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

Eduardo Cenci†

July 22, 2021

Abstract

This paper investigates the components and the recent evolution of the wage premium of current and return migrants within Brazil. Using cross-sectional data from repeated household surveys in 2004–2014, I find that the wages of internal migrants are about 12% higher than the wages of non-migrants. For return migrants, wages are 9% higher on average. I also find that the wage premium for migrants decreased during this period, while the wage premium for return migrants increased. Using longitudinal data from linked employer-employee datasets in 2005–2015, I find a 5–10% wage premium for both migrants and return migrants in panel regressions with individual fixed effects for a subsample of formal sector workers. Restricting the sample to those who move at some point in the panel, I find no wage premium associated with the current migrant status and a 4% wage penalty associated with returning. I explore different regression specifications, subsamples, and an instrumental variables strategy based on past migration rates to discuss the role of self-selection, place-specific effects, and learning on these wage premia. My results suggest that the self-selection of internal migrants in Brazil is based more on absolute advantage (migrants earn more in any location) than comparative advantage (migrants earn more in a specific location). My results are also consistent with learning impacting post-migration earnings regardless of a migrant's location.

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

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

*I 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), and the Universidade Católica de Brasília for valuable comments. I acknowledge support from Brazil's Institute for Applied Economic Research (IPEA) where part of this project was developed. Any errors are my own.

†University of Nebraska at Omaha (contact: ecenci@unomaha.edu).

1 Introduction

The existence of a wage premium for migrants is an empirical regularity well-documented in the literature, being observed in different settings. Such a premium appears for both international and domestic migrants in both developed or developing countries. The earnings of people who departed a given location are often higher than the earnings of those who stayed. In some cases, the earnings of migrants are also higher than the earnings of comparable individuals at the destination (Borjas, 1987; Chiswick, 1999; Clemens, Montenegro and Pritchett, 2009). Explanations for this difference, the migrant wage premium, include positive self-selection of migrants (the more skilled and more likely to succeed migrate), place premium (higher productivity at the destination), survival bias (only successful migrants are observed), and learning (migrants acquire new skills at the destination or improve their existing skills).[1]

Most empirical documentation on migrant wage premia comes from the international migration literature, which often focuses on migration from poor countries to rich ones. Nonetheless, migrant wage premia are also documented for internal migrants, i.e., for people living and working within a free-movement zone in a different location than the one they originated.[2] Examples of such free-movement zones are the European Union and the ECOWAS in West Africa; and, more commonly, "continent-sized" countries like the US, China, and Brazil. Wage differentials of internal migrants observed in the U.S. are smaller than those observed for international migrants, but just as consistently observed (Card, 2009; Clemens et al., 2009; Kennan and Walker, 2011; Kennan, 2013).

The wage premium of return migrants, however, is less documented. It is also more likely to be small or even negative. Looking at foreign-born migrants in the US, Borjas and Bratsberg (1996) argue that the less able migrants tend to return, reinforcing the positive self-selection of those who stay. This negative self-selection of return migrants would then imply a negative wage premium. After discounting place specific effects, the observed earnings of return migrants would be lower than those of current migrants but potentially still higher than the earnings of those who have never migrated (Co, Gang and Yun, 2000). Similar arguments can be made for internal migration, for people moving to different locations within a country and then back to their original location.

This paper uses two large datasets with information on labor earnings and the mobility of Brazilian individuals. I use cross-sectional data from repeated household

[1] See, for example, D'Costa and Overman (2014) or Roca and Puga (2017).

[2] A migrant's origin can be defined as his birthplace, the location where he is first observed in the data, or his location in a previous period (generally 5 or 10 years ago). In this paper, I start using birthplace and switch to location five years ago when moving from cross-sectional to longitudinal data because my longitudinal data does not provide information on workers' birthplace.

surveys in 2004–2014 and longitudinal data from linked employer-employee datasets in 2005–2015 to investigate the composition and evolution of wage differentials for migrants and return migrants in recent years. In particular, I investigate whether return migrants keep part of their wage premium after returning, thus remaining on the right tail of the residual earnings distribution in their original states.[3] Additionally, I try to disentangle the components of the wage premium to identify how much of it is explained by self-selection, place-specific factors, and learning. I use state of residence fixed effects to proxy for the place-specific factors that influence the wage premium. I use an Instrumental Variables (IV) strategy and panel regressions with individual fixed effects in an attempt to circumvent self-selection biases. The analysis of learning done in this paper is simple, but it complements the more robust evidence I gather on self-selection.[4] Finally, the large period covered by the data allows an analysis of how the wage premium of migrants and return migrants might have changed with the recent evolution of Brazilian labor markets.

Brazil is suitable for this type of investigation for a variety of reasons. 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.[5] There has been little international migration both in and out of Brazil in the past decades, which allows me to focus on national workers only. Brazilian household surveys provide a rich set of information on labor and demographics at the individual and household levels for a large period. Brazilian linked employer-employee data, albeit not having information on households and covering only formal sector employees, also provides rich individual-level information. And its longitudinal structure allows the use of individual fixed effects, an important advantage in labor studies because it allows the researcher to control for time-invariant unobserved characteristics. Finally, prior to the recession starting circa 2015, Brazil experienced steady economic growth, inequality reduction, and strengthening of its labor markets and regional integration. The great recession in 2008–2009 did not disrupt Brazilian labor markets as it did in the US and Europe, so my analysis benefits from some "stability" as well. Moreover, the context of economic growth and inequality reduction may help explain the evolution of wage premia observed in this data, particularly for return migrants who were returning to fast-growing states.

[3]Specifically, I want to verify whether return migrants earn more than their non-migrant counterparts at home after controlling for observed characteristics such as demographics and labor occupations, as well as location fixed effects.

[4]An investigation of learning while migrating within Brazil can start with replicating Roca and Puga (2017) with data from Brazil's linked employer-employee dataset (RAIS), which arguably has fewer limitations than the Spanish data used by those authors.

[5]I treat the Federal District, whose population surpasses that of many other Brazilian states, as a state in this study.

I attempt to address the self-selection of migrants in two ways. First, when using cross-sectional data from household surveys, I use an IV strategy based on historical migration rates. This strategy is similar to the approach used in many investigations of Mexican migration to the US (McKenzie and Rapoport, 2007, 2010; Woodruff and Zenteno, 2007; Kaestner and Malamud, 2014). I construct past emigration rates for each Brazilian state using data from the 1970 Census and use these rates as a proxy for migrant networks. Such networks would reduce migration costs and increase potential returns, thus helping predict the current migration status of individuals born in each state. I use a similar measure for return migration, constructing return migration rates for each Brazilian state in 1970.

Second, when using longitudinal data from linked employer-employee datasets, I include individual fixed effects that help account for the individual's unobserved "ability" that is constant over time. I also identify the individuals who move to a different location at some point in the period analyzed, the "movers." Then, restricting my analysis to this subsample of self-selected individuals, I investigate how their wages change with their status of current migrant, return migrant, or non-migrant.[6]

Investigating the returns to migration in labor earnings involves dealing with multiple layers of self-selection. First, there is the self-selection into migrating. Then, there is the self-selection into working and earning an income.[7] Investigating the returns to return migration is even more complicated: people self-select into migrating first, and later, conditional on that first choice, they select into returning or staying (Batista, McIndoe-Calder and Vicente, 2017). The analysis of labor outcomes for return migrants will also involve the self-selection into working or not, which is present for both migrants and return migrants. Quiñones and Barham (2018) refer to this multi-stage process of endogenous selection as "compound-selection" (p.5), which in my case involves the following steps: the decision to migrate, the length of migration (which, when not indefinite, implies returning), and the decision to participate in the labor market while migrating and/or upon returning. The first and the last dimension, decision to migrate and to participate in the labor market, apply to migrants as well. The second dimension, limiting the length of migration and thus returning, applies to return migrants only.

Given the compound selection of migrants and return migrants, other than using an IV strategy and longitudinal data analysis to address selection into migrating/returning, I use different specifications and subsamples to address selection into working and earning

[6]The non-migrant status, in this case, refers to the periods in which the mover has not yet migrated or not yet migrated and returned.

[7]There is also self-selection into particular destinations and in the formal vs. informal sector. The latter is important for my analysis of longitudinal data, which includes only formal sector employees.

labor income. In most of my analysis, I focus on the subsample of workers with a positive labor income and use the log wage as the main outcome of interest. However, I also run my main regression for a full sample of all workers, those employed or not, with and without labor earnings. In the full sample, I use either the IHS transformation of wage (to deal with zeros in the outcome) or an indicator for having positive labor earnings. By using this second outcome in a linear probability model, I shed light on the extensive margin of the migrant wage premium. I complement my analysis by focusing on subsamples of individuals who are more or less likely to self-select into working (e.g., males vs. females, young vs. older individuals, and private vs. public sector workers).

I find consistent evidence of positive self-selection of both current and return internal migrants in Brazil and suggestive evidence of learning impacting post-migration earnings regardless of a migrant's location. Using cross-sectional data from repeated household surveys in 2004–2014, I find that the wages of internal migrants are about 12% higher than the wages of non-migrants. For return migrants, wages are 9% higher on average. Using longitudinal data from linked employer-employee datasets in 2005–2015, I find a 5–10% wage premium for both migrants and return migrants in panel regressions with individual fixed effects for a subsample of formal sector workers. Restricting the sample to those who move at some point in the panel (movers), I find no wage premium associated with the current migrant status and a 4% wage penalty associated with returning. My results suggest that the self-selection of internal migrants in Brazil is based more on absolute advantage (migrants earn more in any location) than comparative advantage (migrants earn more in a specific location).

This study contributes to two strands of the literature. First, I add to the evidence on the effects of return migration on labor earnings (Co et al., 2000; Lacuesta, 2010; Wahba, 2015). My discussion of the wage premium components—learning, location-specific factors, and self-selection—relates most closely to Roca and Puga (2017). My analysis using longitudinal data, in which migration is defined as labor mobility across states and worker's self-selection is partially accounted for with the inclusion of fixed effects, relates to the investigation of sectoral mobility in Alvarez (2020) and Hamory, Kleemans, Li and Miguel (2020). And my investigation of how much of the wage premium remains after returning and my attempts to address the compound selection of returning after migrating relates to Batista et al. (2017) and Quiñones and Barham (2018).

Second, I contribute to the understanding of recent changes in Brazilian labor markets in a context of inequality reduction and sustained economic growth (until 2015). I find that the wage premium for migrants decreased during this period, while the wage premium for return migrants increased. Many studies have looked at internal migration in Brazil. Old studies focused on development, poverty, and inequality (Sahota, 1968;

Yap, 1976; Tannen, 1992), while more recent works focused on labor market adjustments following trade shocks, in particular liberalization in the 90s (Aguayo-Tellez, Muendler and Poole, 2010; Krishna, Poole and Senses, 2014; Dix-Carneiro and Kovak, 2017). Most of these studies use the Brazilian linked employer-employee data and thus are limited to the formal sector.[8] Migration in these studies is seen as a "control," a way to arbitrage for wage differentials between regions—a topic well discussed by Kovak (2013)—not so much as the "treatment:" the focus is not the returns to migration. Some works look at the specific issue of returns to migration, wage differentials, inequality, and its impacts on regional development (Ferreira and Santos, 2007; Avelino, 2010; Fally, Paillacar and Terra, 2010), but they ignore return migration and the evolution of wage premia by analyzing just one year of data and not differentiating between current and return migrants.

The rest of the paper is organized as follows. Section 2 presents the data used in this study, shows some descriptive statistics, and discusses the definition of migrants and the potential instruments for migration. Section 3 presents the empirical methodology and the identification strategy. Section 4 shows the results of OLS regressions using cross-sectional and longitudinal data, results of IV regressions, analysis of subsamples, and robustness checks. Section 5 concludes.

2 Data

2.1 Data sources

I use two main data sources in this study. The first is PNAD (Portuguese acronym for *Pesquisa Nacional por Amostra de Domicílios*), a Brazilian household survey with detailed information on households and individuals. PNAD is administered every year since 1976 by the Brazilian statistical office, the *Instituto Brasileiro de Geografia e Estatística* (IBGE), except in census years (1980, 1991, 2000, and 2010) when a similar questionnaire is applied for a sample of the population. The PNAD survey provides a long set of repeated cross-sections with a rich body of information on demographic characteristics, migration history, and labor outcomes of individuals. As such, it is particularly suitable for an investigation of the evolution of the relationship between internal migration and labor outcomes.

Starting in 1992, both the PNAD and the Brazilian census have detailed information on migration, and starting in 2004, PNAD has a full coverage of rural and urban households in all regions of Brazil.[9] In 2015, the PNAD survey was adjusted to be administered on

[8]Exceptions are Gonzaga, Menezes Filho and Terra (2006) and Paz (2014), which use the same household survey data I use, for periods of time covering 1981–2001 and 1989–2001, respectively.

[9]For logistical reasons, IBGE did not survey rural households in the states of the North region (except

a rolling basis instead of annually (its name changed to PNAD *Contínua*). Also in 2015, Brazil entered a political and economic crisis that persists to this day. For these reasons, I restricted my analysis using PNAD to the 2004–2014 period. The census year of 2010 is left out for now but will be incorporated in future versions of this study.[10]

The primary unit of observation in the PNAD is the household, and every individual in the household is surveyed. The survey is representative at the national and state levels, surveying more than 380,000 individuals in over 115,000 households each year. After restricting the data to the 2004–2014 period, I do some exclusions pertinent to an investigation on migration and labor outcomes like this one. First, I exclude all observations that did not answer the migration portion of the questionnaire or that reported "foreign country," "Brazil," or "NA" for the state of birth. Then, I exclude all individuals who were less than 16 or more than 70 years old at the time of the survey. These exclusions drop around 27% of individuals but less than 0.3% of households because international migration is negligible in Brazil, and the age exclusion is unlikely to drop household heads. In most of my analysis, I use a subsample of individuals with positive earnings, which corresponds to 60% of the full sample. This subsample has non-missing values for log wage, industry, and type of employment.

The second main data source I use this study is RAIS (Portuguese acronym for *Relação Anual de Informações Sociais*), an administrative dataset compiled by the Brazilian Ministry of Labor based on payroll reports that firms in Brazil are required to file every December. Using unique identifiers for firms and individuals, one can link data over the years, creating a rich linked employer-employee dataset. RAIS contains demographic characteristics and information on the job (including wage) and employer of virtually every individual formally employed in Brazil.

I use a four-period balanced panel of individuals observed in the RAIS data in 2000, 2005, 2010, and 2015 to construct migrant definitions based on the state in which workers are first observed because RAIS does not contain information on workers' birthplace. I then discard the first year (2000), in which there are no migrants by definition. With this restriction, I obtain a panel of individuals whose location five years ago I always observe. I then use this information to assign migrant status indicators for each individual-year observation.

Choosing a four-period balanced panel simplifies data handling and ensures minimal comparability with the periods and definitions I obtain using PNAD data. After

Tocantins) until 2004.

[10] Adding census data to the investigation will allow me to cover a longer period, going back at least to 2000 and possibly 1991. Census data will also allow a more detailed investigation of inter-municipality migration (intra-state), expanding the current analysis focused on inter-state migration.

imposing perfect balance in my panel of selected years, I still have millions of observations left. Therefore, I select a 10% subsample of individuals for convenience. On this subsample, I impose similar age and nationality restrictions to those I imposed on my PNAD sample. Finally, I identify a subsample of individuals who were young (age 16–24) in 2000, when I first observe them in my four-period balanced panel. I use this subsample as a robustness check because an individual's location at a young age is more likely to be his birthplace and, therefore, more comparable to migrant definitions based on the state of birth I construct with PNAD data.

2.2 Descriptive statistics

The outcomes of interest in this study are the hourly labor earnings (main dependent variable) and an indicator for non-zero earnings (extensive margin). The main explanatory variable is the current migrant status (the individual is in a state different than his state of birth) and the return migrant status (the individual is in his state of birth but has lived in a different state in the past). Later in the study, I use different definitions of migration based on the location five years ago instead of the state of birth. These are more comparable to the migration definitions I use in my longitudinal data analysis with RAIS data. I use other information such as gender, color/race, age, and years of schooling as controls.[11]

Migration studies define who is a migrant in many ways. Some studies observe households at the origin, in sending regions. They use the household as the unit of analysis and classify them as a "migrant" if one of its members has ever moved or currently lives elsewhere (McKenzie and Rapoport, 2007, 2010; Bryan, Chowdhury and Mobarak, 2014). I, instead, use individuals who I observe at their destinations as the main unit of analysis. Using households as the unit of analysis could lead to double counting in my context because any household member who moves to a different state becomes a new household in the PNAD data. Moreover, because I observe individuals at their destinations, my information on migration length, which I use in some additional exercises, is distinct for migrants and return migrants. PNAD asks how long an individual has lived at the place they are being surveyed. Therefore, for migrants, the answer to this question tells the length of migration, but for return migrants, it tells how long it has been since the individual has returned.

Table 1 below summarizes the information on the main samples and subsamples

[11]IBGE surveys, like the PNAD and the Census, ask respondents to choose one of five race categories: black, white, mixed, yellow, and indigenous. The yellow category, originally intended for those of East Asian descent, is seldom chosen. Asian-Brazilians often choose the mixed ("*pardo*") category. IBGE uses the term "color/race," which I refer to as simply "race" following most of the literature in the social sciences. The Ministry of Labor, which administers the RAIS dataset, uses the same definition and categories for race.

used in my analysis. I show the total number of observations, the number of individuals per year, the share of those individuals defined as current (inter-state) migrants, return migrants, and stayers (or non-migrants). In the top panel, I show information from different samples extracted from the PNAD data. The different migration indicators and their corresponding shares are defined based on the state of birth. In each year, I observe a different cohort of individuals and assign them a migrant status if their current state of residence is different than their state of birth, a return migrant status if their current and birth state is the same but they declare having lived in a different state in the past, and a non-migrant status (stayer) otherwise. The top panel shows information for the full sample in the last and first years of the PNAD data (2014 and 2004) and all years pooled together (2004–2014). Lastly, it shows the information for the "positive wage" subsamples in the same periods.

Table 1: Summary of samples

Sample	Period	Observations		Shares		
		Total	Per year	Migrants	Returnees	Stayers
Panel A: PNAD (household survey data)						
All	2014	256,260	256,260	18.4%	8.6%	73.0%
All	2004	265,710	265,710	19.7%	8.2%	72.0%
All	2004–2014	2,644,155	264,416	18.9%	8.8%	72.3%
Wage > 0	2004	157,113	157,113	20.7%	8.8%	70.6%
Wage > 0	2014	156,248	156,248	19.1%	9.2%	71.6%
Wage > 0	2004–2014	1,590,215	159,022	19.8%	9.3%	70.9%
Panel B: RAIS (linked employer-employee data)						
All	2015	491,465	491,465	6.9%	2.5%	90.7%
All	2005–2015	1,474,395	491,465	5.9%	1.2%	90.7%
Age 16–24 in 2000	2005–2015	397,815	132,605	7.3%	1.6%	88.0%

In the bottom panel, I show information from different samples extracted from my four-period balanced panel formed with RAIS data. I observe each individual in the sample four times in the 2000–2015 period and assign individuals a migrant status in any year in which their current state of residence/work is different than their state in the first year (2000). I assign a return migrant status if the current and first state of an individual is the same but this individual has been in a different state in some previous years (2005 or 2010). By definition, I can assign a return migrant status only to individuals in the last two years (2010 and 2015). Moreover, no individual can be assigned a migrant or return migrant in the first year (2000), so I remove it from the analysis. Finally, I assign each individual a stayer status if they are never observed moving to a different state across the four periods in my balanced panel. The bottom panel shows information for a single year

(2015), the full sample in the three-period (2005–2015) balanced panel, and the "young" subsample of the three-period balanced panel, which contains only individuals 16–24 years old in 2000.

The share of migrants, return migrants, and non-migrants in the PNAD data is remarkably consistent across the years and samples. Around 20% of the individuals are observed in a different state than the one they were born. Another 10% of individuals are observed in their state of birth after having lived in a different state at some point in the past, and the remaining 70% have never moved out of their states of birth.[12] In the RAIS data, on the other hand, the migrant shares are smaller: around 6–7% of the observations are of individuals currently in a state different than the one in which they were first observed in 2000, and around 1.5–2.5% of the observations are of individuals who returned to their "original" states from a different one. Finally, we observe that around 90% of the individuals in the RAIS data do not move to a different state in the four years I use in my initial panel.[13]

Table 2 below shows summary statistics and differences in means for the three main groups of individuals in my analysis. For conciseness, I show information only for the full sample of PNAD in 2014. I show statistics and differences for the outcomes of interest and demographic characteristics used as controls in the main regressions. The means are reported for all observations and the groups of migrants, return migrants, and stayers. The differences in means are reported for migrants vs. stayers, return migrants vs. stayers, and migrants vs. return migrants. The total number of observations in each group in the full and positive wage samples is shown at the bottom of the table. In the appendix, I show a complement of this table, with summary statistics for work and household characteristics included as extra controls in some specifications.

In the first rows of Table 2, we observe that migrants, both current and return migrants, earn more and are more likely to have positive wages than non-migrants. The log wage of migrants is 0.13 log points higher than the log wage of stayers. This difference is even higher for returnees vs. stayers: 0.16 log points. A similar pattern appears for wages measured by the IHS transformation and the probability of having a positive wage: the difference is positive in both comparisons of migrants and stayers but more so for the return migrants. It follows that the return migrants earn more and are more likely to have positive wages than the current migrants. This difference goes contrary to my main results and economic intuition if we believe that current migrants are always positively

[12]These figures corroborate a remark made by Kovak (2013), who points that "Brazil's population is quite mobile" with rates of inter-state migration "similar to those in the United States" (p.1972).

[13]Because the stayer definition in the longitudinal data is based on the individual's entire history instead of the individual-year observation, the percentages in the bottom panel do not add up to 100%.

Table 2: Summary statistics and differences in means using data from PNAD 2014

Variables	Means				Differences					
	All	Migrants	Returnees	Stayers	M - S		R - S		M - R	
Outcomes of interest										
Log of wage	2.25	2.34	2.36	2.21	0.13	***	0.16	***	-0.03	***
IHS of wage	1.80	1.93	2.00	1.74	0.19	***	0.25	***	-0.07	***
Wage > 0	0.61	0.63	0.65	0.60	0.04	***	0.05	***	-0.02	***
Individual characteristics										
Age (years)	38.88	43.06	41.53	37.51	5.55	***	4.02	***	1.53	***
Schooling (years)	8.44	7.98	8.80	8.52	-0.53	***	0.28	***	-0.81	***
Female	0.52	0.52	0.49	0.52	-0.01	***	-0.04	***	0.03	***
Race: mixed	0.47	0.49	0.46	0.47	0.02	***	-0.01	*	0.02	***
Race: white	0.42	0.41	0.44	0.42	-0.01	***	0.02	***	-0.02	***
Race: black	0.10	0.09	0.09	0.10	-0.01	***	-0.01	***	0.00	
Observations	256,260	47,106	22,135	187,019						
Obs. w/ wage > 0	156,248	29,901	14,402	111,945						

Notes: Stars denote: * p<0.10; ** p<0.05; *** p<0.01.

self-selected. In contrast, return migrants can be positively or negatively self-selected.

The answer to this apparent puzzle is in the second set of rows in which I show summary statistics for age, education, gender, and race. While current migrants are less educated than stayers, the opposite is true for return migrants: they are more educated than stayers (0.28 more years of schooling) and, consequently, more educated than current migrants (0.81 more years of schooling).[14] In most wage regressions, I include years of education as controls and find that conditional on schooling, the average wage of migrants is higher than the average wage of return migrants. Finally, I note large differences in age (migrants and return migrants are 4–5 years older than stayers) and statistically significant but small differences in gender and race.

2.3 Past migration rates

My IV strategy use past rates of migration and return migration in Brazilian states in 1970 as instruments for the present migrant status of individuals born in these states.[15] To obtain the past rates of migration and returning migration, I use data from the 1970 census. I calculate the emigration rate of each state in 1970 by dividing the total number of internal migrants in the country born in a given state by the population born in that

[14]The fact that current migrants have higher wages than stayers despite having fewer years of education is indicative of the positive self-selection of internal migrants in Brazil.

[15]I interact past migration rates with individuals' age, gender, and education to increase the variation of the instrument at the individual level. See section 3.2 for more details.

state. The 1970 emigration rate in the state of São Paulo (SP), for example, is

$$\text{Emigration rate in SP} = \frac{\text{\# individuals whose birth state = SP and residence} \neq \text{SP}}{\text{\# individuals whose birth state = SP}}. \tag{1}$$

I calculate a similar measure for return migration by dividing the total number of return migrants living in a given state in 1970 by the population born in that state (by definition, return migrants were born in the state they live in). I then interact the historical rates of emigration and return migration with individual-level characteristics—age, age squared, female dummy, and non-white dummy—to gain variation at the individual level. I use both measures as excluded instruments in the first stage of individual-level IV regressions that include the migrant and return migrant indicators as endogenous variables.

3 Empirical strategy

3.1 Main regressions

A key innovation of my study is to differentiate between current and return internal migrants. I include an indicator for each migrant type in my main regression equation, obtaining coefficients that show differentials—in wages and other outcomes—for current and return internal migrants relative to non-migrants In Brazil. I focus on inter-state migration, so I include state fixed effects for both the origin (birth) and destination (residence) of individuals. These fixed effects control for unobserved factors that could influence the outcomes, such as the quality of education in an individual's origin or agglomeration effects at an individual's destination. I am parsimonious with my choice of individual-level controls, choosing characteristics that are not likely affected by the migration status, like age, gender, and race. I also include years of education given its importance for determining labor outcomes and because most individuals in my sample are adults beyond their school years.[16]

When using cross-sectional data from PNAD in a single year, my main regression equation is

$$y_i = \beta_0 + \beta_M M_i + \beta_R R_i + X_i' \lambda + F_{b(i)} + F_{r(i)} + \varepsilon_i, \tag{2}$$

where y_i is the outcome of interest for individual i: his log wage, IHS transformation

[16]Inter-state migration for educational purposes is not common in Brazil. In 2001, the only year in which PNAD asks the reason for moving, less than 4% of the migrants said education. Family (53%) and work (23%) were the most cited reasons.

of wage, or a non-zero wage indicator. The indicators M_i and R_i correspond to the migrant and return migrant statuses, respectively. The term X_i is a vector of individual-level characteristics (age, age squared, an indicator for female, years of schooling, and race/color categorical dummies). In some specifications, I expand this vector to include work and household characteristics as well. The terms $F_{b(i)}$ and $F_{b(i)}$ denote fixed effects for the individual's state of birth and residence.

The coefficient β_M represents the average difference in the log hourly wage of current migrants compared to non-migrants. Similarly, β_R represents the difference for a return migrant compared to a non-migrant. When both are positive, $\beta_R - \beta_M$ shows the average depreciation of the wage premium for return migrants, i.e., how much is lost by a migrant after returning. A similar interpretation holds for different outcomes, such as the positive wage indicator.

When using additional years of PNAD data, I add time subscripts to all variables and time fixed effects (F_t) to the regression equation in (2) obtaining:

$$y_{it} = \alpha_0 + \alpha_M M_{it} + \alpha_R R_{it} + X_{it}'\delta + F_{b(it)} + F_{r(it)} + F_t + \epsilon_{it}. \tag{3}$$

In some specifications, I interact the migrant indicators (and controls) with a year dummy to recover the change in the wage premia between two particular years. Alternatively, I add time trends interacted with the migrant indicators to see how the wage premia evolved over the 2004–2014 period.

In one exercise, I use a version of equation (2) in which I interact the migrant indicators with categorical dummies for the duration of residence, which represents years since moving or returning in my context. This regression allows me to investigate heterogeneity in terms of duration of residence, giving me measures of migration that are comparable to the definition I can use in the longitudinal analysis (migration defined with respect to the residence in the last 5–10 years). It also sheds light on the role of learning and adaptation in the determination of the wage premium of migrants and return migrants in Brazil.

When using longitudinal data from RAIS, I can run pooled OLS or panel regressions with individual fixed effects. In the first case, the regression equation is similar to the one in equation (3); the main difference is how the migrant indicators are defined and the fact that I use state of origin in 2000 instead of state of birth. In the second case, the main regression equation includes an individual fixed effect (F_i):

$$y_{it} = \gamma_0 + \gamma_M M_{it} + \gamma_R R_{it} + X_{it}'\zeta + F_{o(it)} + F_{r(it)} + F_t + F_i + \nu_{it}. \tag{4}$$

The term $F_{o(it)}$, which denotes the fixed effect for the state of origin in 2000, replaces the fixed effect for the state of birth $F_{b(it)}$.

In both cross-sectional and longitudinal analyses, I perform regressions in different subsamples to gain additional insights or ensure comparability across regression exercises. For example, in one regression, I restrict the sample to males only, who are more likely to be the head of the household in my data and, therefore, more likely to be the individual making the migration decision (as opposed to being a trailing spouse or family member). In another regression, I restrict the PNAD sample to formal sector employees only, which are the only type of worker I observe in the RAIS data. I also run regressions for subsamples of young individuals (ages 16 to 24), for whom the state in which they are first observed and the state of birth are more likely to coincide. Finally, in the longitudinal analysis only, I restrict the sample to individuals who move to a different state at some point in the four years I select from the 2000–2015 period. I use this sample of movers—entirely formed by self-selected individuals—to investigate how the wage of a self-selected mover changes when he becomes, in fact, a migrant or return migrant (as opposed to a future migrant who has not moved yet).

3.2 Instrumental variables

The migrant indicators in my regression equations are endogenous. I would be mistaken if I interpreted the coefficients from these regressions as causal effects rather than wage differentials formed by different components, including self-selection based on unobserved characteristics. My regressions with individual fixed effects and the regressions in which I restrict the sample to movers can address part of the self-selection bias, the one based on absolute advantage. i.e., the selection of individuals who perform better in any location. The selection based on the comparative advantage of individuals who perform better in a specific location remains.

To address the bias arising from the selection on comparative advantage, I turn to an IV strategy that uses past migration rates in an individual's state of birth to predict his migration or return migration status in the present. Conditional on satisfying the exclusion restrictions and having a robust first stage, the IV results can be interpreted as the effect of migration (or return migration) on the wages of those individuals who were induced to migrate (or return) by the existence of migrant networks proxied by past migration rates.

I use the past emigration and return migration rates presented in Section 2.3 as excluded instruments in two different specifications. First, I run two-stage regressions using equation (2), including both migration indicators in the second stage. This specification uses two sets of variables as excluded instruments in the first stage: past emigration rates and their interaction with individual-level characteristics and past rates of returning migration and their interactions. Second, I run two-stage regressions using a single endogenous regressor: the migrant indicator. This specification has the disadvantage of ignoring returning migration but benefits from the simplicity of having just one endogenous variable.

The exclusion restriction requires that the likelihood of migrating, but not the outcome of interest, for an individual living in state r in year t and born in state b in $t-16$ or earlier will be influenced by the 1970 emigration rate of state b and its interactions with the selected characteristics of this individual (age, gender, and race). Similarly, the exclusion restriction for return migration requires that conditional on having ever migrated, the likelihood of returning for an individual living in state b in year t (and born in that same state in $t-16$ or earlier) will be influenced by the 1970 rate of return migration in state b and its interactions, but the outcome of interest of this individual will not be affected by these rates and interactions except via his return migration status. Because all instruments are assigned according to the individuals' state of birth, I cannot include state of birth fixed effects in IV regressions.[17]

4 Results

4.1 Main results with cross-sectional data

Table 3 below shows estimates for the wage premium of migrant and return migrants in 2014 identified out of cross-sectional variation in PNAD data. The table illustrates how the estimates change with the inclusion of controls and fixed effects. I start with no controls in column (1) and progressively add controls in columns (2), (3), and (4). Because work and household controls may be correlated with the migration indicators, I do not carry them to the latter specifications in which I add fixed effects for state of residence (column 5) and birth (column 6). Other than reporting the coefficients for the two migrant indicators, I also report the difference between them.

The estimates in column (1) indicate that the wage of current migrants was 0.13 log points (approximately 14%) higher than the wage of non-migrants on average in Brazil in 2014. This difference was higher for return migrants (0.16 log points), but the two coefficients are not statically different from each other (difference and standard errors shown in the third row). As anticipated in the discussion of Table 2, the positive difference between the wage premium of return and current migrants is probably due to return

[17] I could add state of birth fixed effects if I used only the interactions but not the historical rate as instruments. Doing so, however, would weaken the first stage.

Table 3: Estimates of the migrant wage premium using data from PNAD 2014: Varying controls

Dependent variable:	Log hourly wage					
	(1)	(2)	(3)	(4)	(5)	(6)
Migrant (M)	0.1294	0.1150	0.1096	0.1031	0.0845	0.1152
	(0.0488)**	(0.0215)***	(0.0185)***	(0.0189)***	(0.0173)***	(0.0111)***
Return migrant (R)	0.1576	0.0762	0.0756	0.0683	0.0860	0.0884
	(0.0216)***	(0.0138)***	(0.0130)***	(0.0126)***	(0.0146)***	(0.0150)***
Difference (R - M)	0.0281	-0.0388	-0.0340	-0.0348	0.0015	-0.0268
	(0.0590)	(0.0297)	(0.0269)	(0.0269)	(0.0256)	(0.0146)*
Adjusted R^2	0.01	0.30	0.36	0.36	0.31	0.31
Observations	156,248	155,793	155,793	155,381	155,793	155,793
Individual controls		Y	Y	Y	Y	Y
Work controls			Y	Y		
Household controls				Y		
State of residence FE					Y	Y
State of birth FE						Y

Notes: The dependent variable is the log of the hourly wage. Individual controls: age, age squared, female indicator, race/color categories, and years of schooling. Work controls: categorical indicators for industry and employment status. Household controls: head of the household indicator, number of members, the share of members aged 16 or more, and rural indicator. Standard errors clustered by state of residence in parentheses. Stars denote: * p<0.10; ** p<0.05; *** p<0.01.

migrants having 0.81 more years of schooling on average than migrants. In effect, the difference turns negative in column (2), in which I add education and individual controls to the regression specification: the coefficients are positive for both types of migrants in column (2), meaning that they earn more than non-migrants, but the highest wage premium accrues now to current migrants (the difference is still not statically significant, however).

Across specifications, the coefficients for the wage premium are always economically and statically significant, and the difference in the third row is negative most of the time. In my preferred specification shown in column (6), in which I use only individual characteristics as controls and include fixed effects for both the state of residence and birth, the difference between the premium of migrants and return migrants becomes statistically significant. Migrants have a wage premium of approximately 12%, whereas the wage premium of return migrants is lower (9%). This suggests that migrants lose about a quarter of their wage premium upon returning.

Our main outcome of interest, log wages, is defined only for individuals with positive wages. The likelihood of having some paid work, however, can be influenced by one's migrant status, just like wages. For example, an individual may move to a different state if he expects not only a higher wage conditional on having work but also a higher chance of finding work, to begin with. In other words, migration can influence both the extensive and the intensive margins of labor earnings.

About 40% of the individuals ages 16–70 in the PNAD data do not report any labor earnings. Also, as seen in Table 2, there are differences in the average proportion of individuals with positive income across the migrant, return migrant, and non-migrant types. Therefore, the extensive margin seems an important margin in this case.

Table 4 below shows estimates for the migrant and return migrant premium on three different outcomes in 2014, again exploring cross-sectional variation in PNAD data. In the first column, I repeat the estimation shown in column (6) of Table 3, in which the outcome is log wage, and only observations with positive wage enter the estimation. In column (2), I run a linear probability model in which the outcome is an indicator for having a non-zero wage. And in column (3), I repeat my main regression specification using the IHS transformation of wage instead of the log.[18] This ensures that I do not drop the observations with zero earnings from the sample.

Table 4: Estimates of the migrant wage premium using data from PNAD 2014: Intensive and extensive margins

Dependent variable:	Log wage (1)	$1\{wage > 0\}$ (2)	IHS of wage (3)
Migrant (M)	0.1152	0.0175	0.1146
	(0.0111)***	(0.0037)***	(0.0134)***
Return migrant (R)	0.0884	0.0113	0.0812
	(0.0150)***	(0.0054)**	(0.0169)***
Difference (R - M)	-0.0268	-0.0062	-0.0334
	(0.0146)*	(0.0059)	(0.0186)*
Adjusted R^2	0.31	0.20	0.27
Observations	155,793	255,482	255,482
Individual controls	Y	Y	Y
State of residence FE	Y	Y	Y
State of birth FE	Y	Y	Y

Notes: The dependent variable is the log of the hourly wage in column (1), an indicator for a positive wage in column (2), and the IHS transformation of the hourly wage in column (3). Individual controls: age, age squared, female indicator, race/color categories, and years of schooling. Standard errors clustered by state of residence in parentheses. Stars denote: * p<0.10; ** p<0.05; *** p<0.01.

[18] The interpretation of the coefficients in my regressions using either the log or the IHS transformation, as an approximation, is the same. In most cases, a percentage change in the IHS-transformed dependent variable due to a discrete change in a dummy variable can be approximated in the same way as a log-transformed dependent variable: $\exp(\hat{\beta}) - 1$ (Bellemare and Wichman, 2020).

I find that the probability of having a non-zero wage is 1.75 percentage points higher for migrants and 1.13 percentage points higher for return migrants on average, with no statistical difference between these estimates. This confirms the intuition that migrants might self-select not only based on their prospect of earning higher wages but also on the prospect of being more likely to find paid work in the new residence. The results in columns (1) and (3) are very similar, which suggests that estimates of the wage premium are not determined by changes in the sample composition when the log transformation drops all observations without a positive wage (39% of the total).

4.2 Additional exercises with cross-sectional data

After having estimated how much more migrants earn compared to non-migrants and how much of this difference they lose upon returning, I investigate how these wage differentials may have changed over time. Table 5 shows estimates for my main specifications in different years and periods. In columns (1) and (2), I show results for 2004 and 2014, the first and the last year in my data, separately. In column (3), I use data from both years and interact the migrant indicators (and controls) with an indicator for 2014. In column (4), I use data from all PNAD years in 2004–2014 and the regression specification with year fixed effects shown in equation (3). In column (5), I add year trends interacted with the migrant indicators to this specification.

The coefficients on the interactions in column (3) show how much the migrant and return migrant premia has changed over the period: a decrease of 0.01 log points in the wage premium of migrants and an increase of 0.03 log points in the wage premium of return migrants. In column (5), we see this evolution through a different lens, observing a negative time trend for the wage premium of migrants and a positive one (albeit not significant) for the return migrants. These results suggest that return migrants kept a higher portion of the wage premium in 2014 upon returning than they did in 2004.

The fact that Brazil has experienced a great reduction in inequality between its regions precisely during the period I analyze suggests these two phenomena can be linked. Migration flows in the country traditionally run from the poor North and Northeast regions to the richer South and Southeast. These traditionally sending regions had relatively higher growth rates over the past two decades compared to the traditionally receiving regions. Other than attracting back their workers, these regions might be offering better opportunities for them to use the skills they learned at their previous destinations or to apply the advantageous labor skills that made them migrate in the first place.

Next, I investigate how the duration of the migration experience relates to the wage premium. The PNAD data tells how long individuals have lived at their current residence

Table 5: Estimates of the migrant wage premium using data from PNAD: Evolution in the 2004–2014 period

Dependent variable:	Log hourly wage				
Sample period:	2004	2014	2004 & 2014	2004–2014	2004–2014
	(1)	(2)	(3)	(4)	(5)
Migrant (M)	0.1286	0.1152	0.1286	0.1198	0.1394
	(0.0101)***	(0.0111)***	(0.0101)***	(0.0105)***	(0.0142)***
Return migrant (R)	0.0549	0.0884	0.0549	0.0735	0.0649
	(0.0126)***	(0.0150)***	(0.0126)***	(0.0119)***	(0.0121)***
Migrant x 2014 year			-0.0133		
			(0.0069)*		
Ret. migrant x 2014 year			0.0335		
			(0.0134)**		
Migrant x Trend					-0.0041
					(0.0013)***
Ret. migrant x Trend					0.0018
					(0.0012)
Adjusted R^2	0.41	0.31	0.42	0.38	0.38
Observations	156,034	155,793	311,827	1,582,925	1,582,925
Individual controls	Y	Y	Y	Y	Y
Year FE			Y	Y	Y
State of residence FE	Y	Y	Y	Y	Y
State of birth FE	Y	Y	Y	Y	Y

Notes: The dependent variable is the log of the hourly wage. Individual controls: age, age squared, female indicator, race/color categories, and years of schooling. There is only one year FE in column (3): the 2014 year indicator. The time trend in column (5) enters the specification only as interactions with the migrant and the return migrant indicators. Standard errors clustered by state of residence in parentheses. Stars denote: * p<0.10; ** p<0.05; *** p<0.01.

(up to 4 years, 5 to 9 years, and 10 or more years).[19] For migrants, duration means years since moving and may serve as a proxy for adaptation and learning at the destination. For return migrants, duration means years since returning and may refer, especially in the first years, to a period of recovery from a negative shock that prompted the return or a period of re-adaptation to the home state.

Table 6 reports the estimates for this exercise. As with Table 4, I report estimates for three outcomes of interest—log wages, a non-zero wage indicator, and the IHS transformation of wages—to verify how the duration of the migration (or return migration) experience may affect labor earnings via the extensive or intensive margins. All speci-

[19]PNAD also has information on the actual number of years reported by individuals instead of these categories. I do not to use the number of years because most of the sample (over 75%) reports living ten years or more at the destinations. Thus, more than 75% of this information would be right censored.

fications include the interaction between the migrant indicators and a category dummy for each duration, including a category for missing information on duration (not shown). Interactions are mutually exclusive: the comparison group (or the omitted category) is always the group of non-migrants, for which the length information is nonexistent by definition.

Table 6: Estimates of the migrant wage premium using data from PNAD 2014: Interactions with the duration of current residence, intensive and extensive margins

Dependent variable:	Log wage (1)	$1\{wage > 0\}$ (2)	IHS of wage (3)
Duration x Migrant (M): years after migrating			
up to 4 years x M	0.1765	0.0047	0.1106
	(0.0187)***	(0.0073)	(0.0253)***
5 to 9 years x M	0.1853	0.0302	0.1989
	(0.0131)***	(0.0076)***	(0.0225)***
10 or more years x M	0.0926	0.0209	0.1113
	(0.0115)***	(0.0043)***	(0.0143)***
Duration x Return migrant (R): years after returning			
up to 4 years x R	0.0890	-0.0134	-0.0010
	(0.0197)***	(0.0114)	(0.0292)
5 to 9 years x R	0.1208	-0.0047	0.0675
	(0.0254)***	(0.0099)	(0.0321)**
10 or more years x R	0.0770	0.0254	0.1169
	(0.0161)***	(0.0049)***	(0.0180)***
Adjusted R^2	0.31	0.20	0.27
Observations	155,793	255,482	255,482
Individual controls	Y	Y	Y
State of residence FE	Y	Y	Y
State of birth FE	Y	Y	Y

Notes: The dependent variable is the log of the hourly wage in column (1), an indicator for a positive wage in column (2), and the IHS transformation of the hourly wage in column (3). Individual controls: age, age squared, female indicator, race/color categories, and years of schooling. The regression also includes migrant indicators (M and R) interacted with an indicator for missing information on duration, which is the case for less than 1% of the observations (coefficients not shown). Standard errors clustered by state of residence in parentheses. Stars denote: * p<0.10; ** p<0.05; *** p<0.01.

The first three rows show coefficients for the interactions between duration and the current migrant indicator. The results suggest an inverted U-shaped relation between duration and the wage premium. The inverted U-shape is particularly noticeable in column (3), in which the IHS transformation in the outcome "mixes" both the intensive and the extensive margins. The estimates in columns (1) and (3) show that the wage premium of current migrants is positively correlated with the migrant status in the first years but increases in the medium term (between 5 and 9 years into the migration spell). It later decreases, possibly because the 10 years or more category encompasses a long period, and some individuals in this group live for so long in their new states that they can hardly be considered migrants anymore.

The estimates in column (2) tell a slightly different but complementary story: the likelihood of having paid work is no different for migrants and non-migrants in the first years but becomes significantly higher for migrants in the medium term (0.03 percentage point more for those living in the new state for between five to nine years). This likelihood remains higher for migrants in the years after (0.02 percentage points more). These results suggest that, despite being positively self-selected, migrants struggle in the first years to find work and fully realize their potential at the new destination. Later on, after many more years, the skills that gave them an advantage in the first place might wear off, and their wage premium is reduced.

For return migrants, the estimates in column (1) also suggest an inverted U-shaped relationship between duration and the wage premium. However, the estimates in columns (2) and (3) suggest that return migrants have a hard time finding employment upon returning. As such, they do not surpass the performance of non-migrants in the labor market of their home states until after five or more years of their return.

The estimates shown in Table 6 are useful for comparisons with estimates obtained with longitudinal data. The migration indicators we can construct using RAIS data refer to the location of workers one or two periods in the past, i.e., five or ten years ago since my balanced panel is spaced by five-year intervals. Therefore, when looking at estimates from regressions using RAIS data, we can keep in mind that they refer to durations of five or ten years, comparable to the duration corresponding to the largest estimates in Table 6.

In Table A2 in the appendix, I show the results for cross-sectional regressions using different subsamples of PNAD in 2014 and one different migration definition. As discussed in the introduction, the migrant wage premium can be explained by several factors. The inclusion of residence fixed effects arguably accounts for location-specific factors and agglomeration effects. The role of self-selection, on the other hand, is harder to measure and control. To shed some light on its importance, I repeat my main regression for a sample of only men. Because men in my sample are almost twice as likely to have positive wage and be the head of households than women, I suspect that they are also more likely to drive migration decisions and to positively self-select. It follows that they should have a higher wage premium than women.[20] This is precisely what I find. The

[20]In my sample, approximately 71% of men have positive wages, compared to 45% of women, and approximately 57% of men are household heads, compared to 27% of women (data from PNAD 2004–2014). In the US, there is also evidence that women are more likely to be a “trailing spouse” or “tied-mover” and

estimate for the wage premium of migrants in column (1) of Table A2, for a sample of only men, is 0.14 log points (compared to 0.12 in the original sample with both men and women). For return migrants, the premium is close to 0.11 log points (compared to 0.09 in the original sample).

I select other subsamples with the goal of increasing comparability between estimates of regressions using PNAD and RAIS data. I restrict the sample to formal sector employees, the only type of workers I observe in RAIS. I find estimates that are similar in magnitude but "flipped." Migrants have a wage premium of 0.09 log points, and return migrants have a wage premium of 0.11 log points (compared to 0.12 and 0.09 before; results shown in column (2) of Table A2). I then restrict the sample to young individuals (ages 16–24) and young employees in the formal sector because I use a subsample of young workers in some regressions with longitudinal data. The estimates are lower than before (around 0.05 log points for both migrant types) but still positive and significant.

In the last two columns of Table A2, I vary migrant definitions. First, I run a regression in which I omit the return migrant indicator, so I can obtain an estimate that is comparable to the IV regressions with a single-regressor (the estimate for the migrant wage premium is 0.11 log points, slightly smaller than the estimate in my preferred specification in Table 3: 0.12 log points). Finally, I repeat this estimation with a single regressor but for a migrant indicator defined with respect to the state of residence five years ago (a similar definition to the one I use in the longitudinal analysis). I find a coefficient of 0.11, very close in magnitude to the ones I obtained in my preferred specification.

4.3 Results with instrumental variables

Table 7 shows results for two different specifications used in my IV regressions. As with Tables 4 and 6, I report estimates for three outcomes of interest: log wages, a non-zero wage indicator, and the IHS transformation of wages. In columns (1) and (2), the specification has both migrant indicators and compares migrants and return migrants to non-migrants. The first stage of those specifications includes the two sets of instruments discussed in the previous sections, one for each endogenous regressor (migrant and return migrant status). In columns (3) and (4), the specification drops the indicator for return migrants (and its corresponding instruments in the first stage). These specifications compare current migrants to everyone else (non-migrants and return migrants). In all cases, I report the test statistics for underidentification and weak identification. The results for the corresponding first-stage regressions are shown in Table A3 in the appendix.

suffer wage losses following a move (Venator, 2020).

Table 7: IV estimates of the migrant wage premium using data from PNAD 2014 (2nd stage): Intensive and extensive margins

Dependent variable:	Log wage (1)	$1\{wage > 0\}$ (2)	IHS of wage (3)	Log wage (4)	$1\{wage > 0\}$ (5)	IHS of wage (6)
Migrant (M)	-0.2633	0.1055	0.3854	0.0002	0.0796	0.3787
	(0.6371)	(0.0951)	(0.3547)	(0.0645)	(0.0634)	(0.3312)
Return migrant (R)	2.0326	-0.0473	-0.4914			
	(1.9764)	(0.7375)	(1.0793)			
Difference (R - M)	2.2959	-0.1528	-0.8768			
	2.3908	0.7979	1.2555			
Underid. (K-P LM Stat)	13.58	11.58	11.58	7.52	10.65	10.65
Weak id. (K-P Wald F Stat)	7.13	4.13	4.13	1.46	6.07	6.07
S-W multivariate F test of excluded instruments (weak identification)						
Migrant (M)	7.22	6.29	6.29	1.46	6.07	6.07
Return migrant (R)	7.98	4.78	4.78			
Observations	155,793	255,484	255,484	155,793	255,484	255,484
Individual controls	Y	Y	Y	Y	Y	Y
State of residence FE	Y	Y	Y	Y	Y	Y

Notes: The dependent variable is the log of the hourly wage in columns (1) and (4), an indicator for a positive wage in columns (2) and (5), and the IHS transformation of the hourly wage in columns (3) and (6). Individual controls: age, age squared, female indicator, non-white indicator, and years of schooling. The excluded variables used in the first stage as instruments for the first endogenous regressor (the migrant indicator M) are the 1970 rate of emigration in the individual's state of birth and the interaction of this rate with some of the individual's characteristics (age, age squared, female indicator, and non-white indicator). The excluded variables for the second endogenous regressor (the return migrant indicator R) are the 1970 rate of return migration in the individual's state of birth and its interactions with the same set of individual characteristics. Standard errors clustered by state of residence in parentheses. Stars denote: * p<0.10; ** p<0.05; *** p<0.01.

The results from my IV regressions are not informative for an investigation of migrants' and return migrants' wage premia. None of the coefficients on the migrant indicators are statistically different from zero. I could interpret these "null results" as evidence that there is no wage premium after self-selection is accounted for by the instruments, no statistically significant wage premia for the individuals induced by the instruments to migrate or return. However, such interpretation requires a robust first stage, precise estimates, and a plausible exclusion restriction. While I can argue in favor of the latter requirement, the first two are clearly not satisfied. The standard errors of all estimates are quite large, and the test statistics for weak identification in the first stage are low (in fact, the coefficients on the instruments in the first stage are mostly not significant). Therefore, I refrain from drawing conclusions about the wage premia for migrants induced to migrate by the availability of migrant networks proxied by past migration rates based on these estimates. Instead, I look at the first stage results (or lack thereof) to gain insights on the viability of using past migration rates as an instrument for present migration in Brazil.

Past migration rates have been used successfully in the context of Mexican-US migration as a proxy for migrant networks, which in turn can predict current migration rates and even individual migrant status. In my investigation of internal inter-state migration in Brazil, however, such an instrument fails to predict individuals' current or return migration statuses.[21] The possible reasons are many, and I discuss a few of them below.

First, by focusing on inter-state migration and using past emigration rates by state, I may be using a level of aggregation that has no meaning to the individuals. In other words, the past migration rates at the municipality or district level may be a better proxy for migrant networks that influence one's migration decisions. Second, past emigration rates may be informative of current migration rates aggregated at the municipality or state levels but not individual migration statuses. The interactions I added to the instrument—of past rates with individual-level characteristics like age, gender, and race—are limited by the number of exogenous characteristics I have in the data. I could gain more variation using education indicators or household characteristics in these interactions, but such variables are affected by the migration and return migration statuses. Third, migrant networks at the origin, even when successfully proxied by an exogenous instrument, may not be a relevant determinant of internal migration in Brazil. That may be due to the multitude of possible destinations, which dilute the importance of the origin-based networks. Using origin-destination pairs in a different design or focusing on a single destination (e.g., the state of São Paulo) may yield better results. Finally, migrant networks might be only marginally relevant as a determinant of internal migration in Brazil. Other determinants such as moving costs, wealth, or risk perceptions may be more relevant in the Brazilian context.

4.4 Results with longitudinal data

In this section, I discuss the results for the analysis using longitudinal data from RAIS. By design, I work with a perfectly balanced panel with three years: 2005, 2010, and 2015. I use the year 2000 to define individuals' origin and then discard this year from my analysis because no individual would be assigned a migrant status in it. I also identify two subsamples in the RAIS data: one formed only by individuals who were 16–24 years old in 2000 and for whom the state in which they work will usually coincide with the state in which they were born, and another subsample formed only by individuals who move to a different state at least once in the 2005–2015 period.[22]

[21] Varying the set of individual characteristics used in interactions or adding state of birth fixed effects do not yield stronger first stage results, nor does using past migration rates from 1980 instead of 1970.

[22] Older workers are twice more likely to be migrants than young workers in my sample of cross-sectional data: 21% of those ages 25 and above live in a state different than their state of birth, compared to 11% of those ages 16 to 24 (data from PNAD 2004–2014).

Table 8 shows the results for regressions using longitudinal data from RAIS. In columns (1) and (2), I keep all workers in the sample. In columns (3) and (4), I restrict the sample to those who were young in 2000. In columns (5) and (6), I restrict the sample to movers, excluding all those who never move to a state different from the one in which I first observed them in 2000. In each subsample, I alternate between specifications. First, I run a pooled OLS regression with individual-level controls, state, and year fixed effects. This is similar to the specification I used in Table 5, column (4) (pooled OLS with repeated cross-section from PNAD 2004–2014). Second, I run a panel regression with individual fixed effects.[23]

Table 8: Estimates of the migrant wage premium using data from RAIS 2005–2015

Dependent variable:	Log hourly wage					
Sample:	All		Age 16–24 in 2000		Movers	
	(1)	(2)	(3)	(4)	(5)	(6)
Migrant (M′)	0.1355	0.0504	0.1444	0.0917	0.0302	-0.0107
	(0.0140)***	(0.0063)***	(0.0126)***	(0.0077)***	(0.0180)	(0.0095)
return migrant (R′)	0.0450	0.0688	0.0639	0.0938	-0.0682	-0.0387
	(0.0164)**	(0.0143)***	(0.0176)***	(0.0138)***	(0.0119)***	(0.0088)***
Difference (R′ - M′)	-0.0904	0.0183	-0.0805	0.0021	-0.0984	-0.0281
	(0.0199)***	(0.0112)	(0.0120)***	(0.0142)	(0.0219)***	(0.0050)***
Adjusted R^2	0.45	0.85	0.43	0.78	0.50	0.83
Observations	1,474,395	1,474,395	397,815	397,815	137,817	137,817
Individual controls	Y	Y*	Y	Y*	Y	Y*
Individual FE		Y		Y		Y
Year FE	Y	Y	Y	Y	Y	Y
Current state FE	Y	Y	Y	Y	Y	Y
State in 2000 FE	Y		Y		Y	

Notes: The dependent variable is the log of the hourly wage. Individual controls: age, age squared, female indicator, race/color categories, and education categories. In specifications with individual fixed effects, only the time-varying education categories are used as controls. The current state refers to the state in which the individual is observed in each year in the 2005–2015 period. The migrant (M′) and return migrant (R′) indicators are defined with respect to the state in which the individual was observed in 2000. The standard errors in parenthesis are clustered by the current state in columns (1), (3), and (5). In columns (2), (4), and (6)—in which individual fixed effects are included—the specifications use a two-way clustering by current state and individual. Stars denote: * p<0.10; ** p<0.05; *** p<0.01.

The estimate for the wage premium of migrants in column (1) (0.135 log points) is close in magnitude to the estimates from the cross-sectional analysis. In fact, it is slightly larger than the coefficient in my preferred specification (0.12 log points). This is consonant with my results in Table 6, which show a larger wage premium for migrants who moved

[23] In these regressions, the state of origin fixed effects and most individual-level controls are subsumed into the individual fixed effects.

between five and ten years ago to their new destination, precisely the time span we have in this longitudinal analysis. The wage premium of the return migrants is smaller (0.045 log points), half the magnitude of the coefficient in my preferred specification using PNAD data (0.09 log points). Again, this is consonant with the results in Table 6, which shows that the wage premium of return migrants takes a few years to show up, especially when considering both the intensive and extensive margins (see column 3 in Table 6). The estimates from the pooled OLS regressions using the young sample are very similar.

In columns (2) and (4), in which I include individual fixed effects, the coefficients change. The wage premium of current migrants drops substantially, from 0.135 to 0.05 log points in the full sample and from 0.145 to 0.09 in the young sample. The coefficient of return migrants increases by 0.02 log points in the full sample and 0.03 in the young sample. In both cases, the wage premium of migrants and return migrants become not statistically different from each other when individual fixed effects are added to the specification (see the third row in columns 2 and 4).

The change in the magnitude of the wage premium that we observe with the inclusion of individual fixed effects may indicate the type of self-selection involved in the individuals' migration decision in my sample. Because the individual fixed effects account for time-invariant unobserved ability, I can rule out—or at least reduce considerably—the role that selection based on absolute advantage (individuals come from the right side of the general ability distribution) plays in determining the wage premia. What is left of the premium may reflect selection based on comparative advantage in the case of migrants (individuals move to the right of the ability distribution upon migrating) or accumulated learning in the case of return migrants.

To further shed light on this discussion of self-selection as a determinant of wage premia, I look at the results for the sample of movers. All individuals in this sample have self-selected into migration at some point and are observed, in different years, either at a different state (as migrants) or at their original state. In the second case, they are observed as non-migrants in the years before their first move or return migrants if they have already moved and returned. If the self-selection of internal migrants in my sample is based on absolute advantage, then the migrant indicator should have no effect on wages for movers since every individual in this sample is self-selected. On the other hand, if self-selection is based on comparative advantage, the coefficient on return migration (precisely, the difference $\beta_{R'} - \beta_{M'}$) should be negative because individuals would lose the benefit of this comparative advantage upon returning. This is what we observe in columns (5) and (6). Movers have a wage penalty upon returning and no significant wage premium while migrating.

Nonetheless, the coefficients for the wage premium of both migrant types across various specifications and samples discussed in this section are positive and large. Only when I restrict the analysis to a sample of self-selected individuals in the longitudinal analysis do I observe results consistent with selection based on comparative advantage. I conclude, therefore, that both types of self-selection are at play in Brazil, but the first, selection based on absolute advantage, is stronger. This type of self-selection, in which the individuals who perform better at home also perform better at the destination, can explain why we consistently observe wage premia for return migrants and even larger premia for current migrants.

5 Conclusion

It is well established in the literature that migrants have higher labor earnings than non-migrants. This fact holds for both international and internal migration, and it is commonly explained by the fact that migration, just like the subsequent participation in the labor force, is an endogenous decision. Individuals decide whether or not to migrate, and therefore only those whose expected gains from this decision are positive will choose to migrate. The positive self-selection of migrants and other factors that can explain their observed wage premia—such as learning, location-specific productivity, and agglomeration effects—have been extensively studied by social scientists.

Less studied is the self-selection of return migrants and how it determines their wage premium (if such a premium exists). return migrants can be positively self-selected when they migrate. Their earnings can also be positively affected by factors directly associated with the destination, such as agglomeration effects. However, their decision to return, other than canceling the premium components linked to the destination, also suggests a second self-selection process. This second self-selection could be negative: conditional on being a migrant, those less likely to succeed at the destination would be the first to return. The two self-selection processes likely go in opposite directions, making predicting the final wage premium of a return migrant a hard task. If, after discounting location-specific factors, the positive effects of the first self-selection are larger than the negative effects of the second self-selection, the final wage premium is positive; otherwise, it is negative. Moreover, other factors such as how much learning at the destination is brought back and successfully applied after returning can also affect the final wage premium of return migrants.

This study compares labor outcomes of current and return internal migrants in Brazil to those of non-migrants over a large period to understand how they differ from each other and how they evolved in recent years. Using cross-sectional data from repeated household surveys in 2004–2014, I find that the wages of internal migrants are about 12% higher than the wages of non-migrants. For return migrants, wages are 9% higher on average. I also find that the wage premium for migrants decreased during this period, while the wage premium for return migrants increased. Using longitudinal data from linked employer-employee datasets in 2005–2015, I find a 5–10% wage premium for both migrants and return migrants in panel regressions with individual fixed effects for a subsample of formal sector workers. Restricting the sample to those who move at some point in the panel (movers), I find no wage premium associated with the current migrant status and a 4% wage penalty associated with returning. Based on these and other results, I argue that the self-selection of internal migrants in Brazil is based more on migrant's absolute advantage than comparative advantage.

My IV results suggest that past migration rates are either not a good proxy for migrant networks in Brazil or that these networks do not influence an individual's migrant status in the present. Future work can either refine the proxies for migrant networks or take an alternative approach, instrumenting for different migration determinants. In this paper, I use longitudinal data to control for time-invariant unobserved characteristics via individual fixed effects and overcome some of the limitations of the cross-sectional analysis even in the absence of an instrument.

Both my cross-sectional and longitudinal analyses can be expanded. The cross-sectional analysis can incorporate data from the Census in 1991, 2000, and 2010. This would extend the period of analysis and, more importantly, allow for a refinement of migrant definitions using rich information at the municipality level, which is available only in the Census. With these refined migrant definitions, one can investigate the wage premium of inter-municipality (intra-state) migrants together with inter-state migrants. More interesting are the opportunities to expand the analysis using longitudinal data from RAIS. In this investigation, I restricted the sample to a perfectly balanced three-period panel to construct migration definitions and do regression exercises that would be easy to compare with results from the cross-sectional analysis. Future work can explore a larger and richer panel, in which a more flexible structure of an unbalanced panel with several years will bring different migrant definitions, variations in duration and number of migration experiences, and variability between migration across state and municipality boundaries.

References

Aguayo-Tellez, E., Muendler, M.-A. and Poole, J. P. (2010), 'Globalization and formal sector migration in Brazil', *World Development* **38**(6), 840–856.

Alvarez, J. A. (2020), 'The agricultural wage gap: Evidence from Brazilian micro-data', *American Economic Journal: Macroeconomics* **12**(1), 153–73.

Avelino, R. R. (2010), 'Self-selection and the impact of migration on earnings', *Brazilian Review of Econometrics* **30**(1), 69–89.

Batista, C., McIndoe-Calder, T. and Vicente, P. C. (2017), 'Return migration, self-selection and entrepreneurship', *Oxford Bulletin of Economics and Statistics* **79**(5), 797–821.

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', *The American Economic Review* **77**(4), 531–553.

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

Bryan, G., Chowdhury, S. and Mobarak, A. M. (2014), 'Underinvestment in a profitable technology: The case of seasonal migration in Bangladesh', *Econometrica* **82**(5), 1671–1748.

Card, D. (2009), 'Immigration and inequality', *American Economic Review* **99**(2), 1–21.

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

Clemens, M. A., Montenegro, C. E. and Pritchett, L. (2009), 'The place premium: Wage differences for identical workers across the US Border', *Working Paper* .

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

D'Costa, S. and Overman, H. G. (2014), 'The urban wage growth premium: Sorting or learning?', *Regional Science and Urban Economics* **48**, 168–179.

Dix-Carneiro, R. and Kovak, B. K. (2017), 'Trade liberalization and regional dynamics', *American Economic Review* **107**(10), 2908–46.

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.

Ferreira, P. C. and Santos, C. A. R. (2007), 'Migração e distribuição regional de renda no Brasil', *Pesquisa e Planejamento Econômico* **37**(3), 405–425.

Gonzaga, G., Menezes Filho, N. and Terra, C. (2006), 'Trade liberalization and the evolution of skill earnings differentials in Brazil', *Journal of International Economics* **68**(2), 345–367.

Hamory, J., Kleemans, M., Li, N. Y. and Miguel, E. (2020), 'Reevaluating agricultural productivity gaps with longitudinal microdata', *Journal of the European Economic Association* **Advance online publication**.

Kaestner, R. and Malamud, O. (2014), 'Self-selection and international migration: New evidence from Mexico', *Review of Economics and Statistics* **96**(1), 78–91.

Kennan, J. (2013), 'Open borders', *Review of Economic Dynamics* **16**(2), L1–L13.

Kennan, J. and Walker, J. R. (2011), 'The effect of expected income on individual migration decisions', *Econometrica* **79**(1), 211–251.

Kovak, B. K. (2013), 'Regional effects of trade reform: What is the correct measure of liberalization?', *American Economic Review* **103**(5), 1960–76.

Krishna, P., Poole, J. P. and Senses, M. Z. (2014), 'Wage effects of trade reform with endogenous worker mobility', *Journal of International Economics* **93**(2), 239–252.

Lacuesta, A. (2010), 'A revision of the self-selection of migrants using returning migrant's earnings', *Annals of Economics and Statistics/Annales d'Économie et de Statistique* pp. 235–259.

McKenzie, D. and Rapoport, H. (2007), 'Network effects and the dynamics of migration and inequality: theory and evidence from mexico', *Journal of Development Economics* **84**(1), 1–24.

McKenzie, D. and Rapoport, H. (2010), 'Self-selection patterns in mexico-us migration: the role of migration networks', *The Review of Economics and Statistics* **92**(4), 811–821.

Paz, L. S. (2014), 'The impacts of trade liberalization on informal labor markets: A theoretical and empirical evaluation of the Brazilian case', *Journal of International Economics* **92**(2), 330–348.

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.

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.

Venator, J. (2020), 'Dual-earner migration decisions, earnings, and unemployment insurance', *Working Paper* .

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

Woodruff, C. and Zenteno, R. (2007), 'Migration networks and microenterprises in Mexico', *Journal of Development Economics* **82**(2), 509–528.

Yap, L. (1976), 'Internal migration and economic development in Brazil', *The Quarterly Journal of Economics* **90**(1), 119–137.

Appendix

Table A1: Summary statistics and differences in means using data from PNAD 2014: work and household controls

Variables	Means				Differences					
	All	Migrants	Returnees	Stayers	M - S		R - S		M - R	
Industry										
Agriculture	0.08	0.09	0.08	0.08	0.00		0.00		0.00	
Manufacturing	0.23	0.25	0.23	0.23	0.02	***	0.00		0.02	***
Comm./Services	0.37	0.38	0.36	0.37	0.01	***	-0.01	***	0.03	***
Social	0.16	0.14	0.17	0.16	-0.02	***	0.00		-0.03	***
Public	0.06	0.06	0.07	0.06	0.00		0.01	***	-0.01	***
Employment status										
Hired formal	0.45	0.45	0.39	0.46	-0.01	*	-0.06	***	0.06	***
Hired informal	0.20	0.18	0.21	0.21	-0.03	***	0.00		-0.03	***
Self-employed	0.23	0.24	0.26	0.22	0.02	***	0.04	***	-0.02	***
Public/military	0.08	0.08	0.09	0.08	0.00		0.01	**	0.00	
Employer	0.04	0.04	0.05	0.03	0.01	***	0.02	***	-0.01	**
Household characteristics										
Head of household	0.41	0.50	0.51	0.38	0.12	***	0.13	***	-0.01	***
Rural location	0.14	0.11	0.12	0.14	-0.03	***	-0.02	***	-0.01	***
Nr. of members	3.37	3.25	3.18	3.43	-0.18	***	-0.25	***	0.07	***
Share adults	0.82	0.82	0.82	0.82	0.00	***	0.00		0.00	
Observations	256,260	47,106	22,135	187,019						
Obs. w/ wage > 0	156,248	29,901	14,402	111,945						

Notes: Stars denote: * p<0.10; ** p<0.05; *** p<0.01.

Table A2: Estimates of the migrant wage premium using data from PNAD 2014: subsamples and different migration definitions

Dependent variable:	Log hourly wage					
Sample restriction:	Males	Formal sector	Age 16–24	Formal + Age	None	None
	(1)	(2)	(3)	(4)	(5)	(6)
Migrant (M)	0.1428	0.0873	0.0398	0.0452	0.1041	
	(0.0133)***	(0.0131)***	(0.0127)***	(0.0131)***	(0.0105)***	
Return migrant (R)	0.1058	0.1116	0.0548	0.0463		
	(0.0176)***	(0.0185)***	(0.0160)***	(0.0131)***		
Migrant (reference: 5 years ago)						0.1114
						(0.0151)***
Difference (R - M)	-0.0370	0.0243	0.0150	0.0010		
	(0.0188)*	(0.0155)	(0.0151)	(0.0196)		
Adjusted R^2	0.31	0.27	0.18	0.13	0.31	0.31
Observations	89,970	70,054	25,265	13,621	155,793	155,793
Individual controls	Y	Y	Y	Y	Y	Y
State of residence FE	Y	Y	Y	Y	Y	Y
State of birth FE	Y	Y	Y	Y	Y	Y

Notes: The dependent variable is the log of the hourly wage. Individual controls: age, age squared, female indicator (dropped in column 1), race/color categories, and years of schooling. Standard errors clustered by state of residence in parentheses. Stars denote: * p<0.10; ** p<0.05; *** p<0.01.

Table A3: IV estimates of the migrant wage premium using data from PNAD 2014 (1st stage): Intensive and extensive margins

Dependent variable:	Migrant indicator		Ret. migrant indicator		Migrant indicator	
	(1)	(2)	(3)	(4)	(5)	(6)
Emigration rate in 1970	0.5994	0.5889	-0.2556	-0.1884	0.6922	0.5891
	(1.0222)	(1.0998)	(0.1982)	(0.2184)	(0.9998)	(0.9996)
Emigr. 1970 x Age	-0.0473	-0.0395	0.0126	0.0088	-0.0429	-0.0321
	(0.0326)	(0.0224)*	(0.0030)***	(0.0038)**	(0.0233)*	(0.0160)**
Emigr. 1970 x Age^2	0.0009	0.0008	-0.0001	-0.0001	0.0008	0.0007
	(0.0005)	(0.0004)*	(0.0001)*	(0.0001)	(0.0004)*	(0.0003)**
Emigr. 1970 x Female	-0.0336	0.0548	-0.0753	-0.0706	0.0111	0.0581
	(0.0299)	(0.0172)***	(0.0285)***	(0.0276)**	(0.0306)	(0.0121)***
Emigr. 1970 x Non-white	0.2911	0.2775	0.0354	0.0207	0.0138	-0.0039
	(0.1417)**	(0.1335)**	(0.0295)	(0.0274)	(0.1833)	(0.1826)
Ret. migration rate in 1970	-0.1570	0.0717	-0.2481	-0.1019		
	(1.6184)	(1.6650)	(0.2526)	(0.2566)		
Ret. migr. 1970 x Age	-0.0087	-0.0169	0.0105	0.0019		
	(0.0307)	(0.0245)	(0.0077)	(0.0086)		
Ret. migr. 1970 x Age^2	0.0001	0.0002	-0.0001	-0.0000		
	(0.0004)	(0.0003)	(0.0001)	(0.0001)		
Ret. migr. 1970 x Female	-0.0975	-0.0095	-0.0176	0.0018		
	(0.0401)**	(0.0246)	(0.0230)	(0.0204)		
Ret. migr. 1970 x Non-white	0.5641	0.5870	0.0781	0.0640		
	(0.2658)**	(0.2447)**	(0.0496)	(0.0426)		
S-W or F test (weak identification)	7.22	6.29	7.98	4.78	1.46	6.07
Observations	155,793	255,484	155,793	255,484	155,793	255,484
Individual controls	Y	Y	Y	Y	Y	Y
State of residence FE	Y	Y	Y	Y	Y	Y

Notes: Column (1) reports first stage estimates for the first endogenous regressor (M) used in column (1) of Table 7. Column (2) reports estimates for the first endogenous regressor used in columns (2) and (3) of Table 7. Columns (3) and (4) follow the same logic, but for the second endogenous regressor: column (3) reports first stage estimates for the second endogenous regressor (R) used in column (1) of Table 7, and column (4) reports estimates used in columns (2) and (3) of Table 7. Finally, column (5) reports first stage estimates for the single endogenous regressor (M) used in column (4) of Table 7, and column (6) reports estimates used in columns (5) and (6) of Table 7. Standard errors clustered by state of residence in parentheses. Stars denote: * $p<0.10$; ** $p<0.05$; *** $p<0.01$.