The Urban Wage Premium Over the Life Cycle and the Big-city Job Ladder*

Thiago Patto†

November 16, 2024

Abstract

I use Brazilian matched employer-employee data to study how city size impacts wage and job profiles, focusing on the influence on wage growth (dynamic effects). I show that dynamic effects fade with age and are concentrated on young college graduates, explaining about 30% of the observed city-size wage gap and 40% of the city-size college premium gap. In contrast, sorting on unobserved characteristics within education groups is of minor importance for understanding differences in wage levels and wage dynamics. I also study the relationship between dynamic effects and firm changes. I find that city size influences wage growth within and between firms, with the latter accounting for about 40% of the total effects experienced by young college graduates. Regarding this group, wage growth effects are stronger for multiple-firm career paths, and the quantity and quality of individuals pursuing these career paths increase with city size.

Keywords: Urban wage premium, Dynamic effects, College premium, Job Ladder

JEL Codes: R12, J31, J61

*I would like to thank Rodrigo Soares, Jorge De la Roca, Daniel da Mata, Rogerio Santarrosa, Fernando Ferreira, Mariaflavia Harari, David Card, Bruno Barsanetti, Gilles Duranton, Fabian Eckert and many seminar participants for helpful comments and suggestions. I also thank the São Paulo State Research Foundation (FAPESP) for their financial support (fellowship processes 2019/26148-0 and 2020/16601-7). All errors remain my own.

†Fundação Getúlio Vargas (FGV), São Paulo, *thiago.patto@fgv.br*

1 Introduction

The relationship between wages and city size is one of the central topics in urban economics, with numerous studies quantifying the urban wage premium.[1] More recently, the availability of longitudinal microdata gave rise to a literature showing that big cities affect not only wage levels but also wage dynamics (Glaeser and Maré, 2001; Baum-Snow and Pavan, 2012; De La Roca and Puga, 2017) — a finding that carries potential implications for our understanding of spatial sorting and migration.[2]

However, despite evidence of its existence, data constraints have prevented a comprehensive description of this phenomenon. For example, it is unclear how the effects of city size on wage growth evolve throughout working life. Given that the typical wage profile is steeper in the first years and flattens at some point (Lagakos et al., 2018), does it mean that city size effects follow a similar trend? Additionally, it remains an open question how differences in individual wage profiles translate into differences in job ladders. While some studies document distinct job mobility patterns in large cities (Wheeler, 2008; Bleakley and Lin, 2012), there is scant evidence regarding its influence on wage growth. This issue is especially relevant as it may provide some information on the mechanisms of agglomeration economies and the specificities of thick local labor markets.

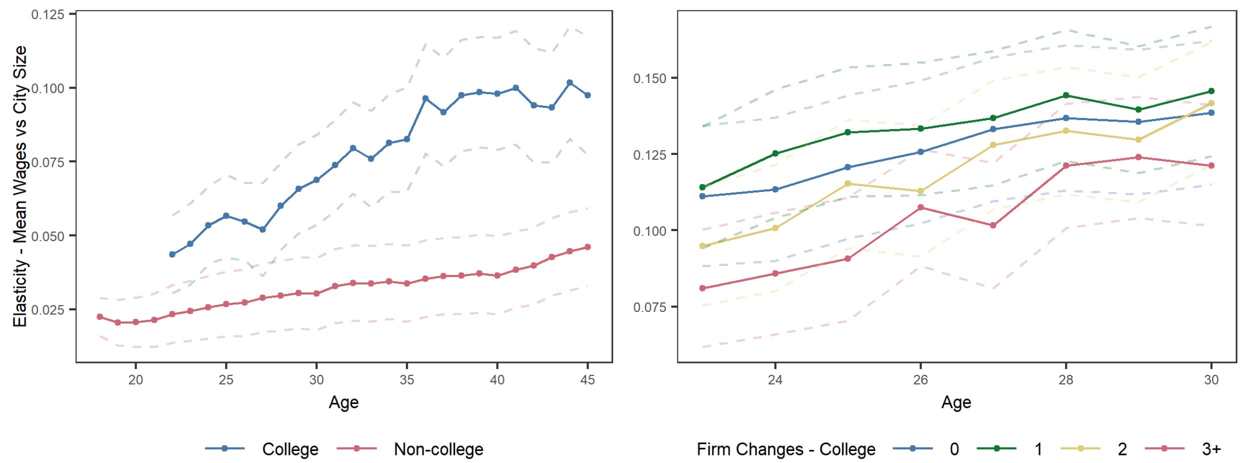
In this paper, I explore a large matched employer-employee database to study how city size shapes the wage and job profiles of college and non-college graduates in Brazil. From a dataset with 158 million observations across 185 urban areas, I perform an empirical analysis that fully leverages differences at the city level while considering heterogeneities over working life. In particular, I examine how city size dynamic effects (henceforth simply dynamic effects or dynamic gains) vary with age and their implications for workers' career paths, with a special focus on the role of firm changes.

A preliminary analysis of the data reveals some interesting facts. The left panel of Figure

[1] See Combes and Gobillon (2015) and Ahfeldt and Pietrostefani (2019) for related surveys.

[2] Papers that discuss this subject include Bilal and Rossi-Hansberg (2021) and De La Roca et al. (2022).

Figure 1: City size elasticity by educational level and age



Notes: Each dot informs the coefficient of a regression of log mean wages on log city size for a specific educational-age group. Data comes from RAIS using the sample described in Section 2. The measure of city size is described in the same section. The dashed lines inform the 95% confidence intervals.

1 plots the coefficients obtained from regressions of log mean wages on log city size for various educational-age groups. Notably, the impact of city size on wage growth appears to be more pronounced for college graduates, with a widening city-size wage gap within this group between the ages of 22 and 35. Over this period, the ratio of elasticities between college and non-college graduates increases roughly from two to three, indicating that differences in wage dynamics are also important for understanding why the college premium is higher in large cities. In Figure D1, I present a similar exercise using data from the American Community Survey (ACS) to show that this pattern is also present in the United States and is not an exclusive feature of Brazil.[3]

The right panel of Figure 1 narrows in on young college graduates, examining wage growth patterns across different career paths for this specific group. For a sample of non-migrants aged 23 to 30, I regress again log mean wages on log city size by age, this time grouping individuals based on the number of firms they worked for during this period. Despite the

[3]There are two methodological differences between the left panel of Figure 1 and Figure D1. First, the ACS does not have enough observations to compute mean wages for each educational-age group by Metropolitan Statistical Area (MSA), so I directly regress individual wages on log population and year dummies. Secondly, Figure D1 uses the MSA log population to measure city size, whereas Figure 1 uses the ”experienced density” concept explained in Section 2.2.

magnitude of the standard errors, differences in the levels and slopes of these curves suggest that: (i) dynamic gains depend on how frequently individuals move between firms, and (ii) city size influences how individuals sort into different career paths. More broadly, it raises the question of whether, in equilibrium, the incentives for job-switching depend on the size of the local labor market, thus shaping the average job ladder in a given location. As a possible implication, two identical individuals working in identical firms may end up with different long-term wages if one has access to more potential employers, even if the firms are highly productive. In other words, city size could promote (or constrain) wage growth through labor market opportunities.

To delve deeper into the issues discussed above, I propose a Mincerian reduced-form approach based on city-specific returns to experience (D'Costa and Overman, 2014; De La Roca and Puga, 2017). Motivated by Figure 1, I allow these returns to vary by age, potentially capturing decaying dynamic effects. The model also includes a standard city static premium and two dimensions of individual heterogeneity that account for the possibility of unobserved worker characteristics affecting both wage levels and wage dynamics.

The empirical analysis starts by estimating the effects of city size on wage growth, measured as the annual variation in wages. It consists of regressing separately, for college and non-college graduates, log wage growth on city-age fixed effects, worker fixed effects and time-varying controls, and then regressing the estimated city-age fixed effects on city size for each age group. This approach yields multiple city-size elasticities that characterize how dynamic effects vary over working life for the two educational groups.[4]

The results show that city size significantly affects the wage growth of young college graduates. There is a positive relationship between city-age fixed effects and city size that fades with age and is especially salient for this educational group. The point estimates suggest they benefit from dynamic gains up to 30 years old, although the magnitude of the standard errors hampers a precise statement about when the effects cease to exist. I also

[4]Following D'Costa and Overman (2014), I perform the estimation using only within-city observations.

estimate a simpler specification with only two age groups - up to 30 years old and above -, and find highly significant city size effects for the younger group. Finally, I show that mean worker fixed effects are weakly correlated with city size, suggesting that selection based on unobserved characteristics is unlikely to contribute to the accelerated wage growth observed in big cities.

In the next step, I evaluate the contribution of dynamic effects to the observed city-size wage gap. Drawing from the conceptual framework proposed, I derive an equation for estimation that resembles the standard static model of the urban wage premium (Glaeser and Maré, 2001; Combes et al., 2008), but in which the dependent variable is the log wage net of all previously accumulated city-specific returns to experience. I calibrate the city-specific parameters using estimates from the wage growth equation. In this approach, the importance of dynamic gains can be assessed by examining how turning these parameters on and off affects estimates of the static urban wage premium and the degree of sorting within educational groups.

I find that subtracting out dynamic effects does not greatly influence the estimated static urban wage premium but significantly affects the degree of sorting on unobserved characteristics among college graduates. For this group, the positive relationship between worker's average fixed effect and city size obtained from a standard static specification disappears once dynamic effects are taken into account. This result indicates that the higher skill level of college graduates in large cities (as measured by their individual fixed effects) is due to large cities "manufacturing" more skilled workers by offering them more valuable experience. From this variation, I estimate that dynamic effects account for about 30% of the college city-size wage gap and 40% of the city-size college premium gap.[5]

In the second part of the paper, I study how dynamic gains impact the job ladder in large cities, beginning with a decomposition of "within-firm" and "between-firm" effects.

[5]This result is in line with De La Roca and Puga (2017), although their analysis does not distinguish between college and non-college graduates. Instead, they examine heterogeneity in dynamic effects based on the individual fixed effects and show that the former increases with unobserved ability.

To do so, I chronologically enumerate workers' firms and modify the wage growth equation described above to incorporate city fixed effects for each progression type (e.g., within the first firm, between the first and second firm, within the second firm, and so on).

The results show that college graduates experience faster wage growth both within and between firms. For those entering the labor market, doubling the city size increases wage growth by 0.9% yearly within the first firm and by 1.1% in the first firm change. Subsequent progressions are also affected, though to a lesser degree. I estimate that between-firm dynamic effects account for about 41% of the predicted gains of college graduates aged between 23 and 30. For non-college graduates between 19 and 26 years old, this fraction is 47%, albeit dynamic effects for this group are significantly lower in general. Interestingly, the higher number of firm changes in big cities enhances the "between" component for both educational groups. I estimate that, within the age intervals mentioned above, the elasticities of firm changes with respect to city size are 0.12 and 0.11 for college and non-college graduates, respectively.

To shed further light on this issue, I also investigate whether firm changes influence dynamic gains in the long run and, if so, whether this correlates with workers' behavior in the labor market. For this purpose, I compare the wage growth of non-migrant college graduates between 23 and 30 years old across different cities and career paths, defined by the number of firm changes observed during this period. In the same way, I also compare the fraction of workers and their composition, summarized by the individual fixed effects estimated in the first part of the paper.

I find that the elasticity of long-term wage growth with respect to city size increases with the number of firm changes, though this relationship is unlikely to be monotonic. Changes in occupation and sector amplify these effects for some career paths and attenuate them for others but do not alter the overall ranking. Moreover, city size influences the composition of workers in each career path. A greater proportion of workers in big cities follow multiple-firm career paths, and these individuals tend to have higher returns to experience.

In summary, this paper argues that, for college graduates, there is an additional benefit to working in big cities that accumulates over time and is closely related to how they move between firms. As suggested by Figure 1, city size affects the wage growth of young college graduates and largely explains key patterns in the data, such as the wage gap and college premium gap by city size. Furthermore, young graduates in big cities experience steeper job ladders in their first years of working life, with more rungs and wider gaps between them. These individuals move more frequently across firms to fully capitalize on the opportunities these locations offer, particularly those who are highly skilled. Overall, the findings in this paper underscore the functioning of thick local labor markets as a crucial element in understanding how worker productivity develops over time in big cities.

This work adds to the literature on wage growth in cities. Starting with Glaeser and Maré (2001), the identification of dynamic agglomeration effects has relied on reduced form estimations (D'Costa and Overman, 2014; De La Roca and Puga, 2017; Frings and Kamb, 2022), structural approaches (Baum-Snow and Pavan, 2012; Martellini, 2021) and more recently on quasi-experimental settings (Eckert et al., 2020). My methodology belongs to the first group but relies on a large longitudinal dataset that allows a more accurate and detailed description of how wage profiles vary with city size.[6] In particular, I extend the approaches of D'Costa and Overman (2014) and De La Roca and Puga (2017) to highlight educational- and age-based heterogeneities and to study differences in the job ladder.

A small group of papers studies firm-worker matching in cities. Some focus on assortative matching (Andersson et al., 2007; Figueiredo et al., 2014; Dauth et al., 2022). Wheeler (2008) and Bleakley and Lin (2012) find that young workers are more likely to change occupations or industries in big cities, but this pattern is reversed for older workers. Wheeler (2006) and Matano and Naticchioni (2016) investigate the relationship between wage growth, job

[6] The dataset size allows me to estimate dynamic effects using a two-step approach, which has been standard in the literature only for static premiums. For instance, De La Roca and Puga (2017) focus on differences in wage dynamics between specific groups of cities separated by size, namely Madrid and Barcelona as one group, Valencia, Sevilla and Zaragoza as another, and the remainder as the "control" group. Baum-Snow and Pavan (2012) separate MSAs into three groups based on size.

transitions and city size, which is one of the goals of this paper. Using a richer data source, I extend their work by exploring heterogeneous effects throughout the life cycle, focusing on both short- and long-run benefits of firm changes.

This paper also relates to studies that discuss spatial differences in the college premium (Moretti, 2013; Lindley and Machin, 2014; Davis and Dingel, 2019). It is widely documented that college premium is larger in big cities. This paper is the first to associate the city-size college premium gap with an asymmetry in how city size affects the life cycle wage dynamics of each educational group.

Finally, I contribute to the literature on agglomeration economies in developing countries (Chauvin et al., 2017; Dingel et al., 2019; Bryan et al., 2019; Combes et al., 2019). While urbanization is a major process in these regions, the articles dedicated to the relationship between wages and cities usually focus on particular problems, such as confronting wages in urban vs. rural areas and agricultural vs. non-agricultural sectors (Hicks et al., 2017; Alvarez, 2020), or the influence of slums on earnings (Marx et al., 2013). The scarcity of panel data makes it more difficult to properly evaluate the effects of city size on wages, especially dynamic effects. I argue that, despite the particularities, the conclusions here can help understand how city size affects wages in developed economies and it can provide a new perspective on previous empirical studies.

The rest of the paper is structured as follows. The next section presents the data, the empirical framework and preliminary estimations. Section 3 presents estimates of dynamic effects and evaluates their importance in explaining patterns in the data. Section 4 investigates how city size affects job and wage ladders. Finally, Section 5 concludes.

2 Data and Methods

2.1 Matched Employer-Employee Data

The analysis uses earnings records from *Relação Anual de Informações Sociais* (RAIS), an administrative matched employer-employee dataset provided by the Brazilian Ministry of Labor. It contains job records on most formal workers in the country, except for domestic workers and other minor employment categories. I use data between 2009 and 2017 for most estimations, sometimes supplementing with data from 2003 to 2008.

The dataset has invariant identifiers of workers and firms that allow us to track them over time. Other information includes job characteristics like monthly earnings, weekly contracted hours, start and end date of the employment spell, occupation and sector, and worker attributes such as age, gender and educational attainment. The wage variable constructed for this analysis is the average monthly earnings adjusted to a 44-hour workweek, and the occupation and sector information are aggregated to one-digit and two-digit levels, respectively.

The study focuses on two educational groups: workers with a college degree or more and workers with a high school degree or less. In RAIS, the infTIn order to simplify the comparison across groups while minimizing measurement errors, I ignore these variations and assign one fixed classification for each worker based on the maximum value observed between 2003 and 2017.[7]

The sample comprises male workers from 2009 to 2017 aged between 18 and 45 among non-college graduates and between 22 and 45 among those with higher education. I construct a yearly panel with one record per worker-year, selecting the job with the highest tenure - or the one with the highest wage in the case of a tie - if there are multiple records for the same worker in a given year. Observations from the public sector were left out to avoid concerns regarding the wage setting of this group, which could confound the conclusions. I also apply

[7]In Online Appendix A.5, I show that the results hold if I use the most frequent value instead.

Table 1: Summary Statistics

Variable	Overall			By city		
	All Workers	College	Non-college	Median	Min	Max
	(1)	(2)	(3)	(4)	(5)	(6)
Observations ('000)	158, 750	25, 839	132, 911	259	7	32, 018
College (%)	16.3	0	0	11.7	4.6	40.3
Non-college (%)	83.7	0	0	88.3	59.7	95.4
Avg. age (years)	30.7	31.7	30.5	30.4	29.2	31.6
Up to 30 yrs old (%)	51.6	46.1	52.6	53.4	46.5	61.3
More than 30 yrs old (%)	48.4	53.9	47.4	46.6	38.7	53.5
Non-movers (%)	75.8	72.2	76.6	72.2	50.8	88
Migrants (%)	24.2	27.8	23.4	27.8	12	49.2
Avg. Wage (R$)	2, 406	5, 284	1, 847	1, 913	1, 203	5, 720
Avg. Wage - Non-movers (R$)	2, 347	5, 202	1, 824	1, 859	1, 140	6, 047
Avg. Wage - Migrants (R$)	2, 593	5, 495	1, 923	2, 042	1, 352	5, 380
(Within-city) Job Changes (%)	16.4	13.8	16.9	13.6	3.2	19.2
Job Changes - Non-movers (%)	17.2	14.7	17.6	14.1	3.5	20
Job Changes - Migrants (%)	13.9	11.3	14.5	12.7	2	18.6
# Workers ('000)	33, 131	4, 702	28, 429	72	3	7, 538

Notes: This table exhibits summary statistics of the RAIS sample used in the analysis. Wages are in 2010 reais.

other minor filters to exclude observations with invalid worker identifiers or some missing information.

Table 1 shows summary statistics of the sample, comprising approximately 158.8 million observations from 33.1 million workers. Those with a college degree have a lower probability of dropping from the dataset due to unemployment or transition to informal work, as they represent about 14% of the pool of individuals but 16.3% of the observations. Higher education seems to considerably affect wages since college graduates are paid more than twice as much as non-college graduates. It is noteworthy that migrants, defined as workers who are observed at least in two different urban areas, tend to have higher levels of education, earn higher wages, and experience fewer within-city job changes compared to non-movers.

2.2 Urban Areas

The definition of urban areas comes from the Brazilian Institute of Geography and Statistics (IBGE), which grouped Brazilian municipalities based on flows to work and school and the contiguity of urban spots (IBGE, 2016a).[8] My study focuses on *Urban Concentrations* (UC), which consist of 185 urban areas with more than 100,000 inhabitants. I deviate slightly from the original IBGE definition to guarantee time-consistent urban areas throughout the period analyzed.[9] Figure 2 shows their location in the Brazilian territory.

According to the 2010 Census, the UCs have a population of 110.5 million inhabitants (60% of Brazil's population). The map shows that urbanization is highly unequal and concentrated in the south and southeast regions, which contain 117 UCs and 74 million inhabitants. The largest urban area is São Paulo, with 37 municipalities and 19.4 million inhabitants. On the other hand, north and center-west regions have only 31 UCs, most of which include the state capital.

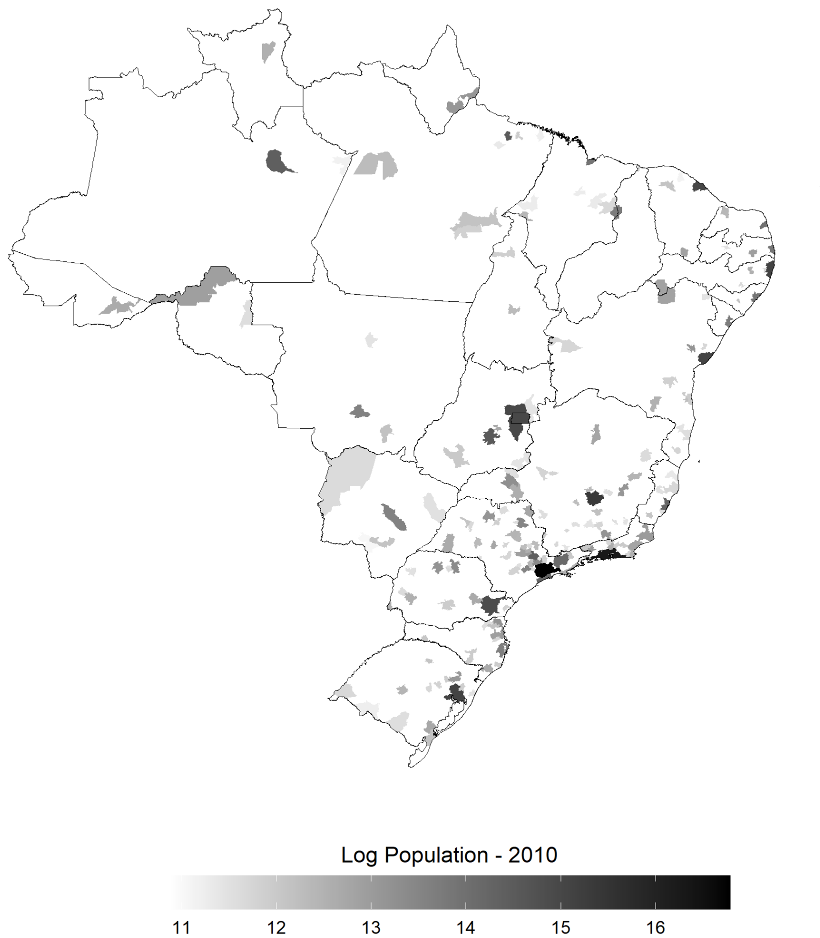
City size is measured by computing the population within a radius of 10 km of the average individual. For this purpose, I use the 1-km-resolution population grid provided by IBGE (2016b). The procedure to construct the city size measure is the following: for each cell located in a given city, I trace a 10 km radius circle around the cell, count the population within that circle, and take the average over all cells of that city weighting by the cell's population. The constructed measure has a correlation of 0.82 with a simple population count.[10]

[8] In this study, IBGE uses Census data on commutes to work and school and Google Earth imagery to identify urban spots.

[9] Some non-UC municipalities were created recently from UC municipalities, so I included them in the respective UC. Only six urban areas were affected by this procedure, and only one had its population increased by more than 10%.

[10] This procedure has the advantage of using the concept of density (more suitable to capture the notion of agglomeration) without relying on municipal boundaries, which can be very different from the actual urban spot. Moreover, it aims to express the level of density perceived by individuals. For a detailed discussion about the advantages of this measure, see Duranton and Puga (2020).

Figure 2: Urban Concentrations



2.3 Empirical Framework

I assume that the log wage of a worker i of educational group s (college or non-college) in any period t can be written as a sum of a city static premium Ψ_c^s and an individual term $h_{i,t}$:

$$\log w_{i,c,t}^s = \Psi_c^s + h_{i,t} \ , \tag{1}$$

where the subscript c represents a function $c(i,t)$ that maps worker i in period t to a specific city c. In Online Appendix E, I show how Equation (1) can be rationalized from a simple partial equilibrium model in which college and non-college labor are different inputs in the production function.[11]

To estimate the relevance of dynamic gains over the life cycle, I propose the following specification for $h_{i,t}$:

$$h^s_{i,t} = \alpha_i + \delta_i e_{it} + \sum_{\tau=1}^{t} \delta^s_{c(i,\tau),a(i,\tau)} + X_{i,t}\beta^s + \epsilon_{i,t} \ . \tag{2}$$

In this expression, α_i is the worker i fixed effect representing her initial unobserved ability. The term e_{it} is the experience that worker i has accumulated up to period t, which means that δ_i represents the individual heterogeneity in returns to experience. The summation term accounts for all for the city dynamic effects accumulated by worker i up to period t. Note that $\delta^s_{c(i,\tau),a(i,\tau)}$ is indexed by the function $a(i,\tau)$, which maps the age a of worker i in period τ. Finally, $X_{i,t}$ is a vector of time-varying individual and job characteristics and $\varepsilon_{i,t}$ is a worker-specific error term.

The combination of (1) and (2) delivers

$$\log w^s_{i,c,t} = \alpha_i + \delta_i e_{it} + \Psi^s_c + \sum_{\tau=1}^{t} \delta^s_{c(i,\tau),a(i,\tau)} + X_{i,t}\beta^s + \epsilon_{i,t} \ . \tag{3}$$

Equation (3) is the main expression of this study, offering several possible explanations to understand why wages are higher in larger cities. These alternatives are of two types. First, there are factors related to the sorting of more skilled workers into larger cities, represented by the terms α_i and δ_i.[12] Secondly, the coefficients Ψ^s_c and $\delta^s_{c,a}$ represent the city effects. While the former represents the "jump" in wages that workers face when moving from one city to another, the latter represents deviations in returns to experience that workers located in a given city jointly experience. An important particularity of the dynamic premium $\delta^s_{c,a}$ is that it is specifically related to the individual and, therefore, has a narrower interpretation

[11] The term Ψ^s_c is assumed to be constant over time for convenience. Given that my main focus is to estimate dynamic effects, variations in city premiums are not an important issue as long as they do not correlate with city size. I discuss this matter in Section 3.1, presenting some evidence that attenuates this concern.

[12] $X_{i,t}$ could also be included in this list, but it is not relevant to the context of this paper.

compared to the static premium (See Online Appendix E for details).[13,14]

2.4 The Urban Wage Premium - Static Model

Before considering Equation (3), it is worth estimating a simpler model to establish some basic facts and discuss possible sources of bias in my empirical strategy. For now I abstract from dynamic effects and individual returns to experience to perform a standard estimation of the urban wage premium

$$\log w_{i,t}^{s} = \alpha_i + \Psi_c^s + X_{i,t}\beta^s + \varepsilon_{i,t}. \tag{4}$$

The analysis of agglomeration patterns throughout this paper follows a two-step procedure in the spirit of Combes et al. (2008). In the context of this section, this means that the urban wage premium is obtained by regressing the estimated city fixed effects from (4) on city size:

$$\Psi_c^s = \gamma_0^s + \gamma^s \log CitySize_c + \nu_c, \tag{5}$$

where ν_c is an error at the city level. The term γ^s is the parameter of interest and represents the elasticity of the city wage premium with respect to city size for educational group s.[15]

This approach hinges on two key identification assumptions that will underpin most of

[13] Duranton and Puga (2004) provide a survey about agglomeration theory.

[14] As shown in Appendix E, Ψ_c^s also embeds the relative supply of college and non-college graduates in an environment with imperfect labor substitution. This relationship can be even more complex in the presence of spillovers from college education. A strand of the literature abstracts from this issue (Glaeser and Maré, 2001; Combes et al., 2008; De La Roca and Puga, 2017). Some papers that have this dimension of analysis include Moretti (2004), Bacolod et al. (2009), Eeckhout et al. (2014) and Combes et al. (2019).

[15] Combes et al. (2008) argue that a two-step approach is more appropriate than including city size directly in the wage equation because it allows distinguishing worker- from city-level shocks, which is useful when discussing identification. They also point to the fact that, in a single-step procedure, standard robust clustering is not enough to avoid biases in standard errors (see their paper for more details). Finally, the two-step procedure is helpful because it can be used to analyze other variables, e.g., regressing average α_i on city size. Doing so makes it easier to compare the contribution of each element to the city-size wage gap.

the estimations in this paper. First, worker migration must be exogenous conditional on the other variables included in Equation (4). While this might be a strong assumption, it allows me to perform a large scale analysis and fully explore city-level variations in the data. In Online Appendix A.1, I delve into this issue in more detail by developing an event study exercise adapted from Card et al. (2013). The result suggests that biases are likely to be small.[16]

Secondly, the estimation of γ^s may also be biased if city size is endogenously determined. For instance, if some city characteristics omitted in the regression simultaneously affect its size and productivity (e.g., location, access to other markets), or if there is reverse causality, i.e., if large cities are a result of high-wage areas attracting workers. To circumvent these concerns, I developed an estimation using instrumental variables presented in Online Appendix A.4.

Table 2 reports the results from Equation (5) for different specifications of Equation (4), whose results are reported in Table B1. In the first row, I estimate the city size elasticity for the non-college sample. Column (1) shows that including only city and year fixed effects yields a static urban wage premium of 0.0267, meaning that doubling the city size would increase the wage premium by 2.67%. This point estimate does not meaningfully change when I add age and tenure polynomials in Column (2) but decreases by about 30% when I include sector and occupation indicators in Column (3), indicating that sectoral composition explains the city size wage gap to some degree. In Columns (4) and (5), I add worker fixed effects to the specifications in Columns (2) and (3), respectively, with minor changes.

Workers with a college degree seem, in principle, to benefit considerably more from city size, as presented in the second row. Column (1) shows an initial static urban wage premium of 0.0607, more than twice what I found for non-college graduates. After including age and tenure in Column (2) and sector/occupation indicators in Column (3), I still obtain elasticities considerably higher than those in the first row. However, Columns (4) and (5) show

[16] Including of worker fixed effects in this setting means that Equation (3) delivers a unique solution only if all cities are connected through migrants. Given the dataset size, this condition is easily achieved.

Table 2: Static Urban Wage Premium - Standard Static Model

	City Static Premium (Ψ_c^s)				
	(1)	(2)	(3)	(4)	(5)
Panel A. Non-college					
Log City Size	0.0267*	0.0251	0.0180	0.0193**	0.0151*
	(0.0155)	(0.0156)	(0.0135)	(0.0092)	(0.0086)
R^2	0.01455	0.01303	0.00928	0.02231	0.01586
Observations	185	185	185	185	185
Panel B. College					
Log City Size	0.0607***	0.0597***	0.0535***	0.0235***	0.0263***
	(0.0198)	(0.0179)	(0.0138)	(0.0071)	(0.0067)
R^2	0.04299	0.05230	0.07348	0.04448	0.06527
Observations	185	185	185	185	185
First Step Variables					
Age, Tenure		✓	✓	✓	✓
Sec & occup indicators			✓		✓
Worker FE				✓	✓

Notes: The dependent variables are city fixed effects estimated from wage regressions according to Equation (4). Each column represents a different specification of the first-step estimation, reported in Table B1. Age and tenure variables include a cubic polynomial. Occupation and sector indicators refer to one-digit and two-digit level information, respectively. Coefficients are reported with robust standard errors in parenthesis. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels.

that adding worker fixed effects substantially reduces the estimated elasticity. Comparing the values in Columns (3) and (5), there is a 50% decrease in the static urban wage premium of college graduates and a 44% decrease in the college/non-college ratio.[17]

There is an ongoing debate in the literature about what fixed effects capture in these estimations. One hypothesis is related to differences in skill formation that arise before labor market entry. For instance, Bosquet and Overman (2019) argue that the intergenerational influence of more skilled parents partially explains the city-size wage gap. Another possibility is that returns to experience are higher in large cities, generating differences in skills that

[17] Since non-college workers represent more than 80% of the pool of individuals (see Table 1), it is possible to use the results of the first row of Table 2 to compare with previous studies that estimate the urban wage premium without splitting the sample by educational attainment. In general, my estimates are of lower magnitude and with lower statistical significance. For instance, Combes et al. (2010) estimates with and without worker fixed effects using French data are 0.051 and 0.033, respectively. De La Roca and Puga (2017)'s equivalent estimates using Spanish data are 0.046 and 0.024. All reported p-values are lower than 1%.

emerge within the labor market. Baum-Snow and Pavan (2012) and De La Roca and Puga (2017) show evidence that corroborates this point of view.

Within this discussion, two important findings emerge from Table 2. First, worker sorting matters only for college graduates. Secondly, the initial disparity in the urban wage premium between college and non-college graduates decreases substantially with the inclusion of fixed effects, indicating that this element is essential to understanding the gap in the urban wage premium between the two educational groups.

3 Dynamic Effects

3.1 Estimation

This section focuses on characterizing dynamic effects. I follow the procedure proposed by D'Costa and Overman (2014), which involves taking the first difference of Equation (3) and dropping observations in which a migration occurs, yielding

$$\Delta \log w_{i,t}^{s} = \delta_i + \delta_{c,a}^{s} + \Delta X_{i,t}\beta^{s} + \Delta\epsilon_{i,t} \ , \tag{6}$$

where $\delta_{c,a}^{s}$ is the dynamic premium of city c for a worker of educational level s and age a between two consecutive periods. Note that excluding migrations from the estimation greatly simplifies the expression.

The term δ_i deserves some discussion. D'Costa and Overman (2014) hypothesize that faster growth in large cities is due to workers sorting based on unobservable characteristics that influence wage dynamics. They show that individual fixed effects attenuate estimates of dynamic effects, so I include this term in the econometric model to account for this possibility.

The estimation of city size effects is performed in two steps, regressing $\delta_{c,a}^{s}$ on log city size separately for each educational-age group:

$$\delta^s_{c,a} = \pi^s_0 + \pi^s_a \log CitySize_c + u_c \ , \tag{7}$$

where u_c is the error term. Now, the parameters of interest are π^s_a, which represent the elasticities of the city dynamic premium with respect to city size for a given educational-age group {s,a}.[18]

Besides delivering a simpler equation, taking the first difference is convenient for other two reasons. First, it releases me from the necessity of knowing the workers' full location history to estimate dynamic effects. Second, it also circumvents the question of whether the experience acquired in one city is uniformly evaluated across locations, which I implicitly assume in Equation (3). The estimation of $\delta^s_{c,a}$ from Equation (6) identifies, strictly speaking, the returns to experience acquired and used in the same city c. On the other hand, identifying dynamic effects using Equation (3) would depend on an extra assumption.[19]

Naturally, migrants are essential to uniquely identify dynamic effects since Equation (6) includes worker fixed effects. It is worth highlighting that I do not exclude migrants from the estimation. If I observe the same in two different cities during the period analyzed, I discard only the observation in which the migration occurs.[20]

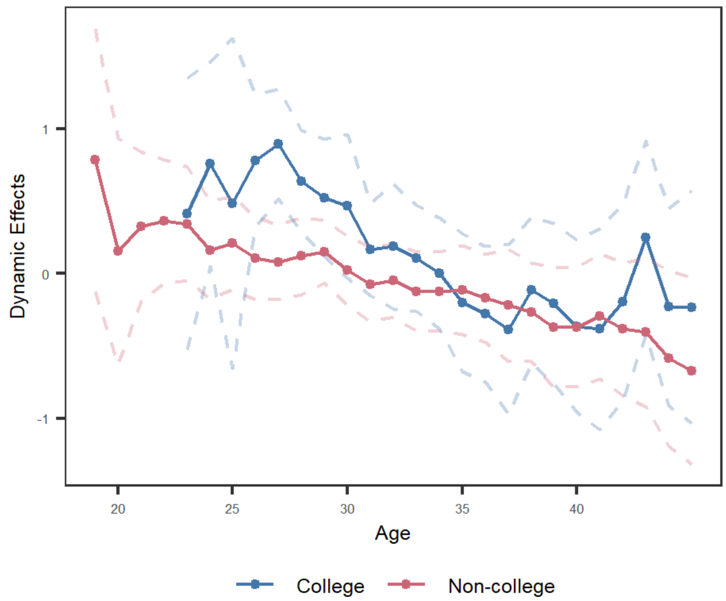
Finally, it might be a concern that π^s_a captures not only dynamic effects but also wage variations at the city level caused by a local shock (e.g., the construction of an airport that increases local productivity). If these shocks correlate with city size, identification may be harmed. To investigate this possibility, I regress annual changes in city-level log outcomes on

[18] The inclusion of city-age fixed effects prevents the use of age controls, so they were excluded from the estimation.

[19] Testing whether the experience acquired in one city has a different evaluation in other locations using Equation (3) would demand the estimation of more parameters. Without any restriction on how they can vary, it means 185 × 185 parameters to be estimated. De La Roca and Puga (2017) estimate the degree of experience portability across Spanish cities and conclude that spatial differences in evaluation are of minor importance. They find no evidence of differential evaluation of experience in big cities and a slight overvaluation of experience acquired in small cities when the worker is in a big city.

[20] The estimation of Equation (6) using migration observations yields very high estimates of dynamic effects because the parameters also capture changes in the static premium. These results were not included in the article but are available upon request.

Figure 3: Dynamic Effects by Age



Notes: Each dot represents the elasticity of log wage growth with respect to city size for a specific educational-age group. Control variables include a quadratic polynomial in tenure and worker fixed effects. The dashed lines inform robust 95% confidence intervals.

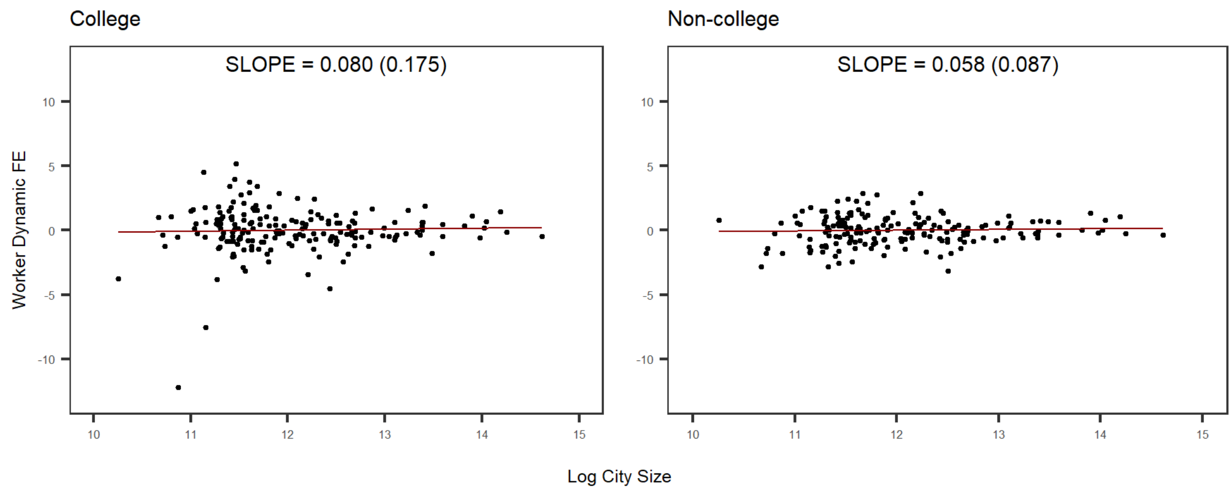
log city size. I construct two variables using data from RAIS: mean wages and employment. The results in Table C1 show that, if anything, city size negatively correlates with these variables, suggesting that estimates of dynamic effects are potentially underestimated.

Figure 3 reports the results from Equation (7). The results of the first-step estimation are presented in Table B2. For better exposition, the dependent variable was multiplied by 100. The figure shows that dynamic gains are decreasing in age and are positive in the first years, especially for college graduates. The point estimates suggest that city size influences workers' wage growth up to 30 years old.

To check if unobserved worker characteristics influence wage growth in large cities, Figure 4 plots the city-average worker fixed effects against log city size. The weak positive relationship for both educational groups indicates that sorting is unlikely to be a relevant factor.

Since the large standard errors in Figure 3 may raise doubts about the existence of

Figure 4: Worker fixed effects and city size - dynamic estimation



Notes: Each dot represents a mean worker fixed effect at the city level subtracted by the sample average. Estimates refer to Column (2) of Table 3. Coefficients are reported with robust standard errors in parenthesis.

dynamic effects, I also estimate a simpler version of Equation (6) that separates observations into two groups:

$$\Delta \log w^s_{i,t} = \delta_i + \delta^s_{c,young} + \delta^s_{c,senior} + \Delta X_{i,t}\beta^s + \Delta\epsilon^{'}_{i,t} \ , \tag{8}$$

where "young" and "senior" refer to workers up to 30 years old and workers over 30 years old, respectively. Then, I regress $\delta^s_{c,young}$ and $\delta^s_{c,senior}$ separately on log city size.

Table 3 exhibits the results for young workers, and the respective first-step results are displayed in Table B3. Column (1) shows no evidence of dynamic effects for non-college graduates, but after including worker fixed effects in column (2), the estimated elasticity jumps to 0.1465. In Column (3), I estimate Equation (6) excluding sector and occupation moves. This procedure aims to control for spatial differences in the frequency of sector/occupation changes, which might correlate with wage growth.[21] In this case, the point estimate reduces to 0.0387, suggesting that large cities make sector and occupation changes slightly more advantageous relative to small cities. Finally, column (4) shows that the estimate remains

[21] Wheeler (2008) and Bleakley and Lin (2012) show that city size increases the probability of a sector/occupation move among young workers but decreases the probability for older workers.

Table 3: Dynamic Effects - Young Workers

	City Dynamic Premium ($\delta^{s}_{c,young}$)			
	(1)	(2)	(3)	(4)
Panel A. Non-college				
Log City Size	-0.0489	0.1465	0.0860	0.0469
	(0.0776)	(0.1055)	(0.1171)	(0.1218)
R^2	0.00230	0.00816	0.00253	0.00074
Observations	185	185	185	185
Panel B. College				
Log City Size	0.4910***	0.5532***	0.4571***	0.4500***
	(0.0945)	(0.1934)	(0.1628)	(0.1621)
R^2	0.13574	0.05468	0.04088	0.04055
Observations	185	185	185	185
First Step Variables				
Age, Tenure	✓	✓	✓	✓
Worker FE		✓	✓	✓
Exclude sec & occup moves			✓	✓
Sec & occup indicators				✓

Notes: The dependent variables are city-age fixed effects estimated from wage growth regressions according to Equation (8). Each column represents a different specification of the first-step estimation, reported in Table B3. "Young" refers to workers up to 30 years old, and "senior" refers to workers over 30. Age and tenure variables include a quadratic polynomial. Occupation and sector indicators refer to one-digit and two-digit level information, respectively. Coefficients are reported with robust standard errors in parenthesis. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels.

fairly stable when I add sector and occupation indicators.

For young college graduates, the effects are more pronounced. The second row displays point estimates that range from 0.4273 to 0.5532. In Column (2), including worker fixed effects increases both the estimated elasticity and the standard error. Column (3) shows that excluding sector and occupation moves decreases the elasticity. Finally, including sector and occupation indicators have a negligible impact. These results sharply contrast with the point estimates obtained for senior workers, displayed in Table 4. For these individuals, there is no evidence of city size affecting wage growth.[22]

[22] Table C2 provides estimates of dynamic effects without age heterogeneities, as proposed by D'Costa and Overman (2014). In line with their results, I find mild evidence of dynamic effects. In particular, the point estimate for college graduates is not significant once worker fixed effects are included, mirroring their findings using a sample of British workers. From this evidence, they conclude that higher wage growth in large cities is due to sorting, even though they also find stronger effects for young workers.

Table 4: Dynamic Urban Wage Premium by Age Group - Senior Workers

	City Dynamic Premium ($\delta^s_{c,senior}$)			
	(1)	(2)	(3)	(4)
Panel A. Non-college				
Log City Size	-0.0890	-0.0759	-0.0278	-0.0653
	(0.0680)	(0.1067)	(0.1287)	(0.1371)
R^2	0.01059	0.00210	0.00021	0.00106
Observations	185	185	185	185
Panel B. College				
Log City Size	0.0470	-0.0362	0.0410	0.0410
	(0.0645)	(0.1683)	(0.1785)	(0.1807)
R^2	0.00282	0.00028	0.00034	0.00034
Observations	185	185	185	185
First Step Variables				
Age, Tenure	✓	✓	✓	✓
Worker FE		✓	✓	✓
Exclude sec & occup moves			✓	✓
Sec & occup indicators				✓

Notes: The dependent variables are city-age fixed effects estimated from wage growth regressions according to Equation (8). Each column represents a different specification of the first-step estimation, reported in Table B3. "Young" refers to workers up to 30 years old, and "senior" refers to workers over 30. Age and tenure variables include a quadratic polynomial. Occupation and sector indicators refer to one-digit and two-digit level information, respectively. Coefficients are reported with robust standard errors in parenthesis. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels.

In summary, the evidence shows that city size affects wage growth only at the beginning of working life, and mainly for those with a college degree. The asymmetry between the two educational groups is consistent with Figure 1 and the hypothesis that the sorting of college graduates reported in Section 2.4 results from larger cities providing more valuable experience to their workers. However, the connection between sorting and dynamic effects, while plausible, cannot be established based on the evidence so far. Moreover, it is still unclear whether dynamic effects are, in fact, significant in explaining the city-size wage gap. I address these issues in the next section.

3.2 Dynamic Effects and the City Size Wage Gap

In order to understand the extent to which dynamic effects contribute to the city-size wage gap, it is necessary to consider all the terms in Equation (3) simultaneously. In order to achieve this, I propose a simpler but feasible approach that makes use of the estimates of the previous section. I rearrange the terms of Equation (3) to obtain

$$\tilde{w}_{i,t}^{s} \equiv \log w_{i,t}^{s} - \hat{\delta}_i e_{it} - \sum_{\tau=1}^{t} \hat{\delta}^{s}_{c(i,\tau),a(i,\tau)} = \Psi_c^s + \alpha_i + X_{i,t}\beta^s + \tilde{\epsilon}_{i,t} \ . \tag{9}$$

In this expression, $\tilde{w}_{i,t}^{s}$ is the adjusted wage net of individual heterogeneity in return to experience and city dynamic effects. I calibrate $\hat{\delta}^{s}_{c,a}$ and $\hat{\delta}_i$ using the parameters estimated from Equation 6, presented in Figures 3 and 4.

The dataset does not provide workers' full location history, but using available records since 2003, it is possible to construct it for a subset of younger individuals. I do so by assuming that college graduates start working at age 22, which means selecting those whose year of birth is from 1981 onwards. For non-college graduates, I assume the beginning of working life is 18 years old, so the cutoff is 1985.

I first analyze the results of regressing the static premium Ψ_c^s on city size under different specifications of Equation (9), which are displayed in Table 5. Estimates of the first-step equation are reported in Tables B4 and B5. In Columns (1), (2) and (3), I use simple log wages as the dependent variable in the first-step equation, whereas in Columns (4), (5) and (6) I use adjusted log wages while keeping the specification of Column (3).[23]

Overall, incorporating dynamic elements does not impact the estimated static premiums significantly. In Column (4), when I incorporate dynamic effects, the static premium slightly increases to 0.0171 and 0.0324 for non-college and college graduates, respectively. When I subtract individual returns to experience from wages, in Column (5), there is a negligible

[23] The results of Columns (1), (2) and (3) differ from those in Table 2 because I use a smaller sample of workers with full location history available.

Table 5: Static Urban Wage Premium

	City Static Premium (Ψ_c^s)					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Non-college						
Log City Size	0.0201	0.0150	0.0151*	0.0169*	0.0148	0.0167*
	(0.0134)	(0.0121)	(0.0090)	(0.0093)	(0.0090)	(0.0094)
R^2	0.01108	0.00802	0.01429	0.01657	0.01369	0.01611
Observations	185	185	185	185	185	185
Panel B. College						
Log City Size	0.0539***	0.0481***	0.0274***	0.0328***	0.0193***	0.0248***
	(0.0166)	(0.0130)	(0.0071)	(0.0076)	(0.0068)	(0.0073)
R^2	0.04747	0.06722	0.06190	0.07295	0.03547	0.04789
Observations	185	185	185	185	185	185
First Step Variables						
Dep. Var.: log wage net of	-	-	-	$\sum_{\tau} \hat{\delta}^s_{c,a(i,\tau)}$	$\hat{\delta}_i e_{it}$	Both
Age, Tenure	✓	✓	✓	✓	✓	✓
Sec & occup indicators		✓	✓	✓	✓	✓
Worker FE			✓	✓	✓	✓

Notes: The dependent variables are city fixed effects estimated from wage regressions according to Equation (9). Each column represents a different specification of the first-step estimation, reported in Tables B5 and B4. Age and tenure variables include a cubic polynomial. Occupation and sector indicators refer to one-digit and two-digit level information, respectively. Coefficients are reported with robust standard errors in parenthesis. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels.

change for non-college graduates but a decrease of about 30% for those with a college degree. Finally, in Column (6), when $\sum_{\tau} \hat{\delta}^s_{c,a}$ and $\hat{\delta}_i e_{it}$ are put together in the model, I get elasticities of 0.0168 and 0.0246 for non-college and college graduates, respectively, which are very close to those in Column (3) when log wages are not adjusted.

However, the conclusions regarding worker fixed effects are quite different, as shown in Table 6. This time, I regress the city-average worker fixed effect, defined as $\bar{\alpha}^{s'}_{c'} \equiv \sum_{c(i,t)=c', s=s'} \alpha_i$, on log city size. Columns (1) to (4) refers to the first-step specifications used in Columns (3) to (6) of Table 5.

For non-college workers, including city dynamic effects do not change the conclusions of the static estimation, although there seems to be a slight negative sorting. However, cor-

Table 6: Unobserved Ability and City Size

	City Average Unobserved Ability ($\bar{\alpha}_c$)			
	(1)	(2)	(3)	(4)
Panel A. Non-college				
Log City Size	0.0011	-0.0028	-0.0012	-0.0064
	(0.0052)	(0.0086)	(0.0106)	(0.0047)
R^2	0.00024	0.00064	0.00008	0.01026
Observations	185	185	185	185
Panel B. College				
Log City Size	0.0255**	0.0075	0.0202*	-0.0004
	(0.0116)	(0.0139)	(0.0112)	(0.0100)
R^2	0.02437	0.00143	0.01405	0.000008
Observations	185	185	185	185
First Step Variables				
Dep. Var.: log wage net of	-	$\sum_{\tau} \hat{\delta}^{s}_{c,a(i,\tau)}$	$\hat{\delta}_i e_{it}$	Both
Age, Tenure	✓	✓	✓	✓
Sec & occup indicators	✓	✓	✓	✓
Worker FE	✓	✓	✓	✓

Notes: The dependent variables are city-level averages of worker fixed effects estimated from wage regressions according to Equation (9). Each column represents a different specification of the first-step estimation, reported in Tables B5 and B4. Age and tenure variables include a cubic polynomial. Occupation and sector indicators refer to one-digit and two-digit level information, respectively. Coefficients are reported with robust standard errors in parenthesis. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels.

recting for dynamic effects significantly impacts the previous estimation for college workers. Now, unobserved worker characteristics become uncorrelated with city size. When I only subtract individual returns to experience from wages, in Column (3), the point estimate remains unchanged compared to Column (1). Finally, when both elements are embedded in the model, I get the same result as in Column (2).

In conclusion, the evidence shows that despite being restricted to young workers, dynamic effects are essential to understanding why the wages of college graduates are higher in large cities. Using previous estimates for a back-of-the-envelope calculation, Columns (3) and (5) of Table 2 show that the reduction in the static premium once individual effects are included is 42% of the estimate in Column (1), which is my baseline measure of the city size wage gap.

If 72% of what the fixed effects are capturing are dynamic effects, as suggested by Columns (1) and (2) of Table 6, it means that 30% of the city-size wage gap is due to dynamic effects. This number is comparable with the static premium obtained in Column (6) of Table 5, which accounts for 40% of the gap.[24]

The estimates further show that dynamic effects impact college graduates more intensively and greatly explain why the college premium is higher in large cities. Using again value from Tables 2 and 6, I infer that 42% of the city-size college premium gap is due to dynamic effects.[25]

To provide more intuition about the implication of these findings, consider four young individuals, two with college degree (c_1 and c_2) and two with high school degree (h_1 and h_2), living in a small city S. In a given year, c_1 and h_1 move to a large city L. Tables 2 and 5 indicate that c_1 will experience a higher wage increase after moving. Moreover, according to Figure 3 and Table 3, c_1 will experience higher wage growth. Assume that i) in S the static premiums and dynamic premiums are equal to zero and ii) in L the dynamic premium is δ for college graduates and the static premiums are Ψ_c and Ψ_h for college and high-school graduates, respectively. In this case, wage gap between c_1 and c_2 will be $\Psi_c + T\delta$, and the wage gap between c_1 and h_1 will be $\Psi_c - \Psi_h + T\delta$, where T is the number of periods since c_1 and h_1 moved to the large city.

Lastly, the results show that sorting on unobservable characteristics is of secondary importance to explain the city-size wage gap in a framework with dynamic effects, corroborating De La Roca and Puga (2017). Professionals like engineers and lawyers have on average the same quality at the beginning, regardless of where they are located. As some of them acquire experience in large cities, their skill level starts do diverge, amplifying the city-size wage gap.

[24] Recall that my initial sample comprises only individuals up to 45 years old, and therefore, the share of young workers who did not accumulate much experience is relatively higher compared to the whole labor force. Because of that, the contribution of dynamic effects may be potentially underestimated.

[25] Columns (3) and (5) of Table 2 show that the reduction in the college/non-college premium ratio due to the inclusion of worker fixed effects is 59% of the initial ratio in Column (1). I assume again that 72% of the fixed effects' impact is due to dynamic effects for college graduates (and virtually zero for non-college graduates). Thus, the contribution of dynamic effects to the city-size college premium gap is approximately 42%.

3.3 Robustness

In this section, I summarize the robustness checks performed to address potential sources of bias. More detailed information is available in the Online Appendix.

In Appendix A.1, I adapt an event-study analysis proposed by Card et al. (2013) to examine whether a model with additive worker and city fixed effects can reasonably capture wage variations when a migration occurs. One particular concern is that migration might be driven by particular good wage offers that do not reflect average city premiums. The test is also suitable for investigating if there is an "Ashenfelter dip" in earnings prior to migration.

The results of this exercise, displayed in Figure A1, suggest that the additive model can be a good approximation. It shows no sign of systematic shocks before the migration. Furthermore, the "jumps" are consistent with the estimated city premiums in relative terms, i.e., migrations from quartiles 1 to quartiles 4 are associated with higher gains than migrations from quartiles 1 to quartile 2. Importantly, the bias tends to be of minor relevance if net migration does not vary with city size, and I show in Figure C1 that this correlation is weak in the data.

In Appendix A.2, I investigate the existence of particularities in the wage dynamics of migrants after moving to a new city. Suppose, as reported by Card et al. (2023), that the wage changes that migrants face after moving from low-wage to high-wage cities are systematically below the difference in the average premiums paid in these cities. In this case, it is reasonable to ask whether this below-average jump is at least partially compensated by a "catch-up" effect in which migrants experience faster wage growth as long as they get paid below the average premium. In the opposite way, the wage growth of migrants after moving from high-wage to low-wage cities may be slower due to a faster depreciation of the human capital acquired in the previous location.

To check these possibilities, I estimate an augmented version of Equation (6) with post-migration indicators for the two years after the migration event, and repeat the process of regressing city fixed effects on log city size for each educational-age group. I create specific indicators to differentiate between moves from low- to high-wage cities and vice-versa and whether this variation is large or small. I also test specific effects for young individuals. Table A1 shows that migrants experience faster wage growth after the migration in most of cases, even when they move from a high-wage city to a low-wage one. It is interesting to observe that non-college graduates also benefit substantially from migration. However, the similarity between Figures A2, A3 and 3 weakens the idea that these effects are driving the results.

Another potential concern is the assumption of uniform experience evaluation across cities, as mentioned in Section 3.1. When computing adjusted wages according to Equation 9, I abstract from the possibility that $\delta^s_{c,a}$ has different values for different cities. To address this issue, I estimate an alternative version of Equation (9) with partial experience portability. I impute a discount λ when the experience acquired in a given city is used in another one. The results in Appendix A.3 show that this modification does not change qualitatively the results, although it increases the relative importance of sorting and decreases the relative importance of the static premium.[26]

Second-step estimations may also be biased if city size is endogenously determined. For instance, if some city characteristics omitted in the regression simultaneously affect its size and productivity (e.g., access to other markets). It is also possible that large cities are a consequence of high-wage areas attracting workers.

In Appendix A.4, I provide alternative estimations instrumenting for city size using lagged population and terrain steepness. The first-stage regression results are given in Table A3, showing that both instruments have high statistical significance. The F-statistic equals 83.742 and easily surpasses all thresholds proposed by Stock and Yogo (2005) (size and relative bias). Table A4 presents the 2SLS estimates and the respective OLS regressions to facilitate comparisons. The first row shows that instrumenting eliminates the static premium

[26]Partial portability is a potential confounder if the discounts in experience evaluation correlate with city size. While the results in Appendix A.2 help to mitigate these concerns, I cannot entirely rule out this possibility.

for non-college graduates and slightly reduces the static premium for college graduates. Regarding dynamic effects, there is a decrease of 70% for young non-college graduates but negligible effects for young college graduates. Overall, the results suggest that endogeneity is a concern only for the less educated workers in this context.[27]

Finally, I replicate the results in Section 3 using an alternative education classification based on the most observed value in RAIS between 2003 and 2017. As shown in Table A5, this definition yields a much lower share of college graduates that does not seem to reflect the population formally employed in Brazil.[28] In Online Appendix A.5, I show that the results are qualitatively similar and quantitatively even more pronounced.

4 Dynamic Effects and the Job Ladder

4.1 Within vs. Between Decomposition

Having characterized how dynamic gains manifest, I now examine what these findings imply for the job ladder in big cities. I start by decomposing city size effects by what I label as progression types. Building on Equation (6), I propose the following specification:

$$\Delta \log w_{i,t}^{s} = \delta_i + \delta_{c,p}^{s} + \Delta X_{i,t}\beta^{s} + \Delta\epsilon_{i,t} \ , \tag{10}$$

where the subscript $p = p(i,t)$ refers to a specific progression in worker's career path. Thus, if the observation is a within-firm progression, I estimate specific city effects for the 1st, 2nd, 3rd and 4th or higher firm. If I observe a worker in different firms between two consecutive

[27] The literature typically finds that instrumenting for city size does not change the results substantially. See Combes et al. (2010) for a discussion. One exception is Combes et al. (2019). They analyze agglomeration gains in China separately for high-skilled natives, low-skilled natives and rural migrants and find that 2SLS estimates are larger than OLS in most cases.

[28] The proportion of the Brazilian workforce with higher education increased from 13.7% in 2012 to 18.5% in 2018. Source: https://repositorio.ipea.gov.br/bitstream/11058/9256/16/cc41_nt_mercado_de_trabalho.pdf. Last accessed in May 23th, 2024.

periods, I estimate city effects for the 1st, 2nd and 3rd or higher firm change. Then, I regress the estimated fixed effects on log city size for each progression type.

Enumerating firms (and firm changes) requires the workers' location history again, so the same criteria and sample restrictions from Section 3.2 apply. Additionally, I drop some cities from the estimation due to a lack of observations for some progressions. The final sample has about 9.4 million observations distributed across 176 cities for college graduates and 15.2 million observations across 183 cities for non-college graduates.[29]

The results in Table 7 indicate that city size influences wage growth both within and between firms for college graduates. Column (1) presents evidence of dynamic effects within the first firm, showing that doubling the city size increases yearly wage growth by 0.9%. Wage growth within the second and third firms also have positive and significant coefficients, although lower in magnitude. Regarding wage growth between firms, Columns (5) to (7) show that a city twice the size results in 1.1% higher wage growth in the first firm change and 0.7% in the second firm change. In contrast, workers without college degree benefit little from city size, as coefficients have lower values and are not statistically significant at 10%.[30]

I show in Table C3 that these results are robust to an alternative classification based on a city-specific job enumeration, i.e., when the count restarts every time the individual moves to a new city. Note also that the elasticities related to firm changes are higher in magnitude than the within-firm effects, suggesting that the results are unlikely to be entirely driven by "hidden" within-firm wage growth. Finally, although the higher variance in elasticities related to firm changes is, to some extent, due to fewer observations within these progression types, it also suggests high heterogeneity in wage growth. While some workers benefit substantially from moving across firms in large cities, others may take no advantage.

The estimates in Table 7 can be used to gauge the quantitative importance of within- ver-

[29]I impose a minimum of 1,500 observations to include the city in the estimation. The results are robust to changing the cutoff value to 1,000 or 2,000. One important caveat is that my definition of a firm change is based on observed changes between consecutive years, which ignores within-year changes. Additionally, I also abstract from differences between firm-to-firm and firm-unemployment-firm transitions.

[30]Table B6 displays the results of the first step estimation. Columns (1) and (3) refer to the content of Table 7, and Columns (2) and (4) refer to the content of Table C3.

Table 7: Dynamic Effects Within and Between Jobs

	Within				Between		
	1st	2nd	3rd	4th+	1st-2nd	2nd-3rd	3rd-4th+
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. Non-college							
Log City Size	0.2304	-0.0066	-0.2138	-0.1829	0.2669	0.0501	-0.5098
	(0.1875)	(0.1522)	(0.1640)	(0.1596)	(0.2755)	(0.2069)	(0.3196)
R^2	0.00662	0.000009	0.00777	0.00515	0.00414	0.00026	0.01322
Observations	183	183	183	183	183	183	183
Panel B. College							
Log City Size	0.9023***	0.4035***	0.4961***	0.3044	1.079***	0.6960**	0.1165
	(0.2084)	(0.1549)	(0.1796)	(0.2077)	(0.3855)	(0.2842)	(0.3332)
R^2	0.06574	0.02441	0.02888	0.00782	0.02884	0.01938	0.00051
Observations	176	176	176	176	176	176	176

Notes: The dependent variables are city-progression fixed effects estimated from wage growth regressions according to Equation (10). Each column exhibits the city-size elasticity for different progression types. Results of the first-step estimation are reported in Table B6. For more details, see Section 4. Controls in the first-step estimations include worker fixed effects and a quadratic polynomial in age and tenure. Coefficients are reported with robust standard errors in parenthesis. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels.

sus between-firm dynamic effects for young individuals, who experience these effects more intensively. Specifically, for college graduates, I compute the expected "within" and "between" wage growth from ages 23 to 30 for each individual-city pair in my sample. For non-college graduates, I analyze the age range between 19 and 26 years old. Subsequently, I take the average for each city and regress these values on log city size.

I present the results of this exercise in Table 8. For college graduates, the elasticity of total expected wage growth with respect to city size is 0.0433, as shown in Column (1), meaning that doubling the city size would increase expected wage growth by 4.3%. In Columns (2) and (3), expected wage growth is split between the two categories considered. The elasticities obtained indicate that, for college graduates, 59% of the total effect is attributable to "within-firm" effects and 41% to "between-firm" effects. The relative contributions are similar for non-college graduates, but the effects are significantly lower in magnitude.

Given that differences in the elasticities depend on the number of "within-firm" and

Table 8: Dynamic Effects Decomposition - Within vs. Between

	Total	Within	Between	Job Changes
	(1)	(2)	(3)	(4)
Panel A. Non-college (19-26 yrs old)				
Log City Size	0.0146***	0.0077**	0.0069***	0.1069***
	(0.0042)	(0.0036)	(0.0024)	(0.0204)
R^2	0.04639	0.01824	0.04135	0.12282
Observations	183	183	183	183
Panel B. College (23-30 yrs old)				
Log City Size	0.0433***	0.0255***	0.0177***	0.1209***
	(0.0056)	(0.0046)	(0.0019)	(0.0125)
R^2	0.20271	0.10784	0.29874	0.35342
Observations	176	176	176	176

Notes: The dependent variables in Columns (1) to (3) are averages at the city level of the predicted dynamic effects for the progressions and age intervals indicated in the table and considering the estimates displayed in Table 7. In Column (4), the dependent variable is the average at the city level of the number of job changes. For more details, see Section 4. Coefficients are reported with robust standard errors in parenthesis. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels.

"between-firm" observations, I also examine whether there is a higher frequency of firm changes in large cities. Following the procedure described above, I compute the average number of firm transitions in each city during the specified age intervals and regress the logarithm of this variable on log city size. Column (4) of Table 8 shows that both educational groups are more likely to change firms during the age intervals analyzed. Doubling the city size would increase the number of firm changes by 12% and 11% for college and non-college graduates, respectively.

4.2 Dynamic Effects and Career Paths

While Table 7 shows that firm changes in large cities are associated with faster wage growth, it does not inform whether working for more than one firm results in higher wages in the long run. To address this question, I propose an empirical analysis comparing similar career paths in cities of different cize. From Equation (3), I derive an expression for the long-run wage growth of a non-migrant:

$$\log w_{i,t+n} - \log w_{i,t} = n_i^* \delta_i + \delta_c^n + \Delta X_{i,t}\beta + \Delta\epsilon_{i,t} \quad , \tag{11}$$

where n is the length of the interval considered and $n_i^* \leqslant n$ is how many times worker i is formally employed between t and $t+n$. $\delta_c^n \equiv \sum_{\tau=t}^{t+n} \delta_{c,a(i,\tau)}$ is the long-run premium of city c for a given n. Motivated by the previous findings, I hypothesize that this term is a function of the number of firms J individual i worked for during the period n: $\delta_c^n = \delta_c^n(J)$. Since this analysis focuses on college graduates, I drop the subscript s from the equation.

Equation (11) clarifies that wage growth in the long-run depends on both location and unobserved personal characteristics. If we abstract from the latter for now, long-run dynamic effects can be estimated using another two-step procedure. First, I run the following wage growth regression:

$$\Delta \log w_i = \delta_{c,J}^{LR} + X_i \Pi + \eta_i \quad , \tag{12}$$

where $\Delta \log w_i$ is the wage growth for a fixed age interval, set to be between 23 and 30 years old, $\delta_{c,J}^{LR}$ is the long-run premium of city c for an individual that worked for J firms during this period, X_i is a vector of individual characteristics - which includes year fixed effects, sector and occupation indicators - and η_i is an error term.[31] After estimating $\delta_{c,J}^{LR}$, I regress them on log city size separately for each J:

$$\delta_{c,J}^{LR} = \lambda_0^J + \lambda^J \log CitySize_c + v_c \quad , \tag{13}$$

where λ^J are the parameters of interest and represent the long-run premium for a career path with J firms. If dynamic effects depend on J, the estimation of Equation (13) should yield different values of λ^J.

Naturally, this procedure does not prevent λ^J from capturing differences in worker com-

[31] I used a sample of non-migrants observed between 2003 and 2017 for this estimation. Cities with too few observations for specific values of J are dropped from the estimation.

position based on unobserved characteristics. In order to disentangle city effects from sorting, I also estimate long-run premiums from adjusted wages net of individual heterogeneity in returns to experience: $\Delta \log w_i - n_i^* \hat{\delta}_i$. I use estimates reported in Figure 4 to calibrate $\hat{\delta}_i$, whereas n_i^* I observe directly from the data.

I also analyze various dimensions of sorting across career paths. I compute the fraction of individuals and average values of $\hat{\delta}_i$, $\hat{\alpha}_i$ and n_i^* for each pair (c, J) and regress them on log city size separately for each J. Differences in the elasticities among different J would inform whether city size affects the composition of workers in each career path.[32]

Table 9 displays estimates of the long-run premium. In Column (1), I estimate this parameter not differentiating by J to serve as a reference, whereas in Columns (2) to (5), I separate the city fixed effects into four categories of J. Each Panel considers a particular specification of Equation (12). In Panel A, Column (1) shows that doubling the city size would increase the wage growth of college graduates between 23 and 30 years old by 1.5% on average. When I split city fixed effects by the number of firms, Columns (2) to (5) show elasticities that range from 0.0076 for workers that stay in the same firm the whole period to 0.0278 for workers that experience two firm changes. Notably, the long-run premium does not increase monotonically with J, suggesting that the relationship between dynamic effects and firm changes has an inverted U-shape form.

In Panel B, I estimate long-run premiums excluding career paths in which the occupation of the individual at the age of 23 is different from that of 30. This sample reduction allows me to investigate to what extent the results in Panel A stem from individuals moving from low- to high-wage occupations more frequently in large cities. Interestingly, the U-shape form becomes more pronounced, with an increase in the long-run premium of the career path with two firm changes and a reduction in the other long-run premiums. In Panel C, I add occupation indicators to control for differences in wage growth across occupations, yielding larger decreases in long-run premiums involving at least one firm change but no

[32] I use estimates reported in Column (4) of Table 6 for $\hat{\alpha}_i$.

Table 9: Long-Run Wage Premium - College Graduates (23-30 yrs old)

	City Premium (δ_c^{LT})	City-career Premium ($\delta_{c,J}^{LT}$)			
# Firms	-	0	1	2	3+
	(1)	(2)	(3)	(4)	(5)
Panel A. Only Year FE					
Log City Size	0.0145**	0.0076	0.0170*	0.0278**	0.0192
	(0.0068)	(0.0067)	(0.0087)	(0.0109)	(0.0123)
R^2	0.02192	0.00609	0.01687	0.03153	0.01144
Observations	161	161	161	161	161
Panel B. Drop Occupation Changes					
Log City Size	0.0119	0.0056	0.0138	0.0320***	0.0173
	(0.0076)	(0.0077)	(0.0096)	(0.0114)	(0.0141)
R^2	0.01219	0.00247	0.00941	0.03997	0.00700
Observations	161	161	161	161	161
Panel C. Drop Occupation Changes + Occupation Dummies					
Log City Size	0.0104	0.0041	0.0075	0.0252**	0.0116
	(0.0078)	(0.0078)	(0.0097)	(0.0112)	(0.0139)
R^2	0.00917	0.00144	0.00279	0.02542	0.00322
Observations	161	161	161	161	161
Panel D. Drop Sector Changes					
Log City Size	0.0128	-0.0033	0.0186*	0.0250**	0.0261**
	(0.0095)	(0.0122)	(0.0097)	(0.0117)	(0.0126)
R^2	0.00883	0.00031	0.01575	0.02218	0.01976
Observations	161	161	161	161	161
Panel E. Drop Sector Changes + Sector Dummies					
Log City Size	0.0142	-0.0023	0.0155*	0.0246**	0.0276**
	(0.0087)	(0.0113)	(0.0090)	(0.0106)	(0.0119)
R^2	0.01327	0.00017	0.01381	0.02589	0.02435
Observations	161	161	161	161	161
Panel F. Only Year FE - Adjusted Wages					
Log City Size	0.0461***	0.0345***	0.0554***	0.0498***	0.0415***
	(0.0110)	(0.0108)	(0.0123)	(0.0147)	(0.0157)
R^2	0.06297	0.03789	0.07245	0.04795	0.02627
Observations	161	161	161	161	161

Notes: The dependent variables are city fixed effects estimated from Equation (12). Column (1) exhibits the city-size elasticity when city premiums are assumed to be constant across career paths, whereas Columns (2) to (5) exhibit city-size elasticities differentiating by the number of firms the individual worked for during the period analyzed. Each panel represents a different specification and sample used in the first-step estimation. In Panel F, wages are adjusted for individual heterogeneity in returns to experience. See Section 4.2 for more details. Coefficients are reported with robust standard errors in parenthesis. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels.

changes in the internal hierarchy.

Panels D and E repeat the procedure of Panels B and C, respectively, this time considering sector differences. Panel D shows that sector changes amplifies the gap between the long-run premium of the single-firm career path and the other long-run premiums. Note also that, in this case, the largest elasticity comes from the long-run premium with three or more firm changes, indicating that moving between firms of the same sector in large cities can be highly beneficial. In Panel E, I add sector dummies and get very similar results.

Panel F shows results using adjusted wages as the dependent variable in Equation (12). Surprisingly, there is a substantial jump in elasticities, revealing that not taking into account unobserved characteristics leads to a strong underestimation of the long-run premium for all career paths. Regarding ranking, the two-firm career path now has the largest long-run premium, but the U-shape form is preserved.

In Table 10, I analyze sorting across career paths. Panel A shows that the elasticity of average $\hat{\delta}_i$ with respect to city size is negative for all J, especially for career paths with less firm changes. Column (1) makes explicit that this sample of non-migrant young workers has, on average, lower individual returns to experience in large cities, constrasting with the findings of Figure 4 when analyzing the full sample. In Panel B, I show that individuals in this sample tend to be formally employed for fewer periods in large cities. Both lower $\hat{\delta}_i$ and n_i^* in large cities align with the higher estimates of the long-run premium using adjusted wages in Panel F of Table 9.

In Panel C, I estimate the elasticity of the fraction of workers in a given career path with respect to city size. There is a clear shift towards career paths with more firm changes. In quantitative terms, the point estimates in Columns (2) to (5) indicate that doubling city size would reduce the share of individuals in single-firm or two-firm career paths by around 6.6%, with a similar increase for career paths with two or more firm changes. Finally, Panel D shows weak evidence that individuals sort into career paths based on initial unobserved ability $\hat{\alpha}_i$.

Table 10: Sorting and Career Paths - College Graduates (23-30 yrs old)

	City	City-career			
# Firms	-	0	1	2	3+
	(1)	(2)	(3)	(4)	(5)
Panel A. Avg. Returns to Experience ($\bar{\delta}_{c,J}$)					
Log City Size	-0.2781**	-0.3236**	-0.2774**	-0.1161	-0.2808
	(0.1281)	(0.1448)	(0.1305)	(0.1565)	(0.1785)
R^2	0.01756	0.02035	0.01593	0.00206	0.00991
Observations	161	161	161	161	161
Panel B. Avg. Number of Observations ($\bar{n}^*_{c,J}$)					
Log City Size	-0.0582***	0.0085	-0.0931***	-0.0467***	-0.0153
	(0.0146)	(0.0068)	(0.0226)	(0.0179)	(0.0126)
R^2	0.07526	0.00718	0.07702	0.03212	0.00714
Observations	161	161	161	161	161
Panel C. Share of Workers					
Log City Size		-0.0513***	-0.0149***	0.0202***	0.0460***
		(0.0066)	(0.0034)	(0.0029)	(0.0042)
R^2		0.23032	0.08765	0.17001	0.40913
Observations		161	161	161	161
Panel D. Avg. Initial Ability ($\bar{\alpha}_{c,J}$)					
Log City Size	-0.0085	0.0149	0.0223	0.0173	0.0109
	(0.0162)	(0.0184)	(0.0166)	(0.0173)	(0.0165)
R^2	0.00165	0.00372	0.01148	0.00566	0.00223
Observations	161	161	161	161	161

Notes: The dependent variables are indicated in the respective panels. Column (1) exhibits the city-size elasticity when the dependent variable is computed at the city level, whereas Columns (2) to (5) exhibit city-size elasticities differentiating by the number of firms the individual worked for during the period analyzed. Coefficients are reported with robust standard errors in parenthesis. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels.

In summary, the findings in Tables 9 and 10 reveal that the magnitude of dynamic effects depends on the chosen career path in the long run. Workers who transition more frequently across firms - in particular within the same sector - receive a larger long-run premium than those who stay in the same firm. Moreover, there is evidence that individuals respond to these incentives by moving across firms more frequently, especially those with a natural tendency to experience higher wage growth, measured by their returns to experience.

Tables C4 and C5 display the results of a similar exercise for (non-migrant) non-college graduates between 19 and 26 years old. City size seems to have a mildly negative effect on these workers in the long run, especially for career paths with at least one firm change. However, this result is reverted when I estimate long-run premiums from adjusted wages. This happens because individuals with higher returns to experience are more likely to stay in the same firm in big cities, even though moving across firms is generally more common, as shown in Panel C of Table C5.

4.3 Discussion - Mechanisms

Since Glaeser and Maré's (2001) seminal paper, there has been a discussion about the nature of dynamic effects. Broadly speaking, the literature provides two interpretations: i) cities may promote better matching between workers and firms, and ii) cities may promote faster human capital accumulation due to a better learning environment. While the former explanation relies on studies connecting firm changes with higher wage growth in big cities (Wheeler, 2006; Bleakley and Lin, 2012; Eckert et al., 2020), the latter finds empirical support in evidence demonstrating that the more valuable experience acquired in a big city does not depreciate once the worker moves to a smaller city or a rural area (Glaeser and Maré, 2001; Baum-Snow and Pavan, 2012; De La Roca and Puga, 2017).

This section presents evidence that aligns with the matching hypothesis, but I do not argue that these mechanisms are exclusionary. Instead, emphasize that wage dynamics in large cities may be connected with the structure of thick local labor markets. While this conjecture is not new, the evidence in the literature is mixed. For instance, some papers in labor economics investigate whether an empirical matching function would have increasing returns to scale. Petrongolo and Pissarides (2001) extensively survey this literature and conclude that a constant returns to scale function has been proven adequate in most cases. Baum-Snow and Pavan (2012) find that search frictions and firm-worker match quality do not vary with city size. There is a small literature that studies assortative matching (Andersson et al., 2007; Figueiredo et al., 2014; Dauth et al., 2022). In special, Dauth et al. (2022) show that not only does the correlation between worker quality and firm quality increase with city size, but this relationship has also strengthened over the last few years.

It is worth highlighting two studies by Bleakley and Lin (2012) and Wheeler (2008), who find that city size affects patterns of sectoral and occupational changes throughout working life. Although their focus is not primarily on wage growth, the studies are possibly the most closely connected to the spirit of this paper since they approach the issue from a life cycle perspective. As the results here show, there is a specific moment and context in which firm changes in large cities lead to faster wage growth.

5 Conclusion

This paper studies the relationship between wages and city size, focusing on wage dynamics. From a rich panel dataset that covers 185 urban areas in Brazil, I explore various sources of heterogeneity to shed light on how large cities impact the wage profiles of different groups of workers, thus providing an accurate description of this phenomenon.

I show that dynamic effects impact mostly young individuals with a college degree, accounting for 30% of their city-size wage gap and 40% of the city-size college premium gap. Despite being present, differences in the static premium among educational groups are less pronounced than what the city-size wage gaps suggest. Moreover, sorting on unobserved characteristics contributes little to wage disparities within educational groups once dynamic effects are considered, corroborating the notion that large cities "manufacture" better workers by providing more valuable experience to them.

In an environment where individuals maximize lifetime utility, one implication of these findings is that college graduates have incentives to engage in more complex migration patterns. De La Roca et al. (2022) explore this possibility by considering a location choice problem with age-specific dynamic benefits that yields migration strategic behaviors. Despite focusing on income shocks, Bilal and Rossi-Hansberg's (2021) model treating location choice as an investment can also offer valuable insights for addressing this issue.

I also use the matched employer-employee data to study the role of firm changes in explaining these patterns. I find two important results regarding college graduates. First, city size influences wage growth within and between firms, with the latter contributing to 40% of the overall dynamic effects. Second, firm changes enhance dynamic effects in the long run, impacting how young individuals sort across career paths. These pieces of evidence reveal both short- and long-run benefits of firm changes and suggest that the structure of thick local labor markets influences wage dynamics in large cities.

References

Ahfeldt, G. and Pietrostefani, E. (2019). The economic effects of density: A synthesis. *Journal of Urban Economics*, 111:93–107.

Alvarez, J. (2020). The agricultural wage gap: Evidence from Brazilian micro-data. *American Economic Journal: Macroeconomics*, 12(1):153–173.

Andersson, F., Burgess, S., and Lane, J. (2007). Cities, matching and the productivity gains of agglomeration. *Journal of Urban Economics*, 61(1):112–128.

Bacolod, M., Blum, B., and Strange, W. (2009). Skills in the city. *Journal of Urban Economics*, 65:136–153.

Baum-Snow, N. and Pavan, R. (2012). Understanding the city size wage gap. *Review of Economic Studies*, 79(1):88–127.

Bilal, A. and Rossi-Hansberg, E. (2021). Location as an asset. *Econometrica*, 89(5):2459–2495.

Bleakley, H. and Lin, J. (2012). Thick-market effects and churning in the labor market: Evidence from US cities. *Journal of Urban Economics*, 72(1):87–103.

Bosquet, C. and Overman, H. (2019). Why does birthplace matter so much? *Journal of Urban Economics*, 110(1):26–34.

Bryan, G., Glaeser, E., and Tsivanidis, N. (2019). Cities in the developing world. *NBER Working Paper no. 26390*.

Card, D., Heining, J., and Kline, P. (2013). Workplace heterogeneity and the rise of west German inequality. *Quarterly Journal of Economics*, 128(3):967–1015.

Card, D., Rothstein, J., and Yi, M. (2023). Location, location, location. *NBER Working Paper no. 31587*.

Chauvin, J., Glaeser, E., Ma, Y., and Tobio, K. (2017). What is different about urbanization in rich and poor countries? cities in Brazil, China, India and the United States. *Journal of Urban Economics*, 98:17–49.

Ciccone, A. and Hall, R. (1996). Productivity and the density of economic activity. *The American Economic Review*, 86:54–70.

Combes, P., Duranton, G., and Gobillon, L. (2008). Spatial wage disparities: Sorting matters. *Journal of Urban Economics*, 63:723–742.

Combes, P., Duranton, G., and Gobillon, L. (2010). Estimating agglomeration effects with history, geology, and worker fixed-effects. In Glaeser, E., editor, *Agglomeration Economics*, pages 15–65. Chicago University, Chicago, IL.

Combes, P., Démurger, S., Li, S., and Wang, J. (2019). Unequal migration and urbanisation gains in China. *Journal of Development Economics*, 142:102328.

Combes, P. and Gobillon, L. (2015). The empirics of agglomeration economies. In Duranton, G., Henderson, J., and Strange, W., editors, *Handbook of Regional and Urban Economics*, volume 5, pages 247–348. North Holland.

Dauth, W., Findeisen, S. Moretti, E., and Suedekum, J. (2022). Matching in cities. *Journal of the European Economic Association*, 20(4):1478—-1521.

Davis, D. and Dingel, J. (2019). A spatial knowledge economy. *American Economic Review*, 109(1):153–170.

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

De La Roca, J., G., O., and Puga, D. (2022). City of dreams. *Journal of the European Economic Association*, 21(2):690—726.

De La Roca, J. and Puga, D. (2017). Learning by working in big cities. *Review of Economic Studies*, 84(1):106–142.

Dingel, J., Miscio, A., and Davis, D. (2019). Cities, lights, and skills in developing economies. *Journal of Urban Economics*.

Duranton, G. and Puga, D. (2004). Micro-foundations of urban agglomeration economies. In Duranton, G., Henderson, J., and Strange, W., editors, *Handbook of Regional and Urban Economics*, volume 4, pages 2063–2117. North Holland.

Duranton, G. and Puga, D. (2020). The economics of urban density. *Journal of Economic Perspectives*, 34(3):3–26.

Eckert, F., Hejlesen, M., and Walsh, C. (2020). The return to big city experience: Evidence from refugees in Denmark. *Working Paper*.

Eeckhout, J., Pinheiro, R., and Schmidheiny, K. (2014). Spatial sorting. *Journal of Political Economy*, 556(3):554–620.

Figueiredo, O., Guimarães, P., and Woodward, D. (2014). Firm–worker matching in industrial clusters. *Journal of Economic Geography*, 14(1):1—-19.

Frings, H. and Kamb, R. (2022). The relative importance of portable and non-portable agglomeration effects for the urban wage premium. *Regional Science and Urban Economics*, 95:103786.

Glaeser, E. and Maré, D. (2001). Cities and skills. *Journal of Labor Economics*, 19(2):316–342.

Hicks, J., Kleemans, M., Li, N., and Miguel, E. (2017). Reevaluating agricultural productivity gaps with longitudinal microdata. *NBER Working Paper no. 23253*.

IBGE (2016a). *Arranjos Populacionais e Concentrações Urbanas no Brasil*. IBGE, Rio de Janeiro, RJ, 2nd edition.

IBGE (2016b). *Grade Estatística*. IBGE, Rio de Janeiro, RJ.

Lagakos, D., Moll, B., Porzio, T., Qian, N., and Schoellman, T. (2018). Life cycle wage growth across countries. *Journal of Political Economy*, 126(2):797–849.

Lindley, J. and Machin, S. (2014). Spatial changes in labour market inequality. *Journal of Urban Economics*, 79:121–138.

Martellini, P. (2021). Local labor markets and aggregate productivity. *Working Paper*.

Marx, B., Stoker, T., and Suri, T. (2013). The economics of slums in the developing world. *Journal of Economic Perspectives*, 27(4):187–210.

Matano, A. and Naticchioni, P. (2016). What drives the urban wage premium? evidence along the wage distribution. *Journal of Regional Science*, 56(2):191–209.

Moretti, E. (2004). Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data. *Journal of Econometrics*, 121:175–212.

Moretti, E. (2013). Real wage inequality. *American Economic Journal: Applied Economics*, 5(1):65–103.

Petrongolo, B. and Pissarides, C. (2001). Looking into the black box: A survey of the matching function. *Journal of Economic Literature*, 116:390–431.

Saiz, A. (2010). The geographic determinants of housing supply. *Quarterly Journal of Economics*, 125(3):1253–1296.

Stock, J. and Yogo, M. (2005). Testing for weak instruments in linear IV regression. In Stock, J. and Andrews, D., editors, *Identification and Inference for Econometric Models: Essays in Honor of Thomas J. Rothenberg*, pages 109–120. Cambridge University Press, Cambridge, UK.

Wheeler, C. (2006). Cities and the growth of wages among young workers: Evidence from the nlsy. *Journal of Urban Economics*, 60:162–184.

Wheeler, C. (2008). Local market scale and the pattern of job changes among young men. *Regional Science and Urban Economics*, 38:101–188.

Appendix for Online Publication

A Additional Empirical Results 2

A.1 Event-Study (Card et al., 2013) 2

A.2 Including Post-migration Indicators 3

A.3 Partial Portability 6

A.4 Instrumenting for City Size 8

A.5 Alternative Education Classification 10

B 1st-step Estimations 15

C Other Tables and Figures 21

D External Validity 27

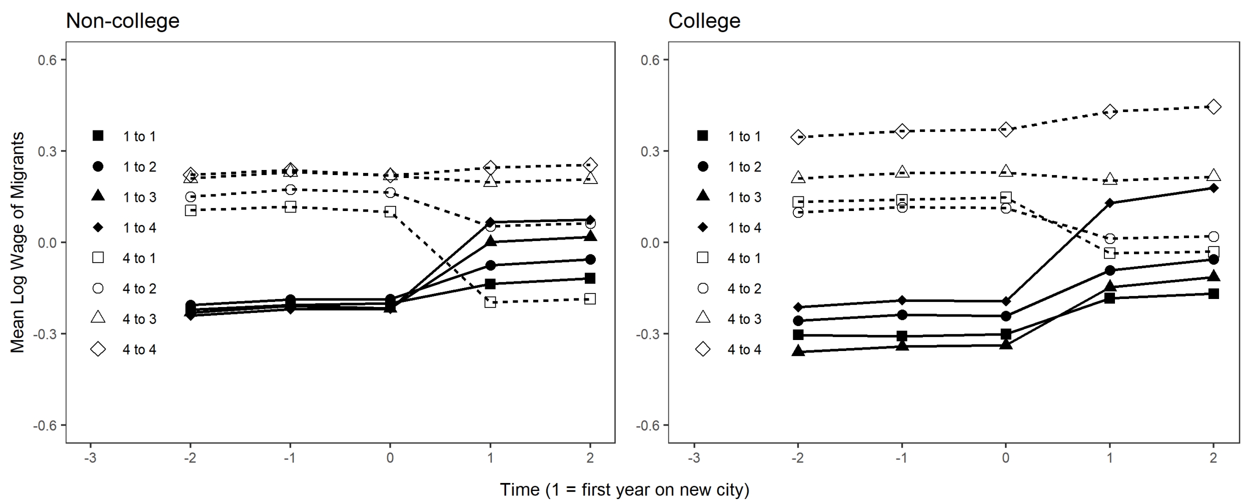
E Theoretical Framework 29

A Additional Empirical Results

A.1 Event-Study (Card et al., 2013)

For each educational group, I separate cities into four quartiles based on the static city premiums estimated in Column (5) of Table 2. Then, migrations are classified into 16 categories based on the quartile before and after the move. I compute mean wages for each category and period relative to moving using events in which I observe migrants for five consecutive periods in the data, three of them prior to the move. The results are displayed in Figure A1. For better exposure, I exhibit only migrations from quartiles 1 and 4.

Figure A1: Event Study - Wages and Migration



Notes: The graph shows mean wages of migrants categorized by quartile of city static premium before and after the migration.

A.2 Including Post-migration Indicators

Departing from Equation (6), I now consider the following specification:

$$\Delta \log w_{i,t}^{s} = \delta_i + \delta_{c,a}^{s} + \sum_{m \in \Omega} \sum_{k=\{1,2\}} D_{i,t,k,m} + \Delta X_{i,t}\beta^{s} + \Delta\epsilon_{i,t} \ , \tag{A.1}$$

where m represents a migration type, k represents event periods relative to t, and $D_{i,t,k,m}$ is an indicator that takes the value one if the migration type is m and if the difference between t and the year of migration is k.[33]. I use the city fixed-effects reported in Column (5) of Table 2 to distribute migrations into types. If the variation in static premiums is positive (negative), the migration is classified as H-L (L-H). If the variation is larger (smaller) in absolute value than the difference between the third and first quartile of static premiums, the migration is labeled as "Large" ("Small"). I also estimate equations with specific indicators for young individuals, who are defined again as those up to 30 years old, respectively.

I report the results of Equation (A.1) in Table A1 and the respective dynamic effects in Figures A2 and A3.

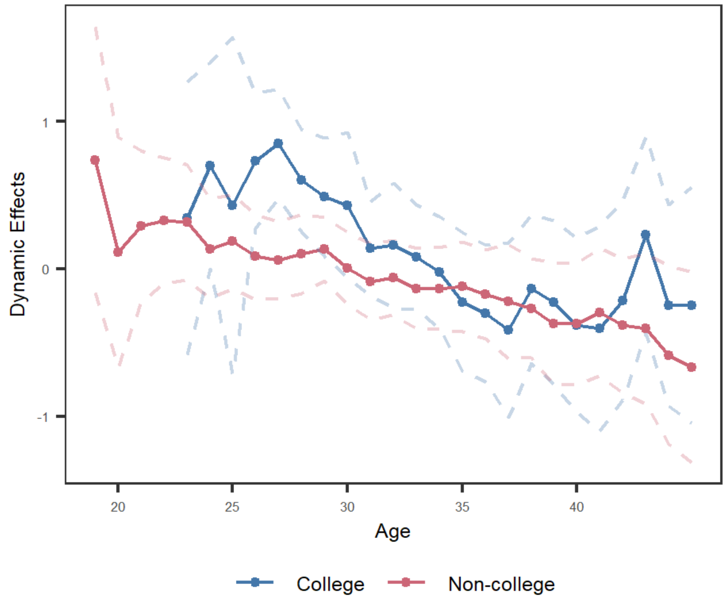
[33] Recall that since migrations are excluded from the estimation, there are no observations in which $k = 0$.

Table A1: Dynamic Effects With Post-migration Indicators - 1st step estimation

	Δ Log Wage $\times 100$			
	College		Non-college	
	(1)	(2)	(3)	(4)
Tenure	-0.5315***	-0.5294***	-3.553***	-3.560***
	(0.2055)	(0.2053)	(0.0432)	(0.0433)
Tenure^2	1.886***	1.883***	0.0414***	0.0414***
	(0.1605)	(0.1603)	(0.0149)	(0.0149)
Low to High, Large, k = 1	1.657***	1.847***	0.3602**	0.5500**
	(0.1163)	(0.1465)	(0.1624)	(0.2276)
Low to High, Large, k = 2	1.357***	1.098***	1.176***	1.114***
	(0.1399)	(0.1706)	(0.2193)	(0.2965)
Low to High, Small, k = 1	2.444***	2.767***	0.8317***	0.8945***
	(0.0722)	(0.0893)	(0.0542)	(0.0735)
Low to High, Small, k = 2	-0.6875***	-1.164***	1.476***	1.138***
	(0.0885)	(0.1058)	(0.0701)	(0.0920)
High to Low, Large, k = 1	0.9051***	0.8314***	0.9173***	0.4028**
	(0.1186)	(0.1462)	(0.1461)	(0.2049)
High to Low, Large, k = 2	1.361***	1.129***	1.269***	0.6718***
	(0.1472)	(0.1747)	(0.1903)	(0.2559)
High to Low, Small, k = 1	0.9482***	0.9605***	0.7928***	0.6555***
	(0.0722)	(0.0884)	(0.0520)	(0.0714)
High to Low, Small, k = 2	1.028***	0.7827***	1.357***	0.7848***
	(0.0874)	(0.1037)	(0.0671)	(0.0884)
Low to High, Large, Young, k = 1		-0.4529**		-0.3624
		(0.2234)		(0.3066)
Low to High, Large, Young, k = 2		0.7035**		0.1095
		(0.2885)		(0.4291)
Low to High, Small, Young, k = 1		-0.8412***		-0.1281
		(0.1450)		(0.1040)
Low to High, Small, Young, k = 2		1.418***		0.7262***
		(0.1902)		(0.1395)
High to Low, Large, Young, k = 1		0.1919		1.018***
		(0.2338)		(0.2844)
High to Low, Large, Young, k = 2		0.6668**		1.284***
		(0.3129)		(0.3779)
High to Low, Small, Young, k = 1		-0.0307		0.2744***
		(0.1450)		(0.1010)
High to Low, Small, Young, k = 2		0.7444***		1.239***
		(0.1889)		(0.1339)
R^2	0.17662	0.17663	0.17726	0.17726
Observations	17,588,577	17,588,577	36,381,355	36,381,355
City-age FE	✓	✓	✓	✓
Worker FE	✓	✓	✓	✓

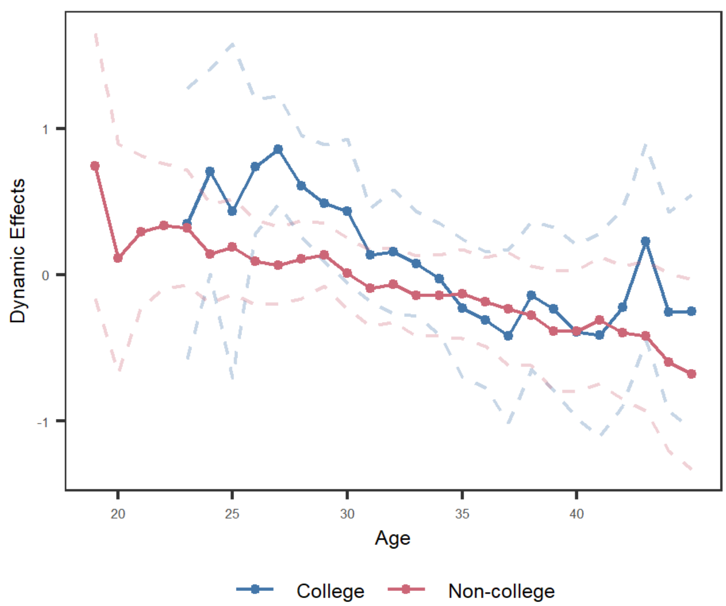
Notes: Dependent variable is the yearly variation in log wage. Each column exhibits the results of estimating Equation (A.1) for a specific educational group. Coefficients are reported with standard errors in parenthesis clustered by worker. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels.

Figure A2: Dynamic Effects with Post-migration Indicators - Specification 1



Notes: Each dot represents the elasticity of log wage growth with respect to city size for a specific educational-age group. Control variables include a quadratic polynomial in tenure, worker fixed effects and post-migration indicators as reported in Columns (1) and (3) of Table A1. The dashed lines inform robust 95% confidence intervals.

Figure A3: Dynamic Effects with Post-migration Indicators - Specification 2



Notes: Each dot represents the elasticity of log wage growth with respect to city size for a specific educational-age group. Control variables include a quadratic polynomial in tenure, worker fixed effects and post-migration indicators as reported in Columns (2) and (4) of Table A1. The dashed lines inform robust 95% confidence intervals.

A.3 Partial Portability

To include partial portability, I now consider a modified version of Equation (9):

$$\tilde{w}_{i,t}^{s} \equiv \log w_{i,t}^{s} - \hat{\delta}_i e_{it} - \sum_{\tau=1}^{t} \lambda_{c(i,\tau),c(i,t)} \hat{\delta}_{c(i,\tau),a(i,\tau)}^{s} = \Psi_c^s + \alpha_i + X_{i,t}\beta^s + \tilde{\epsilon}_{i,t} \ . \tag{A.2}$$

In this expression, $\lambda_{c(i,\tau),c(i,t)} \in [0,1]$ is a term that governs how the experience acquired in city τ is evaluated in the city c where the individual is currently located. I assume that this term has the following structure:

$$\lambda_{c(i,\tau),c(i,t)} = \begin{cases} 1 & \text{if } c(i,\tau) = c(i,t) \\ z & \text{if } c(i,\tau) \neq c(i,t), z < 1 \ . \end{cases} \tag{A.3}$$

In Table A2, I repeat the estimations exhibited in Column (6) of Table 5 and Column (4) of Table 6 for the sample of college graduates considering multiple values of z. Column (1) works as a reference and replicates the results obtained when assuming full portability.

Table A2: Partial Experience Portability - College Graduates

Portability Level (z)	100%	90%	80%	70%	50%	20%
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Sorting ($\bar{\alpha}_c$)						
Log City Size	-0.0004	0.0007	0.0017	0.0028	0.0049	0.0080
	(0.0100)	(0.0099)	(0.0098)	(0.0098)	(0.0097)	(0.0097)
R^2	0.000008	0.00002	0.00015	0.00041	0.00126	0.00333
Observations	185	185	185	185	185	185
Panel B. City Static Premium (Ψ_c^s)						
Log City Size	0.0248***	0.0236***	0.0225***	0.0214***	0.0191***	0.0157**
	(0.0073)	(0.0072)	(0.0070)	(0.0069)	(0.0066)	(0.0065)
R^2	0.04789	0.04605	0.04391	0.04149	0.03582	0.02588
Observations	185	185	185	185	185	185

Notes: The dependent variables are indicated in the respective panels and are obtained from wage regressions according to Equation (A.2). In all columns, the control variables of the first-step estimation include a cubic polynomial in age and tenure, sector and occupation indicators and worker fixed effects. Each column represents an estimation for a different value of z. Coefficients are reported with robust standard errors in parenthesis. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels.

A.4 Instrumenting for City Size

Based on Ciccone and Hall (1996), I instrument current city size using historical urban population data from the 1940 Census. At that time, more than 70% of the population lived in rural areas. The conjecture is that past determinants of population are unlikely to be relevant nowadays. Nevertheless, the population patterns would persist due to non-related factors such as the durability of the urban infrastructure.

I take some simplifying measures to overcome practical difficulties. First, I use a simple urban population count to measure city size since it is the best available information at the municipal level. Secondly, because many municipalities were created from others between 1940 and 2010, I developed a method to infer the spatial distribution of urban population in 1940 using 2010's territorial division. I construct minimum comparable areas, compute the urban population share of each municipality within these areas in 2010 and assume that this share remained constant over the period.

Based on De La Roca and Puga (2017) and Saiz (2010), I also instrument city size using the amount of land with steepness higher than 16% within 15 km of the city centroid. I construct this variable using Global Agroecological Zones (GAEZ) data from FAO.[34] Saiz (2010) shows evidence that residential construction is significantly constrained in these areas. In this case, the exclusion restriction assumption is that geographical constraints on land supply would only affect wages through its influence on city size.

Table A3 displays the results of the first-stage regression, and Table presents A4 presents the 2SLS estimates. The OLS estimates of the static urban wage premium refer to Column (6) of Table 5, while the OLS estimates of dynamic effects refer to Column (2) of Table 3.

[34]GAEZ provides raster data of 5 arc-minute resolution (approximately 9.3 x 9.3 km at the Equator line).

Table A3: 2SLS Estimation - First Stage

	Log City Size
Log Urban Population in 1940	0.4302***
	(0.0484)
% of Terrain Slope > 16%	−0.0012***
	(0.0002)
R^2	0.41796
Observations	185
F-Statistic	65.346

Notes: Coefficients are reported with robust standard errors in parenthesis. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels.

Table A4: 2SLS Estimation - Static Urban Wage Premium and Dynamic Effects

	Non-college		College	
	OLS	*2SLS*	*OLS*	*2SLS*
	(1)	(2)	(3)	(4)
Panel A. Static Urban Wage Premium				
Log City Size	0.0167*	0.0059	0.0248***	0.0219*
	(0.0094)	(0.0126)	(0.0073)	(0.0117)
R^2	0.01611	0.00934	0.04789	0.04725
Observations	185	185	185	185
Panel B. Dynamic Effects - Young				
Log City Size	0.1465	0.0443	0.5532***	0.5762***
	(0.1055)	(0.1627)	(0.1934)	(0.2138)
R^2	0.00816	0.00419	0.05468	0.05458
Observations	185	185	185	185
First Step Variables				
Age, Tenure	✓	✓	✓	✓
Sec & occup indicators	✓	✓	✓	✓
Worker FE	✓	✓	✓	✓

Notes: OLS results refer to Column (3) of Table 2 and Column (2) of Table 3. Coefficients are reported with robust standard errors in parenthesis. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels.

A.5 Alternative Education Classification

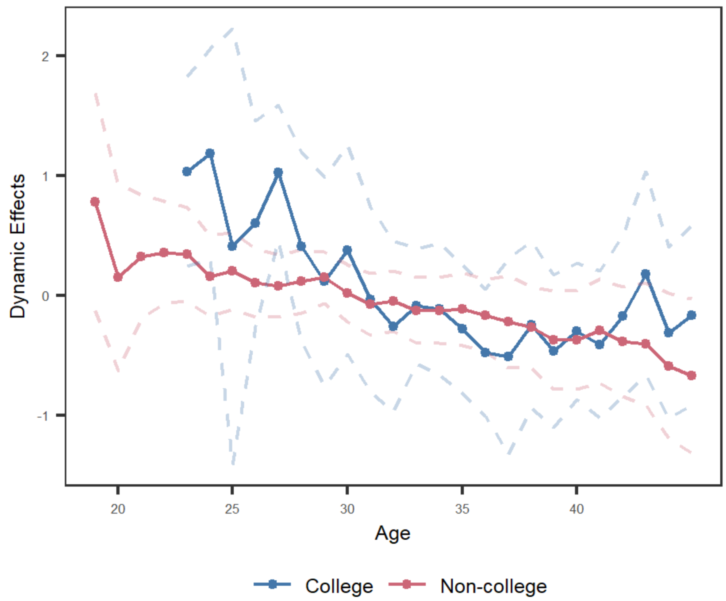
In this section, I replicate the estimations presented in Figure 3 and Tables 3, 5 and 6 using an education classification based on the most observed value between 2003 and 2017.

Table A5: Summary Statistics - Alternative Education Classification

	Overall			By city		
Variable	All Workers	College	Non-college	Median	Min	Max
	(1)	(2)	(3)	(4)	(5)	(6)
Observations ('000)	158,791	12,459	146,332	259	7	32,044
College (%)	7.8	0	0	5	1.5	12.3
Non-college (%)	92.2	0	0	95	87.7	98.5
Avg. age (years)	30.7	32.6	30.5	30.4	29.2	31.6
Up to 30 yrs old (%)	51.6	40.3	52.5	53.3	46.5	61.3
More than 30 yrs old (%)	48.4	59.7	47.5	46.7	38.7	53.5
Non-movers (%)	75.8	74.2	76	72.2	50.8	88
Migrants (%)	24.2	25.8	24	27.8	12	49.2
Avg. Wage (R$)	2,409	7,566	1,970	1,913	1,203	5,721
Avg. Wage - Non-movers (R$)	2,350	7,314	1,937	1,859	1,140	6,047
Avg. Wage - Migrants (R$)	2,596	8,289	2,075	2,042	1,352	5,381
(Within-city) Job Changes (%)	16.4	10.3	16.9	13.6	3.2	19.2
Job Changes - Non-movers (%)	17.2	11	17.7	14.1	3.5	20
Job Changes - Migrants (%)	13.9	8.3	14.4	12.7	2	18.6
# Workers ('000)	33,139	2,411	30,728	72	3	7,543

Notes: This table exhibits summary statistics of the RAIS sample using an alternative education classification. See the text for more details. Wages are in 2010 reais.

Figure A4: Dynamic Effects by Age - Alternative Education Classification



Notes: Each dot represents the elasticity of log wage growth with respect to city size for a specific educational-age group. Control variables include a quadratic polynomial in tenure and worker fixed effects. The dashed lines inform robust 95% confidence intervals.

Table A6: Dynamic Effects - Young College Workers - Alternative Education Classification

	City Dynamic Premium ($\delta^s_{c,young} \times 100$)			
	(1)	(2)	(3)	(4)
Panel A. Non-college				
Log City Size	-0.0167	0.1159	0.0242	-0.0124
	(0.0774)	(0.1388)	(0.1779)	(0.1830)
R^2	0.00026	0.00403	0.00015	0.00004
Observations	185	185	185	185
Panel B. College				
Log City Size	0.7158***	0.5005	0.6225***	0.6076***
	(0.1251)	(0.3168)	(0.1674)	(0.1642)
R^2	0.17793	0.02832	0.05786	0.05743
Observations	185	185	185	185
First Step Variables				
Age, Tenure	✓	✓	✓	✓
Worker FE		✓	✓	✓
Exclude sec & occup moves			✓	✓
Sec & occup indicators				✓

Notes: The dependent variables are city-age fixed effects estimated from wage growth regressions according to Equation (8). Each column represents a different specification of the first-step estimation. "Young" refers to workers up to 30 years old, and "senior" refers to workers over 30. Age and tenure variables include a quadratic polynomial. Occupation and sector indicators refer to one-digit and two-digit level information, respectively. Coefficients are reported with robust standard errors in parenthesis. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels.

Table A7: Static Urban Wage Premium - Alternative Education Classification

	City Static Premium (Ψ_c^s)					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Non-college						
Log City Size	0.0236*	0.0180	0.0171*	0.0186**	0.0185**	0.0200**
	(0.0140)	(0.0123)	(0.0088)	(0.0091)	(0.0091)	(0.0091)
R^2	0.01440	0.01117	0.01868	0.02040	0.02127	0.02337
Observations	185	185	185	185	185	185
Panel B. College						
Log City Size	0.0640***	0.0563***	0.0258***	0.0324***	0.0167*	0.0232**
	(0.0189)	(0.0137)	(0.0081)	(0.0085)	(0.0086)	(0.0092)
R^2	0.05222	0.08531	0.04604	0.06271	0.01878	0.03187
Observations	185	185	185	185	185	185
First Step Variables						
Dep. Var.: log wage net of	-	-	-	$\sum_\tau \hat{\delta}^s_{c,a(i,\tau)}$	$\hat{\delta}_i e_{it}$	Both
Age, Tenure	✓	✓	✓	✓	✓	✓
Sec & occup indicators		✓	✓	✓	✓	✓
Worker FE			✓	✓	✓	✓

Notes: The dependent variables are city fixed effects estimated from wage regressions according to Equation (9). Each column represents a different specification of the first-step estimation. Age and tenure variables include a cubic polynomial. Occupation and sector indicators refer to one-digit and two-digit level information, respectively. Coefficients are reported with robust standard errors in parenthesis. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels.

Table A8: Unobserved ability and city size - Alternative Education Classification

	City average unobserved ability ($\bar{\alpha}_c$)			
	(1)	(2)	(3)	(4)
Panel A. Non-college				
Log City Size	0.0026	0.0013	-0.0046	-0.0074
	(0.0059)	(0.0078)	(0.0107)	(0.0046)
R^2	0.00105	0.00014	0.00116	0.01435
Observations	185	185	185	185
Panel B. College				
Log City Size	0.0378***	0.0181	0.0226*	0.0015
	(0.0141)	(0.0209)	(0.0135)	(0.0104)
R^2	0.03483	0.00477	0.01472	0.00009
Observations	185	185	185	185
First Step Variables				
Dep. Var.: log wage net of	-	$\sum_{\tau} \hat{\delta}^s_{c,a(i,\tau)}$	$\hat{\delta}_i e_{it}$	Both
Age, Tenure	✓	✓	✓	✓
Sec & occup indicators	✓	✓	✓	✓
Worker FE	✓	✓	✓	✓

Notes: The dependent variables are city-level averages of worker fixed effects estimated from wage regressions according to Equation (9). Each column represents a different specification of the first-step estimation. Age and tenure variables include a cubic polynomial. Occupation and sector indicators refer to one-digit and two-digit level information, respectively. Coefficients are reported with robust standard errors in parenthesis. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels.

B 1st-step Estimations

Table B1: Standard Static Model - 1st step estimation

	Log Wage				
	(1)	(2)	(3)	(4)	(5)
Panel A. Non-college					
Age		0.0012	0.0114***		
		(0.0017)	(0.0014)		
Age^2		-0.0198***	-0.0230***	-0.0046***	-0.0045***
		(0.0013)	(0.0011)	(0.00006)	(0.00006)
Age^3		0.0268***	0.0182***	0.0002***	0.0001***
		(0.0002)	(0.0002)	(0.000002)	(0.000002)
Tenure		0.4479***	0.3824***	0.1537***	0.1474***
		(0.0244)	(0.0208)	(0.0033)	(0.0030)
Tenure^2		-0.0331*	-0.0279*	-0.0066**	-0.0061**
		(0.0178)	(0.0150)	(0.0032)	(0.0029)
Tenure^3		0.0002*	0.0002*	0.00004*	0.00004*
		(0.0001)	(0.0001)	(0.00002)	(0.00002)
R^2	0.10992	0.25217	0.35765	0.77171	0.77797
Observations	53,169,047	53,169,047	53,169,047	53,169,047	53,169,047
Panel A. College					
Age		0.1462***	0.1293***		
		(0.0014)	(0.0011)		
Age^2		-0.0109***	-0.0659***	-0.0132***	-0.0127***
		(0.0019)	(0.0015)	(0.0001)	(0.0001)
Age^3		0.0783***	0.0227***	0.0003***	0.0002***
		(0.0007)	(0.0005)	(0.000004)	(0.000004)
Tenure		1.133***	0.7503***	0.3310***	0.3157***
		(0.0061)	(0.0032)	(0.0012)	(0.0011)
Tenure^2		-0.4440***	-0.2896***	-0.1453***	-0.1373***
		(0.0091)	(0.0044)	(0.0015)	(0.0013)
Tenure^3		0.0402***	0.0243***	0.0115***	0.0107***
		(0.0030)	(0.0014)	(0.0005)	(0.0004)
R^2	0.08248	0.29560	0.51956	0.88850	0.89402
Observations	25,838,673	25,838,673	25,838,673	25,838,673	25,838,673
City FE	✓	✓	✓	✓	✓
Sec & occup indicators			✓		✓
Worker FE				✓	✓

Notes: Dependent variable is the log wage. Each column exhibits the results of estimating a particular version of Equation (4). Occupation and sector indicators refer to one-digit and two-digit level information, respectively. All specifications include year fixed-effects. Coefficients are reported with standard errors in parenthesis clustered by worker. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels.

Table B2: Dynamic Effects by Age - 1st step estimation

	Δ Log Wage ×100	
	College	Non-college
	(1)	(2)
Tenure	-0.8531***	-3.643***
	(0.2107)	(0.0430)
Tenure^2	2.000***	0.0421***
	(0.1662)	(0.0150)
R^2	0.17651	0.17722
Observations	17,588,577	36,381,355
City-age FE	✓	✓
Worker FE	✓	✓

Notes: Dependent variable is the yearly variation in log wage. Each column exhibits the results of estimating Equation (6) for a specific educational group. Coefficients are reported with standard errors in parenthesis clustered by worker. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels.

Table B3: Dynamic Effects by Age Group - 1st step estimation

	Δ Log Wage $\times 100$			
	(1)	(2)	(3)	(4)
Panel A. Non-college				
Age	-0.9546***			
	(0.0142)			
Age^2	0.2141***	0.0679***	-0.0189***	-0.0096***
	(0.0077)	(0.0018)	(0.0026)	(0.0026)
Tenure	0.5736***	-3.800***	-8.508***	-8.616***
	(0.0114)	(0.0430)	(0.1797)	(0.1824)
Tenure^2	-0.0054***	0.0435***	4.107***	4.165***
	(0.0018)	(0.0153)	(0.1578)	(0.1602)
R^2	0.01095	0.18178	0.33537	0.33564
Observations	36,381,355	36,381,355	27,785,836	27,785,836
Panel B. College				
Age	-2.059***			
	(0.0248)			
Age^2	1.205***	0.1913***	0.1003***	0.1206***
	(0.0230)	(0.0045)	(0.0051)	(0.0051)
Tenure	2.315***	-1.021***	-4.728***	-4.681***
	(0.0535)	(0.2141)	(0.2678)	(0.2688)
Tenure^2	-1.044***	2.060***	2.963***	2.965***
	(0.0283)	(0.1690)	(0.2019)	(0.2025)
R^2	0.01527	0.18045	0.31059	0.31096
Observations	17,588,577	17,588,577	13,513,300	13,513,300
City-age FE	✓	✓	✓	✓
Worker FE		✓	✓	✓
Exclude sec & occup moves			✓	✓
Sec & occup indicators				✓

Notes: Dependent variable is the yearly variation in log wage. Each column exhibits the results of estimating a particular version of Equation (8). Occupation and sector indicators refer to one-digit and two-digit level information, respectively. Coefficients are reported with standard errors in parenthesis clustered by worker. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels.

Table B4: Full Model Non-college Graduates - 1st step estimation

	Log Wage net of...					
	Nothing			$\sum_{\tau} \hat{\delta}^{s}_{c,a}$	$\hat{\delta}_i e_{it}$	Both
	(1)	(2)	(3)	(4)	(5)	(6)
Age	-0.0321**	-0.0930***				
	(0.0134)	(0.0127)				
Age^2	-0.0599***	-0.0986***	-0.0052***	0.0548***	-0.0596***	-0.0010
	(0.0086)	(0.0082)	(0.0008)	(0.0008)	(0.0010)	(0.0010)
Age^3	0.0139***	0.0011	0.0002***	0.0003***	0.000005	0.0002***
	(0.0018)	(0.0017)	(0.00002)	(0.00002)	(0.00002)	(0.00002)
Tenure	0.5275***	0.4594***	0.0867***	0.0618***	-0.0452***	-0.0657***
	(0.0040)	(0.0048)	(0.0128)	(0.0153)	(0.0096)	(0.0074)
Tenure^2	-0.0286***	-0.0151*	0.0854***	0.1064***	-0.0690***	-0.0505***
	(0.0067)	(0.0080)	(0.0237)	(0.0282)	(0.0175)	(0.0135)
Tenure^3	0.0003***	0.0001	-0.0010***	-0.0012***	0.0008***	0.0006***
	(0.00007)	(0.00009)	(0.0003)	(0.0003)	(0.0002)	(0.0001)
R^2	0.23049	0.31188	0.69232	0.66615	0.79332	0.79547
Observations	23,210,815	23,210,815	23,210,815	23,210,815	21,661,508	21,661,508
Sec & occup indicators		✓	✓	✓	✓	✓
City FE	✓	✓	✓	✓	✓	✓
Worker FE			✓	✓	✓	✓

Notes: Dependent variable is the log wage subtracted by the term indicated at the top of the columns. Each column exhibits the results of estimating a particular version of Equation (9). Occupation and sector information are aggregated to one-digit and two-digit level, respectively. Coefficients are reported with standard errors in parenthesis clustered by worker. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels.

Table B5: Full Model College Graduates - 1st step estimation

	Log Wage net of...					
	Nothing			$\sum_{\tau} \hat{\delta}^{s}_{c,a}$	$\hat{\delta}_i e_{it}$	Both
	(1)	(2)	(3)	(4)	(5)	(6)
Age	0.5995***	0.2727***				
	(0.0160)	(0.0141)				
Age^2	0.4383***	0.0565***	-0.0190***	0.0465***	-0.0678***	-0.0026**
	(0.0138)	(0.0122)	(0.0009)	(0.0010)	(0.0010)	(0.0010)
Age^3	0.2172***	0.0636***	0.00008***	-0.0004***	0.0005***	0.00002
	(0.0038)	(0.0034)	(0.00003)	(0.00003)	(0.00003)	(0.00003)
Tenure	1.289***	0.8351***	0.3529***	0.3719***	0.0360***	0.0548***
	(0.0031)	(0.0026)	(0.0016)	(0.0016)	(0.0015)	(0.0015)
Tenure^2	-0.6907***	-0.4279***	-0.2224***	-0.2090***	-0.0191***	-0.0067***
	(0.0040)	(0.0034)	(0.0018)	(0.0018)	(0.0015)	(0.0015)
Tenure^3	0.0532***	0.0326***	0.0169***	0.0156***	0.0012***	-0.00002
	(0.0008)	(0.0005)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
R^2	0.23180	0.47164	0.86101	0.87694	0.89371	0.89113
Observations	13,172,224	13,172,224	13,172,224	13,172,224	12,699,042	12,699,042
Sec & occup indicators		✓	✓	✓	✓	✓
City FE	✓	✓	✓	✓	✓	✓
Worker FE			✓	✓	✓	✓

Notes: Dependent variable is the log wage subtracted by the term indicated at the top of the columns.Each column exhibits the results of estimating a particular version of Equation (9). Occupation and sector information are aggregated to one-digit and two-digit level, respectively. Coefficients are reported with standard errors in parenthesis clustered by worker. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels.

Table B6: Dynamic Effects by Progression Type - 1st step estimation

	Δ Log Wage ×100			
	College		Non-college	
	(1)	(2)	(3)	(4)
Age^2	0.5523***	0.2779***	0.2608***	0.0326***
	(0.0115)	(0.0110)	(0.0115)	(0.0109)
Tenure	4.849***	1.600***	-3.423***	-5.994***
	(0.3255)	(0.2505)	(0.7750)	(0.6463)
$Tenure^2$	0.6106*	0.5547**	3.233***	2.886***
	(0.3337)	(0.2462)	(1.145)	(0.9344)
R^2	0.18302	0.18225	0.20020	0.19963
Observations	9,362,828	9,362,828	15,225,261	15,225,261
Worker FE	✓	✓	✓	✓
City-progression FE (general count)	✓		✓	
City-progression FE (count by city)		✓		✓

Notes: Dependent variable is the yearly variation in log wage. Each column exhibits the results of estimating a particular version of Equation (10) for different educational groups. Coefficients are reported with standard errors in parenthesis clustered by worker. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels.

C Other Tables and Figures

Table C1: City size and changes in city-level variables

	Δ Log Avg. Wage		Δ Log Employment	
	College	Non-college	College	Non-college
	(1)	(2)	(3)	(4)
Log City Size	-0.1710	-0.2248***	-0.1126	-0.6141**
	(0.1348)	(0.0807)	(0.1752)	(0.2637)
R^2	0.17747	0.36708	0.24419	0.47678
Observations	1,480	1,480	1,480	1,480

Notes: Each column exhibits the city-size elasticity of changes in average wages and formal employment for different educational groups. Coefficients are reported with robust standard errors in parenthesis. All regressions include year fixed-effects/ *, ** and *** indicate statistical significance at the 1, 5 and 10% levels.

Table C2: Dynamic Effects - no Age Heterogeneity

	City dynamic premium (δ_c^s)			
	(1)	(2)	(3)	(4)
Panel A. Non-college				
Log City Size	-0.0674	0.0495	0.0324	-0.0070
	(0.0684)	(0.0936)	(0.1145)	(0.1214)
R^2	0.00618	0.00121	0.00038	0.00002
Observations	185	185	185	185
Panel B. College				
Log City Size	0.2257***	0.2055	0.2140	0.2088
	(0.0594)	(0.1651)	(0.1631)	(0.1645)
R^2	0.06701	0.01091	0.01106	0.01044
Observations	185	185	185	185
First Step Variables				
Age, Tenure,	✓	✓	✓	✓
Worker FE		✓	✓	✓
Exclude sec & occup moves			✓	✓
Sec & occup indicators				✓

Notes: The dependent variables are city fixed effects estimated from wage growth regressions according to Equation (8), abstracting from age heterogeneities. Each column represents a different specification of the first-step estimation (not reported). Age and tenure variables include a quadratic polynomial. Occupation and sector indicators refer to one-digit and two-digit level information, respectively. Coefficients are reported with robust standard errors in parenthesis. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels.

Table C3: Dynamic Effects Within and Between Jobs - Count by City

	Within				Between		
	1st	2nd	3rd	4th+	1st-2nd	2nd-3rd	3rd-4th+
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. Non-college							
Log City Size	0.0309	-0.1031	-0.1195	-0.2296	0.0864	-0.0044	-0.1763
	(0.1461)	(0.1668)	(0.1952)	(0.3110)	(0.2304)	(0.2496)	(0.3132)
R^2	0.00022	0.00180	0.00163	0.00267	0.00066	0.000001	0.00144
Observations	183	183	183	183	183	183	183
Panel B. College							
Log City Size	0.5694***	0.4528**	0.5342**	0.2258	0.9447***	0.5048	0.6029
	(0.1469)	(0.1963)	(0.2622)	(0.4139)	(0.3108)	(0.3472)	(0.5355)
R^2	0.05885	0.01849	0.01614	0.00141	0.03111	0.00779	0.00495
Observations	176	176	176	176	176	176	176

Notes: The dependent variables are city-progression fixed effects estimated from wage growth regressions according to Equation (10). Each column exhibits the city-size elasticity for different progression types. For more details, see Section 4. Controls in the first-step estimations include worker fixed effects and a quadratic polynomial in age and tenure. Coefficients are reported with robust standard errors in parenthesis. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels.

Table C4: Long-Run Wage Premium - Non-college Graduates (19-26 yrs old)

	City premium (δ_c^{LT})	City-career premium ($\delta_{c,J}^{LT}$)			
# Firms	-	0	1	2	3+
	(1)	(2)	(3)	(4)	(5)
Panel A. Only Year FE					
Log City Size	-0.0112	0.0044	-0.0062	-0.0078	-0.0109
	(0.0070)	(0.0077)	(0.0087)	(0.0080)	(0.0075)
R^2	0.01124	0.00122	0.00253	0.00435	0.00873
Observations	171	171	171	171	171
Panel B. Drop Occupation Changes					
Log City Size	-0.0161**	0.0017	-0.0085	-0.0079	-0.0171*
	(0.0080)	(0.0088)	(0.0095)	(0.0089)	(0.0088)
R^2	0.01890	0.00015	0.00382	0.00354	0.01711
Observations	171	171	171	171	171
Panel C. Drop Occupation Changes + Occupation Dummies					
Log City Size	-0.0151**	-0.0020	-0.0095	-0.0077	-0.0158*
	(0.0076)	(0.0083)	(0.0089)	(0.0085)	(0.0085)
R^2	0.01847	0.00022	0.00544	0.00371	0.01583
Observations	171	171	171	171	171
Panel D. Drop Sector Changes					
Log City Size	-0.0159*	-0.0095	-0.0098	-0.0098	-0.0147*
	(0.0082)	(0.0147)	(0.0099)	(0.0087)	(0.0079)
R^2	0.01723	0.00179	0.00478	0.00582	0.01380
Observations	171	170	171	171	171
Panel E. Drop Sector Changes + Sector Dummies					
Log City Size	-0.0119*	-0.0115	-0.0063	-0.0086	-0.0125*
	(0.0067)	(0.0149)	(0.0080)	(0.0070)	(0.0069)
R^2	0.01476	0.00266	0.00307	0.00681	0.01289
Observations	171	170	171	171	171
Panel F. Only Year FE - Adjusted Wages					
Log City Size	0.0147	0.0037	0.0103	0.0138	0.0130
	(0.0123)	(0.0134)	(0.0127)	(0.0128)	(0.0127)
R^2	0.00576	0.00031	0.00275	0.00478	0.00421
Observations	171	171	171	171	161

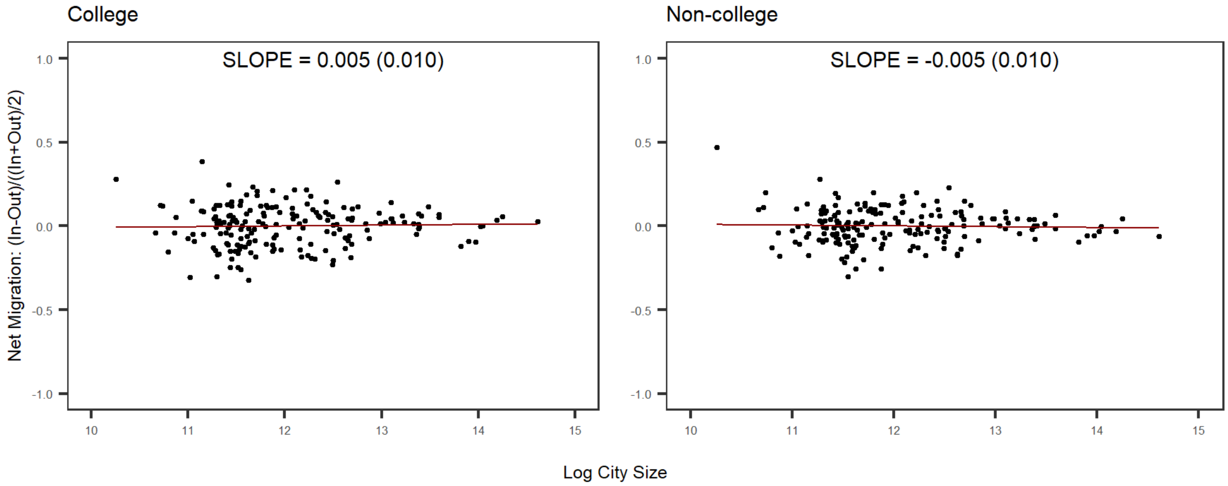
Notes: The dependent variables are city fixed effects estimated from Equation (12). Column (1) exhibits the city-size elasticity when city premiums are assumed to be constant across career paths, whereas Columns (2) to (5) exhibit city-size elasticities differentiating by the number of firms the individual worked for during the period analyzed. Each panel represents a different specification and sample used in the first-step estimation. In Panel F, wages are adjusted for individual heterogeneity in returns to experience. See Section 4.2 for more details. Coefficients are reported with robust standard errors in parenthesis. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels.

Table C5: Sorting and Career Paths - Non-college Graduates (19-26 yrs old)

	City	City-career			
# Firms	-	0	1	2	3+
	(1)	(2)	(3)	(4)	(5)
Panel A. Avg. Returns to Experience ($\bar{\delta}_{c,J}$)					
Log City Size	-0.3306***	0.0332	-0.1912	-0.2680**	-0.3090**
	(0.1184)	(0.1531)	(0.1418)	(0.1256)	(0.1191)
R^2	0.03071	0.00023	0.00887	0.01604	0.02445
Observations	171	171	171	171	171
Panel B. Avg. Number of Observations ($\bar{n}^*_{c,J}$)					
Log City Size	-0.0326*	0.0259**	-0.0398	-0.0563***	-0.0164
	(0.0178)	(0.0107)	(0.0270)	(0.0189)	(0.0140)
R^2	0.01685	0.02465	0.01052	0.04061	0.00581
Observations	171	171	171	171	171
Panel C. Share of Individuals					
Log City Size		-0.0264***	-0.0208***	0.0033	0.0440***
		(0.0038)	(0.0040)	(0.0020)	(0.0069)
R^2		0.14626	0.11697	0.01165	0.16433
Observations		171	171	171	171
Panel D. Avg. Initial Ability ($\bar{\alpha}_{c,J}$)					
Log City Size	0.0128	0.0138	0.0118	0.0156*	0.0145
	(0.0080)	(0.0097)	(0.0085)	(0.0082)	(0.0088)
R^2	0.01434	0.01094	0.01110	0.01853	0.01346
Observations	171	171	171	171	171

Notes: The dependent variables are indicated in the respective panels. Column (1) exhibits the city-size elasticity when the dependent variable is computed at the city level, whereas Columns (2) to (5) exhibit city-size elasticities differentiating by the number of firms the individual worked for during the period analyzed. Coefficients are reported with robust standard errors in parenthesis. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels.

Figure C1: Net migration and city size



Notes: Each dot represents the average net migration for a city, computed as $NetMigr = \frac{InMigr - OutMigr}{(InMigr + OutMigr)/2}$. Coefficients are reported with robust standard errors in parenthesis.

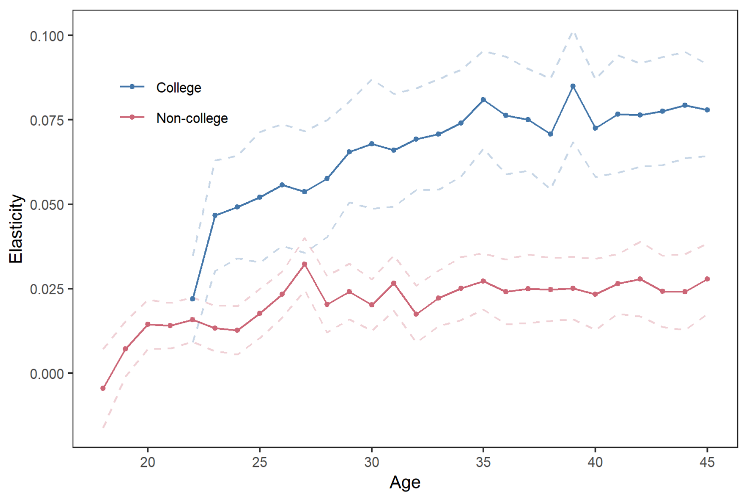
D External Validity

A relevant question is whether the conclusions of this study can be extended to other countries, particularly developed economies. Specifically for the US, I replicate the procedure in Figure 1 to assess how wages vary with city size for different educational-age groups in different Metropolitan Statistical Areas (MSA). For this purpose, I use data from the American Community Survey (ACS) between 2010 and 2017. Since there are not enough observations to compute mean wages for each educational-age group in each MSA, I regress wages directly on log city size, measured as the MSA population.

Figure D1 shows a very similar pattern, with city-size elasticities increasing modestly over working life for non-college graduates but substantially for college graduates. Most of the growth occurs between 22 and 35 years old, similar to Figure 1. There seems to be a relatively higher disparity between the urban wage premium of young college and non-college graduates. Overall, this evidence suggests that the context in the United States is qualitatively similar to Brazil.

It is also worth to mention studies performing cross-country comparisons. Chauvin et al. (2017) studies Brazil, China, India and the United States and find that Brazil has a relatively weaker relationship between income and density. They also show a stronger relationship between income and the share of adults with a college degree, suggesting that non-college graduates in Brazil benefit relatively less from city size but relatively more from the presence of college graduates. Regarding wage dynamics, Lagakos et al. (2018) show that the wage profile of Brazilian workers is relatively steeper compared to other developing economies but falls slightly behind developed economies.

Figure D1: City size elasticity by educational level and age - ACS



Notes: Each dot informs the coefficient of a regression of log individual wages on log MSA population for a specific educational-age group. Data is from the American Community Survey between 2010 and 2017. The dashed lines inform the 95% confidence intervals.

E Theoretical Framework

This section describes how Equation 1 can be obtained from a standard partial equilibrium model. Consider a set of cities indexed by c, each one producing a certain amount of a homogeneous good Y_c according to the following Cobb-Douglas technology

$$Y_c = L_c^{\gamma} K_c^{1-\gamma} \ , \tag{E.1}$$

where L_c and K_c are labor and non-labor inputs and $0 < \gamma < 1$. In this expression, labor is a CES aggregation of high- and low-skill labor. Each group has a productivity A_c^s, $s = \{h, l\}$, which is city-specific, according to the following equation:

$$L_c = \left[(A_c^h L_c^h)^{\rho} + (A_c^l L_c^l)^{\rho} \right]^{1/\rho} \ ,$$

where L_c^s is the total labor of skill group s in city c and $0 < \rho < 1$. Furthermore, workers are heterogeneous based on their personal productivity. Each labor input is thus computed in efficiency units:

$$L_c^s = \sum_{i \in S} z_i l_i \ .$$

Here z_i represents the worker's i productivity and l_i is the number of hours worked. Profit maximization yields an expression for the wage $w_{i,c}^s$ of worker i of skill group s in city c:

$$w_{i,c}^s = \frac{\gamma(1-\gamma)^{\frac{1-\gamma}{\gamma}} p_c^{\frac{1}{\gamma}}}{r_c^{\frac{1-\gamma}{\gamma}}} \left[\sum_{s'} (A_c^{s'} L_c^{s'})^{\rho} \right]^{\frac{1-\gamma}{\gamma}} \frac{(A_c^s)^{\rho}}{(L_c^s)^{1-\rho}} z_i \equiv B_c^s z_i \ ,$$

where r_c is the price of non-labor factors, and p_c is the output price. Hence, the equilibrium wages are a combination of two factors: one associated with city-group variables (B_c^s) and the other associated with personal productivity (z_i).

By taking the natural logarithm on both sides and assuming that this relationship holds in every period, the previous expression can be written as

$$\log w^s_{i,c,t} = \Psi^s_c + h_{i,t} \ ,$$

where Ψ and h are the log counterparts of B and z, respectively.