Income Changes after Inter-city Migration

Eduardo Lora

CID Research Fellow and Graduate Student
Working Paper No. 128
October 2020

Working Papers
Center for International Development
at Harvard University

Income changes after inter-city migration

Eduardo Lora

Senior fellow at the Center for International Development (CID), Harvard University and Associate Researcher at Research in Spatial Economics (RiSE), Universidad Eafit

Abstract

Using panel data for workers who change jobs, changes in several labor outcomes after inter-city migration are estimated by comparing workers in similar circumstances who move to a new city –the treatment group—with those who stay in the same city –the control group. After matching the two groups using Mahalanobis distances over a wide range of covariates, the methodology of "difference-in-difference treatment effects on the treated" is used to estimate changes after migration. On average, migrants experience income gains but their dedication to formal employment becomes shorter. Income changes are very heterogenous, with low-wage workers and those formerly employed by small firms experiencing larger and more sustained gains. The propensity to migrate by groups of sex, age, wage level, initial dedication, initial firm size and size of city of origin is significantly and directly correlated with the expected cumulative income gains of migration, and inversely with the uncertainty of such gains.

JEL codes: J31, J61, J81

Key words: matched employer-employee panel data, diff-in-diff treatment effects, migration risks, migration determinants, Colombia.

1. Introduction

Although international migration has been a major topic in the development literature (see the reviews by Borjas, 1989 and Castles et al, 2013), migration across regions or cities within developing countries has received much less attention. The focus of the internal migration development literature has been the movements of peoples from the the rural to the urban areas (Todaro, 1980; Lucas 1997 and 2015), rather than across urban areas, in spite of the fact that, as economic development takes hold and urbanization increases, migration movements between cities become more important than those from rural to urban areas (Zelinsky, 1971). To the best of my knowledge, there is not a single work on wage changes after inter-city migration within a developing country. This paper is an attempt to start filling this void.

The main objective of this paper is to measure wage changes after inter-city migration in a developing economy. In addition to wage changes, some measures of earning risks are considered. The identification strategy is to focus only on workers who change jobs, some of whom move to a different city (the treatment group), while the rest remain in the same city (the control group). Although I make no claim of causality, I try to minimize the possible biases that result from the fact that migrants self-select by matching migrants with non-migrants with whom they share a large set of characteristics before their change of job (in addition to their sex and age): the year of change, their wage level, their dedication to their job, the size of the firm that was employing them, their recent history of wage adjustments and the type of city where they were working. In order to balance the treatment and control groups over this set of variables I use Mahalanobis distances. Then, I use the methodology of "treatment effects on the treated" for the outcomes of interest in levels and in differences (diff-in-diff). I make extensive use of the latter to explore the heterogeneity of the labor outcomes of interest by categorical groups (of sex, age, initial income, initial dedication, initial firm size, and city sizes of origin and destination). Finally, I use these computations to study the correlation between the probabilities to migrate by group and the changes in wages and risks after migration. I find evidence in support of the hypothesis that migration decisions depend positively on the expectation and negatively on the uncertainty of income gains.

The rest of the paper is organized as follows: section 2 discusses its contribution to the literature; section 3 presents the dataset and some preliminary concepts; section 4 defines the treatment and control groups; section 5 defines and briefly describes the outcomes; section 6 provides additional details about the estimation strategy; section 7 presents the Mahalanobis distance matching results; section 8 focuses on the mean and the variance of the diff-in-diff treatment effects on the migrants; section 9 discusses the heterogeneity of income changes by groups of workers; section 10 presents the estimation results for the other outcomes considered; in section 11 the hypothesis that migration probabilities depend on expected income gains and risks is tested; and section 12 concludes.

2. Contribution to the literature

This paper contributes most directly to the literature on the so-called "returns to migration" (more strictly, income changes after migration). The early literature on migration returns for the US (Bartel, 1979; Nakosteen and Zimmer, 1980) directly compared wage growth of migrants and non-migrants one year after migration to estimate the contemporary returns to migration. Since no distinction was made between those who changed and those who did not change jobs, the estimates confounded the wage gains associated to migration with those associated to the change of job. The inconsistencies of these earlier studies were partly corrected in a second wave of studies (Hunt and Kau, 1985; Borjas, Bronars, and Trejo, 1992; Yankow, 1999), which focused on longer-term earning changes, but still failed to clearly isolate the part of income changes associated with job changes from the part associated with migration itself. Yankow (2003) was the first to tackle the issue. Using matched employer-employee longitudinal data for job switchers, some of whom changed location, he was able to compare the wage growth patterns of migrants and non-migrants in similar job conditions. His results indicate that the years preceding migration, the wage growth of migrants with more than 12 years of education is similar to that of non-migrants with similar education. After migrating, these workers experience significant income gains. After five plus years, their wages peak more than 11 percent higher than the non-migrant wages, and the cumulative income change over the five year period is 43 percent. In contrast, migrants with less education do not experience any statistically significant income improvement with respect to non-migrants. However, migration provides these workers with a means to restoring their wages to prior levels since a large number of migrations undertaken by low-education workers seem to be prompted by negative wage shocks.

The confounding issue between job change and location change can be alternatively tackled by comparing, not the earning patterns of job switchers who migrate with those who do not migrate, but by comparing the earning patterns of migrants and non-migrants who *do not change job.* Drawing upon the spatial dimension of internal labor markets in firms with multiple establishments in Portugal, Tavares, Carneiro and Varejao (2018) estimate that the contemporaneous wage premium associated with migration is around 3 percentage points. Fehn and Frings (2018) find also that the contemporaneous wage increase of individuals who migrate across regions to new jobs is 3 percentage points higher than that of individuals that change jobs but do not change region. Behind such gain is the fact that regionally mobile workers end up in job matches of higher quality.

The empirical literature on "returns" to internal migration has given very scant consideration to the possibility that income changes may vary across groups of workers who differ not just in their level of education and their age, but in other aspects of their work situation (such as their income and the size of their firms) and in their location before and after they migrate. The only distinction in location before and after migrating that has been made in the internal migration development literature is between rural and urban areas (a salient topic in the early development literature, see Harris and Todaro, 1970; Todaro 1980; Lucas, 1997).

The possibility that larger cities may offer better labor opportunities to internal migrants in developing countries has not been assessed empirically.

Conspicuously absent in the empirical literature on the topic of internal migration is earning risks (and not just unemployment risks as in Harris and Todaro, 1970). Although earning risks probably play an important role in the decision to migrate and where to migrate (as first stated by Stark and Levhari, 1982), to the best of my knowledge no published work has measured such risks or tested the hypothesis empirically for internal migration decisions in either developed or developing economies.

3. Data sources

My main data source is a matched employer-employee panel of formal workers in Colombia between 2008 and 2016. These data are compiled by the Ministry of Health and Social Protection, where employers and self-employed workers must report on a monthly basis the mandated contributions to the social security system. It is known by the acronym PILA (*Planilla Integrada de Aportes Laborales*). By construction, the dataset covers only "formal workers", that is, those that either are employed by a firm that abides by the labor code or that are self-employed and contribute to the social security system. Therefore, it excludes "informal workers", meaning by that those workers who directly or through their firms do not make the mandated contributions to the social security system.

I use a 10 percent sample of all the workers that have at least one registry in PILA over the 2008-2016 period (the full yearly dataset has 73,4 million observations for 16,1 million unique individuals, which exceeds the computation capabilities at my disposal). For the purposes of this paper, the relevant information by worker taken from PILA is the following:

- Sex and age.
- Monthly "wage base" in current pesos. For part-time or temporary workers who do not work a full month, the wage base is the full-time monthly equivalent wage. For each worker, I use only the last month of each calendar year with information.
- The number of weeks the worker contributed to the social security system each year.
- The municipality where the firm is located (which may not be the municipality where the worker resides, which is not reported).

To be able to take into account firm features in the matching process, I make use of the following yearly data at the firm level (computed from the *whole* PILA database, not the 10 percent sample):

- The number of workers employed by the firm in the municipality. This variable allows me to compute the firm size as well as its rate of employment growth.
- The number of workers who keep their job in the firm between the previous and the current year. This allows me to compute the firm's rotation index.
- The sex-composition and the average age of the firm's employees.

In order to compare wages across time periods and locations for workers of different wage levels, I deflate the monthly wage base of each worker by a Consumer

Price Index that varies in three dimensions: time, location and wage group. The time *variation* is taken directly from the official CPI, which is produced by the National Statistical Office (DANE) for 23 cities. The *level* differentiation by location and wage group (which is time-invariant) is computed based on the difference for each wage level between the housing rental of the location and that of Bogotá. The source of this information is DANE's National Survey of Household Budgets of July 2016-July 2017 (see further details in Appendix).

Since my focus is inter-city migration, a clear definition of what constitutes a "city" is in order because many large cities span over the political boundaries of more than one municipality (the main reason why workers may not reside in the same municipality where they work). To that purpose, I use the algorithm developed by Duranton (2015) to sequentially aggregate municipalities into "metropolitan areas" when at least 10 percent of the labor force of a municipality regularly commutes to work in the aggregation of municipalities. Using data for 2008, the algorithm generates 19 metropolitan areas that consist of 2 or more municipalities, comprising a total of 115 municipalities. Similar to the standards of the US Office of Management and Budget (OMB) for metropolitan area delineations[1], I add to these 19 urban areas another 43 individual municipalities that have populations above 50,000 inhabitants, for a total of 62 cities.

Given its relatively large number of cities, Colombia is a fitting case to study inter-city migration patterns. In a comparison of 96 developed and developing

[1] http://www.census.gov/population/metro/data/metrodef.html.

countries, Colombia is positioned close to the media in the intensity of domestic mobility, measured by changes of address (Bell et al 2015). In a raking of 47 countries ordered by the degree to which their internal migration movements redistribute their populations spatially, Colombia is in position 34th (Rees et al 2016).

4. Treatment and control groups

In order to define the "treatment" and "control" groups, I consider five possible states to which a waged worker may transition between periods t and t+1: (a) "no change", that is, same job, same city; (b) "new firm, new city"; (c) "new firm, same city"; (d) "same firm, new city"; and (e) "out of formality", which is a short-hand for the two following states: new job in a location other than the 62 cities considered and no formal employment (either in any of the 62 cities or in other locations, i.e. no PILA registry).

On average, 57.2% of all formal workers stay in year t+1 in the same job and city they were in year t, 5.4% move to a new job in another city, 20.5% change job but stay in the same city, 0.9% move in the same job to a new city, and 16.1% move out of formality. As Table 1 indicates, transition probabilities are not uniform by categorical groups. The categorical variables considered are sex, age group, initial income group (defined by multiples of the minimum wage), initial number of weeks in formal employment, initial firm size group and size group of the city of origin.

Men transition more frequently than women to all states other than "no change", young workers transition more frequently than their seniors to new jobs (either in the same or in a new city). Lower wage workers make more frequent transitions across firms than workers of higher incomes. The same is valid for those who work less than 40 weeks per year compared to those that work more weeks. Workers in firms smaller than 10 workers transition less frequently to other states than employees of larger firms. By city of origin, transitioning to new jobs in a different city is more likely among workers of smaller cities, while finding a new job in the same city is more likely among those in larger cities (see the cities included in each size category at the bottom of Table 1).

Table 1. Transition probabilities by groups

	No change	New firm, new city	New firm, same city	Same firm, new city	Out of formality	Total
All formal workers	57.2%	5.4%	20.5%	0.9%	16.1%	100.0%
By sex						
Men	55.0%	6.5%	21.3%	0.9%	16.5%	100.0%
Women	60.6%	3.7%	19.3%	0.8%	15.8%	100.0%
By age group						
20 to 29 years	46.7%	7.0%	27.2%	0.8%	18.6%	100.0%
30 to 39 years	59.5%	5.5%	20.3%	1.0%	14.0%	100.0%
40 to 49 years	66.8%	4.0%	15.5%	0.9%	13.0%	100.0%
50 to 54 years	70.3%	3.3%	12.6%	0.9%	13.1%	100.0%
By initial income group (times the minimum wage)						
Less than 1.25	48.5%	5.6%	23.1%	0.6%	22.4%	100.0%
Between 1.25 and 2.5	61.6%	5.2%	20.8%	0.9%	11.7%	100.0%
Between 2.5 and 5	69.2%	5.8%	14.9%	1.4%	9.1%	100.0%
More than 5	75.9%	3.8%	12.1%	1.6%	6.8%	100.0%
By initial number of weeks in formal employment						
Less than 25	25.5%	8.7%	24.8%	0.6%	41.0%	100.0%
Between 25 and 40	49.1%	7.3%	27.5%	0.9%	16.0%	100.0%
More than 40	76.1%	3.2%	16.3%	1.0%	3.7%	100.0%
By initial firm size group						
Less than 10 workers	61.7%	3.9%	15.3%	1.0%	18.4%	100.0%
Between 10 and 25 workers	52.4%	5.8%	21.4%	0.4%	20.3%	100.0%
Between 25 and 100 workers	51.3%	6.6%	22.3%	0.6%	19.4%	100.0%
More than 100 workers	58.2%	5.5%	21.8%	1.0%	13.7%	100.0%
By city of origin [1/]						
Three largest cities	59.4%	3.0%	23.3%	0.4%	14.1%	100.0%
Seven medium-size cities	55.9%	4.6%	22.0%	0.6%	17.1%	100.0%
Remaining 52 cities	53.3%	11.0%	13.6%	2.2%	20.1%	100.0%

Source: own calculations from Colombia's Ministry of Health's PILA. Averages of yearly transition probabilities between 2008 and 2016.

[1/] The three largest cities are Bogotá, Medellín and Cali. The seven medium-size cities are Barranquilla, Bucaramanga, Cartagena, Cucutá, Ibagué, Manizales and Pereira. They include the municipalities that form part of their metropolitan area using Duranton's (2015) algorithm.

In the remainder of this paper I compare workers who transition from their current job and city to "new firm, new city" –the treatment group—with those who transition to "new firm, same city" –the control group. I refer to the last year before one of the two transitions as the "base year" (t_0) for all my comparisons. However, not all the individuals that make either transition are included in the comparisons. I keep in the sample only those workers who stayed in the same job during at least three years previous to the change of job. This in order to match workers with similar job stories in the treatment and the control groups. After the first change of job is observed, workers may change job again, but I exclude from the control group those who subsequently change city. In the treatment group, workers may change city more than once (including going back to the initial city, which occurs relatively frequently; see Prieto Curiel, et al, 2020). To look for adequate matches between observations in both groups (see section 5), I keep all the observations by worker as long as he/she is observed every subsequent year (although not necessarily every month of the year). Therefore, workers included in the matching process have a minimum of 4 years of observation (3 in the initial job and at least one in the new job) and a maximum of 9 continuous years of observation (the latter dictated by the length of my panel). Table 2 presents the composition of the treatment and control groups (before matching) thus defined by categories.

Table 2. Treatment and control groups
(total number of individuals before matching)

	Workers that change firm	Of whom: change city (Treatment)		Of whom: do not change city (Control)	
	Number	Number	Percentage	Number	Percentage
All formal workers	421,924	57,360	13.6%	364,564	86.4%
By sex					
Men	247,083	39,710	16.1%	207,373	83.9%
Women	174,841	17,650	10.1%	157,191	89.9%
By age group					
20 to 29 years	175,613	25,433	14.5%	150,180	85.5%
30 to 39 years	98,390	14,349	14.6%	84,041	85.4%
40 to 49 years	60,103	7,138	11.9%	52,965	88.1%
50 to 54 years	16,057	1,875	11.7%	14,182	88.3%
By initial income group (times the minimum wage)					
Less than 1.25	228,958	28,143	12.3%	200,815	87.7%
Between 1.25 and 2.5	132,337	17,183	13.0%	115,154	87.0%
Between 2.5 and 5	38,984	7,768	19.9%	31,216	80.1%
More than 5	21,645	4,266	19.7%	17,379	80.3%
By initial number of weeks in formal employment					
Less than 25	150,520	24,886	16.5%	125,634	83.5%
Between 25 and 40	89,438	12,848	14.4%	76,590	85.6%
More than 40	181,966	19,626	10.8%	162,340	89.2%
By initial firm size group					
Less than 10 workers	69,061	8,541	12.4%	60,520	87.6%
Between 10 and 25 workers	37,683	5,058	13.4%	32,625	86.6%
Between 25 and 100 workers	70,754	10,689	15.1%	60,065	84.9%
More than 100 workers	244,426	33,072	13.5%	211,354	86.5%
By city of origin					
Three largest cities	274,727	24,979	9.1%	249,748	90.9%
Seven medium-size cities	88,998	15,690	17.6%	73,308	82.4%
Remaining 52 cities	58,199	16,691	28.7%	41,508	71.3%

Source: own calculations from Ministry of Health's PILA. "Control" includes all the individuals that during 2008-2016 changed firm at least once after having worked in a formal job in the last three years and never changed city. "Treatment" includes all those that changed firm and city the same year after having a formal job in the last three years, irrespective of any further changes across firms or cities they may have had subsequently. Both the control and the treatment groups exclude individuals that did not have *any* formal employment during a full calendar year.

5. Outcomes definition and descriptive statistics

Although my main interest is income changes after migration, this is not the only aspect of workers' lives that the decision to migrate may affect. Others include the dedication to formal employment and the stability of earnings.

Therefore, I consider several outcomes that can be observed at the individual level:

(1) *real wage growth rate* r_t, defined as the annual growth rate of log real wages lnw since "base year" t_0:

$$r_{t_i} = (lnw_{t_i} - lnw_{t_0})/(t_i - t_o)$$

(2) *yearly weeks of dedication to formal employment* since t_0:

$$f = \sum_{t_0}^{t_i} weeks_t /(t_i - t_o)$$

(3) the *standard deviation of the real wage growth rate, sr:*

$$sr = \sqrt{\frac{\sum(r_{t_i - t_0} - \bar{r})^2}{t_i - t_0 - 1}}$$

where $\bar{r}$ is the average of r_{t_i}.

Notice that these variables can also be computed for the base year, where r_{t_0} is real wage growth in year t_0 (with respect to the previous year), f_0 is weeks of dedication in t_0, and sr_0 is the standard deviation of real wage growth the last 2 or 3 years until t_0 (depending on the availability of information). The 3 outcomes in differences are therefore $dr_{t_i} = r_{t_i} - r_{t_0}$, $df = f - f_0$ and $sr = sr - sr_0$.

Table 3. Real wage growth rate of treatment and control groups before matching (yearly since base year)

Number of years since base year	Treatment group			Control group		
	Number of observations	Level Mean	Std. Dev.	Number of observations	Level Mean	Std. Dev.
1 year	57,360	0.063	0.468	364,564	0.038	0.364
2 years	54,116	0.064	0.250	283,683	0.051	0.205
3 years	44,313	0.062	0.178	226,342	0.053	0.147
4 years	34,436	0.060	0.141	179,457	0.052	0.118
5 years	25,260	0.056	0.118	137,981	0.050	0.099
6 years	16,528	0.051	0.103	94,666	0.046	0.086
	Number of observations	Difference Mean	Std. Dev.	Number of observations	Difference Mean	Std. Dev.
1 year	38,199	0.003	0.629	227,506	-0.026	0.480
2 years	38,061	-0.002	0.456	187,420	-0.011	0.347
3 years	32,002	-0.008	0.411	156,755	-0.010	0.309
4 years	25,119	-0.012	0.392	128,847	-0.008	0.295
5 years	18,576	-0.015	0.386	102,967	-0.010	0.288
6 years	12,239	0.009	0.207	74,371	-0.008	0.282

Source: own calculations from Ministry of Health's PILA. Includes all the observations, even those that lack information of balancing variables.

The top panel of Table 3 shows number of observations and mean real wage growth rates of the treatment and control groups (before matching) by number of years since base year. The rapid decline in the number of observations, in both the treatment and the control groups, is due mainly to the limited length of my panel and the high attrition rates that result from workers leaving the formal labor market. Mean wage growth rates are always higher in the treatment than in the control group and both decline slightly with the number of years since change of firm. The standard deviation of this first outcome is very large in the first year and declines markedly thereafter. Since the standard deviation is several times the average real wage growth rate, it is implied that many workers face income losses, especially in the first years after they change jobs, either in the same or in a new city. The bottom panel shows the same outcome in differences with respect to the base year. Now the treatment and control groups show closer mean values but larger standard deviations.

Table 4. Other outcomes in the treatment and control groups before matching

Outcome	Treatment group			Control group		
	Number of observations	Level mean	Std. Dev.	Observations	Level mean	Std. Dev.
Yearly weeks in formal employment since job change	146,109	32.388	13.526	386,588	35.945	12.761
Standard deviation of yearly wage growth since job change	55,909	0.122	0.136	283,811	0.103	0.113
	Number of observations	Difference mean	Std. Dev.	Observations	Difference mean	Std. Dev.
Yearly weeks in formal employment since job change	146,109	3.726	11.945	386,588	3.880	11.790
Standard deviation of yearly wage growth since job change	24,142	0.020	0.166	99,661	0.039	0.133

Source: own calculations from Ministry of Health's PILA.

Table 4 presents descriptive statistics for outcomes (2) and (3). Both the level and the difference with respect to the base year of the means of yearly weeks in formal employment after change of job are smaller in the treatment than in the control group. However, while the mean level of the standard deviation of real wage growth is larger in the treatment than in the control group, the mean *difference* is smaller in the treatment than in the control group.

6. Further details of the estimation procedure

Estimating the treatment causal effects of migration can only be done with observational data, which is not straightforward because workers who migrate are not a random sample of the universe of workers. They differ not only in their personal characteristics and their work circumstances but in their inclination to migrate, which is not independent of such characteristics and circumstances. Therefore, I make no claim of causality. However, to minimize the ensuing potential biases, I take several actions. First, as explained in section 4, I include in the treatment and control groups only individuals who had stable formal employment and change to a new firm, which may or may not be in the same city. To be included in either group, workers must have spent at least 3 years in their job before changing to a new job.

Second, I select a list of covariates aimed at capturing other personal and job circumstances that may affect either the decision to migrate or the outcomes of migrating and, presumably, are not affected by it. More specifically, I use the following list of categorical and continuous variables:

- t_0, that is, the last year before the change of job (categorical)
- Sex (categorical)
- Age at t_0 (continuous)
- Income level (number of times the minimum wage, in logs) at t_0 (continuous)

- Weeks worked in t_0 (discrete, in deciles)
- Real wage growth rate between years t_{-2} and t_{-1} (continuous)
- Real wage growth rate between years t_{-1} and t_0 (continuous)
- Size of city of origin (categorical, see categories in Table 1)
- Initial firm size (in logs) at t_0 (continuous)
- Initial firm employment growth between t_{-1} and t_0 (continuous)
- Initial firm rotation index between t_{-1} and t_0 (continuous).

Arguably, other variables may affect the decision to migrate and/or its outcomes. In particular, the level of education and the occupation of the individual should ideally be included in the list of covariates, but they are not collected in the PILA.

Third, I use Mahalanobis distances over the whole list of covariates to match the treatment and control groups (Leuven and Sianesi 2003). My choice of Mahalanobis distance matching, MDM, was dictated by the results of numerous trials with varieties of propensity score matching, PSM (with and without caliper, Kernel, etc.), which did not produce good balance results. Even worse, many of the trials exhibited the "propensity matching paradox" described by King et al (2011), where imbalances for some variables increase when pruning more observations. I also tried coarsened exact matching (Iacus, King and Porro 2012), which did not exhibit this pattern, but which led to substantially more pruning than MDM.[2] Therefore, following King and Nielsen (2019), I discarded PSM to use MDM with caliper. I

[2] Results available upon request.

choose a different caliper in each estimation in order to balance every variable in both groups. Then, I compute average treatment effects on the treated in two ways: using the absolute values of each of the outcomes and using the differences of each of the outcomes with respect to the base year. Finally, the latter treatment effects on the treated (diff-in-diff TET) of each of the outcomes are averaged over a range of groups (by sex and categories of age, initial income, initial dedication to work, initial firm size and city sizes of origin and destination) to assess their heterogeneity and to test whether they correlate with the probabilities to migrate of the groups.

7. Balancing the treatment and control groups

Although the matching technique does not rest on propensity scores, for illustrative purposes Probit regression results are presented in Table 5. The regression includes all the treatment and control variables with complete data before matching. With only two exceptions, the balancing variables are significantly correlated with the probability to migrate. However, the pseudo R-squared is low, reflecting poor prediction capacity: just 8,5% of the treatment cases are correctly predicted with a probability of at least 50%.

Table 5. Probit regression of migration on list of balancing variables
(Treatment: 1, Control: 0)

Explanatory variables	Outcome: yearly wage growth since migration					
	Coefficient	Std. Error	z	P|z|	[95% confidence	interval]
t0 = 2010	-0.3853	0.0145	-26.55	0	-0.4137	-0.3569
t0 = 2011	-0.2873	0.0146	-19.68	0	-0.3159	-0.2587
t0 = 2012	-0.1882	0.0151	-12.46	0	-0.2178	-0.1586
t0 = 2013	-0.0914	0.0151	-6.04	0	-0.1211	-0.0617
t0 = 2014	-0.0165	0.0151	-1.09	0.275	-0.0462	0.0132
Sex (male=1)	0.3067	0.0088	34.91	0	0.2895	0.3240
Age	-0.0133	0.0005	-29.4	0	-0.0142	-0.0125
Initial income (log of times the min wage)	0.2717	0.0068	39.72	0	0.2583	0.2851
Weeks of work (deciles)	-0.1028	0.0019	-54.76	0	-0.1065	-0.0991
Real wage growth between t-1 and t0	-0.0849	0.0213	-3.99	0	-0.1265	-0.0432
Real wage growth between t-2 and t-1	-0.0554	0.0356	-1.55	0.12	-0.1252	0.0144
Origin: large city (dummy)	-0.7719	0.0112	-69.2	0	-0.7937	-0.7500
Origin: medium ciity (dummy)	-0.3168	0.0129	-24.65	0	-0.3420	-0.2916
Initial firm size (log)	-0.0083	0.0018	-4.71	0	-0.0117	-0.0048
Initial firm annual employment growth	0.2632	0.0129	20.34	0	0.2379	0.2886
Initial firm rotation index	0.5882	0.0207	28.36	0	0.5475	0.6289
Constant	0.4156	0.0260	16.01	0	0.3647	0.4664
Number of observations	145,626, of which 23,906 are treatment, and 121,720 control.					
Pseudo R squared	0.1142					

Source: computations with Ministry of Health's PILA. Omitted year for t0 is 2015. Omitted city size category is small.

Table 6. Tests of differences between treatment and control means and variances

Balancing variables	Mean			t-test		Variance ratio
	Treatment	Control	% bias	t	p>|t|	V(T)/V(C)
	Before matching					
t0 = 2010	0.1632	0.1690	-1.4	-1.71	0.088	.
t0 = 2011	0.1716	0.1664	1.3	1.5	0.133	.
t0 = 2012	0.1573	0.1524	1.3	1.47	0.143	.
t0 = 2013	0.1659	0.1649	0.3	0.3	0.768	.
t0 = 2014	0.1689	0.1717	-0.8	-0.83	0.408	.
t0 = 2015	0.1732	0.1755	-0.7	-0.68	0.5	.
Sex (male=1)	0.6983	0.6949	0.7	0.8	0.426	.
Age	35.1650	35.1930	-0.3	-0.32	0.749	0.96
Initial income (log of times the min wage)	0.6840	0.6831	0.1	0.14	0.887	0.94
Weeks of work (deciles)	6.1121	6.0584	2.4	2.41	0.016	0.96
Real wage growth between t-1 and t0	0.0387	0.0391	-0.2	-0.17	0.865	1.45
Real wage growth between t-2 and t-1	0.0504	0.0498	0.3	0.36	0.716	1.57
Origin: large city (dummy)	0.4614	0.4569	0.9	0.97	0.331	.
Origin: medium ciity (dummy)	0.2722	0.2713	0.2	0.23	0.821	.
Origin: small city (dummy)	0.2664	0.2718	-1.4	-1.32	0.187	.
Initial firm size (log)	5.1889	5.1627	1	1.09	0.277	1.03
Initial firm annual employment growth	-0.0145	-0.0287	3.8	4.43	0	2.11
Initial firm rotation index	0.3518	0.3443	3.2	3.4	0.001	1.25
Number of observations	23,906	21,720				
	After matching with Mahalanobis distances and caliper=0.7					
t0 = 2010	0.171	0.171	0	0	1	.
t0 = 2011	0.192	0.192	0	0	1	.
t0 = 2012	0.145	0.145	0	0	1	.
t0 = 2013	0.158	0.158	0	0	1	.
t0 = 2014	0.165	0.165	0	0	1	.
t0 = 2015	0.168	0.168	0	0	1	.
Sex (male=1)	0.681	0.681	0	0	1	.
Age	34.246	34.266	-0.2	-0.16	0.877	1.01
Initial income (log of times the min wage)	0.496	0.484	1.7	1.25	0.211	1.01
Weeks of work (deciles)	6.921	6.962	-1.8	-1.31	0.192	1.02
Real wage growth between t-1 and t0	0.031	0.030	0.3	0.36	0.716	1.09
Real wage growth between t-2 and t-1	0.036	0.035	0.5	0.72	0.471	1.11
Origin: large city (dummy)	0.583	0.583	0	0	1	.
Origin: medium ciity (dummy)	0.226	0.226	0	0	1	.
Origin: small city (dummy)	0.191	0.191	0	0	1	.
Initial firm size (log)	5.281	5.249	1.2	0.76	0.445	1
Initial firm annual employment growth	0.000	-0.005	1.3	1.47	0.141	1.04
Initial firm rotation index	0.296	0.295	0.8	0.57	0.571	1.03
Number of observations (common support)	15,110	21,720				

Source: computations with Ministry of Health's PILA.

Table 6 presents tests of differences between the treatment and control groups for all the covariates included in Table 5. Before matching (upper panel), several covariates are significantly different between the treatment and control groups. After matching with Mahalanobis distances and caliper (0.7), all the t-stats of mean differences are low (the highest is 1.47) and the variance ratios are moderate (1.11 is the highest), indicating that the two groups are balanced. As a result of the procedure, about a third of the observations that belong to the treatment group are discarded. Although balancing results are acceptable for all the observed variables, imbalances among not observables cannot be ruled out.[3]

Since the sample of individuals vary with the number of years since change of firm (as shown in Table 3), the previous tests are valid only for the outcomes that can be measured the first year after migration. They also become invalid if the sample changes as a result of lack of information on the outcomes and/or the covariates. Table 7 presents a summary of T- tests of differences between treatment and control groups that are valid for the corresponding samples of the 3 outcomes in differences. I will use these results of the matching exercises also when the outcomes are not measured in differences but in levels since this allows me to directly compare the results.

[3] This is the case, for instance, of the variable "share of men in initial firm", which is not included among the balancing variables: although the bias between the control and the treatment variables is reduced 40% when pruning the treatment sample with the MDM, the difference is still statistically significantly.

Table 7. T-tests of differences before and after matching

Balancing variables	Number of years since migration						Other outcomes
	t=1	t=2	t=3	t=4	t=5	t=6	
All observations with complete information needed for outcomes in differences (before mathing)							
Year 2010	-1.71	-0.22	-1.37	-1.09	-0.91	.	-1.69
Year 2011	1.5	0.15	0.57	0.86	0.91	.	0.39
Year 2012	1.47	0.83	0.36	0.23	.	.	1.22
Year 2013	0.3	-0.02	0.47	.	.	.	0.22
Year 2014	-0.83	-0.73	.	.	.	.	-0.09
Year 2015	-0.68	.	.	.	.	.	.
Sex (male=1)	0.8	0.9	-0.1	0.66	0.6	0.71	0.46
Age	-0.32	0.51	-1.48	0.97	-0.43	**-2.28**	0.03
Initial income (log of times the min wage)	0.14	-0.18	1.29	**2.33**	1.31	0.6	1.29
Weeks of work (deciles)	**2.41**	1.6	**2.43**	0.84	0.32	0.81	1.42
Real wage growth between t-1 and t0	-0.17	0.88	0.05	0.45	1.81	-0.29	-0.62
Real wage growth between t-2 and t-1	0.36	-0.07	0.03	0.81	1.61	0	-0.3
Origin: large city (dummy)	0.97	1.03	1.12	1.28	-0.19	0.28	1.58
Origin: medium ciity (dummy)	0.23	1.15	0.74	-1.22	**1.91**	0.86	1.09
Origin: small city (dummy)	-1.32	-2.3	**-2.01**	-0.31	-1.58	-1.15	**-2.87**
Initial firm size (log)	1.09	1.64	0.71	-0.23	-1.07	-0.99	-0.21
Initial firm annual employment growth	**4.43**	**3.62**	**3.59**	1.44	-0.24	0.47	**3.26**
Initial firm rotation index	**3.4**	**4.15**	**2.4**	-0.17	0.71	0.93	**2.75**
Observations: treatment	23,906	23,861	19,275	14,155	9,209	4,343	24,105
Observations: control	121,720	99,515	80,839	62,799	44,654	23,393	99,553
Common support observations for outcomes in differences (after matching with Mahalanobis distances)							
Year 2010	0	0	0	0	0	.	0
Year 2011	0	0	0	0	0	.	0
Year 2012	0	0	0	0	.	.	0
Year 2013	0	0	0	.	.	.	0
Year 2014	0	0	.	.	.	.	0
Year 2015	0	.	.	.	.	.	.
Sex (male=1)	0	0	0	0	0	0	0
Age	-0.16	-0.05	-0.14	0.14	0.02	0.15	0.05
Initial income (log of times the min wage)	1.25	1.53	1.3	1.13	0.92	0.83	1.19
Weeks of work (deciles)	-1.31	-1.56	-1.64	-1.5	-1.36	-1.02	-1.31
Real wage growth between t-1 and t0	0.36	0.52	0.41	0.65	0.41	-0.28	0.47
Real wage growth between t-2 and t-1	0.72	1	0.79	0.77	0.6	0.17	0.98
Origin: large city (dummy)	0	0	0	0	0	0	0
Origin: medium ciity (dummy)	0	0	0	0	0	0	0
Origin: small city (dummy)	0	0	0	0	0	0	0
Initial firm size (log)	0.76	0.6	0.66	0.49	0.2	-0.12	0.51
Initial firm annual employment growth	1.47	1.32	1.21	0.98	0.73	0.04	1.23
Initial firm rotation index	0.57	0.86	0.87	0.61	0.19	0.03	0.47
Memo: caliper	0.7	0.8	0.9	1	1.1	1.2	0.7
Observations: treatment	8,796	10,475	9,454	7,647	5,447	2,748	9,417
Observations: control	121,720	99,515	80,839	62,799	44,654	23,393	99,553

Note: calculations with PILA data. Numbers in bold characters are significantly different from zero with 95% confidence.

8. Yearly real wage growth after migration

Table 8 shows the average treatment effects on the treated for the first outcome, wage growth, calculated with the 3 alternative methods: before matching in level means, after Mahalanobis distance matching in level means and after MDM in differences (diff-in-diff). With each method a separate (cross-section) estimation is performed for each period t_i since the year of firm change ($1 \leq i \leq 6$). Thus, each line of the table comes from a different estimation.

Table 8. Income changes after migration (average treatment effects on the treated)

Number of years since migration	Coefficient	AI Robust Std. Err.	z	P>|z|	[95% Conf. Interval]	
Before matching, absolute means (observations with complete information of matching variables)						
1 year	0.1038	0.0050	20.66	0	0.0939	0.1136
2 years	0.0413	0.0027	15.32	0	0.0360	0.0466
3 years	0.0255	0.0021	12.4	0	0.0215	0.0296
4 years	0.0151	0.0018	8.3	0	0.0116	0.0187
5 years	0.0120	0.0019	6.33	0	0.0083	0.0157
6 years	0.0069	0.0023	3.01	0.001	0.0024	0.0114
After matching with Mahalanobis distances and caliper, level means						
1 year	0.0465	0.0061	7.6	0.000	0.0345	0.0585
2 years	0.0196	0.0031	6.27	0.000	0.0135	0.0257
3 years	0.0101	0.0023	4.36	0.000	0.0055	0.0146
4 years	0.0077	0.0021	3.74	0.000	0.0037	0.0118
5 years	0.0059	0.0021	2.84	0.002	0.0018	0.0099
6 years	0.0073	0.0025	2.89	0.002	0.0024	0.0122
After matching with Mahalanobis distances and caliper, diff-in-diff						
1 year	0.0458	0.0064	7.11	0.000	0.0331	0.0584
2 years	0.0186	0.0039	4.82	0.000	0.0110	0.0261
3 years	0.0092	0.0034	2.7	0.003	0.0025	0.0158
4 years	0.0060	0.0037	1.65	0.049	-0.0011	0.0132
5 years	0.0045	0.0043	1.07	0.142	-0.0038	0.0129
6 years	0.0087	0.0061	1.41	0.793	-0.0033	0.0207

Source: own calculations from Colombia's Ministry of Health's PILA.

The rough estimate before matching suggests that migration is associated with an increase of over 10% in real wages the first year since migration. After six years, the increase is still significant but substantially lower. After matching, the changes in level means are more moderate –starting at 4,6%–but still significant throughout the whole 6-year period. When measured in differences with respect to the base year, the changes are very similar but become not statistically significant in the last years. Figure 1 is a graphical representation of the implied behavior of the annual real wage growth rate and the *cumulative* income gains, based on the diff-in-diff results. Although the mid-point estimates indicate that the cumulative gains of the migrants increase steadily, and even accelerate towards the end of the observation period, the 95% confidence interval suggests a wide range of variation around the mean.

Fig. 1 Yearly and cumulative income changes after migration
(point estimates and 95% confidence intervals)

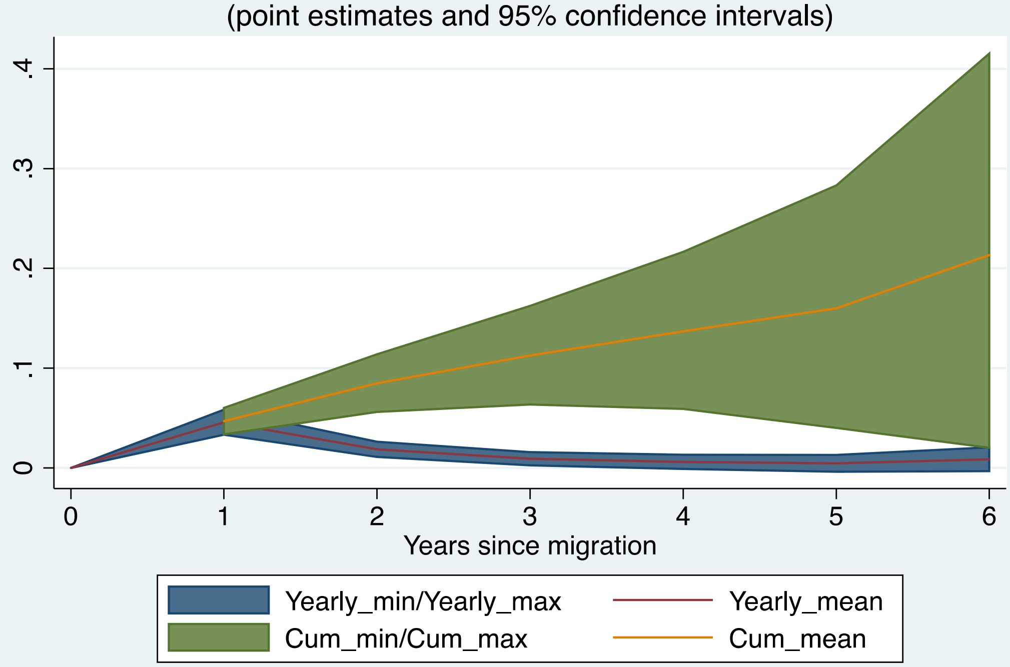
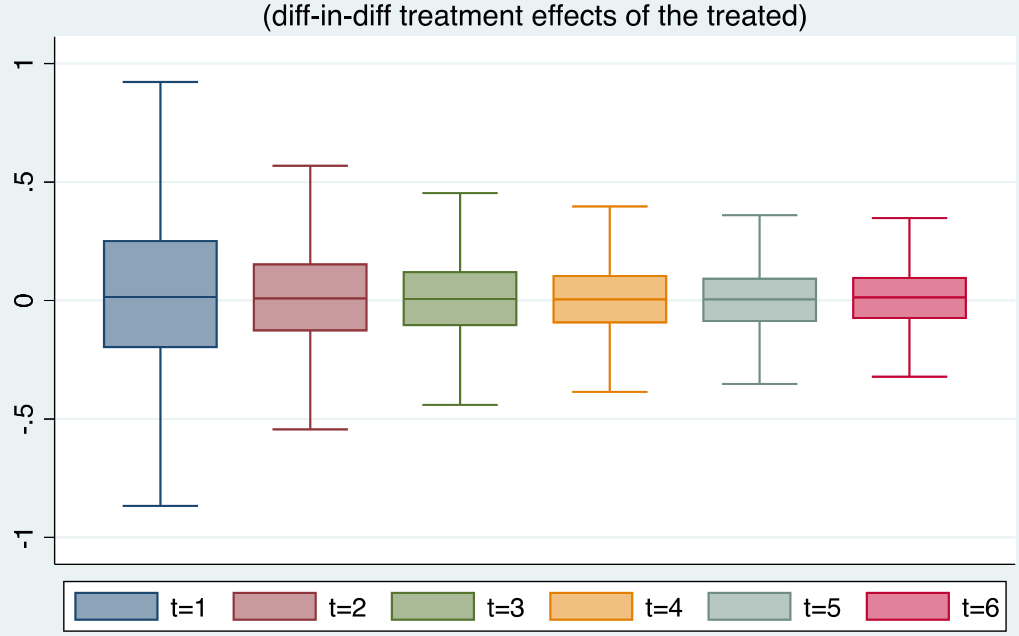


Figure 2 shows the ranges of variation of the yearly income changes, instead of the cumulative gains (based, again, in the diff-in-diff results). The boxes show the 75th, 50th and 25th percentiles. Notice that the medians (50th percentile) are all positive but very close to zero (from 1,71% in the first year to 1,3% in the sixth year), implying that in nearly half of the treatment observations income changes after migration are negative (46% of the observations in t_1 and 43.3% in t_6). The ranges of variation of income changes are wide but declining as time advances.

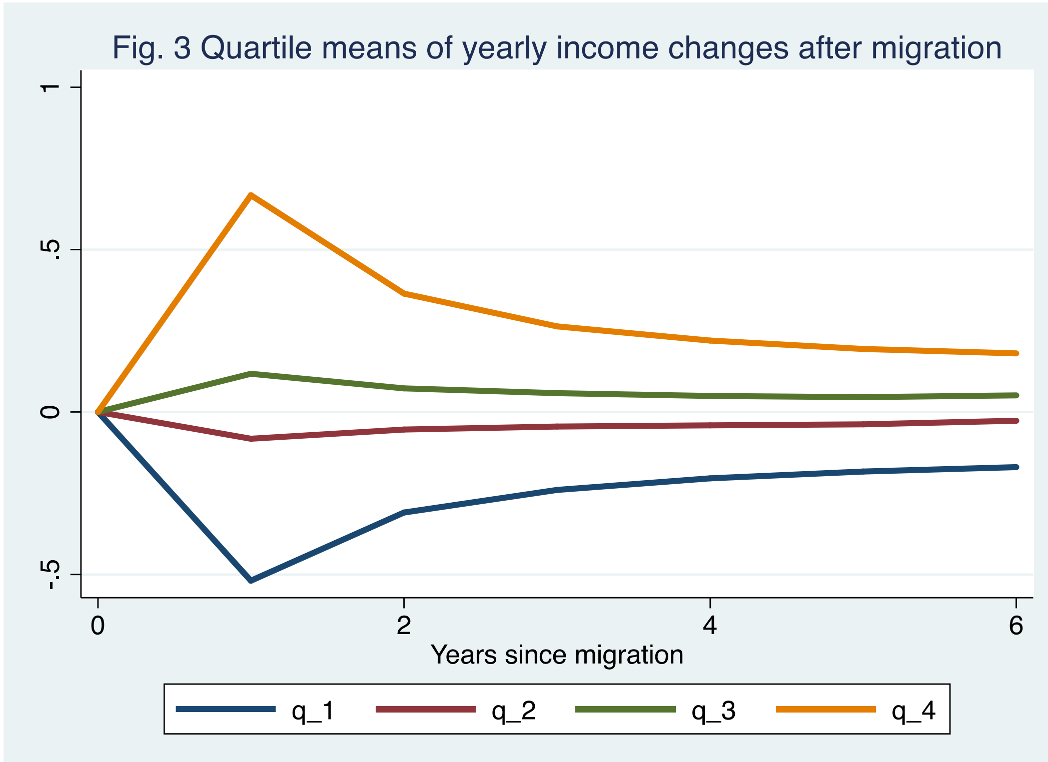
Mean income changes after migration by quartiles are shown in Figure 3. The remarkable symmetry around zero indicates that the income risks of migration are high in relation to the mean income change, especially in the first years. As time advances, quarterly means tend to converge, but at a slow rate. Still at year 6, the 4-quartile means are significantly different from zero.

Fig. 2 Range of variation of yearly income changes after migration

(diff-in-diff treatment effects of the treated)



Boxes show 75th, 50th and 25th percentiles. Outliers are excluded.



9. Income changes after migration by categories

I now use the estimated diff-in-diff treatment effects on the treated to compute averages by groups. Figure 4 presents income changes the first year of migration (t_1) for the same categories in Tables 1 and 2. Within each category, income changes are highest for men, for workers in their 40s, for those whose initial wages were less than 1.25 times the minimum wage, for those that had a formal job during less than 25 weeks in the base year, and for workers who were at firms sized 10 to 25. Notice that in no category are the changes monotonic.

Fig. 4 Income changes after migration (diff-in-diff, first year) by sex and categories of age, initial income, initial dedication and initial firm size

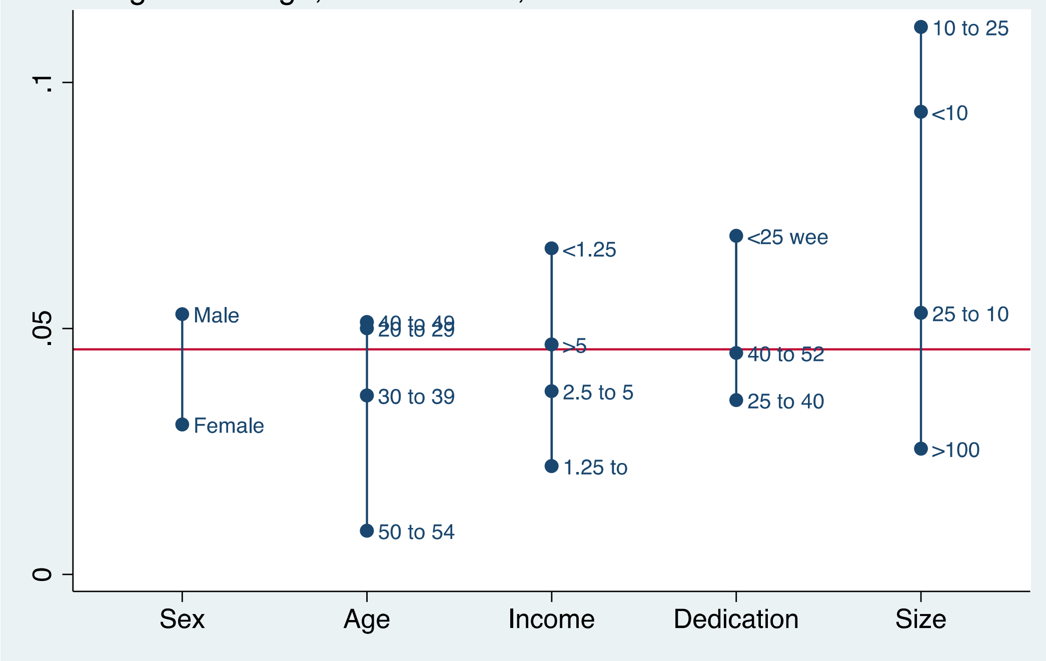


Figure 5 makes similar comparisons by city sizes of origin and destination. On average, migrants from medium and small cities experience real wage growth increases larger than the average of all migrants. The counterpart of this is that migrants from the large to the medium- or small sized cities experience smaller gains.

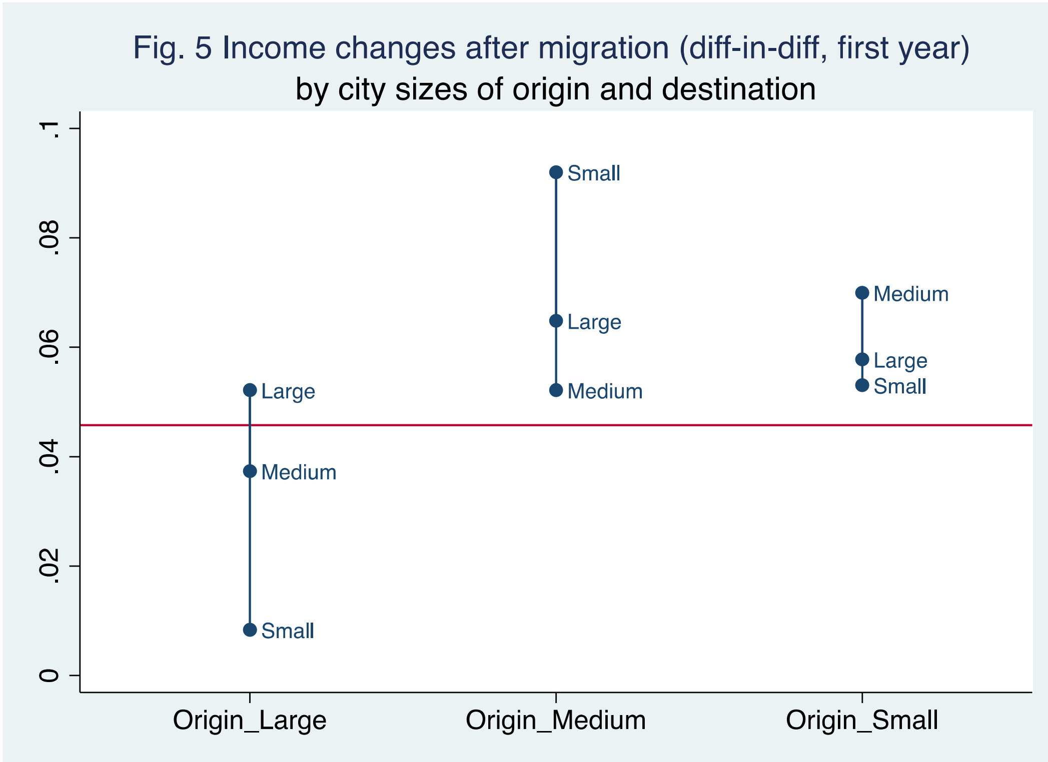


Table 9 shows the average diff-in-diff treatment effects on the treated by the number of years since migration (t_i) and by categories. The first column contains the same information of Figure 4. The final column shows the cumulative income change at year 6. Income changes are significantly different for zero for most groups in the first 2 years, but not in subsequent years. They are significant throughout the whole 6-year period only for workers with initial incomes lower than 1.25 times the minimum wage and for workers who were occupied in firms sized 25 to 100. Cumulative income gains after 6 years present a high level of heterogeneity across groups within each category. On average, male migrants experience income gains larger than female migrants (but the difference is statistically significant only the first year). Those with initial incomes below 1.25 times the minimum wage gain over 55%, dwarfing the changes of all other income groups (which are not statistically significant). Migrants formerly employed by firms with more than 100 employees gain just 9%, while all others gain more than 40% (the difference is statistically significant).

Table 9. Yearly income changes after migration by groups and years since migration (diff-in-diff average treatment effects on the treated after matching with Mahalanobis distances)

	Number of years since migration						Cumulative income gain after six years (percent of base income)
	1	2	3	4	5	6	
All formal workers	**0.046**	**0.019**	**0.009**	**0.006**	*0.005*	*0.009*	21.3%
By sex							
Men	**0.053**	**0.019**	**0.008**	**0.006**	*0.004*	**0.012**	23.8%
Women	**0.030**	**0.018**	**0.011**	*0.006*	*0.005*	*0.001*	15.8%
By age group							
20 to 29 years	**0.050**	**0.012**	*0.005*	*0.002*	*0.003*	*0.008*	16.8%
30 to 39 years	**0.036**	**0.020**	**0.013**	**0.007**	*0.002*	*0.000*	15.3%
40 to 49 years	**0.051**	**0.021**	**0.012**	*0.007*	*0.008*	**0.020**	32.5%
50 to 54 years	*0.009*	*0.023*	*0.015*	*0.016*	**0.032**	*0.020*	46.6%
By initial income group (times the minimum wage)							
Less than 1.25	**0.066**	**0.037**	**0.024**	**0.020**	**0.019**	**0.023**	55.4%
Between 1.25 and 2.5	**0.022**	*0.001*	*-0.005*	**-0.007**	*-0.005*	*0.002*	-3.4%
Between 2.5 and 5	*0.037*	*0.007*	*0.010*	*0.008*	*-0.006*	*0.008*	14.4%
More than 5	*0.047*	*0.021*	*-0.004*	*-0.010*	*-0.001*	**-0.027**	-10.9%
By initial number of weeks in formal employment							
Less than 25	**0.069**	**0.028**	**0.013**	*0.004*	*0.006*	*0.014*	30.1%
Between 25 and 40	**0.035**	**0.012**	*0.008*	**0.010**	*0.006*	**0.018**	30.7%
More than 40	**0.045**	**0.019**	**0.009**	**0.005**	*0.004*	*0.005*	19.6%
By initial firm size group							
Less than 10 workers	**0.094**	**0.044**	*0.012*	**0.015**	**0.016**	*0.016*	47.9%
Between 10 and 25 workers	**0.111**	**0.038**	**0.027**	**0.020**	**0.020**	*0.010*	53.1%
Between 25 and 100 workers	**0.053**	**0.023**	**0.020**	**0.017**	**0.015**	**0.016**	41.7%
More than 100 workers	**0.026**	**0.010**	*0.005*	*0.001*	*-0.001*	*0.005*	9.0%

Source: calculations from Ministry of Health's PILA. Returns in bold characters are significatly different from zero with 95% confidence, those in italics are not.

Table 10. Yearly income gains after migration by cities of origin and destination, and years since migration (diff-in-diff average treatment effects on the treated after matching with Mahalanobis distances)

	Number of years since migration						Cumulative income gain after six years (percent of base income)
	1	2	3	4	5	6	
By cities of origin							
Three largest cities	**0.035**	**0.009**	*0.004*	*0.000*	*-0.001*	*0.004*	8.4%
Seven medium-size cities	**0.065**	**0.040**	**0.021**	**0.017**	**0.013**	**0.017**	46.5%
Remaining 52 small cities	**0.058**	**0.031**	**0.017**	**0.017**	**0.014**	**0.017**	42.5%
By cities of destination							
Three largest cities	**0.056**	**0.022**	**0.009**	**0.006**	*0.004*	**0.008**	22.1%
Seven medium-size cities	**0.047**	**0.016**	*0.001*	*-0.001*	*-0.002*	*0.002*	8.1%
Remaining 52 small cities	**0.031**	**0.014**	**0.015**	**0.012**	**0.011**	**0.017**	31.8%
By origin and destination							
Origin: three largest cities							
Destination:							
Three largest cities	**0.052**	**0.012**	*0.003*	*-0.002*	*-0.001*	*0.003*	9.4%
Seven medium-size cities	**0.037**	*0.007*	*0.000*	*-0.002*	*-0.002*	*-0.003*	1.3%
Remaining 52 small cities	*0.008*	*0.001*	*0.009*	*0.009*	*0.002*	*0.011*	15.0%
Origin: seven medium-size cities							
Destination:							
Three largest cities	**0.065**	**0.064**	**0.048**	**0.041**	**0.028**	**0.033**	90.6%
Seven medium-size cities	**0.052**	**0.019**	*-0.005*	*-0.005*	*-0.006*	*0.000*	2.4%
Remaining 52 small cities	**0.092**	**0.061**	**0.061**	**0.048**	**0.060**	**0.051**	134.6%
Origin: 52 small cities							
Destination:							
Three largest cities	**0.058**	**0.048**	**0.021**	**0.035**	**0.022**	**0.021**	62.2%
Seven medium-size cities	**0.070**	*0.037*	**0.038**	**0.024**	*0.023*	*0.025*	64.9%
Remaining 52 small cities	**0.053**	**0.018**	*0.010*	*0.005*	*0.008*	*0.013*	26.4%

Source: calculations from Ministry of Health's PILA. Returns in bold characters are significatly different from zero with 95% confidence, those in italics are not.

Table 10 presents the results by categories of cities of origin and destination. There is also a lot of heterogeneity in this respect. By origin, income changes are consistently positive and significant throughout the 6 years for mid-sized and small cities. By destination, they are so for the smaller cities. By origin-destination, they are consistently positive and significant until the sixth year in a few cases only: from medium to large cities, from medium to small cities and from small to large cities. *Cumulative* income gains at year 6 are vastly different across the origin-destination matrix. With a cumulative income gain of 134% the payoff of moving from a mid-sized city to a small city is quite remarkable compared with the average income change after migration (21% as shown in Table 9). At the other extreme, moving from a large to a medium-sized city is associated with a cumulative income gain of just 1%, which is not statistically significant.

10. Dedication to formal employment and real wage growth stability after migration

I use the same three methods (before matching, after MDM in levels and after MDM in differences) to study the extent to which migration is associated with changes in the other outcomes mentioned above, namely yearly weeks in formal employment and standard deviation of real wage growth. These are intended as measures of the risks that migration entails. Given the definition of these variables (see section 4), the time dimension t is not present.

Table 11. Changes after migration in weeks of formal employment and income variability (average treatment effects on the treated)

Outcome	Coefficient	AI Robust Std. Err.	z	P>|z|	[95% Conf. Interval]	
Before matching, absolute means (observations with complete information of matching variables)						
Yearly weeks in formal employment	-1.739	0.102	-17.09	0.000	-1.939	-1.540
Standard deviation of yearly wage growth since job change	0.004	0.001	2.41	0.008	0.001	0.006
After Mahalanobis distance matching, in levels						
Yearly weeks in formal employment	-1.624	0.132	-12.35	0.000	-1.882	-1.366
Standard deviation of yearly wage growth since job change	0.006	0.002	3.53	0.000	0.003	0.010
After Mahalanobis distance matching, diff-in-diff						
Yearly weeks in formal employment	-1.366	0.144	-9.5	0.000	-1.648	-1.084
Standard deviation of yearly wage growth since job change	0.000	0.002	-0.12	0.452	-0.004	0.003

Source: own calculations from Colombia's Ministry of Health's PILA.

Table 11 indicates that migrants suffer a reduction of more than one yearly week of dedication to their formal jobs. The reduction is consistently significant across the three methods. On average, migrants do not experience any substantial change in the stability of real wage growth (although the coefficient is significantly different from zero before matching and with MDM in levels, it completely vanishes in the diff-in-diff measure). The loss of dedication to formal employment is statistically significant across all the categorical groups considered, except in the case of migrants from small to medium-sized or large cities (see Tables 12 and 13). Real wage growth becomes more volatile for those migrants with initial incomes below 1.25 times the minimum wage and for those who were employed by firms of less than 15 workers before migrating. Recall that these same groups of migrants are the biggest winners in terms of income. Analogously, wage increases become less volatile after migration among workers with initial wages between 1.25 and 2.5 times the minimum wage and among those who were working in firms with more than 100 employees, which are among the groups with the smallest income gains. By cities of origin and destination, real wage growth changes become more volatile for migrants from medium-sized to large or medium-sized cities, and from small to medium-sized cities.

Table 12. Changes after migration in weeks in formal employment and standard deviation of wage growth by sex and categories of age, initial income, initial dedication and initial firm size
(diff-in-diff average treatment effects on the treated)

	Yearly weeks in formal employment	Standard deviation of real wage growth
All formal workers	**-1.366**	*0.000*
By sex		
Men	**-1.486**	*0.001*
Women	**-1.105**	*-0.003*
By age group		
20 to 29 years	**-1.361**	*0.000*
30 to 39 years	**-1.173**	*-0.004*
40 to 49 years	**-1.327**	*0.004*
50 to 54 years	**-2.474**	*0.012*
By initial income group (times the minimum wage)		
Less than 1.25	**-1.791**	**0.006**
Between 1.25 and 2.5	**-1.098**	**-0.005**
Between 2.5 and 5	**-0.855**	*-0.010*
More than 5	**-0.849**	*0.001*
By initial number of weeks in formal employment		
Less than 25	**-0.808**	**-0.010**
Between 25 and 40	**-1.040**	*-0.005*
More than 40	**-1.569**	*0.003*
By initial firm size group		
Less than 10 workers	**-1.144**	**0.013**
Between 10 and 25 workers	**-2.392**	**0.022**
Between 25 and 100 workers	**-1.744**	*0.006*
More than 100 workers	**-1.239**	**-0.006**

Source: calculations from Ministry of Health's PILA. Returns in bold characters are significatly different from zero with 95% confidence, those in italics are not.

Table 13. Changes after migration in weeks in formal employment and standard deviation of wage growth by city sizes of origin and destination (diff-in-diff average treatment effects on the treated)

	Yearly weeks in formal employment	Standard deviation of real wage growth
By cities of origin		
Three largest cities	**-1.342**	**-0.007**
Seven medium-size cities	**-1.956**	**0.018**
Remaining 52 small cities	**-0.792**	**0.009**
By cities of destination		
Three largest cities	**-1.062**	**-0.009**
Seven medium-size cities	**-2.174**	**0.017**
Remaining 52 small cities	**-1.423**	*0.006*
By origin and destination		
Origin: three largest cities		
Destination:		
Three largest cities	**-1.310**	**-0.014**
Seven medium-size cities	**-1.909**	*0.012*
Remaining 52 small cities	**-1.136**	*0.003*
Origin: seven medium-size cities		
Destination:		
Three largest cities	**-0.801**	**0.020**
Seven medium-size cities	**-2.829**	**0.017**
Remaining 52 small cities	**-1.474**	*0.015*
Origin: 52 small cities		
Destination:		
Three largest cities	*0.552*	*0.001*
Seven medium-size cities	*-0.145*	**0.042**
Remaining 52 small cities	**-1.914**	*0.007*

Source: calculations from Ministry of Health's PILA. Returns in bold characters are significatly different from zero with 95% confidence, those in italics are not.

11. Do income changes after migration correlate with migration probabilities?

If migration were a fully informed rational decision whose only reason were to maximize the individual's utility function, migration probabilities should correlate positively with the expected income gains and negatively with earnings risks. These assumptions may not be entirely valid: migration decisions face information limitations and biases (Banerjee and Duflo, 2019), and many other variables apart from pecuniary calculations may enter the decision because of the diversity of reasons why individuals move (Lucas, 2015).

In order to test these hypotheses, I calculate migration probabilities by groups that combine the categories used previously. With 2 sexes, 4 age categories, 4 initial income categories, 3 initial dedication categories, 4 initial firm size categories and 3 categories of cities of origin, there are 1,152 groups (12 of which have no observations). For each of these groups and each of the outcomes with information I compute the median diff-in-diff treatment effects. However, only 502 groups enter the regression on the first-year outcomes (and fewer the regressions on later outcomes) due to lack of observations to estimate their corresponding treatment effects.

Table 14 shows weighted least square regression results for all the groups with information (the weights are the total number of observations of the groups). I include as regressors the changes after migration in: (a) cumulative income (at the first, the third and the sixth years, depending on the regression), (b) yearly weeks in formal employment, (c) the standard deviation of wage growth and (d) the *within group* standard deviation of the corresponding incomes gains (at the first, the third and the sixth years, respectively). Regressor (a) is a measure of expected income *gain*, regressor (b) is a measure of expect job duration *gain*, regressor (c) in a measure of income *risk*, and regressor (d) is a measure on income *uncertainty.* In addition, all the regressions include a constant, a dummy for female groups and a set of dummies for the 4 age groups.

Table 14. Regressions of probability of migration on income gains and earning risks

Regression number	Explanatory variables: average treatment effects on the treated by groups	Coefficient	Std. Error	t	P>|t|	Adj. R-squared	Number of groups
1	Income gain **first year**	0.0217	0.0090	2.41	0.017	0.212	502
	Yearly weeks in formal employment	0.0006	0.0004	1.42	0.157		
	Standard deviation of real wage growth	-0.0341	0.0292	-1.17	0.243		
	Standard deviation of group's income gains until **first year**	-0.0245	0.0076	-3.22	0.034		
2	Income gain until **third year**	0.0042	0.0025	1.66	0.098	0.2261	407
	Yearly weeks in formal employment	0.0015	0.0005	3.04	0.003		
	Standard deviation of real wage growth	-0.0256	0.0389	-0.66	0.51		
	Standard deviation of group's income gains until **third year**	-0.0093	0.0029	-3.21	0.001		
3	Income gain until **sixth year**	0.0014	0.0011	1.21	0.227	0.2329	158
	Yearly weeks in formal employment	0.0025	0.0012	2.06	0.041		
	Standard deviation of real wage growth	0.0432	0.0732	0.59	0.555		
	Standard deviation of group's income gains until **sixth year**	-0.0037	0.0014	-2.58	0.011		

Source: own calculations from Colombia's Ministry of Health's PILA. The dependent variable is the median of yearly transition probabilities (between 2008 and 2016) from being employed in a formal firm to moving to a new firm in a different city. Migration probabilities are computed for groups of individuals by sex (2), age category (4), initial income category (4), initial dedication category (3), initial firm size category (4) and initial city size category (3), for a total of up to 1152 groups (however, a group may not have observations with complete data). The explanatory variables shown in the table are median values computed from the diff-in-diff treatment effects on the treated in each group. Each regression includes, in addition to the explanatory variables specified in the table, a constant, a dummy for female groups and a set of dummies for the age groups. The method of estimation is weighted least squares, using as weights the population in each group.

The regressions suggest that migration probabilities are directly related to expected gains and negatively related to income risk and uncertainty. Migration probabilities are not significantly correlated with wage variability as measured by the standard deviation of wage growth (alternative measures of earning risks were considered with similarly weak results). Although not shown in the table, the dummy of female groups and the set of dummies of age categories are always strongly significant.

Table 14. Regressions of probability of migration on income gains and earning risks

Regression number	Explanatory variables: average treatment effects on the treated by groups	Coefficient	Std. Error	t	P>|t|	Adj. R-squared	Number of groups
1	Income gain **first year**	0.0217	0.0090	2.41	0.017	0.212	502
	Yearly weeks in formal employment	0.0006	0.0004	1.42	0.157		
	Standard deviation of real wage growth	-0.0341	0.0292	-1.17	0.243		
	Standard deviation of group's income gains until **first year**	-0.0245	0.0076	-3.22	0.034		
2	Income gain until **third year**	0.0042	0.0025	1.66	0.098	0.2261	407
	Yearly weeks in formal employment	0.0015	0.0005	3.04	0.003		
	Standard deviation of real wage growth	-0.0256	0.0389	-0.66	0.51		
	Standard deviation of group's income gains until **third year**	-0.0093	0.0029	-3.21	0.001		
3	Income gain until **sixth year**	0.0014	0.0011	1.21	0.227	0.2329	158
	Yearly weeks in formal employment	0.0025	0.0012	2.06	0.041		
	Standard deviation of real wage growth	0.0432	0.0732	0.59	0.555		
	Standard deviation of group's income gains until **sixth year**	-0.0037	0.0014	-2.58	0.011		

Source: own calculations from Colombia's Ministry of Health's PILA. The dependent variable is the median of yearly transition probabilities (between 2008 and 2016) from being employed in a formal firm to moving to a new firm in a different city. Migration probabilities are computed for groups of individuals by sex (2), age category (4), initial income category (4), initial dedication category (3), initial firm size category (4) and initial city size category (3), for a total of up to 1152 groups (however, a group may not have observations with complete data). The explanatory variables shown in the table are median values computed from the diff-in-diff treatment effects on the treated in each group. Each regression includes, in addition to the explanatory variables specified in the table, a constant, a dummy for female groups and a set of dummies for the age groups. The method of estimation is weighted least squares, using as weights the population in each group.

The size of the coefficients indicate that migration probabilities are highly sensitive to expected income gains, especially short-run ones (according to regression 1, the relevant coefficient is 0.0217). One additional standard deviation (across groups) of first year expected income gains raises the average probability of migration of all groups by 0.54 pp, which is a moderate increase with respect to the base probability (5.4%, see Table 1). One additional standard deviation of uncertainty (as measured by the standard deviation of the same short-run income gains within the groups, whose coefficient is -0.0245) reduces the average probability of migration by 0.69 pp.

12. Conclusions

In this paper, I have used social security administration data for Colombia between 2008 and 2016 to estimate changes after inter-city migration of three labor outcomes: real wage growth, dedication to formal employment and real wage growth stability. To evaluate how inter-city migration is associated with each of these outcomes, I define treatment and control groups. The treatment group is the set of workers that change job and city, while the control group is the set of workers that change job but do not change city. Although I make no claim of causality, in order to reduce potential biases, I balance the treatment and control groups over a wide range of covariates (using Mahalanobis distances). I compute differences-in-differences treatment effects on the treated up to 6 years after they migrate. For the whole set of treated individuals migration is significantly associated with higher annual real wage growth, but only until the fourth year since the change of jobs, suggesting that, on average, income changes after migration are transient.

However, there is a lot of heterogeneity across groups of workers by sex, age, initial income level, initial formal work dedication, initial firm size, and by the size of the cities of origin and destination. The largest cumulative income changes after migration (as percent of initial income) accrue to workers with initial wages below 1.25 times the minimum wage, to workers initially employed by firms with fewer than 100 employees, and to workers who move from medium or small cities. Several groups of migrant workers experience on average negligible (i.e. less than 10 percent their initial income) or even negative cumulative income gains 6 years after migrating. These include workers with initial incomes between 1.25 and 2.5 or more than 5 times the minimum wage, those who were working in firms with more than 100 employees, and those who migrate from one of the 3 largest cities or to one of the 7 medium-sized cities.

Migration is associated with changes in other observable labor outcomes apart from income amounts. For almost every group considered, the number of yearly weeks in formal employment after the initial change of job is reduced for migrants (the only exception being the group of migrants from small to large cities). The relationship between migration and the variability of inter-annual wage changes is more differentiated by groups and nil on average.

Migration probabilities (by groups that combine the categories of sex, age, initial income, initial formal work dedication, initial firm size and type of city of origin) are directly and significantly correlated with the expected short-run income gains that migration can bring on average to the migrants of each group, but negatively and significantly correlated with the uncertainty of such gains in the short- and mid-run.

Acknowledgments

Valuable suggestions from Matias Busso, Johanna Fajardo, Arlén Guarín, Carlos Alberto Medina, Leonardo Morales, Neave O'Clery, Christian Posso, Rafael Prieto Curiel and Mauricio Quiñones Domínguez are acknowledged. This paper also benefitted from many comments by other participants at seminars held at Banco de la República (Medellín), Inter-American Development Bank (Washington, DC), Universidad del Norte (Barranquilla) and Universidad Eafit (Medellín).

Competing interests: no competing interests to declare.
Funding: This research was part of the "Urban PEAK" Program, with the valuable support of *UKRI's Global Challenge, Research Fund* (Grant reference: ES/P011055/1).
Availability of data: the micro (administrative) database (*PILA*) was provided by the Colombian Ministry of Health for exclusive use in the "Urban Peak" Program and cannot be transferred or used for any other purpose. Aggregated variables are available upon request.
Ethics approval and consent to participate do not apply.

References

Banerjee, Abhijit V. and Esther Duflo. *Good Economics for Hard Times.* Hachette Book Group Inc., Public Affairs, New York (2019): chapter 2: "From the mouth of the shark".

Bartel, Ann P. "The migration decision: what role does job mobility play?" *The American Economic Review* 69, no. 5 (1979): 775-786.

Bell, Martin, Elin Charles-Edwards, Philipp Ueffing, John Stillwell, Marek Kupiszewski, and Dorota Kupiszewska. "Internal migration and development: comparing migration intensities around the world." *Population and Development Review* 41, no. 1 (2015): 33-58.

Borjas, George J. "Economic theory and international migration." *International migration review* 23, no. 3 (1989): 457-485.

Borjas, George J., Stephen G. Bronars, and Stephen J. Trejo. "Self-selection and internal migration in the United States." *Journal of urban Economics* 32, no. 2 (1992): 159-185.

Castles, Stephen, Hein De Haas, and Mark J. Miller. *The age of migration: International population movements in the modern world*. Macmillan International Higher Education. (2013).

Duranton, Gilles. "Delineating metropolitan areas: Measuring spatial labor market networks through commuting patterns." In *the Economics of interfirm Networks*, pp. 107-133. Springer, Tokyo, 2015.

Fehn, Rebecca, and Hanna Frings. "Decomposing the Returns to Regional Mobility." (2018). Econstor.

Harris, John R., and Michael P. Todaro. "Migration, unemployment and development: a two-sector analysis." *The American economic review* 60, no. 1 (1970): 126-142.

Hunt, Janet C., and James B. Kau. "Migration and wage growth: a human capital approach." *Southern Economic Journal* (1985): 697-710.

Iacus, Stefano M., Gary King, and Giuseppe Porro. "Causal inference without balance checking: Coarsened exact matching." *Political analysis* 20, no. 1 (2012): 1-24.

King, Gary, and Richard Nielsen. "Why propensity scores should not be used for matching." *Political Analysis* 27, no. 4 (2019): 435-454.

King, Gary, Richard Nielsen, Carter Coberley, James E. Pope, and Aaron Wells. "Comparative effectiveness of matching methods for causal inference." *Unpublished manuscript* 15 (2011): 41.

Leuven, Edwin, and Barbara Sianesi. "PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing." (2003).

Lucas, Robert EB. "Internal migration in developing economies: An overview." *KNOMAD's Working Paper* 6 (2015).

Nakosteen, Robert A., and Michael Zimmer. "Migration and income: the question of self-selection." *Southern Economic Journal* (1980): 840-851.

Polachek, Solomon, and Francis Horvath. "The life cycle hypothesis approach to migration." *Research in labor economics* (1977): 103-149.

Prieto Curiel, Rafael, Mauricio Quiñones Domínguez, Neave O'Clery and Eduardo Lora. "Modelling internal migration on a complex network". RiSE, Universidad EAFIT (2020).

Rees, Philip, Martin Bell, Marek Kupiszewski, Dorota Kupiszewski, Philipp Ueffing, Aude Bernard, Elin Charles-Edwards and John Stullwell. "The impact of internal migration on population redistribution: an international comparison", *Population, Space and Place*, Vol. 23, issue 6 (2016). DOI: 10.1002/psp.2036

Stark, Oded, and David Levhari. "On migration and risk in LDCs." *Economic development and cultural change* 31, no. 1 (1982): 191-196.

Tavares, Marisa, Anabea Carneiro and José Verajão. "The spatial dimension of internal labor markets". *Journal of Regional Science*, 58 (2018): 181-123.

Todaro, Michael. "Internal migration in developing countries: a survey." In *Population and economic change in developing countries*, pp. 361-402. University of Chicago Press, 1980.

Yankow, Jeffrey J. "The wage dynamics of internal migration within the United States." *Eastern Economic Journal* 25, no. 3 (1999): 265-278.

Yankow, Jeffrey J. "Migration, job change, and wage growth: a new perspective on the pecuniary return to geographic mobility." *Journal of Regional Science* 43, no. 3 (2003): 483-516.

Zelinsky, Wilbur. "The hypothesis of the mobility transition." *Geographical review* (1971): 219-249.

Appendix. Computing a Consumer Price Index by city and wage group

The official Consumer Price Index produced by the National Statistical Office (DANE) provides information about price *changes,* but not about prices levels, for 23 cities. Comparing price *levels* across cities and across social strata would require information on the composition of the basket of goods by stratum in each city, as well as their corresponding prices. Since this is not possible, I compute a CPI that *within each stratum* differentiates across cities. I assume that housing rents are the only price that differs across strata *within each city*.

Housing rents by stratum and city are computed from DANE's National Survey of Household Budgets of July 2016-July 2017 in the following way. First, I classify all the households whose main provider is a formal worker (who contributes to the social security system) in four strata, based on the income of the main provider (the groups are the same I use throughout the paper: less than 1.25 times the minimum wage, between 1.25 and 2.5, between 2.5 and 5, and more than 5 times the minimum wage). Second, with the data of the households that pay rent, I compute the average rent paid by household in each stratum *s* in each city *c* ($rent_{s,c,2016}$), as well as the share of that expenditure in "total current monetary expenditure" (i.e. consumption) of the household ($share_{s,c,2016}$). Third, using Bogotá as 100 for 2016, for each stratum and each city, I compute the corresponding CPI as:

$$CPI_{s,c,2016} = \left(rent_{s,c,2016}/rent_{s,Bogotá,2016}\right) \times share_{s,c,2016} + (1 - share_{s,c,2016})$$

Finally, I compute the CPI by stratum and city by year between 2008 and 2016 as:

$$CPI_{s,c,t} = CPI_{s,c,2016} * CPI_{c,t}/CPI_{c,2016}$$

where the series $CPI_{c,t}$ is the official CPI for city *c.*

Since the official CPI covers only 23 cities, I use the average of the 13 smallest cities to calculate the CPI of the remaining 39 cities in my database. I follow a similar approach to compute the rent and the share variables for cities not represented in the National Survey of Household Budgets.

Notice that the CPI thus computed allows me to deflate the wages of each individual depending on the city where he/she lives, once the individual is classified in one of the strata according to his/her wage at the base year. It would not allow me to compare real incomes across individuals that belong to different strata, but this is not required to perform the rest of the calculations in the paper.