The wage premium of migrants and return migrants: Internal migration in Brazil*

Eduardo Cenci James Hamlette†

April 15, 2025

[click here for the most recent version]

Abstract

We study the wage premium of internal migrants in Brazil, exploring different migrant definitions in large cross-sectional and longitudinal datasets. By differentiating between current and return migration both across and within states, we shed light on the composition of the 9-10% wage premium of interstate migrants. Our findings are consistent with a model where migrants (i) are positively self-selected based on ability, (ii) have a person-specific matching with their locations, and (iii) face migration costs increasing with distance. Panel regressions with individual fixed effects suggest that around 5/6 of the wage premium of current migrants is due to unobserved ability and 1/6 to the person-place matching. Additional results show that internal migrants are more likely to have positive labor income and work longer hours, and the difference in the premia of current and return migrants—the returning penalty— shrank to near zero between 2000 and 2010.

JEL Classification: J24, J31, J61, O15, R23

Keywords: internal migration, return migration, wage premium, self-selection

*Corresponding author: Eduardo Cenci (educenci@gmail.com). We thank Jennifer Alix-Garcia, Bradford Barham, Joshua Deutschmann, Paul Dower, Esteban Quiñones, Laura Schechter, Jeffrey Smith, Emilia Tjernström, and seminar participants at the University of Wisconsin-Madison, the Institute for Applied Economic Research (IPEA), the Universidade Católica de Brasília, and CLEAR/FGV-SP for their valuable comments and feedback. Eduardo Cenci acknowledges support from Brazil's Institute for Applied Economic Research, where part of this project was developed. Any errors are our own. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Cleveland or the Board of Governors of the Federal Reserve System.

†Federal Reserve Bank of Cleveland.

1 Introduction

Understanding the determinants and consequences of the decision to migrate is central in the migration literature. Migrants move for many reasons, including work, family, education, and others. Economic migrants are those who move for work or other economic opportunities (Chiswick, 1999). Their decisions are largely driven by the expected earnings or wages in the new location. And in general, migrants earn more than non-migrants (Borjas, 1987). This premium has been documented for both international and domestic (internal) migration. Explanations for the migrant premium include positive self-selection (the more skilled and more likely to succeed migrate), a place premium (migrants move to high-productivity places), a positive person-place matching (individual-specific skills are better rewarded in a particular location), and learning (migrants acquire new skills at the destination or improve their existing skills).

A less studied but related topic is the wage premiums of return migrants, those who eventually move back to their place of origin.[1] The decision to move back incorporates repeated self-selection and other challenges that make it hard to study (Wahba, 2015).[2] Nonetheless, many papers have looked into this topic and found that return migrants earn similar or higher wages than non-migrants (Co, Gang and Yun, 2000; Lacuesta, 2010; Baptista, Campos and Rigotti, 2017). Current and return migrants are both likely positively self-selected, but their wage premia have different components. For example, return migrants enjoy neither the place premium nor the potentially positive matching with a new location that currents migration do. On the other hand, return migrants do not suffer the cost of being away from home.

We study the wage premium of internal migrants in Brazil between 2000 and 2010 leveraging information in large cross-sectional and longitudinal datasets to differentiate between current and return migration and between migration across and within state borders. Our analysis can thus compare the different migrant types among themselves and against non-migrants. In our main specification using cross-sectional census data in 2010, we find that the wage of current interstate migrants is approximately 10% higher than the wage of non-migrants. The wage of return migrants is 9% higher, and the difference between current and return migrants not significant. Additional results exploring different labor outcomes show that the lack of a significant "return penalty" is due to current migrant working longer hours and to the evolution of the migrant

[1] The place of origin can be either the place of birth or where an individual is first observed in the data. We explore both definitions in our empirical analysis.

[2] Quiñones and Barham (2018) refer to the multiple layers of self-selection involved in the return migration decision as "compound selection."

wage premia in Brazil between 2000 and 2010 (the return penalty was larger in 2000). Using matched employer-employee data between 2000 and 2010 we obtain a panel of all formal workers in Brazil and use it to run regressions with individual fixed effects. These suggest that around 5/6 of the wage premium of current migrants is due to unobserved ability and 1/6 to the person-place matching. In general, our findings are consistent with a model where migrants (i) are positively self-selected based on ability, (ii) have a person-specific matching with their locations, and (iii) face migration costs increasing with distance.

Brazil is a suitable setting for an investigation of the components of migrant wage premia for current and return internal migrants. First, it is a free movement zone formed by twenty-seven states with substantial socio-economic diversity between them, a large population, and considerable movement of people between states.[3] There has been little international migration both in and out of Brazil in the past decades, which allows us to focus on national workers only. Brazilian data, especially the extended questionnaire of its population censuses and the administrative dataset compiled by its Ministry of Labor, the RAIS dataset, has detailed information on internal migration, labor outcomes, and demographic characteristics for a very large number of individuals in multiple years. Finally, prior to the recession starting circa 2015, Brazil experienced steady economic growth, inequality reduction, and strengthening of its labor markets and regional integration.[4]

Our paper contributes to the broad literature on the labor outcomes of internal migrants. Many studies have documented the existence and discussed the determinants of the wage premia for internal migrants in different settings. For example, Kennan and Walker (2011) study internal migration in the US, and Avelino (2010) in Brazil. Many of these studies do not explicitly consider return migration, however. Notable exceptions are Co *et al.* (2000), Junge, Revilla Diez and Schätzl (2015), Ho and Turk-Ariss (2018), and El-Mallakh and Wahba (2021), which also look at non-wage outcomes of return migrants such as education and occupational mobility. Other studies have focused on the costs and barriers to internal migration. Examples are Kone, Liu, Mattoo, Ozden and Sharma (2018), and Lagakos, Marshall, Mobarak, Vernot and Waugh (2020). The latter is an example of a large literature focused on internal rural-urban migration.[5]

Our study innovates by exploring the richness of the data available in Brazil to

[3]More precisely, Brazil has 26 states and the Federal District, thus having 27 "federative units". We use the term states for simplicity.

[4]The great recession in 2008-2009 did not disrupt Brazilian labor markets as it did in the US and Europe, so our analysis benefits from some stability during that period.

[5]See Lagakos (2020) for a review of this literature.

compare different migration definitions and leverage the use of both cross-sectional and longitudinal analyses to uncover theoretical components of the wage premium in empirical analyses. Our empirical strategy has similarities with the one Hamory, Kleemans, Li and Miguel (2021) used to study rural-urban migration in Indonesia and Kenya in that we also leverage the use of individual fixed effects and different migration types. Our theoretical framework, on the other hand, has similarities with many other papers that model the migration decision as a balance between expected returns and costs associated with migrating (e.g., Kennan and Walker (2011), Roca and Puga (2017), and Lagakos, Mobarak and Waugh (2023)).

We contribute, in particular, to the literature on internal migration in Brazil. Earlier studies have found internal migration in Brazil to be highly responsive to earnings differentials (Sahota, 1968; Yap, 1976; Tannen, 1992). Other studies have looked more closely at the self-selection and wage differentials of migrants in Brazil, as well as the consequences of internal migration for inequality, earnings, and regional development (Santos and Ferreira, 2007; Avelino, 2010; Fally, Paillacar and Terra, 2010; Amaral, Rios-Neto and Potter, 2016). But these studies have largely ignored return migration and recent evolutions in the migrant wage premia. Other studies have looked at the relationship between internal migration in Brazil and education (Brotherhood, Ferreira and Santos, 2019), rural-urban migration (Alvarez, 2020), and climate change (Oliveira and Pereda, 2020), informality (Corbi, Ferraz and Narita, 2025), city size (Egger, 2021), and crime (Egger, 2022), among other topics related to internal migration.

The rest of the paper is organized as follows. In section 2, we present a simplified model of the wage components of migrants and non-migrants, the determinants of migration decisions, and our empirical strategy. In section 3, we discuss our data sources, samples, and descriptive statistics. In section 4, we show the results of our analysis using both cross-sectional and longitudinal data. Section 5 concludes.

2 Model and Empirical Strategy

We first present a model of wage determination and location choices. The model illustrates the different migrant definitions in our study and the possible wage differences between them. We then discuss the regression equations, estimations, and variable choices that further refine these migrant definitions and their associated wage differentials. We close the section by discussing possible extensions of our model and empirical strategy.

2.1 Model: Timing, productivity, and wages

We consider a model of location choices and three periods in the life of working-age adults. In the first, individuals are born and schooled in their place of origin, j_0, which we call home. In the second, they start working and decide, at any moment, to migrate to a different place $j \neq j_0$, or remain at home. In the third period, those who have migrated decide whether to return home or remain in place j for the rest of their working life. Individuals, therefore, belong to one of three mutually exclusive migrant types or statuses. They are a Current Migrant (M) if they live in a place different than home, a Return Migrant (R) if they live at home but have lived in a different place before, and a Non-Migrant (N) if they have never left home.

Each place in our model is a competitive labor market. The wage of an individual reflects her labor productivity, which is a function of individual characteristics, place characteristics, and the matching between individual and place. Specifically, the wage of individual i living and working at place j can be written as $y_{ij} = f(a_i, X_i, b_j, m_{ij})$. In this function, a_i represents unobserved ability, and X_i is a vector of observed characteristics like gender, age, and education. The term b_j is a place effect common to all individuals; it reflects differences in local infrastructure, amenities, agglomeration effects, and other factors. Finally, m_{ij} is the matching between individual and place, which captures how an individual's ability and observed characteristics are differently applied and rewarded in different labor markets. We normalize the matching at home, making $m_{ij_0} = 0$.

Taking one step further, we assume that the individual's wage function is additively separable:

$$y_{ij} = b_j + X_i^{'}\gamma + \nu_{ij} \qquad \nu_{ij} = a_i + m_{ij}. \tag{1}$$

This additive form allows us to estimate the model using linear regressions and to easily illustrate migration decisions, the different migrant types generated by these decisions, and possible differences in wages among migrant types.

2.2 Migration and return decisions

An individual moves out of home and becomes a migrant if the expected difference between her wage at home and the best destination is greater than the costs associated with moving. Let C_{ij} represent the migration costs for individual i going from their home to a new place j. Then, an individual becomes a migrant, $M_i = 1$, if $\mathrm{E}(y_{ij} - y_{ij_0}) > C_{ij}$.

Migration costs have a temporary and a permanent part. The one-time, temporary part includes traveling and adaptation costs, denoted by T_{j,j_0}, which we assume to be

the same for all individuals but different and symmetrical for every origin-destination pair. [6] The permanent cost is the cost of being away from home, denoted by A_i, which we assume to be individual-specific. Both costs can include monetary payments and disutility.[7]

We assume that individuals know their ability, a_i, and the cost of being away from home, A_i. Individuals also know the place effect at home and other places, and the traveling costs. That is, they know b_{j_0} and b_{j_0} and T_{j,j_0} for all $j \neq j_0$. Before moving, individuals have expectations over their matching with the new place, m_{ij}, but do not know it for sure.[8]

With the above assumptions, the migration decision can be written as

$$M_i = 1 \Leftrightarrow \mathrm{E}(y_{ij} - y_{ij_0}) > C_{ij} \Leftrightarrow (b_j - b_{j_0}) + \mathrm{E}(m_{ij}) > T_{j,j_0} + A_i \tag{2}$$

After moving, migrants learn their matching with the new place, m_{ij}, and return home if $y_{ij_0} - y_{ij} > C_{ij}$. The return decision can be written as

$$R_i = 1 \Leftrightarrow A_i - (b_j - b_{j_0}) - m_{ij} > T_{j,j_0}, \tag{3}$$

where the left-hand side of the inequality emphasizes the fact that, upon returning, return migrants "recover" their cost of being away from home and lose both their matching with the previous place and the (possibly positive) difference between the place factors at home and the previous place.

From the migration and return decisions, it follows that returning in our model is driven solely by differences in the expected and realized values of the person-place matching. In particular, the difference between these values, $\mathrm{E}(m_{ij}) - m_{ij}$, must be positive (expectation better than reality).[9]

[6] The cost of moving from home to place j is the same as returning home from place j. That is $T_{j,j_0} = T_{j_0,j}$.

[7] Temporary costs may include traveling and moving expenses plus the cost of searching for a new house and job in the new place. Permanent costs may include the cost of living differentials and the disutility of being distant from one's family and friends.

[8] Alternatively, one could assume that individuals know their matching but have only an expectation over A_i before moving. Both assumptions, not knowing m_{ij} or A_i *ex ante*, are plausible in our context and have the same model implications, but adding both is redundant, so we keep only one for simplicity.

[9] In fact, it has to be more than twice the value of the temporary moving costs. Indeed, by adding the last inequality in equations (2) and (3) one gets $\mathrm{E}(m_{ij}) - m_{ij} > 2T_{j,j_0}$.

2.3 Wage premia

Consider three individuals: a current migrant, a return migrant, and a non-migrant, and recall our additively separable wage function in equation (1). For simplicity, assume that these individuals are comparable in all observable dimensions so we can ignore the term $X_i'\gamma$. Assume also that all three were born in the same place, j_0, and that the two migrants moved to place, j_1. In this simplified context, the following exercise illustrates potential explanations for existing wage premia for current and return migrants, and possible differences between these premia.

Let $i = M, R, N$ be the individual subscripts identifying the current migrant, return migrant, and non-migrant, respectively. Recall that we normalize the matching at home so that $m_{ij_0} = 0$. Wages in a given year are, thus, $y_M = b_{j_1} + a_M + m_{Mj_1}$ for the current migrant, $y_R = b_{j_0} + a_R$ for the return migrant, and $y_N = b_{j_0} + a_N$ for the non-migrant.

By taking the difference between these wages we get the migrant premium, $y_M - y_N$, the return premium, $y_R - y_N$, and the (negative of) the return penalty, $y_M - y_R$. And by taking a closer look at each component on the right-hand side of these differences, we get possible explanations for them. For example, the migrant and return premia can be written as

$$y_M - y_N = (b_{j_1} - b_{j_0}) + (a_M - a_N) + m_{Mj_1} \tag{4}$$

and

$$y_R - y_N = a_R - a_N. \tag{5}$$

Equation (4) gives three non-conflicting explanations for a positive difference between the average wage of current migrants and non-migrants: (i) migrants move to places that are more productive than home, $(b_{j_1} - b_{j_0}) > 0$, (ii) migrants are positively self-selected, $a_M - a_N > 0$, and (iii) the matching of migrants with the new place is better than their matching with home, $m_{Mj_1} > 0$. Equation (5), on the other hand, provides only one explanation for a positive difference between the average wage of return migrants and non-migrants: return migrants are positively self-selected, $a_R - a_N > 0$.[10]

Finally, we can open the difference between the wages of current and return migrants, the negative of the return penalty, rewriting it as

$$y_M - y_R = (b_{j_1} - b_{j_0}) + (a_M - a_R) + m_{Mj_1}. \tag{6}$$

[10]Positive self-selection on unobserved ability, $a_R - a_N > 0$, as an explanation for the return wage premium, is equivalent to relaxing the assumption that the matching at home is equal to zero and letting return migrants have a better matching with home than non-migrants, $m_{Rj_0} - m_{Nj_0} > 0$.

The possible explanations given by equation (6) are the same as in equation (4), with the exception that the positive self-selection now considers the difference in ability between current and return migrants, and not between migrants and non-migrants. Our empirical strategy includes both destination and origin fixed effects, so the difference in place effects is effectively removed from our premia estimates. Therefore, the focus of our discussion of explanations is on positive self-selection on unobserved ability and matching.

2.4 Regression equations and identification

Our empirical analysis uses standard OLS regressions and cross-sectional data (mainly census data from 2010) with an extended analysis using individual longitudinal data (matched employer-employee data from RAIS, described in the next section). Our main regression equations resemble the equation (1) in our model.

When using cross-sectional data, we run

$$y_{ij} = \alpha M_i + \beta R_i + X_i^{'}\gamma + b_j + \epsilon_{ij} \tag{7}$$

where y_{ij} is the outcome of interest for individual i living in place j (wage, earnings, hours, or a positive earnings indicator);[11] M_i and R_i are the mutually exclusive indicators for current and return migrants, respectively; X_i is a vector of covariates including individual characteristics (age, age squared, female, education, and race), a constant, and a place of birth fixed effect; b_j is the place of residence fixed effect; and ϵ_{ij} is the error term.

In one extension, we include additional mutually exclusive migrant indicators that differentiate between the state and municipality levels in the basic regression specification shown in equation (7). In another extension, we include a second period by stacking cross sections using 2000 and 2010 census data and year indicators interacted or not with the migrant indicators.

In all cases, the coefficients of interest are α and β, which show how the outcome of interest differs for migrants and return migrants relative to the omitted group of non-migrants. We are also interested in the difference between these coefficients, $\alpha - \beta$, which shows how the outcomes of current and return migrants differ. Naturally, unobserved characteristics of the individuals are correlated with the migrant and return

[11] When varying the outcome of interest, we assume that the same theoretical framework used for wages applies to earnings, hours worked, and the positive earnings indicator. That is, all individual labor outcomes can be written as the additively separable function combining individual and place characteristics shown in equation (1).

indicators. That is, $E(\varepsilon_{ij} \mid M_i, R_i, X_i, b_i) \neq 0$. Therefore, equation (7) does not identify causal effects of migration statuses on outcomes. Rather, it shows associations in the data, which we interpret in light of our theoretical framework and empirical context to discuss possible causal mechanisms behind the coefficients. As such, the empirical findings in our analysis are correlational only, not causal.

When using longitudinal data, we run

$$y_{ijt} = \pi M_{it} + \rho R_{it} + X_{it}^{'}\delta + a_i + b_j + d_t + \epsilon_{ijt} \tag{8}$$

where y_{ijt}, M_{it}, R_{it}, and X_{it} are the time-variant versions of the outcome, current and return migrant indicators, and vector of covariates defined before. In addition to the fixed effect for place of residence, b_j, this specification also has fixed effects for individual, a_i, and year, d_t. The coefficients of interest are π and ρ, which show how the outcome of interest changes when an individual changes their status from non-migrant to current migrant or return migrant within a span of 10 years (the length of our panel).

Including individual fixed effects allows us to control for time-invariant unobserved characteristics that would be otherwise included in the error term. Still, the remaining error is correlated with the migrant indicator and, again, we cannot identify causal effects of changes in migrant status on changes in the outcome of interest, only correlations. Therefore, throughout the paper, we refrain from making causal claims and discuss our evidence and model only within the limitations of studies using observational data and correlational analysis.

2.5 Variable choices and alternative definitions

Different choices and data constraints are involved in the construction of our variables. Some of these choices directly shape the migrant definitions and the associated wage premia we investigate. Here we discuss the most consequential of these choices.

Recall that our model considers three periods. Individuals are born in the first and start working in the second. At any point in the second period, individuals can move to a different location. Those who move may return to their original place for the rest of their working life. The time between returning and retiring defines the third period for individuals who move and return; for the rest, periods two and three are the same. When considering prime age workers, one can think of period one going from ages 0 to 24, period two starting at age 25, and period three happening at any point between ages 25 and 54. At age 55, individuals stop working and moving. Alternatively, one can think of period one ending between ages 16 and 21, and periods two and three ending

between ages 60 and 70. Our empirical analysis focuses on prime age individuals but also looks at other age groups.

Our model considers individuals moving between places, or labor markets. In the context of Brazil, one can think of states as places and the state of birth as home. Alternatively, the municipality can delimit a place, and home can be the place where an individual has lived for a long period (ten or more years, for example). Our empirical analysis uses state and birthplace as the default definitions for places and home, but also discusses and investigates other definitions.[12]

In some empirical exercises, we consider an extended set of migrant definitions. First, by differentiating between interstate and intrastate migrants. In this case, individuals are placed into five non-overlapping migrant statuses (current interstate, return interstate, current intrastate, return intrastate, and non-migrant). Interstate migrants are defined as before, and part of the individuals formerly defined as non-migrants are now defined as intrastate migrants if they have moved between municipalities within their state of birth.[13] Second, we ignore the state level and consider only municipality borders to define migrant types. In this case, individuals are again placed into three non-overlapping migrant categories (current, return, and non-migrants), which encompass both inter- and intra-state migration. Migration costs and distances are generally higher across state borders than municipality borders. Therefore, comparing the wage premia for these different migrant definitions helps us understand the relationship between migration costs and wage premia (see Table 4 in the Results section).

As for home (the place of origin), we could define it, alternatively, as the place where the individual was living ten years ago, or the place where the individual was first observed in the data (we don't know the place where the individuals acquired most of their education or spend their formative years). We use the first of these alternative definitions in an additional exercise of our cross-sectional analysis (see Table 6 in the Results section) and the second in our longitudinal analysis because our panel data has no information on an individual's birthplace.[14]

[12]In 2010, Brazil had 5,560 municipalities distributed across 27 Federative Units (26 states plus the Federal District). Other definitions could use administrative levels smaller than state but bigger than municipality to define a place (Brazil's 137 mesoregions or 558 microregions), or to consider municipalities only without differentiating between state and municipality migrants. While interesting, these alternative comparisons do not provide significant insights and are left out of our paper.

[13]Individuals are current intrastate migrants if they live in their state of birth but not their municipality of birth and have never lived in a different state, and return intrastate migrants if they live at their municipality of birth but have lived in a different municipality within their state of birth before. Individuals who have never left their municipality of birth are non-migrants. See Table A5 in the appendix for a cross-tabulation with the number of individuals in each of these definitions in 2010.

[14]We differentiate between migrants living for less than ten years in a place or for ten years or more

2.6 Extensions and limitations

When using cross-sectional analysis, we cannot separate the effects of ability and matching components. Therefore, we consider an extension of our model that can be estimated using longitudinal data and considers individuals who are observed in multiple periods. Such a framework allows the inclusion of individual fixed effects that absorb any individual time-invariant components of the wage function, in particular, unobserved ability, a_i. For simplicity and data limitations, we consider only current migrants and non-migrants in this extension. Estimates of the migrant wage premium in this extension, if they still appear, are explained solely by the matching component.

Our model does not consider learning in the sense of human capital accumulation (increases in unobserved ability) or an improvement of the matching between individual and place, both of which would change with the duration of the migration experience. This would require us to observe the duration of the migration experience (years away from home) for both current and return migrants. Our data, however, has this information only for current migrants. For return migrants, our data informs years after returning. Moreover, there is no evidence that the migrant premium increases with the duration of the migration experience for current migrants. If anything, the premium decreases over time (see Figure 2 in the Results section). The only learning allowed in our model is the update in the expectation regarding the value of the matching at the new place. A more precise investigation of learning is beyond the scope of our paper.[15] Finally, we also do not consider physical capital accumulation in our model, since we do not observe assets, wealth, or savings in the data.

3 Data

3.1 Data Sources

We use two main data sources in this study. The first is the publicly available, extended questionnaire of the Brazilian census in 2000 and 2010 administered by the Brazilian statistical office, the Instituto Brasileiro de Geografia e Estatística (IBGE).[16] This cross-sectional data is representative at the municipality level and provides information

and show how their wage premia may differ in these additional empirical exercises. We also restrict the migrant definition to only those living for less than ten years in a place, which is equivalent to defining "home" as the place where an individual lived ten years ago.

[15] Other migration papers discuss learning, capital accumulation, and the evolution of wage differences associated with migrant statuses. See, for example, Roca and Puga (2017) in the context of Spain and Patto (2024) in the context of Brazil.

[16] The microdata for the last population census in Brazil, conducted in 2022, is not yet available.

on individuals and households, including detailed information on current and past location of residence, place of birth, demographic characteristics, and labor outcomes. Brazilian census data is suitable for a detailed investigation of the relationship between different migration statuses and labor outcomes. It allows us to differentiate between current and return, and interstate and intrastate migration. It also allows us to observe various labor outcomes at the extensive and intensive margins (labor force participation, hours worked, earnings, wages, occupation, and more.

Census data is collected for all individuals in a household. Therefore, we can use individuals as our unit of analysis while controlling for household characteristics. The original data contain approximately 20 million observations in 2000 and 20.4 million in 2010, which corresponds to approximately 11–12% of the total number of census respondents in both years. These are Brazilian citizens who were asked to answer long-form versions of the census.[17] In our main analysis, we drop the foreign-born and those who did not respond to the migration section of the questionnaire. We then restrict the sample to prime age individuals (those 25–54 years old). Because our main outcome of interest is the hourly wage, we restrict the sample to those with positive labor income in most analyses. Finally, we drop observations with extreme values for earnings, wages, and hours worked (below and above percentiles 0.1 and 99.9).

Our second main data source is the RAIS data.[18] This is an administrative dataset compiled by the Ministry of Labor in Brazil. All firms in the country are required to file detailed payroll reports in December of each year. Therefore, with access to the restricted, identified version of this data, one can create a linked employer-employee dataset that includes several characteristics of individuals, the establishments where they work, and their jobs.[19] Importantly, we observe the location (municipality) of the establishments where individuals worked over the years in the RAIS data and can infer migration statuses from this information. And we can leverage the panel structure of the sample we constructed to add individual fixed effects to our regressions.

A limitation of RAIS data is that it only has information on individuals formally employed (around 53% of the prime-age workers in 2010). Therefore, we use it for supplementary analysis and rely on census data for the main parts of our investigation.

[17]The procedure by IBGE is random sampling at the municipality level. If there is more than one individual in a household sampled, everyone in that household was given the long-form version of the census survey.

[18]Relação Anual de Informações Sociais is the Portuguese acronym.

[19]We apply to the RAIS data the same restrictions applied to the census data: we drop all foreign-born and keep only individuals 25–54 years old. Because RAIS data come from payroll reports, only individuals with positive labor income are included, so we do not use it to look at labor force participation, only wages.

Moreover, RAIS data does not inform an individual's place of birth, just their place of work. Thus, the migration definition we use while working with RAIS data is different from the one we use in the main analysis. Instead of using an individual's place of birth as their origin, we use their place of work/residence when they first appear in the panel. Finally, since we use RAIS data from 2000–2010 only to match the period covered by our census data, our migration definitions consider only those moving within the last 10 years. Fortunately, similar migration definitions can be constructed with census data, so we can reasonably compare results from our two main data sources.

3.2 Descriptive Statistics

Table 1 shows summary statistics for the 2010 census cross-section—it gives an overview of the sample and how some variables are constructed (i.e., indicators, categorical variables, or level variables). Almost a quarter of the 8.5 million prime-age adults in our 2010 census sample had an interstate migration experience: 16% are current migrants and 6% are return migrants.[20] We have plenty of observations to explore variations in labor outcomes compared to non-migrants (78% of the sample). The demographic characteristics are, not surprisingly, representative of the Brazilian population in 2010: around half of the observations are female (51%) and white (48%), and 21% live in rural areas. The average age of 38 years is a reflection of our age restriction. Around two-thirds of the sample reported positive earnings in 2010 and earned, on average, R$ 2,454 per month.[21] More than half of the individuals with positive labor income are formally employed in either the private sector (45%), or the public sector and military (7%).

[Table 1: Summary Statistics, 2010]

Most variables included as controls in our regressions are shown in the table (age, female, rural, race, and education). Not shown are race and education categories with very few observations (e.g., education unknown and the Asian and indigenous race categories). For descriptive purposes only, we show the variable Years of Education in this table together with some aggregate categorical variables based on the highest education attained. These categories consider education "up to" the level in their

[20]The shares of current and return migrants in a sample restricted to those with positive labor income are similar: 17% and 6%, respectively.

[21]Earnings and wages are in reais (R$), the Brazilian currency, in real values of July 2023 when US$1.00 ≈ R$5.50. Earnings are reported per month and hours worked per week. We divide monthly earnings by hours multiplied by 4.35 to obtain the hourly wage.

label.[22] In our regressions, however, we use more detailed indicators for education categories, which correspond more directly to information in the census data.[23]

In Table 2, we show differences in means by three migrant statuses: current interstate migrant, return interstate migrant, and non-migrant (a status that includes intrastate migrants). The first three columns show the average for each group; the last three show the differences between current migrants and non-migrants, return migrants and non-migrants, and current and return migrants. All non-zero differences are statistically significant at 1%.

[Table 2: Differences in Means by Migrant Status, 2010]

There are no noticeable differences in gender, age, or race. Migrants (current and return) are just slightly older than non-migrants. Current and return migrants are much less likely to be rural, and more likely to be the head of their households. These differences are in line with the idea that rural-to-urban migration is more prevalent, and that people who migrate often form a new household and are more likely to be its head.

Work is a primary reason for internal migration in Brazil. Accordingly, migrants, both current and return, are more likely to have positive earnings than non-migrants (and current migrants more so than return migrants). Current migrants are also more likely to be formally employed in the formal sector. Return migrants, on the other hand, are more likely to be employers (perhaps because, upon returning, some migrants move to self-employment and entrepreneurship). We conduct an additional exercise by regressing an indicator for each employment status on the migrant indicators and controls, and observe correlations between migrant and employment statuses that are in line with the difference in means shown here. Results for this exercise are shown in Table A7.

The monthly earnings of current and return migrants are 20-24% higher than those of non-migrants. The return migrants have slightly higher earnings than current migrants (4%) but are also more educated. They have approximately 2/3 of a year of education more than current migrants and non-migrants, and are much more likely to have a college degree or higher. Current migrants work longer hours than non-migrants and return migrants, which can explain why their wage, while larger than that of

[22] For example, the Some High School category includes individuals with an education level up to high school incomplete, Some College considers education up to incomplete college (including high school graduates), and College considers a college degree or higher.

[23] Using years of education and its square as regression controls instead of education categories does not change our results substantially.

non-migrants, is smaller than the wage of return migrants. In logs, both earnings and wages of migrants are larger than those of non-migrants, and there are no substantial differences between current and return.

Due to positive self-selection, one would expect migrants to have higher earnings and wages than non-migrants, as observed in Table 2. The difference in earnings and wages between current and return migrants, however, is less clear *ex ante* and is subject to the influence of differences in education and hours worked. We explore the question of whether current and return migrants have different labor outcomes in more detail, using linear regressions and exploring different outcomes and specifications in the results section.

In the appendix, we show the same type of information as in Tables 1 and 2, summary statistics and differences in means, for census data in 2000 and RAIS data in 2010 (see Tables A1 to A4). And in Table A5 we provide detailed information (number of individuals and percentages) for each of our migrant definitions using census data from 2010. This table also highlights when our different migration definitions overlap or not.

Finally, to give a sense of the geographic distribution of internal migrants in Brazil, Figure 1 shows the state-level shares of current and return migrants in a map of the country in 2010. The left panel shows percentages of current interstate migrants, i.e., the share of state residents born in a different state. The states along the agricultural frontier in the west and north of Brazil have the highest shares of current migrants in their populations. The right panel shows percentages of return interstate migrants, i.e., the share of residents born in the state who have lived in a different state in the past. Some states in migrant-sending parts of the northeast and one state in the south (Paraná, PR) have the highest shares of return migrants in their populations. In the appendix, we show two complementary pairs of maps. In Figure A1, we show state-level percentages of inter-state emigrants (the share of the population born in the state living in a different state) and the non-migrants. In Figure A2, we show a similar percentage of current and return migrants at the municipality-level. Finally, in Table A6, we show the percentages of interstate immigrants, emigrants, return migrants, and non-migrants along with the population residing and born in each state.

[Figure 1: Migrant Shares by State: Current and Return Migrants]

4 Results

4.1 Primary Specifications

Table 3 shows the results from regressions of the log hourly wage on two non-overlapping migrant indicators (current and return) plus an increasing set of controls and fixed effects. It also shows the difference between the two coefficients of interest and the corresponding standard error. All regressions use census data in 2010 and follow the basic specification shown in equation (8). We discuss all coefficients in log points, rounding them up to two decimals.[24]

[Table 3: Wage Premium of Current and Return Migrants]

The results in the column (1) come from specifications with no covariates, just the two migrant indicators. Both indicators are associated with a wage that is 15 log points higher and are no different from each other. In column (2), we add our basic set of covariates as regression controls. The coefficient for current migrants remains while that of return migrants drops to 8 log points. Accordingly, their difference rises to 7 log points. Part of the change might come from the fact that we control for education categories, and return migrants are more educated on average. In column (3), we add state of birth fixed effects, and both coefficients rise (perhaps because place of birth serves as a proxy for the quality of the education individuals received when young). In column (4), we add state of residence fixed effects, which captures any place effect common to all individuals (the term b_j in our model). In line with the idea that migrants, on average, move to places with positive place effects, the coefficients on current migration drop by half, from 20 to 10 log points, while the return migrant coefficient remains at 9 log points. Their difference of 1 log point is not statistically significant.

Finally, in column (5), we add other covariates that extend our basic set. The coefficients drop and their difference rises, but only slightly. Because many of the covariates in the extended set are arguably endogenous to the migration decision (e.g., head of household status and occupation and industry categories), we stick to our more conservative basic set and adopt the regression specification in column (4) as our preferred one. Throughout the paper we refer to the results in column (4) as our benchmark results.

[24] Many of our coefficients are in the 10-20 log points range, so their corresponding value in percentage points, given by $\exp(\hat{\beta}) - 1$ is slightly larger. For coefficients in the 0-10 log points range, the percentage points are close.

The coefficients in our preferred specification show a wage premium of approximately 10% for current migrants and 9% for return migrants, respectively. The difference between the two coefficients is not statistically significant and, arguably, not economically significant either. We find this finding puzzling because, unlike return migrants, current migrants face the cost of being away from home and should, in theory, demand higher wages to compensate for such a cost. We investigate this puzzle further in another set of regressions. But first, we investigate the extent to which using different migrant definitions would change our benchmark result.

Table 4 shows the results for regression specifications that include the basic set of covariates plus fixed effects for state of birth and state of residence. We vary the number and type of migrant indicators used in the different regressions, thus varying also the comparison (omitted) group of non-migrants in each case. In column (1), we include only the current migrant indicator that considers only interstate migration. The comparison group thus includes return interstate migrants, all intrastate migrants, and those who have never migrated within or across states. The regression shows a wage premium of 9 log points. Adding the return migrant indicator in column (2) increases the current migrant premium to 10 log points, in line with the idea that we now have a "cleaner" comparison group that does not include return migrants—individuals who may be positively self-selected since they have migrated in the past.[25]

[Table 4: Varying Migrant Definitions using State or Municipality Levels]

In column (3), we add an indicator for current and return intrastate migrants, thus completely cleaning the comparison group of any individual who has migrated. Only those who have always lived in their municipality of birth are left. The wage premium of current and return interstate migrants increases by 2 log points, and there are premia associated with intrastate migrant indicators (5 log points for current, and 3 for return). These results are in line with the idea that all kinds of migrants, both interstate and intrastate, are positively self-selected individuals and thus have a wage premium over non-migrants whether they are current or return migrants. They are also in line with migration costs that increase with distance and the crossing of state borders. Interstate migrants travel farther, on average, than intrastate migrants, and incur on costs associated with living in a different state (e.g. adapting to a different legislation and culture, registering with public agencies, etc.). They must, therefore, receive higher wages, all else equal, to compensate for these higher migration costs. In fact, the premium of interstate migrants, both current and return, is on average

[25]The results in column (2) in Table 4 mirror those of column (4) in Table 3, our benchmark.

7 log points higher than that of intrastate migrants. Finally, in column (4), we use only a different set of definitions in which current and return migrants are defined considering only municipality borders (i.e., combining inter and intrastate migration).[26] The coefficients for current and return migrants thus defined are indistinguishable from each other, and their magnitudes lie between those of interstate and intrastate migrants in the previous column.

From these exercises, we conclude that our results do not change qualitatively when intrastate migrants are considered. Therefore, for simplicity, we consider only interstate migration in our main analyses.

Having settled on which migrant definitions to focus on, we now return to investigating possible explanations for the non-significant difference in the wage premia of current and return migrants in our benchmark result. We do so by investigating other labor outcomes that shed light on labor supply decisions at the extensive and intensive margins. Table 5 shows the results of this exercise. In all columns, we use the same migrant indicators, controls, and fixed effects of our preferred specification and vary only the outcome of interest. To help interpret coefficients, we show the means of each dependent variable.

[Table 5: Extensive and Intensive Margins: Wages and Labor Supply by Migrant Status]

In column (1), we use the log hourly wage, thus reproducing the benchmark result. In column (2), we use hours worked per week and verify that current migrants work almost one full hour per week more than the return migrants and non-migrants (2% of the mean value of 41.3 hours/week worked). This difference can explain why current migrants have the same wage premium as return migrants despite facing higher migration costs: they compensate by working longer hours and, thus, earning more. In fact, that is what we observe in column (3) when we use the log of monthly labor earnings as the dependent variable. Again, both migrant types have a premium, but the earnings premium of current migrants (12.6 log points) is significantly higher than the premium of return migrants (7.4 log points), both in the statistical and economic senses.

In columns (1) to (3), our choice of dependent variable determines a regression sample that includes only individuals with positive labor income (two-thirds of our sample). In column (4), we use the IHS transformation of the hourly wage instead of the log so the observations with no wage (value equal to zero) do not drop from the

[26]Appendix Table A5 helps compare our different migrant definitions and observing when they overlap or not.

regression. The interpretation of the coefficients on the IHS transformed variable is similar to that of a log-transformed one if the number of zeros is small (Bellemare and Wichman, 2020). In our sample, however, one-third of the individuals have no wage. The coefficient on the IHS transformed wage, then, mixes the intensive and extensive margins (conditional on having a wage, migrants earn more, but they are also more likely to have a wage at all). To investigate how much the extensive margin matters, we use an indicator for having positive labor income as the dependent variable in column (5). The results show that current migrants are 2.2 percentage points (3.3%) more likely to have positive labor income than non-migrants. Return migrants are also more likely to have positive income (0.8 percentage points), but less so than current migrants (1.4 percentage points).

We draw a few interesting conclusions from this exercise. First, current migrants are more likely to have positive labor income and work longer hours. These two facts combine to give them higher earnings and compensate for the cost of being away from home. Second, when we look only at the hourly wage, a better proxy for labor productivity, both current and return migrants have comparable wage premia. This suggests they have a similar degree of positive self-selection and that the person-place matching explains little of the wage premia of internal migrants in Brazil.

4.2 Evolution of Migrant Wage Premia

In regressions shown in Tables 3 to 5, we used different definitions of migrants but kept the state of birth as the initial location. While this is our preferred choice, it is valid to explore alternative ways to define migrants' origin. In particular, alternative definitions of origin that we can use in our longitudinal data, which do not inform the individual's place of birth. Additionally, we would like to verify how our results change when we use migrant definitions that do not include individuals who may have migrated as children and thus neither selected themselves into migration nor have lived their formative years in a different location. In other words, we want to distinguish between recent and old movements, and at the same time obtain a definition of origin that can be used in our 11-year panel data. We do so by differentiating between migrants who live for less than 10 years or for 10 years or more in their current residence. Given our age restriction, migrants who live in a given state for less than 10 years have migrated or returned when they were no younger than 16 years old. We exclude anyone who may have migrated as a child and, therefore, obtain a cleaner comparison group.

Table 6 shows the results for this exercise. In column (1), we reproduce our

benchmark result. In column (2), we split current and return migrants into two groups: those living in the state for less than 10 years, and those living in the state for 10 years or more.[27] As before, the migration indicators defined this way are mutually exclusive, so we can include all four indicators in our regression and obtain a coefficient of interest for each. Finally, in column (3), we repeat the previous exercise ignoring migrants, whether current or return, who have lived for 10 years or more in their current residence. All regression specifications use the same set of controls and fixed effects of our preferred specification.

[Table 6: Evolution of the Migrant Wage Premium: Years since Migrating or Returning]

The results in column (2) show that the wage premium of current migrants who have moved to a new state less than 10 years ago is more than twice as high as the premium of those who have lived in their new state for 10 years or more. In column (3), when only recent migrants are considered, the wage premium is more than 3 log points higher than our benchmark result. For return migrants, on the other hand, we observe much smaller differences (1 log point or less). These results are in line with the idea discussed above that recent current migrants are more likely to have been the migration decision-maker (and thus be positively self-selected). They are also in line with the idea that migration costs—the cost of being away from home, in particular—can be higher for recent migrants (after 10 years, maybe the new state becomes "home").

The results in columns (2) and (3) of Table 6 also provide a useful comparison for the results we will see later in the paper using longitudinal data from RAIS. In the RAIS sample, individuals are defined as migrants if their place of residence is different from the place in which they are first observed in the data. Since the data covers only 11 years (2000 to 2010), all migrants defined this way have lived in their new residence for 10 years or less.

In Figure 2, we show the results of a regression specification that is similar but more detailed than the one used in column (2) of Table 6. In this exercise, we include an indicator for migrants living for 10 years or more (10+) as before, but split the group of recent migrants into nine different indicators, one for each number of residence years. This is done for both current and return migrants.

[27] Note that years of residence means years since migrating for current migrants, but years since returning for return migrants. We cannot compare current and return migrants with the same length of the migration experience (years migrating) because we do not have this information for return migrants.

The pattern of wage premiums across the years since migrating back home for return migrants is stagnant and, if anything, decreasing slightly. For current migrants, their average wage premium decreases over time. This finding contrasts with other papers in the literature, which find that migrants earn more as they live for longer periods in the new residence due to learning and adapting.

[Figure 2: Evolution of the Migrant Wage Premium: Years since Migrating or Returning]

Next, we investigate how the wage premium of current and return migrants may have changed over the period covered by our data, from 2000 to 2010. Table 7 shows the results for this exercise. As in previous tables, column (1) reproduces our benchmark result, and all specifications include the same set of controls and fixed effects as before. The main difference is the inclusion of time fixed effects (an indicator for the year 2010) when data from both census years are pooled together. In column (2), we run our preferred specification for the year 2000, and in column (3) we pool data from both years and leverage the coefficient on the indicator for 2010 (not shown) to calculate the wage premium for each migrant type in each year, as well as their differences.[28] In columns (4) to (6), we repeat the same specification as in column (3) for different outcomes of interest: hours worked, log of monthly earnings, and the indicator for having positive labor earnings.

[Table 7: Evolution of the Migrant Wage Premium from 2000 to 2010]

By comparing the results in columns (1), (2), and (3), we observe that the current migrant premium has decreased by 2 log points on average from 2000 to 2010, while the return migrant premium has increased by the same magnitude. Accordingly, the difference between the current and return migrant premia has shrunken by 4 log points. The results in columns (4) and (5) show that this shrinking in the "returning penalty" comes from a decrease of 3 log points in the earnings premium for current migrants (the return earnings premium did not change) combined with an increase, relative to return migrants, of 0.09 hours worked for current migrants (hours worked decreased for both types, and but less so for current migrants). That is, relative to return migrants, current migrants earn less and work longer than return migrants in 2010 compared to 2000.

[28] Ignoring controls and state of birth and residence fixed effects, the regression specification used in columns (3) to (6) is $y_{it} = \alpha_{00} M_{it} + \beta_{00} R_{it} + \alpha_{10}(M_{it} \times Year10) + \beta_{10}(R_{it} \times Year10) + \gamma Year10 + \mu_{it}$, where M_{it} and R_{it} are the migrant dummies and the $Year10$ is the indicator for 2010. The coefficients α_{00} and β_{00} give the wage premium for current and return migrants, respectively, in 2000. The sums $(\alpha_{00} + \beta_{10})$ and $(\beta_{00} + \beta_{10})$ give the wage premium for current and return migrants in 2010.

Therefore, their relative migrant wage premium has decreased. Both migrant types, however, still earn more, work longer hours, and have higher wages than non-migrants. Finally, in column (6), we observe that the higher likelihood of having positive labor income, which we have observed for both migrant types in column (5) of Table 5, was not present in 2000. An investigation of the reasons behind the evolution of the different migrant wage premia between 2000 and 2010, documented here, is beyond the scope of our study.

4.3 Results using Alternative Samples

Our final exercise with cross-sectional data from the census in 2010 repeats the main regression specifications for a few selected subsamples. Results are shown in Table 8.

[Table 8: Heterogeneity in the Migrant Wage Premium: Alternative Sample Restrictions]

In column (1), we examine how robust our results are for less restrictive age restrictions, including all individuals aged 16 and above in the sample. Compared to our benchmark result, there is only a slight decrease of 0.7 log points in the current migrant wage premium (9.4 versus 10.1 log points). The return migrant premium does not change. In column (2), we restrict the sample to (prime-age) males only, a restriction that was common in earlier migration studies either due to data limitations or to abstract away from labor supply decisions that differ between males and females. In line with the hypothesis that men are more likely to be the decision-maker in joint migrations and, therefore, more likely to be positively self-selected, we observe that the wage premia for both current and return migrants is approximately 2 log points higher in the male subsample than the main sample (specifically, 2.3 log points higher for current migrants and 1.7 log points for return migrants, so the gap ion their premia increases and becomes statistically significant).

In column (3), we restrict the sample to college-educated individuals and find, again, higher premia for both migrant types. Here, however, the increase in the wage premium for return migrants is stronger, 5.5 log points versus 2.7 log points for current migrants, and the sign of the current-return premium difference flips: college-educated return migrants have a higher wage premium than college-educated current migrants. An investigation of the detailed reasons for this result is beyond the scope of our study, but it is likely due to individuals who migrated for college returning home.

Finally, in column (4), we restrict the sample to individuals formally employed in the private and public sectors or the military. We observe a smaller wage premium for

current migrants (8 log points versus 10 log points in the full sample) and a negative but not significant return penalty. Like some of the results in Table 6, the results for the formally employed subsample provide a useful comparison for results using longitudinal data from RAIS, which has only formally employed individuals observed over a period of 11 years.

4.4 Longitudinal Analysis

This section presents and discusses the results of our regressions using data from RAIS. The panel structure of the data allows us to include individual fixed effects and thus control for any time-invariant individual characteristics such as unobserved ability.[29] In our model, individuals self-select into migration based on their time-invariant ability (the term a_i) among other factors like expectations over their person-place matching. Individual fixed effects, therefore, capture the importance of unobserved ability in explaining the migrant wage premium and help. Any remaining premium must be due to matching or other factors not contemplated in our model (like learning).

Our longitudinal data has only formally employed individuals and uses migrant definitions based on the place an individual first appears in the data. Effectively, this considers only individuals residing for less than 10 years in the current state as migrants. Our best comparison results in the cross-sectional analysis are in column (2) in Table 6, which considers only migrants with less than 10 years of residence, and column (4) in Table 8, which considers only formally employed individuals. For return migrants, these are very close to our benchmark result. For current migrants, they bound the benchmark above and below (the current migrant wage premium is 16 log points for those with less than 10 years of residence, 8 log points for the formally employed, and 10 log points in the benchmark). Therefore, our benchmark results, though pertaining to a different sample and migrant definitions, provide an adequate comparison for the longitudinal analysis results, too.

Table 9 shows our main results. In column (1), we use a cross-section from the RAIS data in 2010. The current migrant wage premium is 12.6 log points, and the return is 13.4 log points. Their difference, of negative 0.9 log points, is statistically significant but arguably not economically significant.[30] The sharp increase in the return migrant premium, compared to the benchmark result of 9 log points, may be due to

[29] In all panel regressions we also include a set of basic covariates similar to the one used in the cross-sectional analysis and fixed effects for state of residence (state of birth is not observed).

[30] Our longitudinal analysis clusters standard errors at the level of the individual (and there are several million of them), not at the state level like our cross-sectional analysis. Therefore, standard errors are smaller and the statistical significance is higher in the longitudinal analysis.

the way return migration is defined in this analysis. Because we must rely on the place where the individual first appears in the data to assign them an origin, a return migrant is someone who (i) was a current migrant in the last ten years, and (ii) appears at least three times in the data (i.e., was formally employed for at least three, potentially non-consecutive, years). The comparison group of current migrants and non-migrants, on the other hand, must appear only two years to be included in our data. It is possible that appearing for more years in the data (being formally employed) correlated with ability or other factors that also impact wage.

[Table 9: Longitudinal Analysis using RAIS data]

In column (2), we use a cross-section from the RAIS data in 2000. Here, the current and return migrant indicators mean future migrant statuses: the individuals will be a current migrant or a return migrant in 2010. That is, we can observe if someone we know will be a migrant in the future already has a wage premium (perhaps because they have higher ability). We find this to be the case, indeed. Future current migrants in 2000 have a 10.5 log points wage premium, suggesting these individuals have higher ability than the non-migrants. Future return migrants also have a wage premium, but the magnitude (and economic significance) is much smaller, just 1.5 log points.

In columns (3) and (4), we use a two-period panel with data from the years 2000 and 2010. The results in column (3) are very similar to column (1). Both migrant types have a wage premium of approximately 13 log points. In column (4), we add individual fixed effects. The wage premium of current migrants drops to 2.6 log points. Curiously, this is approximately the difference between the premium of current migrants and future current migrants in the 2010 and 2000 cross-sections. Such a result suggests that approximately 5/6 of the current migrant wage premium is due to positive self-selection (higher unobserved ability) and 1/6 due to a positive person-place matching. In columns (5) and (6), we repeat the same exercise but using data from all years between 2000 and 2010 in a multi-year panel. The results for current migrants are qualitatively the same.

The results for return migrants are harder to interpret. In column (4), they have no wage premium, which suggests the wage premium in column (1) was all positive self-selection and no matching. But the result in column (2) shows only a modest wage premium for future return migrants. Moreover, in column (6), when more years and observations are included, return migrants show the same wage premium of current migrants, 2.7 log points. Part of this discrepancy may come from data limitations that make the definition of return migrant in the longitudinal data more different and special, as discussed above. Part may be due to factors not accounted for in our model,

like learning.

In Table A7 in the appendix, we repeat the exercises in columns (1), (3), and (4) using alternative sets of migrant definitions. We reduce the set of migrant definitions to include only current interstate migrants in some specifications and expand it in others, including both inter and intrastate current and return migrants. Results for interstate migrants are qualitatively similar, especially for current migrants. Results for intrastate migrants are all negative.

5 Conclusion

We study the wage premia of current and return migrants in Brazil and the evolution of these premia over the 2000-2010 period. Following the migration literature, we assume that individuals migrate when their difference in expected earnings between home and the destination exceeds migration costs. We develop a simple model where the migrant wage premium can reflect positive self-selection on unobserved ability, a place premium common to all individuals, and a person-specific matching with the new residence. We then perform a series of regression exercises exploring the richness of our data to compare different migrant statuses and investigate how these statuses associate with various labor outcomes, in particular, wages.

In our preferred regression specification using census data from 2010, we find that the hourly wage of current interstate migrants is 10% higher on average than the wage of non-migrants, while the wage of return migrants is 9% higher on average. We then investigate the extensive and intensive margins of labor supply as well as the evolution of the migrant wage premia for current and return migrants to understand the apparent lack of a returning penalty for returning migrants in Brazil. We verify that current migrants are more likely to have positive income and work longer hours. Moreover, we verify that the return penalty was larger in 2000 but has shrunk in 2010. Our cross-sectional results suggest that current and return migrants are similarly positively self-selected, and our longitudinal analysis shows that, at least for current migrants, positive self-selection does not explain all the wage premium. Approximately 1/6 of the premium is due to a person-place matching or other factors not accounted for in our model. Our exercises using municipality-level definitions of migrants do not change the findings we observe for interstate migrants but corroborate the notion that migration costs increase with distance.

Due to data and methodological limitations, we do not investigate how learning and capital accumulation explain the size and evolution of migrant wage premia in Brazil.

The role of learning and how it affects the evolution of the person-place matching, in particular, is a question that should be addressed in future research. We are also limited by the data in the period we study. Unfortunately, microdata for the last Brazilian census conducted in 2022 was not available at the time of our study. It would be interesting to extend our analysis to include these data when available and examine how internal migration patterns and the wage premia of current and return migrants have evolved in Brazil since 2010.

References

Alvarez, J. A. (2020). The Agricultural Wage Gap: Evidence from Brazilian Micro-data. *American Economic Journal: Macroeconomics*, **12** (1), 153–173.

Amaral, E. F., Rios-Neto, E. L. and Potter, J. E. (2016). The influence of internal migration on male earnings in Brazil, 1970-2000. *Migration and Development*, **5** (1), 55–78.

Avelino, R. R. G. (2010). Self-Selection and the Impact of Migration on Earnings. *Brazilian Review of Econometrics*, **30** (1), 69–89.

Baptista, E. A., Campos, J. and Rigotti, J. I. R. (2017). Return Migration in Brazil. *Mercator*, **16** (4), 1–18.

Bellemare, M. F. and Wichman, C. J. (2020). Elasticities and the inverse hyperbolic sine transformation. *Oxford Bulletin of Economics and Statistics*, **82** (1), 50–61.

Borjas, G. J. (1987). Self-Selection and the Earnings of Immigrants. *American Economic Review*, **77** (4), 531–553.

Brotherhood, L. M., Ferreira, P. C. and Santos, C. (2019). Education Quality and Returns to Schooling: Evidence from Migrants in Brazil. *Economic Development and Cultural Change*, **67** (3), 439–459.

Chiswick, B. R. (1999). Are Immigrants Favorably Self-Selected? *American Economic Review*, **89** (2), 181–185.

Co, C. Y., Gang, I. N. and Yun, M.-S. (2000). Returns to Returning. *Journal of Population Economics*, **13** (1), 57–79.

Corbi, R. B., Ferraz, T. and Narita, R. (2025). Internal migration and labor market adjustments in the presence of non-wage compensation. *Working Paper*.

Egger, E.-M. (2021). Migrating out of mega-cities: Evidence from Brazil. *IZA Journal of Development and Migration*, **12** (1).

— (2022). Internal Migration and Crime in Brazil. *Economic Development and Cultural Change*, **71** (1), 223–259.

El-Mallakh, N. and Wahba, J. (2021). Upward or downward: Occupational mobility and return migration. *World Development*, **137**, 105203.

Fally, T., Paillacar, R. and Terra, C. (2010). Economic geography and wages in Brazil: Evidence from micro-data. *Journal of Development Economics*, **91** (1), 155–168.

Hamory, J., Kleemans, M., Li, N. Y. and Miguel, E. (2021). Reevaluating

Agricultural Productivity Gaps with Longitudinal Microdata. *Journal of the European Economic Association*, **19** (3), 1522–1555.

Ho, G. and Turk-Ariss, R. (2018). The labor market integration of migrants in Europe: New evidence from micro data. *IMF Working Paper No. 2018/232*.

Junge, V., Revilla Diez, J. and Schätzl, L. (2015). Determinants and Consequences of Internal Return Migration in Thailand and Vietnam. *World Development*, **71**, 94–106.

Kennan, J. and Walker, J. R. (2011). The Effect of Expected Income on Individual Migration Decisions. *Econometrica*, **79** (1), 211–251.

Kone, Z. L., Liu, M., Mattoo, A., Ozden, C. and Sharma, S. (2018). Internal borders and migration in India. *Journal of Economic Geography*, **18** (4), 729–759.

Lacuesta, A. (2010). A Revision of the Self-selection of Migrants Using Returning Migrant's Earnings. *Annals of Economics and Statistics*, **1** (97/98), 235–259.

Lagakos, D. (2020). Urban-rural gaps in the developing world: Does internal migration offer opportunities? *Journal of Economic Perspectives*, **34** (3), 174–92.

—, Marshall, S., Mobarak, A. M., Vernot, C. and Waugh, M. E. (2020). Migration costs and observational returns to migration in the developing world. *Journal of Monetary Economics*, **113** (C), 138–154.

—, Mobarak, A. M. and Waugh, M. E. (2023). The welfare effects of encouraging rural–urban migration. *Econometrica*, **91** (3), 803–837.

Oliveira, J. and Pereda, P. (2020). The impact of climate change on internal migration in Brazil. *Journal of Environmental Economics and Management*, **103**, 102340.

Patto, T. (2024). The urban wage premium over the life cycle and the big-city job ladder. *Working Paper*.

Quiñones, E. J. and Barham, B. L. (2018). Endogenous selection, migration and occupation outcomes for rural southern Mexicans. *Working Paper*.

Roca, J. D. L. and Puga, D. (2017). Learning by Working in Big Cities. *The Review of Economic Studies*, **84** (1), 106–142.

Sahota, G. S. (1968). An Economic Analysis of Internal Migration in Brazil. *Journal of Political Economy*, **76** (2), 218–245.

Santos, C. and Ferreira, P. C. (2007). Migração e Distribuição Regional de Renda no Brasil. *Pesquisa e Planejamento Econômico*, **37** (3), 405–426.

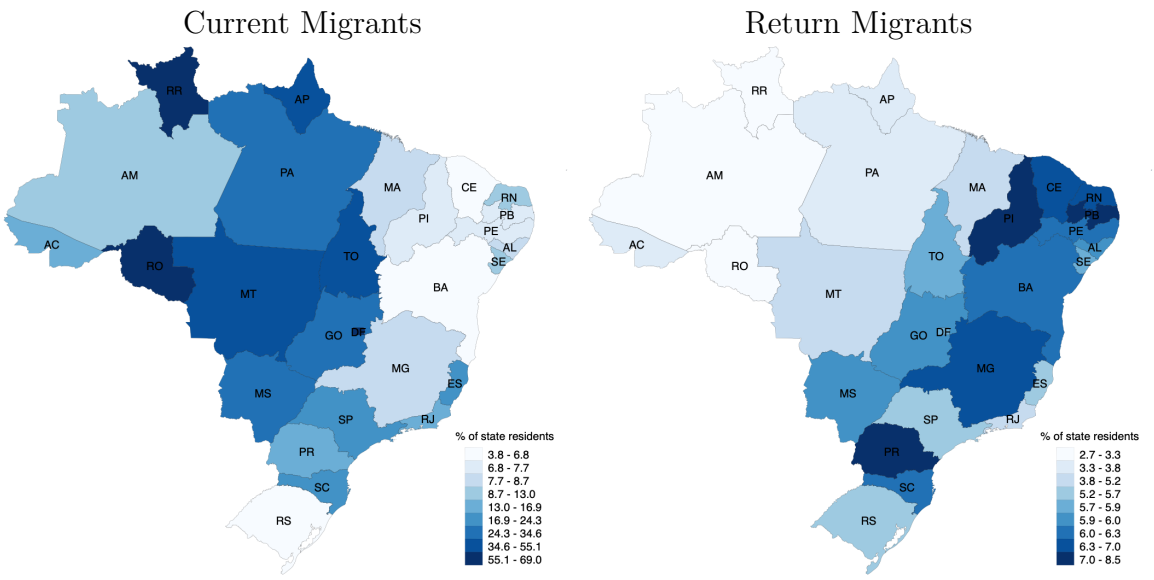
Tannen, M. B. (1992). Migration from the Northeast to the Southeast in Brazil: Do Migrants Succeed? *Review of Urban & Regional Development Studies*, **4** (1), 32–49.

Wahba, J. (2015). Selection, selection, selection: the impact of return migration. *Journal of Population Economics*, **28** (3), 535–563.

Yap, L. (1976). Internal Migration and Economic Development in Brazil. *The Quarterly Journal of Economics*, **90** (1), 119.

Tables & Figures

Figure 1: Migrant Shares by State: Current and Return Migrants



Notes: Data from the Brazilian population census in 2010. Sample restricted to individuals ages 25–54. The left map shows the share of current interstate immigrants: individuals living in the state but born elsewhere. The right map shows the share of return interstate migrants: individuals born and living in the state who have lived in a different state. Both maps show percentages with the total number of migrants in the numerator and the number of individuals living in the state in the denominator.

Table 1: Summary Statistics: 2010

	Mean	SD	Min.	Max.	N(millions)
Current Migrant	0.16	0.37	0	1	8.54
Return Migrant	0.06	0.24	0	1	8.54
Age	38.08	8.53	25	54	8.54
Female	0.51	0.50	0	1	8.54
Rural	0.21	0.41	0	1	8.54
Household Head	0.45	0.50	0	1	8.54
Household Members	2.86	1.41	1	60	8.54
Years of Education	8.26	4.72	0	22	8.54
Education: None	0.06	0.22	0	1	8.54
Education: Basic	0.43	0.49	0	1	8.54
Education: Some High School	0.16	0.37	0	1	8.54
Education: Some College	0.26	0.44	0	1	8.54
Education: College	0.10	0.30	0	1	8.54
Race: White	0.48	0.50	0	1	8.54
Race: Black	0.08	0.27	0	1	8.54
Race: Mixed	0.42	0.49	0	1	8.54
Employment: Private Formal	0.45	0.50	0	1	8.54
Employment: Public/Military	0.07	0.25	0	1	8.54
Employment: Private Informal	0.21	0.41	0	1	8.54
Employment: Self-Employed	0.24	0.43	0	1	8.54
Employment: Employer	0.02	0.14	0	1	8.54
Positive Earnings	0.67	0.47	0	1	8.54
Earnings/Month (R$)	2,454	3,114	45	42,794	8.54
Hours Worked/Week	40.74	14.54	2	125	8.54
Wage/Hour (R$)	16.09	28.00	0.12	3,954.83	8.54
Log Earnings	7.40	0.87	3.81	10.66	8.54
Log Wage	2.29	0.91	-2.15	8.28	8.54

Notes: Data from the Brazilian population census in 2010. Sample restricted to individuals ages 25–54.

Table 2: Difference in Means by Migrant Status: 2010

	Non-Migrants (N)	Current Mig. (M)	Return Mig. (R)	M - N	R - N	M - R
Age	37.79	39.16	38.89	1.36	1.09	0.27
Female	0.51	0.50	0.52	0.00	0.01	-0.01
Rural	0.23	0.14	0.15	-0.09	-0.08	-0.01
Household Head	0.43	0.49	0.49	0.06	0.06	0.00
Household Members	2.90	2.74	2.68	-0.15	-0.22	0.06
Years of Education	8.22	8.23	8.89	0.01	0.67	-0.65
Educ: None	0.05	0.05	0.03	0.00	-0.02	0.02
Educ: Basic	0.43	0.44	0.40	0.01	-0.02	0.03
Educ: Some High School	0.16	0.17	0.16	0.01	0.01	0.00
Educ: Some College	0.26	0.25	0.26	-0.02	0.00	-0.02
Educ: College	0.10	0.10	0.13	0.00	0.03	-0.03
Race: White	0.48	0.46	0.51	-0.02	0.03	-0.05
Race: Black	0.08	0.08	0.07	0.00	-0.01	0.01
Race: Mixed	0.42	0.45	0.41	0.03	-0.02	0.05
Empl: Private Formal	0.44	0.48	0.42	0.04	-0.02	0.06
Empl: Public/Military	0.07	0.06	0.08	-0.01	0.01	-0.01
Empl: Private Informal	0.21	0.19	0.21	-0.02	0.00	-0.02
Empl: Self-Employed	0.24	0.23	0.25	-0.01	0.01	-0.02
Empl: Employer	0.02	0.02	0.03	0.00	0.01	-0.01
Positive Earnings	0.66	0.71	0.69	0.05	0.03	0.02
Earnings/Month (R$)	2,340	2,806	2,903	466	562	-96
Hours Worked/Week	40.49	41.93	40.58	1.44	0.09	1.35
Wage/Hour (R$)	15.41	17.93	19.57	2.53	4.17	-1.64
Log Earnings	7.36	7.54	7.49	0.18	0.13	0.05
Log Wage	2.26	2.41	2.41	0.15	0.15	0.00
Observations	6,663,311	1,371,882	505,425			

Notes: Data from the Brazilian population census in 2010. Sample restricted to individuals ages 25–54. All non-zero differences are statistically significant at 1% or less.

Table 3: Wage Premium of Current and Return Migrants

Dependent Variable:	Log Hourly Wage				
	(1)	(2)	(3)	(4)	(5)
Current Migrant	0.149	0.147	0.196	0.101	0.088
	(0.073)*	(0.026)***	(0.032)***	(0.009)***	(0.008)***
Return Migrant	0.147	0.076	0.086	0.087	0.070
	(0.014)***	(0.010)***	(0.009)***	(0.009)***	(0.007)***
Current – Return	0.002	0.070	0.110	0.014	0.018
	(0.066)	(0.033)**	(0.025)***	(0.011)	(0.010)*
Adj. R^2	0.00	0.30	0.32	0.33	0.39
Observations	5,702,626	5,702,626	5,702,626	5,702,626	5,702,626
Covariates		Basic	Basic	Basic	Extended
State Fixed Effects			Birth	Birth & Resid.	Birth & Resid.

Notes: The dependent variable is the log of hourly wage (considering only the person's main work). Basic covariates are age, age squared, indicators for female and rural, education categories, and race categories. The extended set of covariates also includes a head-of-household indicator, the number of household members, a student indicator, occupation categories, industry categories, and employment categories. Standard errors in parentheses clustered by state of residence. Stars denote: * p<0.10; ** p<0.05; *** p<0.01.

Table 4: Varying Migrant Definitions using State or Municipality Levels

Dependent Variable:	Log Hourly Wage			
	(1)	(2)	(3)	(4)
Current Migrant (interstate)	0.094	0.101	0.119	
	(0.009)***	(0.009)***	(0.011)***	
Return Migrant (interstate)		0.087	0.105	
		(0.009)***	(0.006)***	
Current Migrant (intrastate)			0.050	
			(0.014)***	
Return Migrant (intrastate)			0.031	
			(0.007)***	
Current Migrant (inter & intrastate)				0.073
				(0.013)***
Return Migrant (inter & intrastate)				0.072
				(0.006)***
Adj. R^2	0.32	0.33	0.33	0.33
Observations	5,702,626	5,702,626	5,702,626	5,702,626

Notes: The dependent variable is the log of hourly wage (considering only the person's main work). The migrant definitions used in each column are mutually exclusive (non-overlapping). For a current interstate migrant, the state of residence and birth are different. For a return interstate migrant, the state of residence and birth are equal, but the individual has lived in a different state before. For a current intrastate migrant, the state of residence and birth are equal, but the municipality of residence and birth are different. And for a return intrastate migrant, the municipality of residence and birth are equal, but the individual has lived in a different municipality within their birth state before. In column (4), the migrant definitions ignore the state level, such that a current migrant is an individual living in a municipality different than their municipality of birth in any state, and a return migrant is an individual living in their municipality of birth that has lived in a different municipality in any state before. All specifications include the basic set of covariates (age, age squared, female indicator, rural indicator, education categories, and race categories), state of residence fixed effects, and state of birth fixed effects. Standard errors in parentheses clustered by state of residence. Stars denote: * p<0.10; ** p<0.05; *** p<0.01.

Table 5: Extensive and Intensive Margins: Wages and Labor Supply by Migrant Status

Dependent Variable:	Log Wage	Hours Worked	Log Earnings	IHS Earnings	Positive Earnings
	(1)	(2)	(3)	(4)	(5)
Current Migrant	0.101	0.962	0.126	0.244	0.022
	(0.009)***	(0.079)***	(0.011)***	(0.027)***	(0.003)***
Return Migrant	0.087	0.118	0.074	0.094	0.008
	(0.009)***	(0.160)	(0.007)***	(0.037)**	(0.005)*
Current – Return	0.014	0.844	0.053	0.150	0.014
	(0.011)	(0.140)***	(0.008)***	(0.024)***	(0.003)***
Mean Dependent Var.	2.29	41.31	7.40	5.41	0.67
Adj. R^2	0.33	0.05	0.40	0.23	0.17
Observations	5,702,626	5,702,626	5,702,626	8,540,616	8,540,616

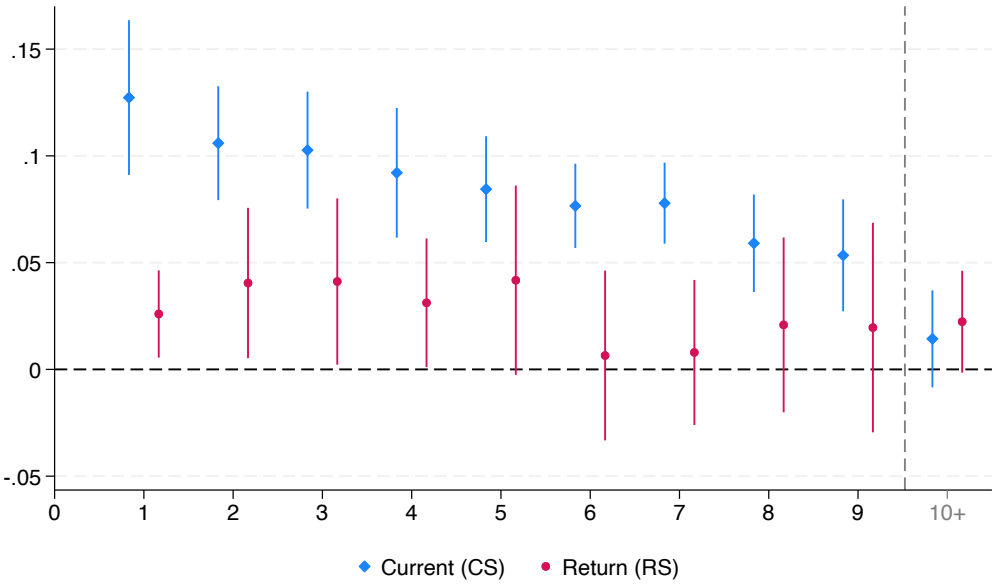
Notes: The dependent variable in all columns considers only the person's main work and is, in column (1), the log of hourly wage; in column (2), the number of hours worked per week; in column (3), the log of monthly earnings, excluding zeros; in column (4), the IHS transformation of monthly earnings, including zeros; and in column (5), an indicator for positive earnings. All specifications include basic covariates (age, age squared, female, rural, education categories, and race categories), state of residence fixed effects, and state of birth fixed effects. Standard errors in parentheses clustered by state of residence. Stars denote: * p<0.10; ** p<0.05; *** p<0.01.

Table 6: Evolution of the Migrant Wage Premium: Years since Migrating or Returning

Dependent Variable:	Log Hourly Wage		
	(1)	(2)	(3)
Current Migrant	0.101		
	(0.009)***		
<10 years in the state		0.164	0.135
		(0.013)***	(0.018)***
10+ years in the state		0.076	
		(0.008)***	
Return Migrant	0.087		
	(0.009)***		
<10 years in the state		0.094	0.083
		(0.013)***	(0.013)***
10+ years in the state		0.083	
		(0.007)***	
Adj. R^2	0.33	0.33	0.32
Observations	5,702,626	5,702,626	5,702,626

Notes: The dependent variable is the log of hourly wage (considering only the person's main work). All specifications include basic covariates (age, age squared, female, rural, education categories, and race categories), state of residence fixed effects, and state of birth fixed effects. Standard errors in parentheses clustered by state of residence. Stars denote: * p<0.10; ** p<0.05; *** p<0.01.

Figure 2: Evolution of the Migrant Wage Premium: Years since Migrating or Returning



Notes: The figure shows, on the vertical axis, the coefficients from a regression of log wage on migrant definitions (Current and Return, interstate) interacted with indicators for number of years the individual has lived in the state (years since migrating, for current migrants, or years since returning, for return migrants). The 10+ category groups together those living for ten years or more in a state. The vertical bars show 95% confidence intervals.

Table 7: Evolution of the Migrant Wage Premium from 2000 to 2010

Dependent Variable:	Log Hourly Wage			Hours	Log Earnings	Pos. Earnings
Census Year:	2010	2000	2000 & 2010		2000 & 2010	
	(1)	(2)	(3)	(4)	(5)	(6)
Current Migrant, 2010	0.101		0.102	0.878	0.122	0.027
	(0.009)***		(0.012)***	(0.141)***	(0.014)***	(0.005)***
Return Migrant, 2010	0.087		0.081	0.137	0.069	0.012
	(0.009)***		(0.011)***	(0.166)	(0.008)***	(0.005)**
Current Migrant, 2000		0.126	0.122	1.251	0.151	0.003
		(0.013)***	(0.015)***	(0.127)***	(0.016)***	(0.003)
Return Migrant, 2000		0.061	0.058	0.600	0.068	-0.001
		(0.006)***	(0.009)***	(0.075)***	(0.010)***	(0.002)
Current – Return, 2010	0.014		0.021	0.741	0.052	0.015
	(0.011)		(0.014)	(0.166)***	(0.012)***	(0.004)***
Current – Return, 2000		0.065	0.065	0.650	0.083	0.005
		(0.010)***	(0.010)***	(0.108)***	(0.010)***	(0.003)
Adj. R^2	0.33	0.45	0.38	0.08	0.43	0.18
Observations	5,702,626	4,714,958	10,417,584	11,129,963	10,447,816	16,179,732

Notes: The dependent variable in all columns considers only the person's main work and is, in columns (1–3), the log of hourly wage; in column (4), the number of hours worked per week; in column (5), the log of monthly earnings, excluding zeros; and in column (6), an indicator for positive earnings. In columns (3–6), the coefficients for Current and Return Migrant in 2010 come from a linear combination of two regression coefficients: the coefficient on the migrant indicator plus the coefficient on the interaction of the migrant and the Year = 2010 indicators. All specifications include basic covariates (age, age squared, female, rural, education categories, and race categories), state of residence fixed effects, and state of birth fixed effects. Standard errors in parentheses clustered by state of residence. Stars denote: * p<0.10; ** p<0.05; *** p<0.01.

Table 8: Heterogeneity in the Migrant Wage Premium: Alternative Sample Restrictions

Dependent Variable:	Log Hourly Wage			
Sample:	Age 16+	Males	College	Formal Emp.
	(1)	(2)	(3)	(4)
Current Migrant	0.094	0.124	0.128	0.082
	(0.010)***	(0.010)***	(0.016)***	(0.010)***
Return Migrant	0.087	0.104	0.142	0.090
	(0.009)***	(0.008)***	(0.013)***	(0.010)***
Current – Return	0.008	0.020	-0.014	-0.008
	(0.011)	(0.010)**	(0.008)*	(0.008)
Adj. R^2	0.32	0.33	0.14	0.33
Observations	8,095,167	3,332,765	2,427,557	3,023,810

Notes: The regression sample in each column includes: in column (1), all individuals ages 16 and above; in column (2), only prime-age males; in column (3), only prime-age individuals with a college education (complete or incomplete); in column (4), only prime-age individuals formally employed (private and public sectors, including military). The dependent variable is the log of hourly wage (considering only the person's main work). All specifications include basic covariates (age, age squared, female, rural, education categories, and race categories), state of residence fixed effects, and state of birth fixed effects. Standard errors in parentheses clustered by state of residence. Stars denote: * p<0.10; ** p<0.05; *** p<0.01.

Table 9: Longitudinal Analysis using RAIS data

Dependent Variable:	Log Hourly Wage					
Year(s):	2010	2000	2000 & 2010		2000 to 2010	
	(1)	(2)	(3)	(4)	(5)	(6)
Current Migrant	0.126	0.105	0.126	0.026	0.140	0.027
	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.000)***	(0.000)***
Return Migrant	0.134	0.015	0.129	0.001	0.156	0.027
	(0.001)***	(0.001)***	(0.001)***	(0.001)	(0.001)***	(0.001)***
Current – Return	-0.009	0.090	-0.003	0.026	-0.016	-0.001
	(0.001)***	(0.002)***	(0.001)***	(0.002)***	0.000	0.000
Adj. R^2	0.43	0.45	0.43	0.33	0.44	0.19
Observations	26.5M	16M	42.5M	42.5M	246.3M	246.3M
Individuals	26.5M	16M	33.7M	33.7M	42.5M	42.5M
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Individual Fixed Effects				Y		Y

Notes: The dependent variable is the log of hourly wage (considering only the person's main work). Both the current and the return migrant definitions consider the state where the individual is first observed in the data as their "home." The sample in all columns includes only individuals observed at least twice in the 2000–2010 period. In columns (1–2), the sample is restricted to individuals formally employed (private and public sectors, including the military) on December 31 of each year (2010 or 2000). In columns (3–4), the sample is restricted to individuals formally employed on December 31 in both 2000 and 2010, and in columns (5–6), to those employed on December 31 in at least two years from 2000 to 2010. In column (2), where only the year 2000 is used, individuals are defined as current or return migrants if they are observed as such in 2010. All specifications include the basic set of covariates (age, age squared, female indicator, education categories, and race categories) and state of residence fixed effects. The number of observations and the number of individuals are shown in millions. Standard errors in parentheses clustered by individual. Stars denote: * p<0.10; ** p<0.05; *** p<0.01.

Appendix

Table A1: Summary Statistics: 2000

	Mean	SD	Min.	Max.	N(millions)
Current Migrant	0.21	0.41	0	1	7.64
Return Migrant	0.10	0.30	0	1	7.64
Age	37.61	8.30	25	54	7.64
Female	0.51	0.50	0	1	7.64
Rural	0.21	0.41	0	1	7.64
Household Head	0.46	0.50	0	1	7.64
Household Members	2.95	1.43	1	60	7.64
Education: None	0.10	0.30	0	1	7.64
Education: Basic	0.44	0.50	0	1	7.64
Education: Some High School	0.21	0.41	0	1	7.64
Education: Some College	0.18	0.38	0	1	7.64
Education: College	0.07	0.25	0	1	7.64
Race: White	0.55	0.50	0	1	7.64
Race: Black	0.07	0.25	0	1	7.64
Race: Asian	0.00	0.06	0	1	7.64
Race: Mixed	0.37	0.48	0	1	7.64
Employment: Private Formal	0.37	0.48	0	1	5.06
Employment: Public/Military	0.28	0.45	0	1	5.06
Employment: Employer	0.03	0.18	0	1	5.06
Employment: Self-Employed	0.26	0.44	0	1	5.06
Positive Earnings	0.62	0.49	0	1	7.64
Earnings/Month (R$)	2,731.72	4,228.73	47	64,143	4.73
Hours Worked/Week	44.11	14.51	4	98	5.04
Wage/Hour (R$)	15.42	26.69	0.11	2,463.48	4.71
Log Earnings	7.34	1.01	3.85	11.07	4.73
Log Wage	2.14	1.02	-2.17	7.81	4.71

Notes: Data from the Brazilian population census in 2000. Sample restricted to individuals ages 25–54.

Table A2: Difference in Means by Migrant Status: 2000

	Non-Migrants (N)	Current Mig. (M)	Return Mig. (R)	M - N	R - N	M - R
Age	37.17	38.48	38.72	1.31	1.55	-0.24
Female	0.51	0.51	0.51	0.00	0.00	0.00
Rural	0.24	0.15	0.16	-0.09	-0.08	-0.01
Household Head	0.44	0.50	0.49	0.06	0.05	0.01
Household Members	2.99	2.85	2.88	-0.14	-0.11	-0.03
Educ: None	0.10	0.10	0.09	0.00	-0.01	0.01
Educ: Basic	0.44	0.46	0.44	0.02	0.00	0.02
Educ: Some High School	0.21	0.22	0.22	0.01	0.01	0.00
Educ: Some College	0.19	0.16	0.18	-0.03	-0.01	-0.02
Educ: College	0.07	0.06	0.07	-0.01	0.00	-0.01
Race: White	0.55	0.54	0.56	-0.01	0.01	-0.02
Race: Black	0.07	0.06	0.06	-0.01	-0.01	0.00
Race: Asian	0.00	0.00	0.01	0.00	-0.01	-0.01
Race: Mixed	0.36	0.39	0.37	0.03	0.01	0.02
Empl: Formal	0.36	0.41	0.36	0.05	0.00	0.05
Empl: Private Informal	0.28	0.27	0.28	-0.01	0.00	-0.01
Empl: Employer	0.26	0.25	0.27	-0.01	0.01	-0.02
Empl: Self-Employed	0.03	0.03	0.04	0.00	0.01	-0.01
Positive Earnings	0.62	0.64	0.62	0.02	0.00	0.02
Earnings/Month (R$)	2,599	3,049	2,942	450	343	107
Hours Worked/Week	43.64	45.56	44.32	1.92	0.68	1.24
Wage/Hour (R$)	14.79	16.78	16.75	1.99	1.96	0.03
Log Earnings	7.29	7.50	7.38	0.21	0.09	0.12
Log Wage	2.10	2.27	2.18	0.17	0.08	0.09
Observations	5,246,971	1,605,195	786,950			

Notes: Data from the Brazilian population census in 2000. Sample restricted to individuals ages 25–54. All non-zero differences are statistically significant at 1% or less, except for the M – R difference in Wage/Hour, which is not statistically significant. Formal employment includes private and public sectors plus the military.

Table A3: Summary Statistics, RAIS data 2010

	Mean	SD	Min.	Max.	N(millions)
Current Migrant	0.06	0.25	0	1	26.5
Return Migrant	0.02	0.14	0	1	26.5
Age	37.86	7.92	25	54	26.5
Female	0.41	0.49	0	1	26.5
Education: Basic	0.17	0.37	0	1	26.5
Education: Some High School	0.20	0.40	0	1	26.5
Education: Some College	0.45	0.50	0	1	26.5
Education: College	0.18	0.38	0	1	26.5
Race: White	0.48	0.50	0	1	26.5
Race: Black	0.04	0.20	0	1	26.5
Race: Mixed	0.22	0.42	0	1	26.5
Race: Ignored	0.25	0.43	0	1	26.5
Earnings/Month (R$)	4,116	5,736	321	160,683	26.5
Hours Worked/Week	41.29	5.52	10	44	26.5
Wage/Hour (R$)	24.08	34.64	1.68	1,916.19	26.5
Log Earnings	7.93	0.78	5.77	11.99	26.5
Log Wage	2.75	0.82	0.52	7.56	26.5

Notes: Data from RAIS in 2010. Sample restricted to individuals ages 25–54.

Table A4: Differences in Means by Migrant Status, RAIS data 2010

	Non-Migrants (N)	Current Mig. (M)	Return Mig. (R)	M - N	R - N	M - R
Age	37.88	37.21	39.25	-0.67	1.37	-2.04
Female	0.42	0.27	0.25	-0.15	-0.17	0.02
Educ: Basic	0.17	0.16	0.16	-0.01	-0.01	0.00
Educ: Some High School	0.20	0.18	0.19	0.02	-0.01	-0.01
Educ: Some College	0.45	0.43	0.44	-0.02	-0.01	-0.01
Educ: College	0.18	0.22	0.21	0.04	0.03	0.01
Race: White	0.48	0.48	0.49	0.00	0.01	-0.01
Race: Black	0.04	0.04	0.05	0.00	0.01	-0.01
Race: Mixed	0.22	0.31	0.30	0.09	0.08	0.01
Race: Ignored	0.26	0.17	0.15	-0.09	-0.11	0.02
Earnings/Month (R$)	3,968	5,702	5,678	1,734	1,710	24
Hours Worked/Week	41.20	42.23	42.29	1.03	1.09	-0.06
Wage/Hour (R$)	23.33	32.18	31.79	8.85	8.46	0.39
Log Earnings	7.91	8.12	8.15	0.21	0.24	-0.03
Log Wage	2.73	2.92	2.94	0.19	0.21	-0.02
Observations	24,265,675	1,714,792	568,547			

Notes: Data from RAIS in 2010. Sample restricted to individuals ages 25–54. All non-zero differences are statistically significant at 1% or less, except for the M – R difference in Earnings/Month, which is significant at 5%.

Table A5: Cross-Tabulation of Migrant Definitions: Number of Individuals in 2010

	Current (inter & intrastate)	Return (inter & intrastate)	All	%
Current (interstate)	1,371,882	0	1,371,882	16.06
Return (interstate)	185,505	319,920	505,425	5.92
Current (intrastate)	2,189,497	0	2,189,497	25.64
Return (intrastate)	0	125,018	125,018	1.46
Non-Migrant (intrastate)	0	0	4,348,796	50.92
Total	3,746,884	444,938	8,540,618	100

Notes: Data from the Brazilian population census in 2010. Sample restricted to individuals ages 25–54. Current interstate migrant: the state of residence and birth are different. Return interstate migrant: the state of residence and birth are equal, but the individual has lived in a different state before. Current intrastate migrant: the state of residence and birth are equal, but the municipality of residence and birth are different. Return intrastate migrant: the municipality of residence and birth are equal, but the individual has lived in a different municipality within their birth state before. The "inter" & "intrastate" definitions consider only the municipality level. In this case, a current migrant is an individual living in a municipality different than their municipality of birth in any state, and a return migrant is an individual living in their municipality of birth that has lived in a different municipality in any state before.

Table A6: Migrant and Employment Statuses

Dependent Variable:	Private Formal	Public/Military	Private Informal	Self-Employed	Employer	All Formal
	(1)	(2)	(3)	(4)	(5)	(6)
Current Migrant	0.015	-0.011	-0.008	-0.002	0.005	0.004
	(0.003)***	(0.002)***	(0.002)***	(0.002)	(0.001)***	(0.003)
Return Migrant	-0.026	-0.005	0.008	0.013	0.007	-0.031
	(0.005)***	(0.002)***	(0.002)***	(0.003)**	(0.001)***	(0.005)***
Current – Return	0.041	-0.005	-0.016	-0.014	-0.002	0.035
	(0.005)***	(0.001)***	(0.002)***	(0.003)***	(0.001)*	(0.005)***
Mean Dependent Var.	0.45	0.07	0.21	0.24	0.02	0.52
Adj. R^2	0.10	0.10	0.08	0.07	0.02	0.12
Observations	5,849,616	5,849,616	5,849,616	5,849,616	5,849,616	5,849,616

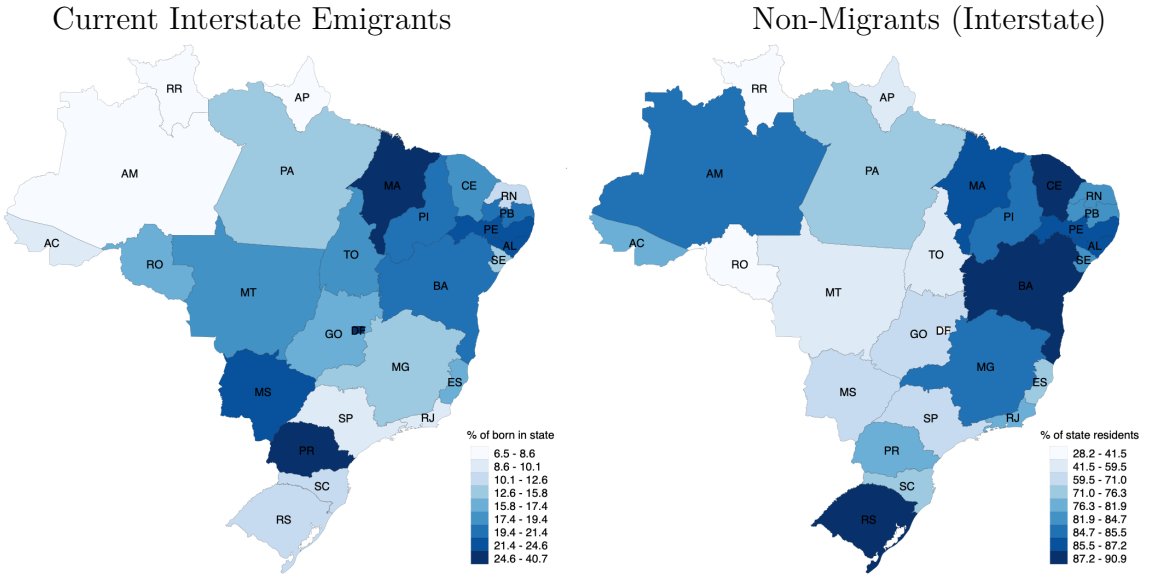
Notes: The dependent variable in each column is an indicator for the employment status in the person's main work. It is, in column (1), private sector, formal; in column (2), public service or military; in column (3), private sector, informal; in column (4), self-employed; in column (5), employer; and in column (6), all formal statuses combined (public, military, and private). All specifications use a linear probability model regression and include basic covariates (age, age squared, female, rural, education categories, and race categories), state of residence fixed effects, and state of birth fixed effects. Standard errors in parentheses clustered by state of residence.

Table A7: Longitudinal Analysis using RAIS data: Alternative Sets of Migrant Definitions

Dependent Variable:	Log Hourly Wage					
Year(s):	2010		2000 & 2010			
	(1)	(2)	(3)	(4)	(5)	(6)
Current Migrant (interstate)	0.122	0.114	0.123	0.112	0.026	0.008
		(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.001)***
Return Migrant (interstate)		0.143		0.001		0.008
		(0.000)***		(0.001)		(0.001)***
Current Migrant (intrastate)		-0.051		0.026		-0.066
		(0.000)***		(0.002)***		(0.001)***
Return Migrant (intrastate)		-0.021		0.33		-0.065
		(0.001)***		(0.002)***		(0.001)***
Adj R^2	0.43	0.43	0.43	0.43	0.33	0.33
Observations	26.5M	26.5M	42.5M	42.5M	42.5M	42.5M
Individuals	26.5M	26.5M	33.7M	33.7M	33.7M	33.7M
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Individual Fixed Effects					Y	Y

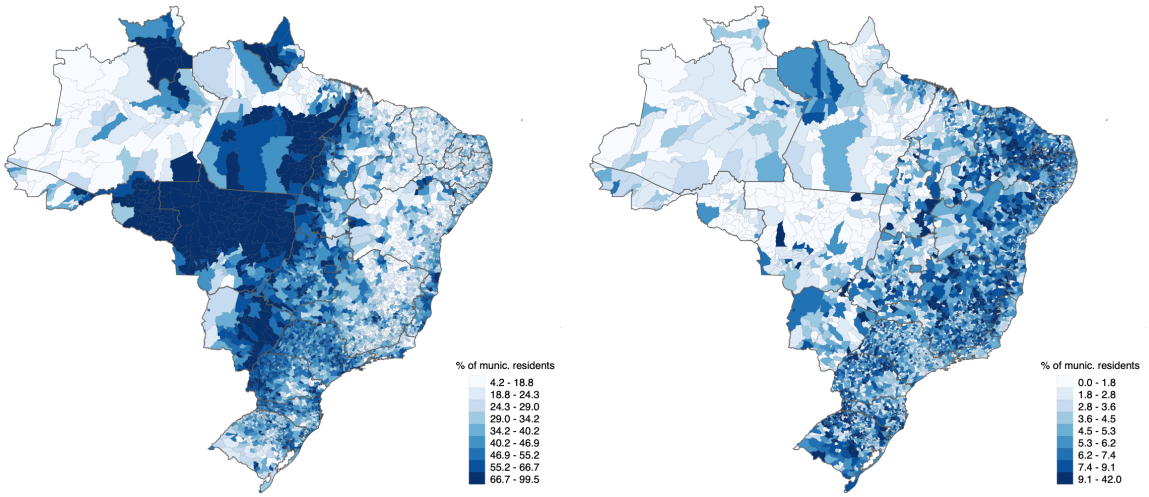
Notes: The dependent variable is the log of hourly wage (considering only the person's main work). Both the current and the return migrant definitions consider the state (or municipality) where the individual is first observed in the data as their home. The sample in all columns includes only individuals observed at least twice in the 2000–2010 period. In columns (1–2), the sample is restricted to individuals formally employed (private and public sectors, including the military) on December 31, 2010, and in columns (3–6), to those formally employed on December 31 in both 2000 and 2010. All specifications include the basic set of covariates (age, age squared, female indicator, education categories, and race categories) and state of residence fixed effects. The number of observations and the number of individuals are shown in millions. Standard errors in parentheses clustered by individual.

Figure A1: Migrant Shares by State: Current Emigrants and Non-Migrants



Notes: Data from the Brazilian population census in 2010. Sample restricted to individuals ages 25–54. The left map shows the share of current interstate emigrants: individuals born in the state but living elsewhere. The right map shows the share of non-migrants: individuals born and living in the state who never lived in a different state. Both maps show percentages with the total number of migrants (or non- migrants) in the numerator. In the first (left), the denominator is the number of individuals born in the state, and in the second (right), the number of individuals living in the state.

Figure A2: Migrant Shares by State: Current Emigrants and Non-Migrants



Notes: Data from the Brazilian population census in 2010. Sample restricted to individuals ages 25–54. The left map shows the share of current interstate emigrants: individuals born in the state but living elsewhere. The right map shows the share of non-migrants: individuals born and living in the state who never lived in a different state. Both maps show percentages with the total number of migrants (or non- migrants) in the numerator. In the first (left), the denominator is the number of individuals born in the state, and in the second (right), the number of individuals living in the state.