 2013).

Following this analysis, the author further investigates how the realized wages instead of expected ones reflect a gain or loss because of migration from metropolitan cities. The author uses counterfactual wages of migrants to compute the return to out-migration in nominal and real terms. A few studies analyze the counterfactual situation of households that had their member not migrated, but not of the migrants themselves (Barham and Boucher, 1998; Rodriguez, 1998; Tunalı, 2000; Adams, 2006; Lokshin et al., 2007; Brown and Jimenez, 2008; Adams and Cuecuecha, 2013).

The author finds that migrants' destination choices differ by education level. Low-educated metropolitan out-migrants prefer destinations where the cost of living is lower, whereas better-educated migrants do not evaluate prices in a significant way. 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 the strongest for the low-skilled workers who would normally experience a decline in nominal wages from leaving cities. Furthermore, out-migrants of all education levels prefer towns closer to their origin and within their own state of birth, which reduces the economic and social costs of moving. In terms of amenities, crime does not appear to matter. Less educated migrants seek better access to health care facilities, whereas highly educated migrants seem to be willing to accept lower quality in health service provision but prefer destinations where education quality is relatively better.

In this article, the author contributes to the related literature in several ways. To the best of the author's 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 The author exploits the detailed information on migration through a unique census survey. These data allow to calculate migration between local labor markets so that the author captures the largest share of labor mobility within Brazil, thereby improving on studies that investigate internal migration only at the regional level (Yap, 1976; Santos and Ferreira, 2007; dos Santos Junior et al., 2005; Lall et al., 2009; Aguayo-Tellez et al., 2010; Fally et al., 2010). This approach also allows testing the importance of secondary towns as potential destinations for migrants not only from rural areas, adding contribution to the recent literature (Christiaensen et al., 2017). Furthermore, by combining the destination choice and counterfactual analysis, the author assesses the ability of migrants to evaluate expected earnings across a large set of possible destinations. These insights into the location choice of workers are relevant for regional planning (Moretti, 2011).

The article is structured as follows. The data used for the empirical analysis are described in Section 2. Descriptive maps, graphs, and tables that explore the nature of migration from

³ One exception is McCormick and Wahba (2005) who analyze migration in and out of big cities in Egypt. However, their sample of migrants moving out of the big cities is 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.

metropolitan to non-metropolitan cities are described in detail in Section 3. Thereafter, the conceptual framework of the destination choice model is discussed, and the results are presented in Section 4. The results of the counterfactual exercise are presented in Section 5. Finally, Section 6 concludes the article.

2 Data

2.1 Data source

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

2.2 Definition of migration

The Census survey from 2010 allows to identify migrants in the sample using the following questions: “Were you born in this municipality?” – to know whether people are living in their birthplace; “When did you move to this municipality?” – provides 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?” – provides the exact origin of migrants. It further asks for the municipality of the current job and the previous job. Migrants are individuals who used to live and work in a different location than the one they are 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 from 25 years to 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 labor market, which means that they are either employed or unemployed but looking for work. The author restricts the sample of migrants to those who moved within the past year, between 2009 and 2010, 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 labor market and economic activities, a bit larger than counties in the US. The author defines 22 of these *microregiões* as metropolitan based on their population size of 1 million and above.⁴ There are 558 *microregiões*, 29 metropolitan, and 529 non-metropolitan *microregiões*. Information on

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

the local characteristics is aggregated to the *microregião* level using individual-level data from the Census survey. The author uses survey weights to obtain local estimates of wages and housing prices measured with the amount of rent per room.⁵ At the level of the federal state, 43% of the metropolitan out-migrants leave their state of birth, the other 57% 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 captures also intra-regional population dynamics, the largest movements in the country.

2.5 Other variables

The author obtains average wages in 2009 from the National Employment Registry (RAIS). Other information on local characteristics is obtained from *Ipeadata*. This is an online data pool provided by Instituto de Pesquisa Economica 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 quality of education and health provision, and homicide rates as a measure of crime.

Quality of education and health is measured using an index, which is 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, and the results of a national education development score. The index for health provision quality comprises the number of pre-natal consultations, deaths due to badly defined causes, and child deaths due to evitable causes. These are measures commonly used in health economics to capture the quality of health care provision (Hanefeld et al., 2017; Donabedian, 1988). Access to health care provision is measured by the number of health care facilities per 100,000 inhabitants.⁶ These data come from the National Health Establishments Registry (CNES).

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

3 Descriptive Statistics

3.1 Patterns of internal migration in Brazil

Figure 1 shows a map of Brazil. The map shows 5 greater regions and the 22 metropolitan cities of Brazil, which are the focus of this analysis.⁷ The metropolitan cities are located mainly along the coast except for the state capitals in the South-eastern region, Goiânia in the Central-West, Manaus in the Amazon, and the national capital Brasília.

Labor migration within Brazil is historically very common and is mainly attributed to socioeconomic differences between regions, underdeveloped rural areas, and several large

⁵ The average number of rooms across localities is six and its standard deviation 0.59 so that there is no concern of systematic differences in the number of rooms across locations.

⁶ Health care facilities included are general hospitals, day hospitals, polyclinics, health point, general emergency, and pharmacies and basic health centers.

⁷ Some of the metropolitan cities comprise more than one *microregião* so that the number of cities is slightly lower than the number of metropolitan *microregiões*.

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

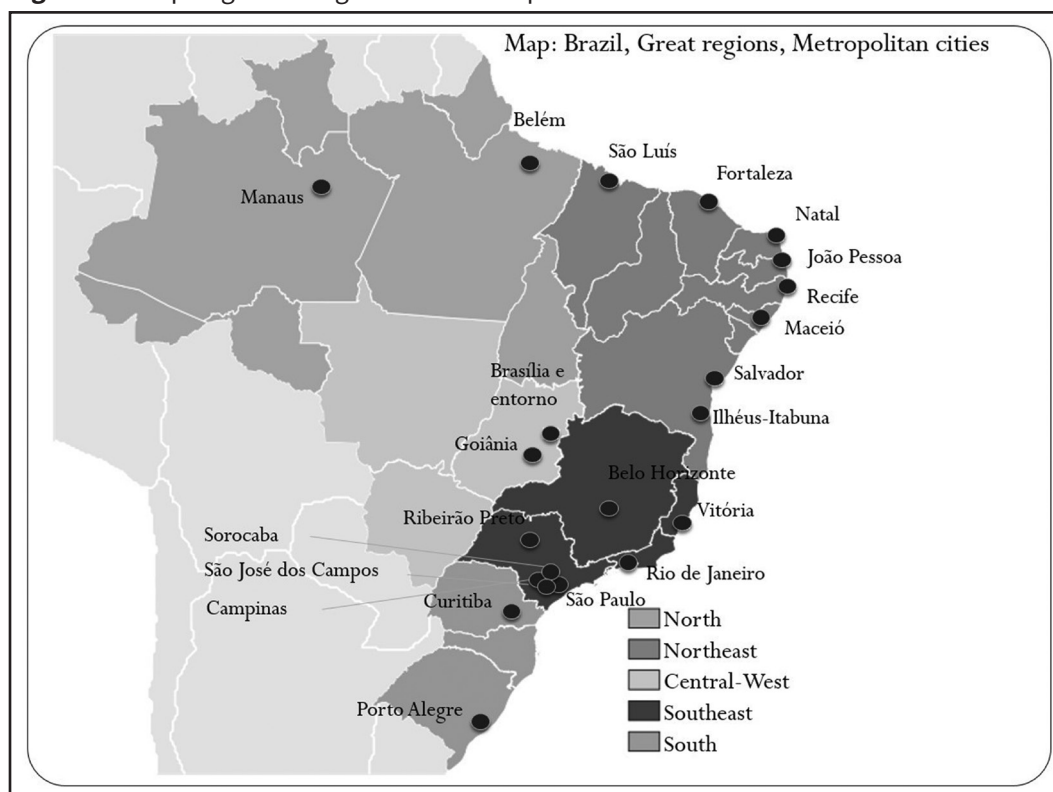


Table 1 Migrants between metropolitan and non-metropolitan *microregiões* between 2009 and 2010

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

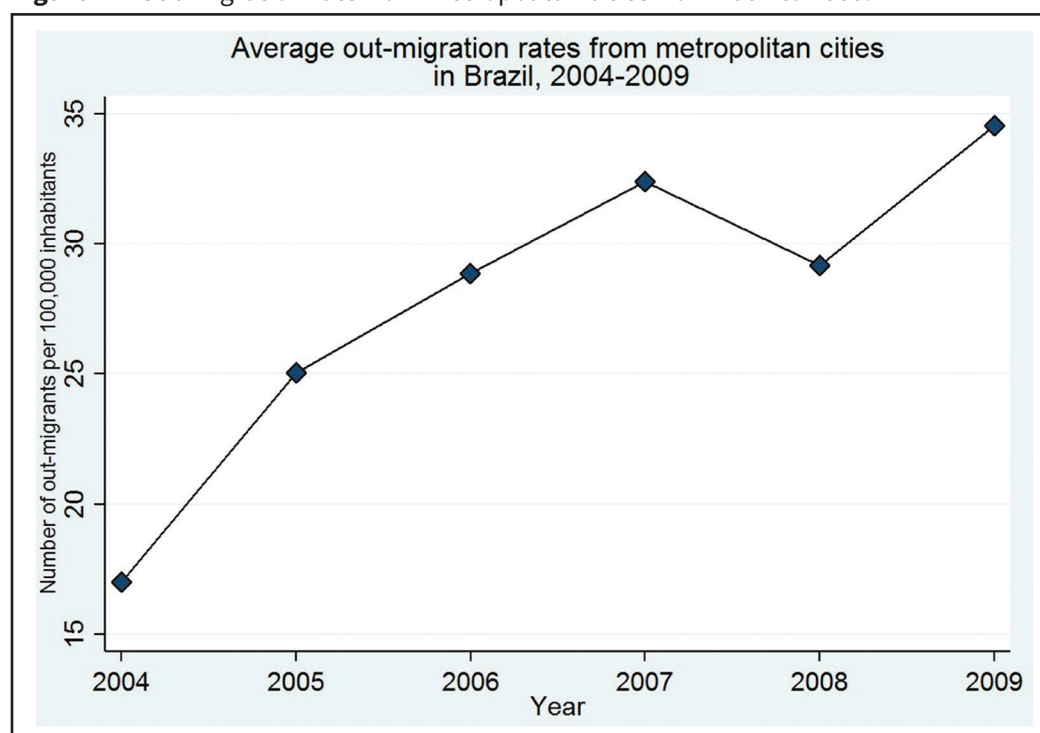
Total *N* = 810,196, using survey weights.

urban centers (Yap, 1976). Recently, migration patterns in Brazil have been changing. Of all Brazilian internal migrants in the year before the Census of 2010, 47% moved between non-metropolitan areas (Table 1). The second-largest movement is into and out of metropolitan cities as well as from and to non-metropolitan *microregiões* comprising around 20% each of all recent migrants, a substantial share of migration in the country. The remaining 12% of migrants move between the metropolises.

The rate of people leaving 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 1 million inhabitants in Brazil from 2004 to 2009.

3.2 Comparing origins and destinations

Table 2 compares metropolitan and non-metropolitan *microregiões* in terms of socioeconomic characteristics. In the second and fourth columns in Table 2, the author also includes the

Figure 2 Out-migration rate from metropolitan cities from 2004 to 2009.**Table 2** Characteristics of metropolitan and non-metropolitan *microregiões* in 2010

	Metropolitan		Non-metropolitan	
	Mean	Coeff. of variation	Mean	Coeff. 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
<i>Share of</i>				
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
<i>Share of workers in</i>				
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
<i>Other measures</i>				
GDP growth 2005–2010	0.16	0.31	0.18	0.79
Health facilities (per 100,000)	16.40	0.42	41.86	0.35
Health quality index (0–1)	0.82	0.09	0.79	0.11
Education quality index (0–1)	0.77	0.14	0.73	0.14
Homicide rate (per 100,000)	38.00	0.54	18.58	0.77

coefficient of variation for the metropolitan and non-metropolitan *microregiões* to illustrate how diverse non-metropolitan areas are.

The 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/sewage, and 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. The author replaced them with 0 due to the way homicides are reported.

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 >50% 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 labor markets are much more formalized in these big cities. The employment share of various sectors is higher for services with 53% in the metropolitan areas and 35% in non-metropolitan *microregiões*. Yet, agriculture in the non-metropolitan areas employs around 30% compared to only 9% 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% of metropolitan residents live with adequate sewage, water, and electricity provision compared to only 36% outside of these cities.⁸ This illustrates the stark spatial inequality not only in economic but also in social aspects.

This gap is similar for 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.

The variation in these characteristics among non-metropolitan *microregiões* is large. The second and fourth columns in Table 1 show the coefficients of variation for metropolitan and non-metropolitan *microregiões*, respectively. It is relatively larger for non-metropolitan areas in almost all categories, such as population, public service worker share, and education quality. This motivates the analysis of the metropolitan out-migrants' destination choice. Labor mobility is expected to respond to this spatial variation of real wages and other socioeconomic 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, the author compares the characteristics of non-metropolitan residents, metropolitan out-migrants, and metropolitan residents. The comparison of these residents with metropolitan out-migrants allows shedding light on the differences between migrants and residents at the origin and destination.⁹

⁸ The definition of which type of sewage, water, and electricity provision is adequate comes from the report on sub-normal agglomerations in Brazil (IBGE, 2010).

⁹ For simplification, in the descriptive analysis, this comparison does not account for the fact that migrants could all be concentrated in a specific sub-set of non-metropolitan *microregiões* so that residents in all non-metropolitan *microregiões* might not represent the exact comparison group of that specific subset.

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
Age	40.25	36.85	40.22
Female	0.41	0.37	0.45
White	0.51	0.51	0.51
<i>Education level</i>			
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

Proportions and means are computed using survey weights.

Table 4 Labor 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
<i>Sector</i>			
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
<i>Industry, ISIC</i>			
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

Proportions and means are computed using survey weights.

Industries include extractive industry processing industry, electricity/gas, sanitation/sewage, 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, and international organizations/foreign institutions.

On average, migrants are 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 with 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 groups.

Table 4 documents the labor market characteristics of out-migrants and residents. Around 12% of workers who left the metropolitan cities for non-metropolitan destinations

are unemployed in contrast to an unemployment rate of only 5% among non-metropolitan residents. This indicates that metropolitan out-migrants are a heterogeneous group and some 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 article aims to disentangle these factors.

More than 60% of non-migrants in metropolitan cities are employed either in the public formal or in the private formal sector, whereas only around 46% 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 with the residents there.

Most migrants work in service sectors. Only a few work in agriculture at the non-metropolitan destinations even though it is the second-largest sector after services in these locations. In metropolitan cities, services are the main sector of employment. 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 article.

4 Destination Choice of Migrants

4.1 Empirical methodology

The empirical analysis focuses on the estimation of the effect of various local attributes on the destination choice of migrants. The analysis is based on a multiple-choice setting presented by McFadden (1974). The empirical application is restricted to those who migrated.¹⁰ As in Fafchamps and Shilpi (2013), the author models destination choice conditional on the individual being a migrant.

¹⁰ The model allows to include also residents in the analysis and assume 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 that some of those who did not move might have done so due to high costs or risk which gives rise to a selection bias in the decision to migrate. By excluding the choice to stay at one's origin, and estimate the destination choice model with migrants only, this specific selection bias does not arise.

Migrants are assumed to choose their location to maximize their utility. Motivated by a random utility model, a migrant i residing in the current location chooses among all possible destinations. Let 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, the author defines $c_j = 0$ if $j = o$. The utility of moving to destination j is assumed to have the following form:

$$U_{ij} = \beta' z_{ij} - c_j + \epsilon_{ij} \quad (1)$$

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

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

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

$$F(\epsilon_{ij}) = \exp(-e^{-\epsilon_{ij}}) \quad (3)$$

The probability of moving to destination j is now modeled 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:

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

This is equal to:

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

Eq. (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} can vary across destinations and also vary across individuals:

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

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 Eq. (5) so that this model automatically controls for any individual-specific factors in the destination choice.¹¹ However, this also implies that the author cannot estimate the effect of such factors, such as the age of the migrant, etc. Hence, the alternative specific conditional logit model takes the following form:

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

¹¹ In some applications, e.g., 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.

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:

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

For the main estimation, there are N observations and regressors for each of the 14,509 metropolitan out-migrants.¹² They choose from $J = 514$ possible non-metropolitan *microregiões* as destinations. Only one of the destinations will have a positive outcome as a chosen destination, i.e., the one observed in the data. This results in 5,730,782 individual-destination observations for the multivariate analysis.

Based on the human capital migration model (Sjaastad, 1962), the destination attributes of interest in this analysis are wages and prices. Observed average wages and prices at *microregião* level in the 2010 data are equilibrium results from local economic dynamics, among other factors people's location choices. The simultaneous inclusion of these variables in the model, therefore, raises endogeneity concerns.

To address this issue, the author computes expected wages instead of applying the observed wages realized by migrants at their destination. Average wages in a location need not reflect the wages a migrant can expect to earn. The author, therefore, predicts expected wages for migrants based on their characteristics and the coefficients from a wage estimation of residents at each location.

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

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

Log hourly wages of individual i in location j are determined by the individual characteristics a_i^j , E_i^j , and 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 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, e.g., in more rural areas, households tend to be larger and female labor force participation lower so that wages would be overestimated in these areas if this was not controlled for. The author uses the survey weights in these regressions to make the estimates representative of the population.

For each *microregião*, the coefficients $\hat{\alpha}_j, \hat{\beta}_j, \hat{\gamma}_j$ 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 $\hat{\delta}_j$:

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

¹² The sample of metropolitan out-migrants is slightly reduced as the author only includes those in the analysis who are matched so that the results are comparable. Those dropped were not matched.

The coefficients of the wage predictions corresponding to Eq. (10) (see Table A8 in Appendix) confirm the relationships documented in the literature: Age has a positive, but diminishing effect on expected wages, women earn less than men, white Brazilians 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, the author showed that migrants differ from the resident population in several observable characteristics. This implies that the expected wage measures used in the analysis so far could be biased by unobservable characteristics. The author thus estimates another measure for expected wages that should reduce the selection bias. The author predicts expected wages in non-metropolitan destinations from a sample of previous migrants from the same origin as the migrant. These migrants have moved more than a year ago to the destinations. They are assumed to be more comparable to migrants than residents in terms of unobservable characteristics specific to migrants, for example, risk-taking preferences.

For the metropolitan origins, the author predicts expected wages based on a matched sample of residents at the origin. The author applies 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 the year at migration), race, education level, marital status, sector of activity, and city of origin. Balance statistics are presented in Tables A1–A3 in Appendix.

With this approach, expected wages are allowed to vary across individuals and locations in the individual choice model. In this setting, the endogeneity of wages would only arise if we believed that the destination choice of an individual systematically determines the expected wage level at all possible destinations also for other individuals. This is not the case as the model controls for individual-specific effects and employs individual-specific expected wages.

This approach does not apply to prices. Living costs vary by destination only and thus could be driven by location choice. The use of an instrumental variable is required that would predict differences in living costs across locations, but not destination choices. The alternative specific conditional logit model does not allow for the application of an instrumental variable approach as it does not have normally distributed error terms. The author thus proposes a reduced form regression using the mean of average monthly municipality wages in the contingent *microregiões* surrounding a given *microregião* as an instrument for local living costs. This approach follows Morten and Oliveira (2018) who used trade-cost weighted Bartik shocks in neighboring locations as an instrument for local prices. The rationale is simply that if productivity in neighboring locations increases, prices for traded goods increase, too. While the authors employed the dynamic approach of Bartik-shocks in wage growth, the author uses simply the average of monthly mean wages of the year before migration (2009). Figure A1 in Appendix shows a strong and positive correlation between these indicators (coeff.: 1.59; SE = 0.06) confirming that this is a valid approach in this setting.

Migration costs are measured by the Euclidean distance between origin and destination in kilometers.¹³ Distances are first computed between municipality pairs and then collapsed to the level of *microregiões* weighting for the inter-municipal migration flow. In cases with zero migration flows between municipality pairs, the author assigned a very small positive value ($0.1e-12$) as migration flow to be used for the weighting to avoid losing municipality pairs.¹⁴ Additionally, the author includes a dummy of whether a migrant moved out of her state of birth to a different state. This captures the 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 also imply the social cost of being farther away from one's family and friends. The author further controls for the local population to test for the different attractiveness of non-metropolitan *microregiões* due to varying population sizes.

4.2 Results

In the specific application of this paper, metropolitan out-migrants choose their destination not only based on destination attributes but also these attributes relative to the attributes of migrants' metropolitan origins. For each location attribute, the author thus computes the difference between the destination and origin, e.g., the expected wages in destination j minus expected wages 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 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, -0.53 compared

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

Difference between destination and origin in	Chosen destination	Alternative destinations	t-statistic, difference in mean
Expected hourly wages (log)	-0.53	-0.61	-24.7
Matched expected wages (log)	-2.73	-2.79	-16.2
Rent per room (log)	-0.55	-0.64	-21.9
IV (wages in neighboring MRs, log)	-0.10	-0.14	-23.5
Population in thousands	-5,605	-6,326	-17.1
Homicide rate	-17.66	-14.11	18.2
Health facilities (per 100,000)	25.49	26.46	8.1
Health provision quality index (0–1)	-0.03	-0.05	-24.0
Education provision quality index (0–1)	-0.00	-0.04	-32.8
Distance to origin (km)	573	1,295	108.6
Other state than origin	0.45	0.92	202.4

¹³ The author also used road travel time based on OpenStreetMap. This measure is positively correlated with Euclidean distance (0.87, significant at 1% level) and results do not change in their sign nor significance. Results are available upon request.

¹⁴ 771 migrant observations are dropped as the coordinates of their municipalities of origin or destination were imprecise. These locations are not systematically located in a specific region or remote location and can thus be assumed random.

to -0.61 . This also applies for selection-corrected expected wages (Matched wages). Similarly, housing prices, 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 proposed instrument for prices, wages in neighboring *microregiões*, reflects the pattern of actual living costs (rent per room).

Chosen destinations are on average larger in population size 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 lower levels of homicide rates in contrast to alternative destinations and metropolitan origins. There are also relatively more health facilities in chosen destinations and health care provision is of better quality. Education service provision is also on average better in the chosen *microregião* than in alternative destinations. These averages are all statistically different between chosen and alternative destinations.

Another clear pattern is revealed with regards to the geographic and social distance of chosen destinations. About 45% of chosen destinations are in a different state than the origin contrasting 92% of the destination alternatives. This is also reflected in a lower average distance of chosen destinations to the migrants' origin. Many of these factors are highly correlated with each other which makes it necessary to apply multivariate analysis to disentangle their influence on the metropolitan out-migrants' destination choice.

Many of these factors are highly correlated with each other which makes it necessary to apply multivariate analysis to disentangle their influence on the metropolitan out-migrants' destination choice.

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 Eq. (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 the author will do later in Section 4.3.

The specifications in columns 1 and 2 of Table 6 are that of only wage and price differences, migration costs in form of distance and controlling for population size. In column 1 of Table 6, the author presents the naive estimators of expected wages and local living costs not accounting for selection bias in wages and endogeneity of prices. Both coefficients are insignificant. Only distance and whether a destination is in the same state significantly predict migrants' destination choice. Migrants prefer destinations that are closer to their origin and within their state of birth. These results remain unchanged across specifications.

Column 2 of Table 6 introduces the selection-corrected expected wages and the instrument for prices, wages in neighboring locations. The coefficients remain insignificant, but the sign for wages switches. This indicates that observable and unobserved characteristics of migrants led to an upward bias in expected wage differences. In column 3 of Table 6, the author then further includes other local amenities. In terms of public service provision, it appears that migrants show preferences for locations with relatively more health facilities, but they accept a relatively lower quality of health service provision. They also significantly prefer locations with relatively higher levels of education provision quality.

Table 6 Destination choice conditional on migration, alternative specific logit

	(1)	(2)	(3)
<i>Wage measure:</i>	Expected wages (log)	Matched expected wages (log)	
<i>Price measure:</i>	Rent per room (log)	Wages in neighboring locations (log)	
<i>Difference in:</i>			
Wages	0.054 (0.175)	-0.041 (0.234)	-0.069 (0.244)
Prices	-0.173 (0.213)	-0.805 (0.730)	-0.822 (0.631)
Population (log)	-0.019 (0.068)	-0.040 (0.064)	0.011 (0.080)
Homicide rate			0.004 (0.004)
Health facilities			0.008*** (0.003)
Health quality index			-1.292* (0.758)
Education quality index			1.569* (0.929)
<i>Destination specific:</i>			
Distance to origin (log)	-0.524*** (0.085)	-0.523*** (0.087)	-0.521*** (0.087)
Other state	-1.800*** (0.265)	-1.850*** (0.258)	-1.853*** (0.250)
Observations	5730782	5730782	5730782
Wald chi2	742	1222	1367
Number of cases	14509	14509	14509
Number of alternatives	514	514	514

Standard errors are clustered at the metropolitan *microregião* of origin. Estimator is alternative specific conditional logit. In each column, the first set of regressors is the difference between destination and origin for each destination alternative. The second set, indicated as *Destination specific*, is measured at destination relative to the origin. Prices are measured with the rent per room. Columns 2 and 3 use expected wage differences based on past migrants at the destination and matched residents at origin as explained in Section 4.1. Significance levels *10%, **5%, ***1%.

In summary, the results of the multivariate analysis did not confirm that wages and prices matter significantly for the destination choice of metropolitan out-migrants. However, physical and social moving costs appear to matter. Finally, the author documents that amenities in terms of public service provision are significantly correlated with the destination choice.

One reason why wages and living costs do not enter significantly could be that they matter only for specific groups of migrants. The author, therefore, has run the analysis for sub-samples of migrants with different levels of education. Results are presented in Table 7. In each column, the author used selection-corrected expected wages and the instrument for prices. Column 1 shows the results for migrants with no or only primary education, column 2 for those with lower secondary education, column 3 for upper secondary, and column 4 for those with higher education.

Table 7 Destination choice conditional on migration by the education of migrant, alternative specific logit

	(1)	(2)	(3)	(4)
Level of education:	None or primary	Lower secondary	Upper secondary	Higher
<i>Difference in:</i>				
Matched expected wages (log)	-0.279 (0.255)	-0.161 (0.328)	-0.023 (0.238)	0.336 (0.238)
Prices (IV)	-1.360* (0.700)	-1.205* (0.691)	-0.686 (0.638)	0.202 (0.563)
Population (log)	-0.028 (0.093)	0.049 (0.112)	0.022 (0.076)	0.036 (0.069)
Homicide rate	0.004 (0.004)	0.003 (0.003)	0.006 (0.004)	0.000 (0.004)
Health facilities	0.009*** (0.004)	0.010** (0.004)	0.008** (0.003)	0.002 (0.003)
Health quality index	-0.787 (0.950)	-0.754 (0.953)	-1.668* (0.886)	-2.184** (0.852)
Education quality index	1.684 (1.128)	1.238 (1.114)	1.977* (1.019)	1.055 (0.693)
<i>Destination specific:</i>				
Distance to origin (log)	-0.496*** (0.086)	-0.574*** (0.081)	-0.515*** (0.102)	-0.538*** (0.104)
Other state	-1.963*** (0.253)	-1.808*** (0.261)	-1.881*** (0.281)	-1.663*** (0.282)
Observations	1871193	954109	1840023	1065457
Wald chi2	765	1255	1658	2850
Number of cases	4835	2425	4598	2651
Number of alternatives	514	514	514	514

Standard errors are clustered at the metropolitan *microregião* of origin. Estimator is alternative specific conditional logit. Expected wage differences are based on past migrants at destination and matched residents at origin as explained in Section 4.1. Prices are measured with average municipality wages from neighboring *microregiões*. In each column, the first set of regressors is the difference between destination and origin for each destination alternative. The second set, indicated as *Destination specific*, is measured at destination relative to the origin.

Significance levels *10%, **5%, ***1%.

Three results can be highlighted in Table 7. First, expected wages do not matter significantly for the destination choice of any migrant group, but living costs do for those migrants with lower skill levels. Low-educated migrants (columns 1 and 2) significantly prefer destinations with relatively lower living costs than alternative locations and their metropolitan origins. If the price difference between a destination option and the origin increases, implying that prices are closer to those of the expensive origins, then migrants are less likely to choose such a destination. This does not apply to migrants with secondary schooling or higher education. Second, lower educated migrants give significant importance to the availability of health facilities (access and quantity) but not to the quality of health service provision or education. More educated migrants show a significant preference for education

quality (column 3) but are willing to accept lower health provision quality. Third, migration costs in terms of geographical and social distance enter the destination choice equally for all education groups.

4.3 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 across variables and sub-samples, we can look at one destination alternative, e.g., the one with a price difference very close to the average price difference to metropolitan origins. The author takes the significant regressors from the full specification with selectivity robust wage measures and the price instrument as in Table 7 by education level and computes 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:

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

where β_x is the coefficient of the destination attribute from the conditional logit estimation (Greene, 2000). The elasticities for significant covariates for each education group are presented in Table 8:

The elasticities reveal that by far the largest effect on migrants' destination choice is that of distance and a migrant leaving her/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 labor mobility (Morten and Oliveira, 2018). Table 7 further showed that lower educated migrants value the presence of health facilities, but the elasticity of this amenity is relatively small compared to those of distance. Of similar size is the elasticity of living costs, which was also only significant for the sample of less educated migrants.

A 10% increase in the price difference between the metropolitan origin of a migrant and this specific location makes it on average 2% more likely to be chosen as a destination by migrants with no or only primary education. The range of living costs differences is around 148% of its mean so that this is not a negligible effect. Effect sizes for education and health quality are very small.

Table 8 Elasticities of significant covariates by sub-sample

Education	No or primary	Lower secondary	Upper secondary	Higher
Distance (log)	-3.4	-4.0	-3.6	-3.7
Other state	-1.8	-1.7	-1.7	-1.5
Health facilities	0.2	0.3	0.2	0.1
Prices (IV)	0.2	0.2	0.1	0.0
Education quality	-0.1	-0.1	-0.1	0.0
Health quality	0.0	0.0	0.1	0.1

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 the author controls for self-selection of migrants. This may be 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 with 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 destination, y_d , and at the origin, y_o :

$$r = y_d - y_o \quad (12)$$

In the empirical application, income is proxied by the log of hourly wages, W .¹⁵ 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

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

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

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

$$ATT = E(Y^1 - Y^0 | M = 1) - E(Y^1 | M = 1) - E(Y^0 | M = 1) \quad (15)$$

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. The author only observes wages for migrants at their destination and non-migrants at the origin. If wages were estimated using 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 where spatial wage differences have been found to reflect variation in labor force composition and industry concentration. In Brazil, the labor 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

¹⁵ The author chooses 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.

(Almeida dos Reis and Paes de Barros, 1991; Foguel et al., 2015; Ferreira et al., 2006). Furthermore, Brazilian workers have shown little mobility across industries so that it seems reasonable to focus on self-selection by location and not by sector (Menezes-Filho and Muendler, 2011; Hering and Paillacar, 2015).

The author, therefore, uses 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 cities. Real wages are computed using the local average rent per room as a 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 the author accounts 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 in Appendix). This indicates an overestimation of wages at origin when not accounting for selection and it suggests that out-migrants are negatively selected from the metropolitan working population.

Table 10 documents heterogeneity in wage returns along with the education level of migrants. The author defines 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. Results show that real wages are higher at the destination than origin for both groups. For highly 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 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***

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

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

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

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

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 with the wage distribution. Figures 3 and 4 show the wage distributions of workers who have moved out of metropolitan areas. 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.¹⁶

In the analysis of the destination choice of migrants, we learned that low-educated metropolitan out-migrants choose between destinations where their gain in lower living costs is the 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.

¹⁶ 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 plots of actual and predicted nominal wages of metropolitan out-migrants with matching.

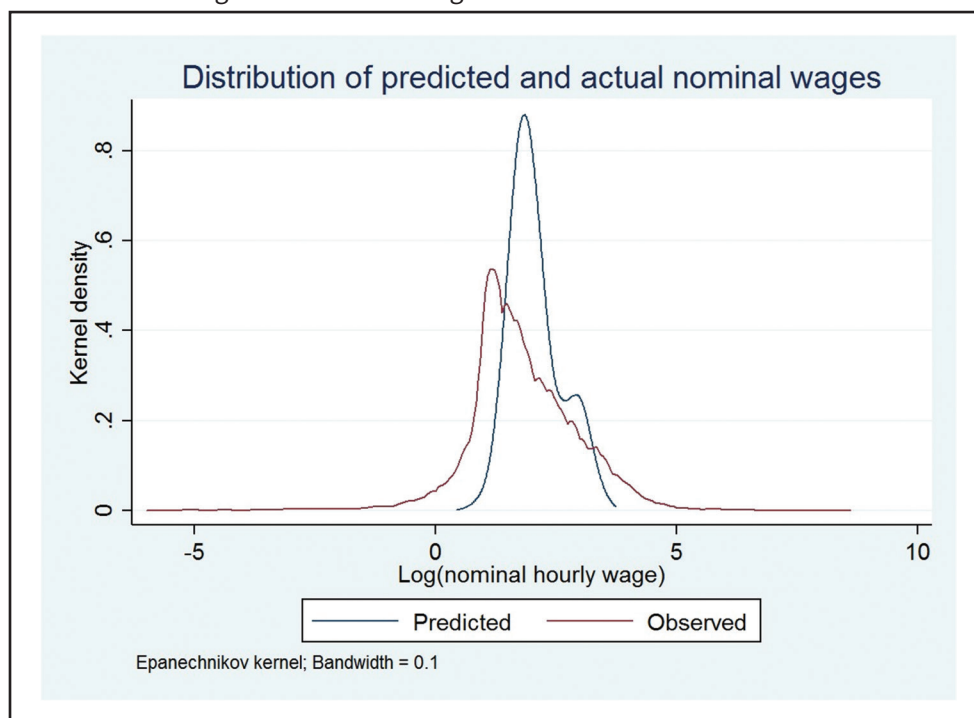
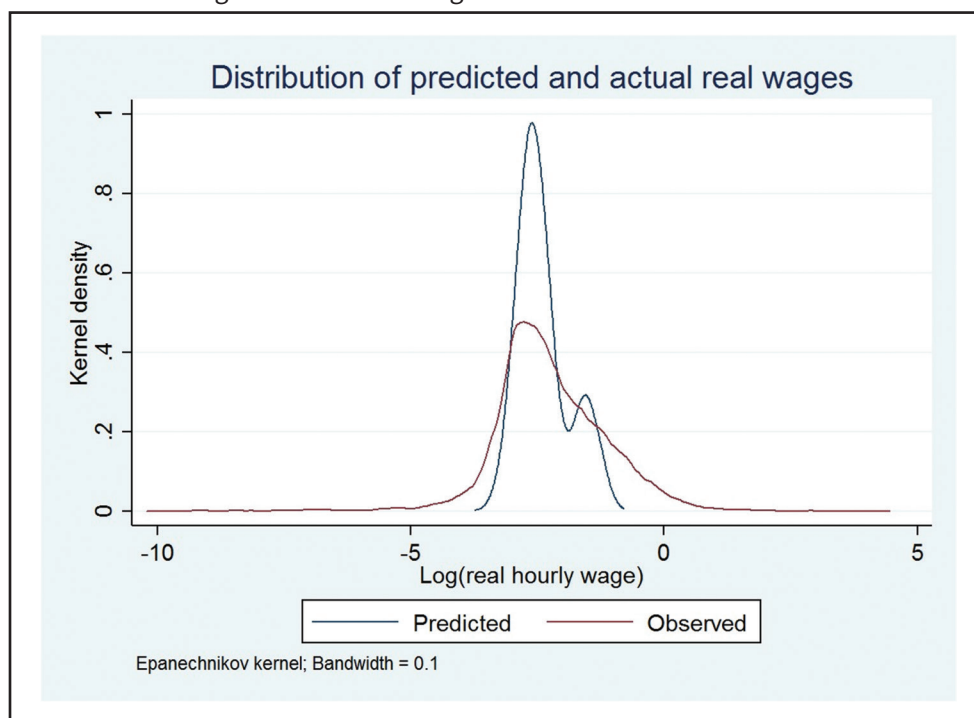


Figure 4 Kernel density plots of actual and predicted real wages of metropolitan out-migrants with matching.



To understand better how migrants evaluate possible destinations, the author contrasts the previous results with those of people moving into metropolitan cities applying the same methodology. Table A9 in 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 have never considered living costs as a significant factor so that they fail to account for them in their destination choice.

5.3 Robustness: Price measures

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 of the unit. The author used these 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), the author constructs 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 sewage, wastewater, and electricity access. With these variables, the author can estimate a hedonic housing price for each location. The author regresses the rent per room on these characteristics weighted by the household survey weight and the author includes a dummy for each *microregião*. These regression results are presented in Table A7 in Appendix. The coefficients of the *microregião* dummies capture any location-specific amenities that contribute to spatial price differences. The author extracts these estimates to construct a location-specific hedonic living cost measure. This variable is independent of differences in housing quality.

The author, therefore, also computes the real wages using the hedonic price as a denominator. Results are presented in Table 11. The wage difference remains positive and statistically significant, but it is smaller by around a third than in the initial results.

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. The author, therefore, conducts the counterfactual analysis using education-group specific living costs as a denominator when computing real wages (see Tables A5 and A6 in Appendix). The author also uses 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 the author uses 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

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

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

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

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 the author did not account for the education-group specific rents. The opposite applies to 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 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 of 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 (Mata et al., 2005; Lall et al., 2009; do Planejamento, 2010).

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 documents that non-metropolitan areas in Brazil are significantly different from the big cities, but expose a large variation among them. Wages are lower than in cities, but so are amenities resulting in lower living costs. The out-migrants face the balancing act to reconcile lower earnings with lower living costs and worse amenities.

The author, therefore, estimates the importance of real wages in the destination choice of metropolitan out-migrants and finds 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 to leaving the city. This finding is especially strong for low-educated workers who would lose from leaving the big cities if only nominal wages were considered and who significantly prefer cheaper locations in their destination choice. Non-metropolitan areas have on the average worse quality of public service provision. Better educated metropolitan out-migrants reveal a preference for education provision over health service emphasizing that preferences vary between amenities and by skill level.

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 (Tunali, 2000; Dahl, 2002; Kennan and Walker, 2011; Moretti, 2011). Furthermore, selection corrected expected wages entered the model of destination choice insignificantly, 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 the author finds a positive return in real wages. These results suggest that high prices are pushing workers out of metropolitan cities.

Declarations

Availability of data and material

The dataset analyzed during the current study is the Censo Demografico 2010 of Brazil, and it is publicly available online at the website of the national statistics office of Brazil IBGE: <https://www.ibge.gov.br/estatisticas/sociais/populacao/22827-censo-2020-censo4.html?=&t=microdados>. The public availability of these data is at the discretion of IBGE. Other complementary data were accessed through the data portal ipeadata.gov.br.

Competing interests

The author declares that there is no competing interests.

Funding

The author grateful for the financial support in the form of a doctoral grant by the Economic and Social Research Council (ESRC) and the research program consortium “Migrating out of Poverty” (U.K. Department for International Development). The latter funding source determined that this study should look at the topic of migration. Neither of these funding sources had any influence on the design, data, analysis, or interpretation of data in this study. The author further acknowledge the funding of open access through the UNU-WIDER ‘Social mobility in the Global South – concepts, measures, and determinants’ project.

Authors' contributions

Not applicable.

Acknowledgements

The author indebted to Andy McKay and Amalavoyal Chari for their constant support throughout this work. Further, the author thanks Corrado Giulietti, Panu Pelkonen, L. Alan Winters, and Michael Keller for helpful comments and also grateful to an anonymous reviewer for the comments that substantially improved this paper. Any errors are mine alone.

References

- Adams, R. H. (2006): Remittances and Poverty in Ghana. Policy Research Working Paper Series. Washington, D.C.: The World Bank.
- Adams, R. H.; A. Cuecuecha (2013): The Impact of Remittances on Investment and Poverty in Ghana. *World Development* 50, 24-40. doi:10.1016/j.worlddev.2013.04.009.
- Aguayo-Tellez, E.; M. A. Muendler; J. P. Poole (2010): Globalization and Formal Sector Migration in Brazil. *World Development* 38, 840-856. doi:10.1016/j.worlddev.2010.02.018.
- Almeida dos Reis, J. G.; R. Paes de Barros (1991): Wage Inequality and the Distribution of Education. *Journal of Development Economics* 36, 117-143. doi:10.1016/0304-3878(91)90007-1.
- Aroca Gonzalez, P.; W. F. Maloney (2005): Migration, Trade, and Foreign Direct Investment in Mexico. Policy Research Working Paper Series 3601. Washington, D.C.: The World Bank.
- Barham, B.; S. Boucher (1998): Migration, Remittances, and Inequality: Estimating the Net Effects of Migration on Income Distribution. *Journal of Development Economics* 55, 307-331. doi:10.1016/S0304-3878(98)90038-4.
- Borjas, G. J. (1987): Self-Selection and the Earnings of Immigrants. *The American Economic Review* 77, 531-553.
- Borjas, G. J., S. G. Bronars; S. J. Trejo (1992): Self-Selection and Internal Migration in the United States. *Journal of Urban Economics* 32, 159-185. doi:10.1016/0094-1190(92)90003-4.
- Brown, R. P.; 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, 547-571. doi:10.1002/jid.
- Bryan, G.; M. Morten (2019): The Aggregate Productivity Effects of Internal Migration: Evidence from Indonesia. *Journal of Political Economy* 127, 2229-2268.
- Chernina, E. M. (2020): The Role of Migration Experience in Migrants' Destination Choice. *IZA Journal of Development and Migration* 11, 1-29. doi:10.2478/izajodm-2020-0001
- Christiaensen, L.; J. De Weerd; R. Kanbur (2017): Cities, Towns, and Poverty: Migration Equilibrium and Income Distribution in a Todaro-type Model with Multiple Destinations. Discussion Paper Series 10692. Bonn: Institute for the Future of Labor (IZA).
- Christiaensen, L.; J. De Weerd; Y. Todo (2013): Urbanization and Poverty Reduction: The Role of Rural Diversification and Secondary Towns. *Agricultural Economics* 44, 435-447. doi:10.1111/agec.12028.

- Cochrane, S. G.; D. R. Vining (1988): Recent Trends in Migration between Core and Peripheral Regions in Developed and Advanced Developing Countries. *International Regional Science Review* 11, 215-243. doi:10.1177/016001768801100301.
- Dahl, G. B. (2002): Mobility and the Return to Education: Testing a Roy Model with Multiple Markets. *Econometrica* 70, 2367-2420.
- Donabedian, A. (1988): The Quality of Care. How Can It Be Assessed? *JAMA* 260, 1743-1748. doi:10.1001/jama.1988.03410120089033.
- do Planejamento, M. (2010): Balanço 4 Anos 2007 - 2010. PAC Programa de Aceleracao do Crescimento. Report. Brasília: Ministerio do Planejamento.
- dos Santos Junior, E. D. R.; N. Menezes-Filho; P. C. Ferreira (2005): Migracao, Selecao e Diferencas Regionais de Renda no Brasil. *Pesquisa e Planejamento Economico* 35, 299-331.
- Fafchamps, M.; F. Shilpi (2013): Determinants of the Choice of Migration Destination. *Oxford Bulletin of Economics and Statistics* 75, 388-409. doi:10.1111/j.1468-0084.2012.00706.x.
- Fally, T.; R. Paillacar; C. Terra (2010): Economic Geography and Wages in Brazil: Evidence from Micro-Data. *Journal of Development Economics* 91, 155-168. doi:10.1016/j.jdeveco.2009.07.005.
- Ferreira, F. H. G.; P. G. Leite; J. A. Litchfield (2006): The Rise and Fall of Brazilian Inequality: 1981-2004. Policy Research Working Paper Series 3867. Washington, D.C.
- Foguel, M. N.; I. Gill; A. R. Mendonc; 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, 329-349.
- Greene, W. (2000): *Econometric Analysis*. 4th ed. London: Prentice-Hall International (UK).
- Grogger, J.; G. H. Hanson (2011): Income Maximization and the Selection and Sorting of International Migrants. *Journal of Development Economics* 95, 42-57. doi:10.1016/j.jdeveco.2010.06.003.
- Ham, J. C.; X. Li; 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, 208-227. doi:10.1016/j.jeconom.2010.12.004.
- Hanefeld, J.; T. Powell-Jackson; D. Balabanova (2017): Understanding and Measuring Quality of Care: Dealing with Complexity. *Bulletin of the World Health Organization* 95, 368-374. doi:10.2471/BLT.16.179309.
- Harris, J. R.; M. P. Todaro (1970): Migration, Unemployment and Development: A Two-Sector Analysis. *American Economic Review* 60, 126-142.
- Hering, L.; R. Paillacar (2015): Does Access to Foreign Markets Shape Internal Migration? Evidence from Brazil. *World Bank Economic Review* 30, 78-103.
- Hicks, J. H.; M. Kleemans; N. Y. Li; E. Miguel (2017): Reevaluating Agricultural Productivity Gaps with Longitudinal Microdata. NBER Working Paper Series 23253. Cambridge, MA: National Bureau of Economic Research.
- IBGE (2010): Censo Demografico 2010: Caracteristicas urbanisticas do entorno dos domicilios. Technical Report. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística - IBGE.
- IBGE (2012): Censo Demografico 2010: Microdados. www.ibge.gov.br.
- Kennan, J.; J. R. Walker (2011): The Effect of Expected Income on Individual Migration Decisions. *Econometrica* 79, 211-251. doi:10.3982/ECTA4657.
- Lall, S. V.; C. Timmins; S. Yu (2009): Connecting Lagging and Leading Regions: The Role of Labor Mobility. Policy Research Working Paper Series 4843. The World Bank.
- Li, C.; J. Gibson (2014): Spatial Price Differences and Inequality in the People's Republic of China: Housing Market Evidence. *Asian Development Review* 31, 92-120.
- Lokshin, M.; M. Bontch-Osmolovski; E. Glinskaya (2007): Work-Related Migration and Poverty Reduction in Nepal. Policy Research Working Paper Series 4231. Washington, D.C.: The World Bank.
- Mata, D. D.; U. Deichmann; J. V. Henderson; S. V. Lall; H. G. Wang (2005): Examining the Growth Patterns of Brazilian Cities. Technical Report 3724. Washington, D.C.: The World Bank.
- McCormick, B.; J. Wahba (2005): Why do the Young and Educated in LDCs Concentrate in Large Cities? Evidence from Migration Data. *Economica* 72, 39-67. doi:10.1111/j.0013-0427.2005.00401.x.
- McFadden, D. (1974): The Measurement of Urban Travel Demand. *Journal of Public Economics* 3, 303-328. doi:10.1108/eb028592.
- Menezes-Filho, N. A.; M. A. Muendler (2011): Labor Reallocation in Response to Trade Reform. NBER Working Paper Series 17372. Cambridge, MA: NBER.
- Moretti, E. (2011): Local Labor Markets, in: *Handbook of Labor Economics*, Volume 4b, Chapter 14. Amsterdam: Elsevier. 1237-1313. doi:10.1016/S0169-7218(11)02412-9.

- Morten, M.; J. Oliveira** (2018): The Effects of Roads on Trade and Migration: Evidence from a Planned Capital City. NBER Working Paper Series 22158. Cambridge, MA: National Bureau of Economic Research.
- Munshi, K.; M. Rosenzweig** (2016): Networks and Misallocation: Insurance, Migration, and the Rural-Urban Wage Gap. *American Economic Review* 106, 46-98. doi:10.1257/aer.20131365.
- Rodriguez, E. R.** (1998): International Migration and Income Distribution in the Philippines. *Economic Development and Cultural Change* 46, 329-350.
- Santos, C.; P. C. Ferreira** (2007): Migracao e distribuicao de renda no Brasil. *Pesquisa e Planejamento Economico* 37, 405-426.
- Sjaastad, L. A.** (1962): The Costs and Returns of Human Migration. *The Journal of Political Economy* 70, 80-93.
- Tunali, I.** (2000): Rationality of Migration. *International Economic Review* 41, 893-920. doi:10.1111/1468-2354.00089.
- Yap, L.** (1976): Internal Migration and Economic Development in Brazil. *The Quarterly Journal of Economics* 90, 119-137. doi:10.2307/1886089.

Appendix

Figure A1 Price instrument (average wages in contiguous *microregiões*) and prices (average municipality rents in *microregião*).

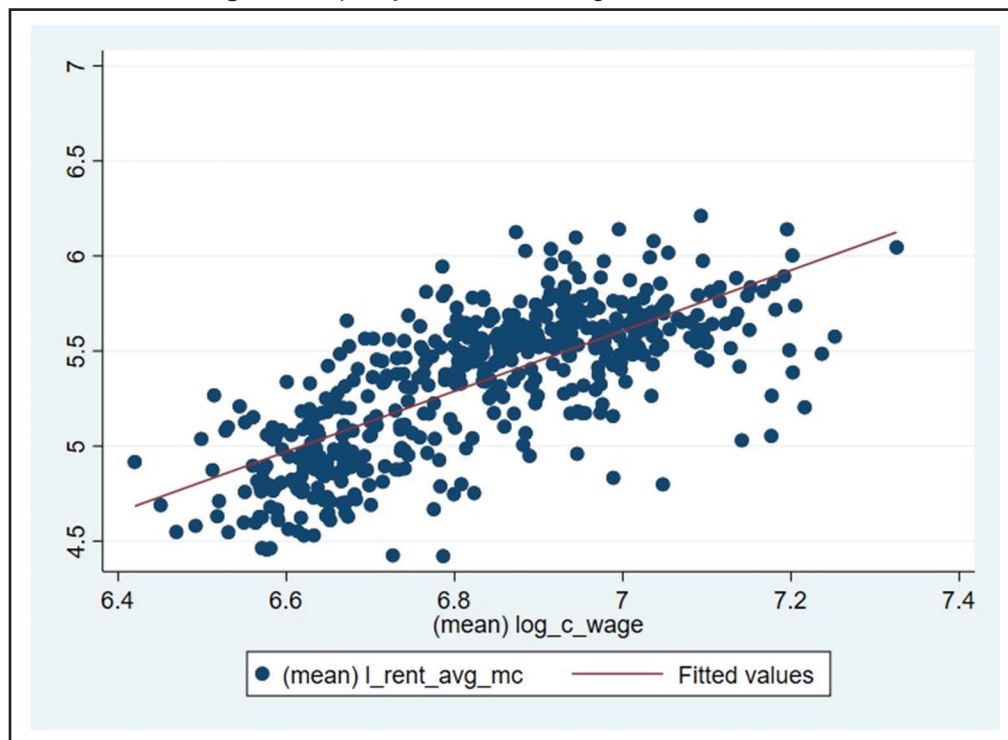


Table A1 Balancing statistics before matching

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

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

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

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)	N	Mean
<i>Hedonic price as denominator</i>		
Observed	14,810	-2.353
Predicted	14,810	0.204
Difference	0.204***	

Significance levels *10%, **5%, ***1% for *t*-test of difference in means between observed and predicted wages.

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

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

Significance levels *10%, **5%, ***1% for *t*-test of difference in means between observed and predicted wages. Predicted wages are based on a matched sample. Rent for the 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, the author also uses the median to aggregate the hedonic housing price measure. These different price measures are used as denominator to compute real hourly wages.

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

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

Significance levels *10%, **5%, ***1% for *t*-test of difference in means between observed and predicted wages. Predicted wages are based on a matched sample. Rent for the 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, the author also uses the median to aggregate the hedonic housing price measure. These different price measures are used as denominator to compute real hourly wages.

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

	log(rent per room)
Urban area	0.256*** (0.005)
<i>Type of dwelling (Base = House)</i>	
Townhouse/condominion	0.146*** (0.003)
Flat	0.396*** (0.002)
Hut	0.196*** (0.006)
<i>Wall material (Base = Bricks coated)</i>	
Bricks not coated	-0.160*** (0.002)
Wood	-0.265*** (0.003)
Plaster coated	-0.461*** (0.015)
Plaster not coated	-0.521*** (0.020)
Wood unprepared	-0.344*** (0.010)
Straw	-0.073 (0.155)
Others	-0.146*** (0.015)
<i>Bathroom (Base = none)</i>	
1	-0.213*** (0.006)
2	-0.095*** (0.006)
3	0.047*** (0.007)
4	0.220*** (0.012)
5	0.355*** (0.027)
6	0.517*** (0.054)
7	0.430*** (0.119)
8	1.046*** (0.237)
9 or more	0.356*** (0.083)

(Continued)

Table A7 Continued

	log(rent per room)
<i>Sanitation (Base = General sanitation network)</i>	
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)
<i>Waste water (Base = General distribution network)</i>	
Well on property	0.007** (0.003)
Well outside property	-0.088*** (0.005)
Carro-pipa	-0.072*** (0.014)
Rainwater cistern	-0.074*** (0.028)
Rain water other	-0.097 (0.068)
Rivers, lakes, etc.	-0.081*** (0.023)
Other	-0.155*** (0.010)
Well in village	0.165** (0.066)
<i>Canalization access (Base = Yes, in min. 1 room)</i>	
Yes, only on the property	-0.052*** (0.004)
No	-0.148*** (0.006)
<i>Garbage collection (Base = Collected directly)</i>	
Collected in collective	-0.054*** (0.002)
Burnt	-0.229*** (0.008)
Buried	-0.017 (0.043)
Tossed in a public area	-0.229*** (0.008)
Tossed in river, lake, or sea	-0.195*** (0.036)

(Continued)

Table A7 Continued

	log(rent per room)
Other	0.005 (0.027)
<i>Electricity provision (base = Yes by the company)</i>	
Yes, other	-0.094*** (0.010)
No electricity	-0.238*** (0.021)
Constant	4.235*** (0.016)
Microregion dummies	Yes
Observations	927,192
R-squared	0.539

Standard errors are robust. Observations are households.

Significance levels *10%, **5%, ***1%.

Table A8 Coefficients and t-statistics of prediction of wages for migrants based on past migrants at the 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
<i>Education (Base = none)</i>		
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	

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

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

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

Significance levels *10%, **5%, ***1% for *t*-test of difference in means between observed and predicted wages. Predicted wages are based on a matched sample. Rent for the 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, the author also uses the median to aggregate the hedonic housing price measure. These different price measures are used as denominator to compute real hourly wages.

Table A10 Variables and data sources

Variable	Description	Source
<i>Variables for descriptive statistics and destination choice model on microregião level</i>		
Wages (IV)	Average monthly wages in neighboring <i>microregião</i>	RAIS*
Housing prices	Average rent on <i>microregião</i> level	Census, IBGE
Education provision quality index	Index from 0 to 1, computed based on: Subscription rate of pre-school children, dropout rate	FIRJAN**
Health provision quality index	Rate of teachers with higher education, average daily teaching hours, results of the IDEB (Indicator of development of education in Brazil)	FIRJAN**
Number of health care facilities	Index from 0 to 1, computed based on: Number of pre-natal consultations, deaths due to mal-defined causes, child-deaths due to evitable causes	CNES***
Homicide rate	Per 100,000 inhabitants; include general hospitals, day hospitals, polyclinics, health point, general emergency, pharmacy, basic health center.	Ipeadata
Distance to the state capital	Per 100,000 inhabitants in 2008	Ipeadata
GDP	Indicator for market access (Fally et al. 2010)	Ipeadata
Distance between origin and destination	Log of GDP in 2009	Census, IBGE
	Geodesic distance as an indicator for fixed moving costs, author's calculation from coordinates	
<i>Additional variables for wage regression, on an individual level</i>		
Partner participation	Dummy whether the partner is working	Census, IBGE
The proportion of children in the 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
<i>Variables for matching, on an individual level</i>		
Age	At the 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

*RAIS (National formal sector firm and employment registry).

**FIRJAN (Industrial federation in Rio de Janeiro state).

***CNES (National Health Establishments Registry).