The Urban Wage Premium Over the Life Cycle

Thiago Patto*

October 4, 2023

Abstract

I use Brazilian matched employer-employee data to study how city size impacts the wage profile of college and non-college graduates, focusing on the influence on wage growth (dynamic effects). The data size allows me to fully explore variations at the city level and estimate city size effects for different educational-age groups. Using a reduced-form approach, I show that the static urban wage premium does not differ significantly across educational groups, but dynamic effects are concentrated on young college graduates and explain 46% of their city-size wage gap. This asymmetry between college and non-college graduates also accounts for 62% of the city-size college premium gap. In contrast, the results indicate that sorting on unobserved characteristics within education groups is of minor importance to understanding both differences in wage levels and wage dynamics. Finally, I present evidence that job transitions have a close connection with wage growth in large cities. Not only do job transitions amplify dynamic effects, but staying in the same job in a large city does not provide any future wage gains. This evidence challenges interpretations of city size effects on wage growth based exclusively on faster human capital accumulation.

Keywords: Agglomeration economies, Urban wage premium, Life cycle, Wage growth, College premium, Job transitions

*Insper, São Paulo, *thiagospms@al.insper.edu.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 more accurate description of this phenomenon. For example, it remains an open question how the relationship between wage growth and city size evolves 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 the same pattern? Additionally, it is unclear how these effects compare in terms of magnitude with the static gains experienced by individuals once they move to a larger city.

This paper overcomes this limitation by analyzing how city-size dynamic effects (henceforth simply dynamic effects or dynamic gains) shape the wage profiles of college and non-college graduates and their importance in explaining both the city-size wage gap and the city-size college premium gap. From Brazilian matched employer-employee data, I construct a dataset with 140 million observations across 185 urban areas that allows me to fully explore differences at the city level while considering life cycle variations.

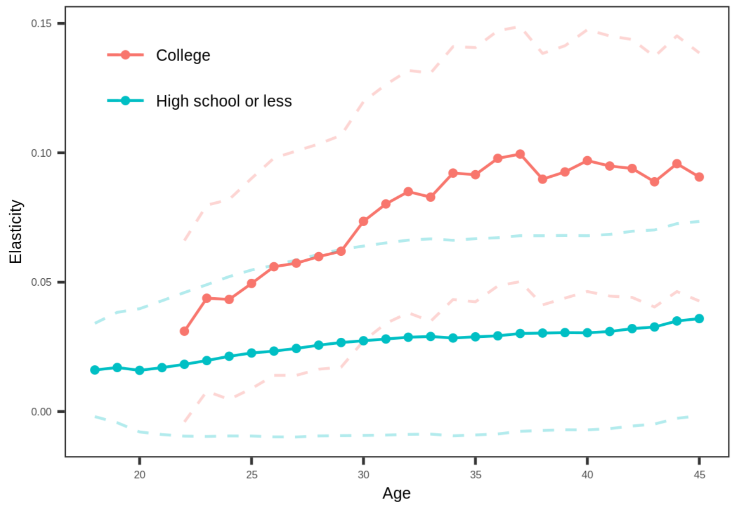
Focusing on city size and wage dynamics is particularly relevant because it can be informative about the nature of agglomeration economies. Unlike the static gains workers experience when moving from a small to a large city, dynamic effects offer a closer look at how geographic proximity impacts individuals. Traditionally, the literature has revolved around two explanations: city size influences wage growth either through improvements in human capital accumulation or improvements in firm-worker matching (the ”learning” and ”matching” hypotheses, respectively). This paper aims to shed some light on this subject by exploring the interplay between wage growth, job transitions and city size.

A preliminary analysis of the data provides some interesting facts. Figure 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 significant for college graduates, indicating a widening city-size wage gap within this group during the age range of 22 to 35 years old. Additionally, Figure 1 shows that the city-size wage gap does not differ substantially between college and non-college graduates in their first years of working life, meaning that the correlation between college premium and city size is initially weak

[1] For surveys, see Combes and Gobillon (2015) and Ahfeldt and Pietrostefani (2019).

[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.

but increases with age.[3]

To illustrate the interaction between wage growth, job transitions and city size, Figure 2 plots the average wage profile of college graduates for different quartiles of city size and different number of job transitions observed during the interval analyzed. A comparison of the curves within each graph shows that wage profiles with job transitions are relatively steeper in large cities. Moreover, by comparing the same curve type in different graphs, we see that wage profiles with no job transitions are relatively similar in all quartiles, whereas those with at least one job transition become steeper as the city size increases.[4]

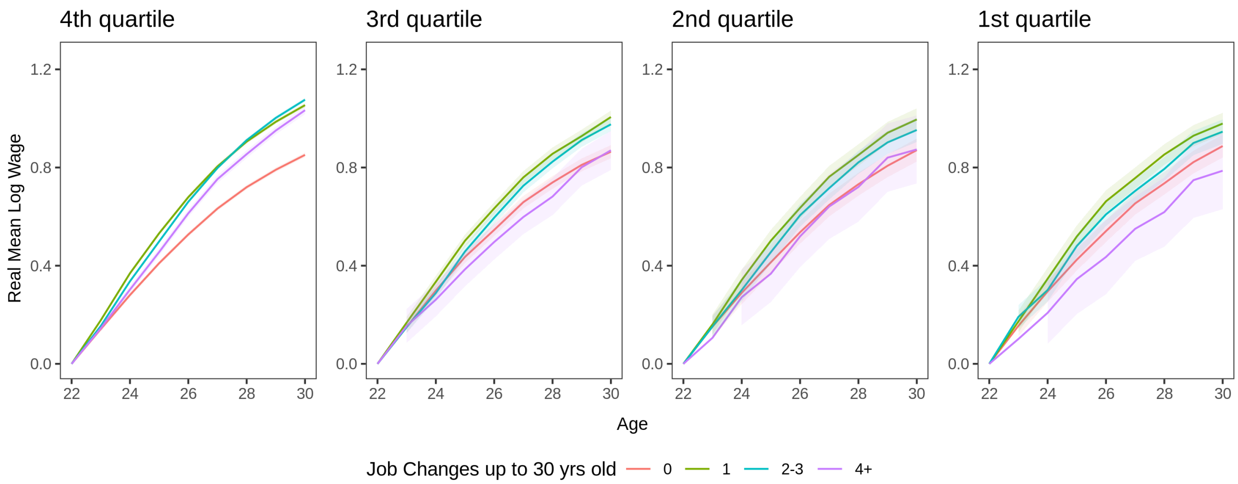
Building on previous literature, I developed a reduced-form approach to delve deeper into the patterns depicted in Figures 1 and 2. I start the empirical analysis with a standard estimation of the urban wage premium that ignores dynamic elements. This procedure helps me construct the arguments presented in this paper and establish a connection with prior studies. I regress log wages on city fixed-effects, worker and job characteristics and then regress the estimated city fixed-effects on log city size, measured as the" experienced density" (Duranton and Puga, 2020). The estimation is performed separately for college and non-college graduates and yields, for each educational group, its static urban wage premium.

One issue that has received particular attention in the literature in recent years is the role of unob-

[3] In Figure 6, I present a similar exercise using data from the American Community Survey (ACS), showing that this pattern is not an exclusive feature of Brazil. There are two important differences between the two figures. First, the ACS does not have enough observations to compute mean wages for each educational-age group by Metropolitan Statistical Area, so I directly regress individual wages on log population and year dummies. Secondly, Figure 6 uses the MSA log population to measure city size, whereas Figure 1 uses the "experienced density" concept explained in Section 2.2.

[4] The analysis considers non-migrants observed in all periods from age 22 to 30 years old.

Figure 2: Wage Profiles, Job Transitions and City Size



Notes: Each curve shows the average log wage profile of college workers who experience different number of job transitions. All curves are subtracted by the average initial log wage to highlight relative differences. The analysis uses observations between 2003 and 2017 from non-migrants observed in all periods from age 22 to 30 years old. Colored shadows indicate 95% confidence intervals.

served worker characteristics in explaining the city-size wage gap. This channel is typically assessed by including worker fixed-effects in a wage regression. Controlling only for observed worker characteristics and sectoral/occupational composition, the elasticities of wage premium with respect to city size obtained are 0.019 and 0.066 for non-college and college workers, respectively. When I add worker fixed-effects, the point estimate slightly decreases to 0.016 for the non-college but drops substantially to 0.022 for college graduates. Intuitively, it means that individual unobserved ability accounts for more than half of the city-size wage gap of workers with a college degree.

This exercise reveals that a relevant portion of the city-size wage gap of college graduates is due to an over-representation of more skilled workers in large cities. However, it says little about how these differences arise. De La Roca and Puga (2017) show evidence that, in this setting, the fixed effects capture the more valuable experience that workers accrue over time in large cities. To put it differently, the fixed effects absorb the accumulated dynamic effects.

I investigate this hypothesis by estimating the effects of city size on wage growth, measured as the annual variation in wages. The strategy is built on D'Costa and Overman (2014) but considers the possibility of dynamic effects varying over the working life, as discussed earlier. It consists of regressing log wage growth on city-age fixed-effects, worker fixed-effects and other time-varying variables and then regressing the estimated fixed effects on city size for each age group. Again, this procedure is performed separately for college and non-college graduates.

The results show that large cities affect the wage growth of young college graduates. There is an initial positive relationship between city-age fixed effects and city size that decays 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 be relevant. In a simpler specification in which city-age heterogeneities are limited to only two age groups - up to 30 years old and above - I find highly significant city-size effects for the younger group.

The next step is to evaluate the relevance of dynamic effects in explaining the city-size wage gap. Based on the conceptual framework proposed, I derive an equation for estimation similar to the static specification, except that now the dependent variable is the log wage net of all previously accumulated city-specific returns to experience. The goal is to analyze to what extent introducing dynamic effects in the econometric model changes the results obtained previously. I use estimates of the wage growth equation to calibrate the parameters and compute this new variable.[5]

I find that once dynamic effects are taken into account, the unobserved ability of college graduates becomes uncorrelated with city size. In other words, large cities have more skilled workers because they "manufacture" them. Moreover, dynamic effects largely explain why big cities are more beneficial for college graduates than their non-college counterparts. A back-of-the-envelope calculation indicates that dynamic effects account for 46% of their city-size wage gap and 62% of the city-size college premium gap.

Finally, I study the relationship between dynamic effects and job transitions. I first investigate the short-run benefits of job transitions by examining how dynamic effects vary within and between firms. Departing from the wage growth regression described above, I change the specification to include city fixed-effects within and between jobs. I also enumerate workers' jobs and estimate specific parameters for each within- or between-job observation to capture decaying effects.

I find that college graduates experience faster wage growth during their initial job and in the first and later job transitions. The estimated elasticities indicate that doubling the city size increases wage growth by 0.6% yearly within the first job, by 1.8% in the first job change and by 2.3% in the third job change onwards. The results do not change qualitatively if the job count is unique or city-specific for migrants.

Another pertinent question is whether these workers accrue gains in the long run, i.e., if wage profiles in large cities are steeper when workers move across firms. In this case, I apply a different methodology focusing on wage growth between 23 and 30 years old. I regress this variable on city fixed-effects interacted with indicators of the number of job transitions during this period. I discuss in the paper the assumptions under which the parameters of interest are identified, which seem to hold in my setting.

The results reveal that the city-size long-term effects increase with the number of job transitions for

[5] In this case and subsequent regressions that investigate the role of job transitions, I use a sub-sample of workers whose location history can be entirely constructed with available data.

college graduates. Interestingly, workers with no job transitions have equally steep wage profiles regardless of the city where they locate, but workers with at least one job transition experience higher wage growth in large cities. This evidence indicates that in order to benefit from being in a large city, workers with a college degree need to transition across jobs.

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 dataset that allows a more accurate description of how wage profiles vary with city size. In particular, I extend the approaches of D'Costa and Overman (2014) and De La Roca and Puga (2017) to highlight age-based heterogeneities and the role of job transitions. Moreover, the two-step analysis focused on city fixed-effects provides a more robust estimation of the city-size effects, which has been standard in the literature for static premiums but not dynamic premiums.[6]

There is a small group of papers that studies firm-worker matching in cities. Some of them 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) approaches the relationship between wage growth, job 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 and focusing on both short- and long-run benefits of job transitions.

This study also relates to the 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 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). Urbanization is a major process in these regions, but 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

[6] 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.

size affects wages in developed economies, and in particular, 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. Section 3 discusses the conceptual framework. Section 4 provides classic estimates of the static urban wage premium and explains how the discussion about sorting fits into this context. Section 5 presents estimates of dynamic effects and how they vary by age and education. Section 6 analyzes the estimates of a model that includes static and dynamic gains and evaluates the relative importance of the latter. Section 7 investigates the role of job transitions to explain the relationship between wage growth and city size. Section 8 presents the robustness checks. Finally, Section 9 concludes.

2 Data

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 the majority of formal workers in the country, with the exception of domestic workers and other minor employment categories. I use data between 2010 and 2017 for the estimations and data from 2003 to 2017 to construct the workers' location history.

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, together with 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 information can change over time for the same worker due to incorrect inputs or upgrades in educational attainment. In 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 most frequent value observed between 2003 and 2017.[7]

The initial sample comprises the universe of male workers from 2010 to 2017 aged between 18 and 45 among non-college graduates and between 22 and 45 among those with higher education. Using this

[7] The same procedure is undertaken to determine workers' year of birth to obtain consistent age information.

Table 1: Summary Statistics

	Overall			By city		
Variable	All Workers	Non-movers	Migrants	Mean	Min	Max
Observations	142,243,066	110,057,061	32,186,005	768,881	7,011	28,522,407
Non-movers				594,903	5,832	23,350,010
Migrants				173,978	1,179	5,172,397
Age (years)	30.7	30.8	30.3	30.4	29.3	31.6
Age $\leqslant$ 30 yrs old (%)	51.1	50.4	53.6	52.9	46.1	60.6
Age > 30 yrs old (%)	48.9	49.6	46.4	47.1	39.4	53.9
High school of less (%)	92.9	93	92.6	95.2	88.8	98.7
College (%)	7.1	7	7.4	4.8	1.3	11.2
Wage (R$)	2,433	2,378	2,621	2,013	1,220	5,786
Wage - high school or less (R$)	2,001	1,969	2,113	1,813	1,150	4,507
Wage - college (R$)	8,081	7,808	8,961	5,755	3,040	16,452
Number of workers	31,846,491	26,070,844	5,775,647	172,143	2,171	6,012,619

information, I construct a yearly panel with one record per worker-year. In the case of multiple records for the same worker in the same year, I choose the job with the highest tenure. If there is a tie, I choose the one with the highest wage (as defined in the above paragraph). 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 other minor filters to exclude observations with invalid worker identifiers or some missing information.

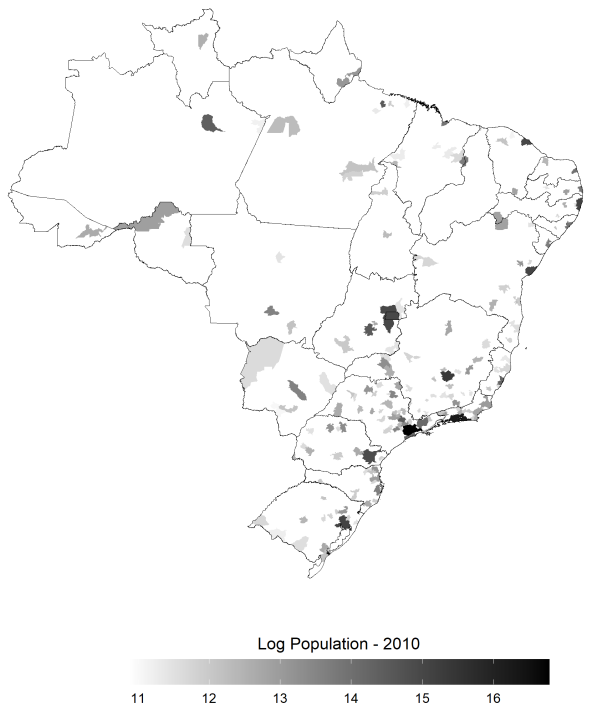
Table 1 shows summary statistics of the final sample, which contains about 142.2 million observations of 31.8 million workers. The table shows some values for the whole workforce and split between migrants — defined as workers who are observed at least in two different urban areas — and non-movers. Note that migrants have a slightly lower chance of dropping from the dataset due to unemployment or a transition to informality, since this group represents 18.2% of all individual identifiers but 22.6% of observations. Only 7.1% of the sample have a college degree, a fraction that does not change significantly among migrants. Higher education seems to have an enormous effect on wages since college graduates are paid, on average, almost four times more than non-college graduates. Among migrants, this ratio is even higher. Overall, the data show that migrants have more stability in the formal labor market, are more educated and earn higher wages.

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-

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

Figure 3: Urban Concentrations



consistent urban areas throughout the period analyzed.[9] Figure 3 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 in Brazil 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. 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).

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

[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%.

city weighting by the cell's population. The constructed measure has a correlation of 0.82 with a simple population count.[10]

3 Methodology

I consider that the log wage of a worker i in any period t can be written as a sum of a city static premium Ψ and an individual term h:

$$\log w^s_{i,c,t} = \Psi^s_c + h_{i,t} \ , \tag{1}$$

where the subscripts s and c represent the educational level of worker i (college or non-college) and a function $c(i,t)$ that maps worker i in period t to a specific city c, respectively. Note that the term Ψ^s_c is assumed to be constant over time. In Appendix B, I show how this expression can be rationalized from a simple partial equilibrium model in which college and non-college labor are different inputs in the production function.

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, so 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 $a(i,\tau)$, which represents the age 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

[10] The main results of this paper using a simple population count are qualitatively the same and quantitatively similar. In general, changing this variable yields slightly higher elasticities of wages with respect to city size, so the proposed measure is, in fact, more conservative.

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.[11] Secondly, the coefficients Ψ_c^s and $\delta_{c,a}^s$ 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 $\delta_{c,a}^s$ is that it is specifically related to the individual and therefore has a narrower interpretation compared to the static premium (See Appendix B for details).[12,13]

4 The Urban Wage Premium - Static Estimation

Before considering Equation (3), it is worth starting with a simpler framework to better connect with the literature. In this section, I abstract from dynamic effects and individual returns to experience and focus on a standard static 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 estimation of agglomeration effects follows a two-step procedure in the spirit of Combes et al. (2008). After estimating city fixed-effects in (4), I regress them on city size separately for each educational group:

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

where ν_c is an error at the city level. The term γ^s is the parameter of interest and informs the elasticity of city premium with respect to its size for educational group s.[14]

The estimation of Equation (4) delivers a unique solution if all cities are connected through migrants. Given the dataset size, this condition is easily achieved (See Table A2 for more information). The identification assumption is that migration across cities is exogenous conditional on the other variables included in the regression. While this is a strong assumption, it allows me to study the urban wage premium on a

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

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

[13] As shown in Appendix B, the term 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).

[14] 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 the estimated 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.

Table 2: Static Urban Wage Premium

	City wage premium (Ψ_c^s)		
	(1)	(2)	(3)
Panel A. Non-college			
Log City Size	0.028*	0.019	0.016*
	(0.016)	(0.014)	(0.008)
R^2	0.015	0.010	0.019
Panel B. College			
Log City Size	0.080***	0.066***	0.022***
	(0.020)	(0.015)	(0.007)
R^2	0.069	0.103	0.042
Observations	185	185	185
First Step Variables			
Age, Tenure	Yes	Yes	Yes
Sec & occup indicators	No	Yes	Yes
Worker FE	No	No	Yes

Notes: Results of the first-step estimation are reported in Table A3. 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.

large scale and fully explore city-level variations in the data. In Section 8.1, I investigate in more detail this assumption by performing an event study adapted from Card et al. (2013). The evidence suggests that biases are likely to be small.

Table 2 reports the results from Equation (5) for different samples. The respective first-step results are exhibited in Table A3. In the first row, I estimate the city size elasticity for a 50% random sample of non-college workers. Column (1) shows that including age and tenure controls yields a static urban wage premium of 0.028. When sector and occupation indicators are added in Column (2), the elasticity drops to 0.019. Finally, Including worker fixed-effects does not change the results meaningfully, as shown by Column (3). Taking the last estimate as an example, an elasticity of 0.016 means that doubling the city size would increase the wage premium by 1.6%.

Workers with a college degree, on the other hand, seem to benefit more from city size. Column (1) shows that controlling for age and tenure delivers a static urban wage premium of 0.080, almost three times what I found for non-college graduates. After including sector and occupation in Column (2), the estimate decreases to 0.066, indicating that sectoral composition explains the city-size wage gap to some degree. Lastly, Column (3) shows that adding worker fixed-effects reduces the elasticity to 0.022, a substantial decline of 67%.[15,16]

[15] The estimates for the whole sample are close to those from the first row since non-college workers represent roughly 90% of the observations (see Table 1). Considering the specifications in columns (1), (2) and (3), the elasticities (standard errors) are respectively 0.029(0.016), 0.021(0.014) and 0.017(0.008).

[16] Compared with previous studies, the estimates are of lower magnitude and with lower statistical significance. For instance, Combes et al. (2010)'s estimates with and without worker fixed-effects using French data are 0.051 and 0.033, respectively.

To stress the importance of worker fixed-effects, Figure A1 plots its city average against log city size for each educational group. As expected, I find a strong positive relationship among college graduates.

There is a debate in the literature about what fixed effects capture in this estimation. 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 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.

In the context of this discussion, two important findings emerge from Table 2. First, the inclusion of worker fixed-effects only matters for college graduates. Secondly, the difference in the urban wage premium of college and non-college graduates initially observed in Column (1) almost disappear with the inclusion of fixed effects, showing that this element is essential to understanding why the relationship between wages and city size is stronger for the first group.[17]

5 Estimating Dynamic Effects

To estimate the city-size effects on wage growth, 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

$$\Delta \log w^s_{i,t} = \delta_i + \delta^s_{c,a} + \Delta X_{i,t}\beta^s + \Delta\epsilon_{i,t} \;, \tag{6}$$

where $\delta^s_{c,a}$ 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 returns to

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%.

[17] As mentioned in footnote 13, it is possible that the elasticities of Table 2 are also capturing labor supply effects related to imperfect substitution between college and non-college labor. Disentangling these two effects is particularly challenging (see Moretti, 2004 for a discussion). Since this paper focuses on studying dynamic effects, I do not dive into this issue. Nevertheless, I report here the results when including the share of college graduates in the second-step regression. This variable is highly significant for non-college graduates, with an elasticity of 0.185 (s.e. 0.025) and an R-squared of 0.29. The city-size elasticity, on the other hand, becomes slightly negative (−0.015, s.e. 0.008). For college graduates, the history is quite different. Including the share of college workers marginally increases the city size elasticity to 0.033 (s.e. 0.008), with an R-squared of 0.08. The college share elasticity is −0.064 (s.e. 0.022). I compute the share of college graduates using the 2010 Census, which also includes workers in the informal sector.

experience. They show that these fixed effects attenuate estimates of dynamic effects. To account for this possibility, I include this term in the specification.

The estimation of city size effects is performed again in two steps, regressing $\delta^s_{c,a}$ 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 allowing for a simpler specification, taking the first difference releases me from the necessity of constructing the workers' location history to estimate dynamic effects. 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 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]

As in the previous section, 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 the same individual locates 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). The identification is not harmed as long as these shocks are uncorrelated with city size. To check if this assumption holds, I regress annual changes in city-level outcomes on log city size. I construct two variables using my dataset: mean log wages and mean log employment. I report these results in Table A1, which shows that city size negatively correlates with these variables. This evidence shows that estimates of dynamic effects are potentially underestimated.

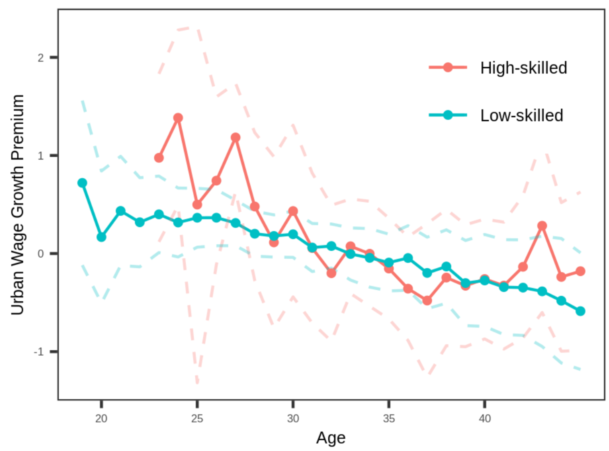
Figure 4 reports the results from Equation (7). For better exposition, the dependent variable was multiplied by 100. Overall, the figure shows that dynamic gains are decreasing in age and positive in the first years for both educational groups, suggesting that city size influences the wage growth of workers up to 30

[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 locates in a big city.

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

Figure 4: Dynamic Effects by Age



Notes: Each dot represents the elasticity of log wage growth with respect to city size for a skill-age group. Control variables include tenure and its square and indicators for the periods 2010, 2011-2014 and 2015-2017. The dashed lines inform robust 95% confidence intervals.

years old. The curve is relatively steeper for college graduates, although less precise. The results of the first-step estimation are presented in Table A4.

To check if unobserved worker characteristics influence wage growth in large cities, Figure A2 plots the city-average worker fixed-effects against log city size. There is a weak positive relationship among college graduates, which suggests that sorting is not relevant for this educational group. For non-college graduates, however, there is a negative correlation, indicating that the less skilled workers of this group are more likely to locate in large cities.

Since the large standard errors in Figure 4 may raise doubts about the existence of 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 A5. Both educational groups seem to benefit to some extent from city size. 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 0.252. 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,

Table 3: Dynamic Effects - Young Workers

	City dynamic premium ($\delta^{s}_{c,young} \times 100$)			
	(1)	(2)	(3)	(4)
Panel A. Non-college				
Log City Size	-0.007	0.252^{**}	0.149	0.154
	(0.081)	(0.100)	(0.134)	(0.135)
R^2	0.00004	0.025	0.007	0.007
Panel B. College				
Log City Size	0.706^{***}	0.565^{*}	0.635^{***}	0.617^{***}
	(0.121)	(0.303)	(0.178)	(0.175)
R^2	0.175	0.035	0.053	0.051
Observations	185	185	185	185
First Step Variables				
Age, Tenure	Yes	Yes	Yes	Yes
Worker FE	No	Yes	Yes	Yes
Exclude sec & occup moves	No	No	Yes	Yes
Sec & occup indicators	No	No	No	Yes

Notes: Results of the first-step estimation are reported in Table A5. "Young" refers to observations of workers up to 30 years old. 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.

which might correlate with wage growth.[21] In this case, the point estimate reduces to 0.149, suggesting that large cities make sector and occupation changes slightly more advantageous relative to small cities. Finally, column (4) shows that the estimate remains stable when I add sector and occupation indicators.

For college graduates, the effects are more pronounced. The second row displays point estimates that range from 0.565 to 0.706. In Column (2), the inclusion of worker fixed-effects attenuates the estimated elasticity and increases the standard error. Column (3) shows that excluding sector and occupation moves now has the opposite effect, with a slight increase in the elasticity. Finally, including sector and occupation indicators have a negligible impact. These results sharply contrast with the point estimates obtained for senior workers, as displayed in Table A6. For these individuals, there is no evidence of city size affecting wage growth.[22]

Taken together, the evidence shows that dynamic effects exist only at the beginning of working life, mainly for those with a college degree. However, it is still unclear whether these effects are, in fact, significant in explaining the city-size wage gap. I address this question in the next section.

[21] As mentioned in the introduction, 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.

[22] Table A7 provides estimates of dynamic effects without age heterogeneities. This is the specification proposed by D'Costa and Overman (2014). In line with their results, I find mild evidence of dynamic effects. In particular, the estimate for college graduates becomes no significant when worker fixed-effects are included, which is what D'Costa and Overman (2014) found using a sample of british workers (college and non-college graduates). 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.

6 Full Estimation - Combining Static and Dynamic Effects

So far, the analysis of dynamic effects was conducted apart from the "static" variables - namely the static premium and the unobserved ability - to emphasize their description and heterogeneities across educational and age groups. However, to understand how these factors relate to each other and their importance in explaining the city-size wage gap, it is necessary to consider all the terms in Equation (3) simultaneously. By estimating α_i and Ψ_c^s in the presence of dynamic effects, I can quantify the contribution of each element and evaluate how omitting dynamic gains introduces bias in the results of Section 4.

In order to achieve this goal, 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, which are presented in Figures 4 and A2.

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 that the beginning of working life is 18 years old, so the year cutoff is 1985.

I first show the results of regressing the static premium Ψ_c^s on city size under different specifications of Equation (9). First-step estimations are reported in Tables A8 and A9. Columns (1), (2) and (3) report estimations that replicate the first three columns of Table 2 using the new sample. Columns (4), (5) and (6) use the same specification as Column (3) but with a modified dependent variable as proposed by Equation (9).

Overall, incorporating dynamic elements does not seem to impact the estimated static premiums significantly. In Column (4), when the estimation is corrected for dynamic effects, the static premium slightly increases to 0.019 and 0.030 for non-college and college graduates, respectively. When individual returns to experience are subtracted from wages, in Column (5), there is no change for non-college graduates but a decrease of about 25% for workers 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 again get elasticities very close to those in Column (3).

Regarding worker fixed-effects, the conclusions are quite different, as shown in Table 5. Now, I regress

Table 4: Static Urban Wage Premium - Full Model

	City wage premium (Ψ_c^s)					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Non-college						
Log City Size	0.023*	0.017	0.016*	0.019**	0.016*	0.018**
	(0.014)	(0.012)	(0.009)	(0.009)	(0.009)	(0.009)
R^2	0.014	0.010	0.017	0.021	0.016	0.019
Panel B. College						
Log City Size	0.065***	0.055***	0.025***	0.030***	0.018**	0.022**
	(0.019)	(0.014)	(0.008)	(0.008)	(0.008)	(0.009)
R^2	0.053	0.081	0.045	0.057	0.022	0.032
Observations	185	185	185	185	185	185
First Step Variables						
Age, Tenure	Yes	Yes	Yes	Yes	Yes	Yes
Sec & occup indicators	No	Yes	Yes	Yes	Yes	Yes
Worker FE	No	No	Yes	Yes	Yes	Yes
log wage net of	-	-	-	$\sum_{\tau} \hat{\delta}_{c,a}^s$	$\hat{\delta}_i e_{it}$	Both

Notes: Results of the first-step estimation are reported in Tables A8 and A9. 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.

the city-average worker fixed-effect, defined as $\bar{\alpha}_{c'}^{s'} \equiv \sum_{c(i,t)=c', s=s'} \alpha_i$, on log city size. Columns (1) to (4) refer to Columns (3) to (6) of Table 4, respectively.

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 in large cities based on the elasticity displayed in Column (2). However, correcting for city 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 included in the model, I get the same results as Column (2).

In conclusion, the results show 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 the estimates of Tables 2 and 5 to perform a back-of-the-envelope calculation, the reduction in the static premium once individual effects are included ($0.066 - 0.022 = 0.044$) is 55% of the estimate initially found in Column (1) of Table 2 (0.08), which is my baseline measure of the city-size wage gap.[23] If 84% of what the fixed effects are capturing are dynamic effects, as suggested by Table 5 ($(0.038 - 0.006)/0.038$), it means that 46% of the city size gap is due to dynamic effects. This number is significantly higher than the static premium obtained in

[23] A simple regression of city log mean wage on log city size delivers the same number.

Table 5: Unobserved ability and city size

	City average unobserved ability ($\bar{\alpha}_c$)			
	(1)	(2)	(3)	(4)
Panel A. Non-college				
Log City Size	0.002	-0.012^{*}	0.008	-0.005
	(0.007)	(0.007)	(0.007)	(0.005)
R^2	0.0003	0.013	0.006	0.007
Panel B. College				
Log City Size	0.038^{***}	0.006	0.038^{***}	0.006
	(0.014)	(0.014)	(0.015)	(0.011)
R^2	0.036	0.001	0.031	0.001
Observations	185	185	185	185
First Step Variables				
log wage net of	-	$\sum_{\tau} \hat{\delta}^s_{c,a}$	$\hat{\delta}_i e_{it}$	Both

Notes: The columns in this table refers respectively to Columns (3) to (6) in Table 4. Coefficients are reported with robust standard errors in parenthesis. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels.

Column (3) of Table 2 (0.022), which accounts for 27.5% of the city-size wage 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. This result contrasts with estimates of the static urban wage premium, which are relatively more similar across educational groups. Using again Tables 2 and 5, I infer that 62% of the city-size college premium gap is due to dynamic effects.[25]

To give 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 4 show that they will get a roughly similar static wage premium after moving. However, according to Figure 4 and Table 3, college graduates in L will experience higher wage growth. If we assume a dynamic premium of δ in the large city, the college premium gap between the two workers in S and the two in L will be $T\delta$, where T is the number of periods since c_1 and h_1 moved to the large city.

Finally, the results show that sorting on unobservable characteristics is of secondary importance to explain the city-size wage gap when dynamic effects are considered, corroborating De La Roca and Puga (2017). Professionals like engineers and teachers have on average the same quality at the beginning, regardless of

[24] Because I restrict the sample to individuals with full location history available, which are relatively younger, the contribution of dynamic effects to the city-size wage gap is potentially underestimated. Note also that the introduction of dynamic effects does not change the conclusions regarding the importance of the static premium, as shown in Table 4.

[25] Table 2 shows that the reduction in college premium due to the inclusion of worker fixed-effects is 73% of the initial number in Column (1) $((\frac{0.066}{0.019} - \frac{0.022}{0.016})/\frac{0.080}{0.028})$. I assume again that 84% of the fixed-effects' impact on Column (3) 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 62%.

where they locate. As they acquire experience in large cities, their skill level increases, amplifying the city-size wage gap.

7 Wage Growth, City Size and Job Transitions

Having characterized how dynamic benefits manifest over the life cycle and how relevant they are, I now turn to the question of why they exist. In general, the literature provides two explanations for this phenomenon. First, cities may promote faster human capital accumulation due to a better learning environment. This interpretation is supported by empirical evidence demonstrating that the more valuable experience acquired in a large 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). Second, cities may promote better matching between workers and firms, an explanation that also has empirical support (Wheeler, 2006; Bleakley and Lin, 2012; Eckert et al., 2020; Dauth et al., 2022).

One possible strategy to confront these two hypotheses is to investigate how city size differentially affects wage growth within and between jobs. Intuitively, dynamic effects observed during job transitions would indicate the presence of externalities related to firm-worker matching. Although this connection certainly has limits, an inquiry of this nature can be informative.

Taking advantage of the matched employer-employee data, I build on Equation (6) to 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 refers to a specific progression worker's trajectory. Thus, if the observation is a within-job progression, I estimate specific city-effects for the 1st, 2nd, 3rd and 4th or higher jobs. If I observe the worker in different jobs between two consecutive periods, I differentiate between the 1st, 2nd and 3rd or higher job transition. To investigate the role of city size on each progression, I regress the estimated fixed effects on log city size for each type of progression.

Enumerating jobs (and job transitions) requires the workers' location history again, so the same criteria and sample restrictions from Section 6 apply. Additionally, I drop some cities from the estimation due to a lack of observations for each progression. I impose a cutoff of 1,500 observations. The final sample has about 3.9 million observations of 148 cities for college graduates and 21.1 million observations of 184 cities for non-college graduates.[26]

[26] The results are robust to change the cutoff value to 1,000 or 2,000. One important caveat is that my definition of a job transition is based on observed changes between consecutive years, which ignores within-year transitions. Furthermore, I also

The results in Table 6 indicate that city size influences both within- and between-firm wage growth of college graduates. Column (1) presents evidence of dynamic effects within the first job, showing that doubling the city size increases yearly wage growth by 0.6%. Later jobs have positive coefficients as well but with high p-values. Regarding job transitions, Columns (5) to (7) show positive effects in the first job transition and from the third onwards. In this case, a city twice the population results in 1.8% higher wage growth in the first job transition and 2.3% in the third job transition onwards.

Table 6: 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	0.102	−0.019	−0.067	−0.579*	0.236	0.053	−0.463
	(0.138)	(0.160)	(0.193)	(0.457)	(0.215)	(0.234)	(0.413)
Obs.	184	184	184	184	184	184	184
R^2	0.003	0.0001	0.001	0.010	0.005	0.0002	0.007
Panel B. College							
Log City Size	0.615***	0.180	0.181	0.886	1.757***	0.187	2.278***
	(0.199)	(0.270)	(0.496)	(0.802)	(0.561)	(0.725)	(0.914)
Obs.	148	148	148	148	148	148	148
R^2	0.048	0.002	0.001	0.006	0.049	0.0003	0.033

Notes: Controls include 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 A10 shows 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. In fact, point estimates tend to be larger in this case. Note also that the magnitude of the elasticities related to job transitions weakens the possibility that the results are driven by "hidden" within-job wage growth. Finally, although the higher variance in job transitions is, to some extent, due to fewer observations to estimate the parameters, it also suggests high heterogeneity in wage growth. While some workers benefit substantially from job transitions in large cities, others may take no advantage.

While Table 6 shows evidence that college graduates in large cities experience higher between-firm wage growth, it remains silent about long-run effects. It is unclear whether transitioning between firms in large cities ultimately translates into higher wages in the long run. To address this question, I return to Equation (3) and derive an expression for the long-term wage growth of a non-migrant:

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

abstract from differences between job-to-job and job-unemployment-job transitions.

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

Equation (11) shows that if n_i^*, δ_i and $\Delta X_{i,t}$ do not vary with city size, it is possible to assess city size effects using another two-step procedure. First, I run the following wage growth regression:

$$\Delta_t \log w_i = \delta_{c,J}^{LT} + \rho_t + \eta_i \quad , \tag{12}$$

where $\Delta_t \log w_i$ is the wage growth for a fixed age interval, set to be between 23 and 30 years old. $\delta_{c,J}^{LT}$ is the city effect for a worker with J job transitions during this period, ρ_t is a year fixed-effect, and η_i is the error term. For this estimation, I use a sample of non-migrants with full employment history available. After estimating $\delta_{c,J}^{LT}$, I regress them on log city size separately for each J:

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

I argue that the conditions to identify λ^J hold in this case. First, Figure A2 shows that the average δ_i is not correlated with city size. Secondly, Table A11 shows that the average number of observations in the dataset between 23 and 30 years old is also not correlated.[27]

Table 7 displays the results for three categories of J, with two key findings. First, city size effects increase with the number of job transitions. Columns (1) to (3) show elasticities that range from 1.294 for workers that stay in the same job the whole period to 5.125 for workers that experience two or more job transitions. Secondly, the effects are mostly on movers, as the coefficient for one-job workers cannot reject the null hypothesis at 10%. In other words, college graduates need to transition between firms in order to benefit from city-size dynamic effects.

In summary, this section shows that job transitions are crucial to understanding city size dynamic effects and suggests that wage dynamics in large cities are connected with how thick local labor markets work. While this conjecture is not new, few empirical studies exist on this topic. For instance, some papers in labor economics investigate whether an empirical matching function would have increasing returns to scale. Petrongolo and Pissarides (2001) provide an extensive survey of 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

[27] By fixing the age interval and estimating city effects for different number of job transitions, I also control for spatial variations in $\Delta X_{i,t}$.

Table 7: City Size, Wage Growth and Job Transitions in the Long-Run

	City long-term premium ($\delta^{LT}_{c,J}$)		
	(1)	(2)	(3)
College			
Log City Size	1.294	3.216***	5.125***
	(0.921)	(1.169)	(1.246)
Obs.	147	147	147
R^2	0.011	0.033	0.094
Job Transitions	0	1	2+

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

studies assortative matching (Andersson et al., 2007; Figueiredo et al., 2014; Dauth et al., 2022). In special, Dauth et al. (2022) show not only that the correlation between worker quality and firm quality increases with city size, but this relationship has also strengthened over the last years.

To conclude, 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. While their focus is not primarily on wage growth, their findings are possibly the most closely connected to this paper since, unlike the other studies mentioned, they approach the issue from a life cycle perspective. As the results here show, there is a specific moment and context in which job transitions in large cities lead to faster wage growth.

8 Robustness Checks

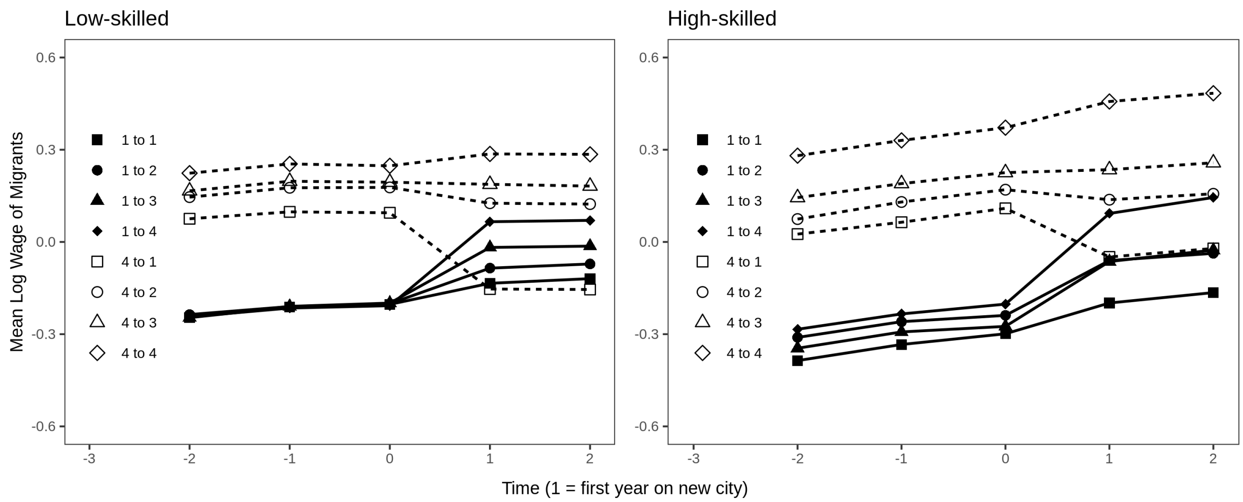
This section addresses possible sources of bias in the estimations reported. I focus on two potential vulnerabilities: the exogenous migration assumption and the possible endogeneity of city size. I also dedicate some lines to discuss to what extent the conclusions of this paper can be extended to other countries.

8.1 Endogeneity in the First-step Equation - Event Study Analysis

In Section 6, I show that not accounting for dynamic effects does not have a meaningful impact on estimates of the static urban wage premium. However, other omitted factors in the wage regression may violate the exogenous migration assumption.

In particular, workers may migrate conditional on particular good wage offers that do not reflect city wage premiums. If this is the case, the wage gains for workers migrating from a low-wage city to a high-wage city will differ in absolute terms from workers moving in the opposite direction. To see that, consider a simple

Figure 5: Event Study - Wages and Migration



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

case of two cities, A and B. B is larger than A and has a wage premium Ψ_B. If migration occurs exogenously, the wage variation for workers moving from A to B and from B to A are Ψ_B and $-\Psi_B$, respectively. However, if migration in either direction only happens if workers obtain an additional gain η, the wage variations would be $\Psi_B + \eta$ in an A-B migration and $-\Psi_B + \eta$ in a B-A migration.

I check this possibility by performing an event study analysis based on Card et al. (2013). For each educational group, I separate cities into four quartiles based on the static city premiums estimated in Column (3) 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 5. For better exposure, I exhibit only migrations from quartiles 1 and 4.

This exercise offers some insights. For instance, it shows no evidence 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. Figure 5 also shows a fair amount of symmetry between moves in opposite directions. For workers with a college degree, the wage gains in a 1-4 migration slightly differ from the wage losses in a 4-1 migration.

In practice, the bias is of minor relevance if the matching component or net migration does not vary with city size, especially because the asymmetries are relatively small compared to the wage variation observed. While verifying the first condition is not possible, I show in Figure A3 that net migration weakly correlates with city size.

Table 8: 2SLS Estimation - First Stage

	Dep. variable:
	Log City Size
Log Urban Population in 1940	0.349^{***}
	(0.032)
% of Terrain Slope > 16%	-0.001^{***}
	(0.0002)
Observations	185
R^2	0.479
F-Statistic	83.742

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

8.2 Endogeneity in the Second-step Equation - IV Estimation

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., location, access to other markets). It is also possible that large cities are a result of high-wage areas attracting workers. To circumvent these concerns, I follow the literature and develop an instrumental variables approach.

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 develop 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 by using Global Agro-ecological Zones (GAEZ) data from FAO.[28] 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.

The first-stage regression results are given in Table 8, showing that both instruments have high statistical

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

Table 9: 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	0.016*	0.022**	0.022***	0.022**
	(0.008)	(0.011)	(0.007)	(0.010)
R^2	0.019	0.016	0.042	0.042
Panel B. Dynamic Effects - Young				
Log City Size	0.252**	0.140	0.565*	0.776***
	(0.100)	(0.154)	(0.303)	(0.267)
R^2	0.025	0.020	0.035	0.030
Observations	185	185	185	185
First Step Variables				
Age, Tenure	Yes	Yes	Yes	Yes
Sec & occup indicators	Yes	Yes	Yes	Yes
Worker FE	Yes	Yes	Yes	Yes

Notes: OLS results refers 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.

significance. The F-statistic equals 83.742 and easily surpasses all thresholds proposed by Stock and Yogo (2005) (size and relative bias).

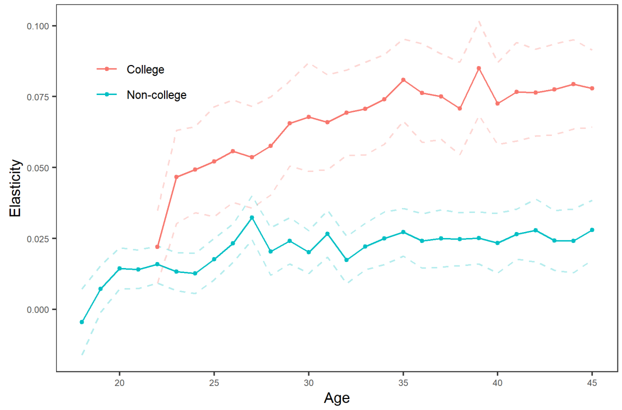
Table 9 presents the 2SLS estimates. To facilitate comparisons, the results of the respective OLS regressions are exhibited in columns (1) and (3). The OLS estimates of the static urban wage premium refer to Column (3) of Table 2, while the OLS estimates of dynamic effects refer to Column (2) of Table 3.

The first row shows that instrumenting has a slight impact on the static premium of non-college graduates but does not change the previous results for college graduates. Regarding dynamic effects, the second row exhibits larger divergences. The 2SLS estimation results in a 44% lower estimate for young non-college graduates and a 37% higher estimate for young college graduates.

Overall, the results in Table 9 suggest that the endogeneity of city size is not a concern in this context, which aligns with what the literature has found.[29] 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.

[29] For a detailed discussion, see Combes et al. (2010).

Figure 6: 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.

8.3 External Validity

A relevant question is whether the conclusions of this study can be extended to other countries, in particular developed economies. Specifically for the US, I replicate the procedure of 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. Because I do not have 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 6 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.

Chauvin et al. (2017) provide cross-section comparisons between Brazil, China, India and the United States and find that Brazil has a relatively weaker relationship between income and density but a stronger relationship between income and the share of adults with a college degree. This finding is consistent with the results I report in footnote 17. These pieces of evidence suggest 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 steep compared to other developing economies but falls slightly behind developed economies.

9 Conclusion

This paper studies the relationship between wages and city size, with a particular focus 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 46% of their city-size wage gap and 62% of the city-size college premium gap. The static urban wage premium, on the other hand, is roughly similar between college and non-college graduates. Moreover, sorting on unobserved characteristics contributes little to wage disparities within educational groups.

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

I also take advantage of the matched employer-employee feature of the data to study the role of job transitions in explaining these patterns. I find two important results. First, wage growth between firms increases with city size. Secondly, wage profiles with more than one job are steeper in large cities. These pieces of evidence reveal both short- and long-run benefits of job transitions and suggest that firm-worker matching is a relevant element to understanding why large cities influence wage trajectories.

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.

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*, 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*, 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

A - Additional Figures and Tables

Table A1: City size and changes in city-level variables

	College		Non-college	
	log av. wage	log obs.	log av. wage	log obs.
	(1)	(2)	(3)	(4)
Log City Size	-0.121	-0.355^{**}	-0.135^{**}	-0.593^{***}
	(0.142)	(0.178)	(0.057)	(0.180)
Observations	1,480	1,480	1,480	1,480
R^2	0.190	0.198	0.582	0.481

Notes: 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 A2: Summary Statistics - Migrations per city

		Average	Min	Max
1	Overall	49,069	475	1,450,190
2	High School or less	45,972	454	1,319,941
3	College	3,097	19	130,249
4	Young	25,530	221	774,758
5	Senior	23,539	254	675,432
6	High School or less - Young	24,446	211	727,908
7	College - Young	1,084	10	46,850
8	High School or less - Senior	21,526	243	592,033
9	College - Senior	2,013	6	83,399

Notes: "Young" refers to workers up to 30 years old, and "Senior" refers to workers with more than 30 years old.

Table A3: Static Urban Wage Premium - 1st step estimation

	Dependent Variable: Log Wage					
	Non-college			College		
	(1)	(2)	(3)	(4)	(5)	(6)
Age	0.156^{***}	0.111^{***}	0.143^{***}	0.568^{***}	0.360^{***}	0.479^{***}
	(0.001)	(0.001)	(0.001)	(0.005)	(0.004)	(0.003)
Age^2	-0.004^{***}	-0.002^{***}	-0.002^{***}	-0.014^{***}	-0.008^{***}	-0.010^{***}
	(0.00002)	(0.00002)	(0.00002)	(0.0002)	(0.0001)	(0.0001)
Age^3	0.00003^{***}	0.00002^{***}	0.00001^{***}	0.0001^{***}	0.0001^{***}	0.0001^{***}
	((0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Tenure	0.062^{***}	0.048^{***}	0.017^{***}	0.093^{***}	0.064^{***}	0.021^{***}
	(0.0002)	(0.0001)	(0.0001)	(0.001)	(0.0005)	(0.0002)
$Tenure^2$	-0.001^{***}	-0.001^{***}	-0.0003^{***}	-0.004^{***}	-0.003^{***}	-0.001^{***}
	(0.00001)	(0.00001)	(0.00001)	(0.0001)	(0.0001)	(0.00002)
$Tenure^3$	0.00000^{***}	0.00000^{***}	0.00000^{***}	0.00004^{***}	0.00002^{***}	0.00001^{***}
	(0.000)	(0.000)	(0.000)	(0.00000)	(0.00000)	(0.00000)
Sec & occup indicators	No	Yes	Yes	No	Yes	Yes
Worker FE	No	No	Yes	No	No	Yes
R^2	0.252	0.381	0.807	0.2934	0.4586	0.9106
Observations	66,061,164	66,061,164	66,061,164	10,468,026	10,468,026	10,468,026

Notes: Dependent variable is monthly log wage adjusted to a 44-hour workweek. Occupation and sector indicators refer to one-digit and two-digit level information, respectively. Coefficients are reported with robust standard errors in parenthesis clustered by worker. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels. Columns (1), (2), (4), (5), (7) and (8) include year indicators, and columns (3), (6) and (9) include indicators for the periods 2010, 2011-2014 and 2015-2017.

Table A4: Dynamic Urban Wage Premium by Age - 1st step estimation

	Dep. Variable: Δ *Log Wage* $\times$ *100*	
	Non-college	College
	(1)	(2)
Tenure	-0.860^{***}	-0.073^{*}
	(0.026)	(0.037)
Tenure^2	0.052^{***}	0.021^{***}
	(0.003)	(0.003)
Worker FE	Yes	Yes
R^2	0.176	0.205
Observations	50,664,886	8,219,724

Notes: Dependent variable is the first difference of monthly log wage adjusted to a 44-hour workweek. Occupation and sector indicators refer to one-digit and two-digit level information, respectively. Coefficients are reported with robust standard errors in parenthesis clustered by worker. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels. Columns (1) and (2) include indicators for the periods 2010, 2011-2014 and 2015-2017.

Table A5: Dynamic Urban Wage Premium by Age Group ("Young" and "Senior") - 1st step estimation

	Dependent Variable: Δ Log Wage × 100							
	Non-college				College			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Age	-0.265^{***}	-0.129^{***}	0.469^{***}	0.371^{***}	$--1.935^{***}$	-3.118^{***}	-1.635^{***}	-1.729^{***}
	(0.005)	(0.018)	(0.017)	(0.017)	(0.026)	(0.059)	(0.061)	(0.062)
Age^2	0.002^{***}	0.001^{***}	-0.003^{***}	-0.002^{***}	0.022^{***}	0.033^{***}	0.018^{***}	0.019^{***}
	(0.0001)	(0.0003)	(0.0002)	(0.0003)	(0.0004)	(0.001)	(0.001)	(0.001)
Tenure	0.147^{***}	-0.860^{***}	-0.758^{***}	-0.771^{***}	0.003	-0.075^{**}	-0.291^{***}	-0.287^{***}
	(0.003)	(0.026)	(0.015)	(0.016)	(0.006)	(0.037)	(0.038)	(0.038)
$Tenure^2$	-0.005^{***}	0.052^{***}	0.037^{***}	0.038^{***}	-0.001^{***}	0.021^{***}	0.025^{***}	0.025^{***}
	(0.0001)	(0.002)	(0.001)	(0.001)	((0.0003)	(0.003)	(0.003)	(0.003)
Worker FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Exclude sec & occup moves	No	No	Yes	Yes	No	No	Yes	Yes
Sec & occup indicators	No	No	No	Yes	No	No	No	Yes
R^2	0.011	0.175	0.337	0.337	0.019	0.204	0.296	0.296
Observations	50,664,886	50,664,886	37,957,331	37,957,331	8,219,724	8,219,724	6,846,833	6,846,833

Notes: Dependent variable is the first difference of monthly log wage adjusted to a 44-hour workweek. Occupation and sector indicators refer to one-digit and two-digit level information, respectively. Coefficients are reported with robust standard errors in parenthesis clustered by worker. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels. Columns (1) and (5) include year indicators, and columns (2), (3), (4), (6), (7) and (8) include indicators for the periods 2010, 2011-2014 and 2015-2017.

Table A6: Dynamic Urban Wage Premium - Senior Workers

	City wage growth premium ($\delta^s_{c,senior}$)			
	(1)	(2)	(3)	(4)
Panel A. Non-college				
Log City Size	−0.085	−0.028	0.002	0.0001
	(0.065)	(0.111)	(0.142)	(0.144)
R^2	0.011	0.0003	0.00000	0.000
Panel B. College				
Log City Size	0.087	−0.112	0.120	0.109
	(0.079)	(0.235)	(0.238)	(0.237)
R^2	0.007	0.002	0.002	0.002
Observations	185	185	185	185
First Step Variables				
Age, Tenure,	Yes	Yes	Yes	Yes
Worker FE	No	Yes	Yes	Yes
Exclude sec & occup moves	No	No	Yes	Yes
Sec & occup indicators	No	No	No	Yes

Notes: Results of the first-step estimation are reported in Table A5. "Senior" refers to observations of workers with more than 30 years old. 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: Dynamic Effects - no Age Heterogeneity

	City dynamic premium ($\delta_c^s \times 100$)			
	(1)	(2)	(3)	(4)
Panel A. Non-college				
Log City Size	-0.041	0.131	0.083	0.083
	(0.067)	(0.091)	(0.130)	(0.131)
R^2	0.002	0.009	0.002	0.002
Panel B. College				
Log City Size	0.307^{***}	0.173	0.330	0.315
	(0.075)	(0.242)	(0.203)	(0.202)
R^2	0.089	0.005	0.017	0.016
Observations	185	185	185	185
First Step Variables				
Age, Tenure	Yes	Yes	Yes	Yes
Worker FE	No	Yes	Yes	Yes
Exclude sec & occup moves	No	No	Yes	Yes
Sec & occup indicators	No	No	No	Yes

Notes: 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 A8: Urban Wage Premium Non-College Workers - Full Model 1st step estimation

	Dependent Variable: 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	0.051^{***}	0.059^{***}	0.057^{***}	0.137^{***}	-0.113^{***}	-0.047^{***}
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
Age^2	0.0003^{***}	-0.0004^{***}	0.002^{***}	-0.002^{***}	0.006^{***}	0.002^{***}
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Age^3	-0.00002^{***}	-0.00001^{***}	-0.00004^{***}	0.00002^{***}	-0.0001^{***}	-0.00005^{***}
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Tenure	0.061^{***}	0.049^{***}	0.010^{***}	0.003^{***}	-0.009^{***}	-0.009^{***}
	(0.0001)	(0.0002)	(0.001)	(0.0002)	(0.0001)	(0.0001)
Tenure^2	-0.001^{***}	-0.0004^{***}	0.0005^{***}	0.002^{***}	-0.001^{***}	-0.001^{***}
	(0.00002)	(0.00004)	(0.0002)	(0.00004)	(0.00002)	(0.00002)
Tenure^3	0.00000^{***}	0.00000^{***}	-0.00000^{***}	-0.00002^{***}	0.00001^{***}	0.00001^{***}
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Sec & occup indicators	No	Yes	Yes	Yes	Yes	Yes
Worker FE	No	No	Yes	Yes	Yes	Yes
R^2	0.233	0.334	0.727	0.718	0.806	0.807
Observations	29,823,579	29,823,579	29,823,579	29,823,579	27,966,939	27,966,939

Notes: Dependent variable is 44-hour weekly log wage subtracted by the term indicated above the columns' numbers, according to Equation (9). Occupation and sector information are aggregated to one-digit and two-digit level, respectively. Coefficients are reported with robust standard errors in parenthesis clustered by worker. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels. Columns (1) and (5) include year indicators, and columns (2), (3), (4), (6), (7) and (8) include indicators for the periods 2010, 2011-2014 and 2015-2017.

Table A9: Urban Wage Premium College Workers - Full Model 1st step estimation

	Dependent Variable: 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	1.124^{***}	0.586^{***}	0.373^{***}	-0.279^{***}	0.641^{***}	-0.023^{*}
	(0.016)	(0.014)	(0.012)	(0.012)	(0.012)	(0.012)
Age^2	-0.033^{***}	-0.016^{***}	-0.006^{***}	0.010^{***}	-0.015^{***}	0.002^{***}
	(0.001)	(0.001)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
Age^3	0.0003^{***}	0.0002^{***}	0.00003^{***}	-0.00001^{***}	0.0001^{***}	-0.00003^{***}
	(0.00001)	(0.00001)	(0.00000)	(0.00001)	(0.00000)	(0.00000)
Tenure	0.110^{***}	0.072^{***}	0.019^{***}	0.017^{***}	-0.004^{***}	-0.006^{***}
	(0.001)	(0.0004)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Tenure^2	-0.006^{***}	-0.004^{***}	-0.002^{***}	-0.002^{***}	0.0005^{***}	0.001^{***}
	(0.0001)	(0.0001)	(0.00003)	(0.00003)	(0.00003)	(0.00003)
Tenure^3	0.00005^{***}	0.00003^{***}	0.00001^{***}	0.00001^{***}	-0.00000^{***}	-0.00001^{***}
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Sec & occup indicators	No	Yes	Yes	Yes	Yes	Yes
Worker FE	No	No	Yes	Yes	Yes	Yes
R^2	0.248	0.442	0.886	0.877	0.903	0.905
Observations	5,060,115	5,060,115	5,060,115	5,060,115	4,869,270	4,869,270

Notes: Dependent variable is 44-hour weekly log wage subtracted by the term indicated above the columns' numbers, according to Equation (9). Occupation and sector information are aggregated to one-digit and two-digit level, respectively. Coefficients are reported with robust standard errors in parenthesis clustered by worker. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels. Columns (1) and (5) include year indicators, and columns (2), (3), (4), (6), (7) and (8) include indicators for the periods 2010, 2011-2014 and 2015-2017.

Table A10: 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	0.293	0.011	-0.138	-0.357^{*}	0.314	0.104	-0.471
	(0.181)	(0.145)	(0.155)	(0.205)	(0.255)	(0.216)	(0.326)
Obs.	184	184	184	184	184	184	184
R^2	0.012	0.00003	0.004	0.014	0.006	0.001	0.011
Panel B. College							
Log City Size	1.005^{***}	0.331	0.333	0.334	2.068^{***}	0.272	1.824^{***}
	(0.245)	(0.224)	(0.282)	(0.366)	(0.599)	(0.609)	(0.606)
Obs.	148	148	148	148	148	148	148
R^2	0.073	0.011	0.007	0.004	0.059	0.001	0.040

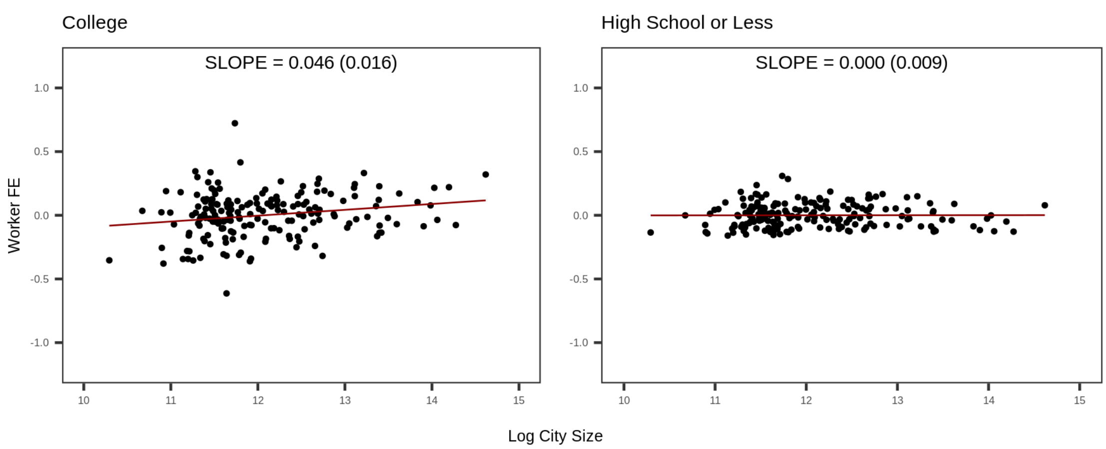
Notes: Dependent variable is the first difference of monthly log wage adjusted to a 44-hour workweek. Coefficients are reported with robust standard errors in parenthesis. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels. All regressions include age and tenure controls (quadratic polynomial) and worker fixed-effects.

Table A11: Formal Employment and City Size

	Avg. number of observations ($\bar{n}_c^*$)		
	(1)	(2)	(3)
College			
Log City Size	0.005	-0.125^{***}	-0.033
	(0.008)	(0.031)	(0.022)
Obs.	147	147	147
R^2	0.001	0.076	0.013
Job Transitions	0	1	2+

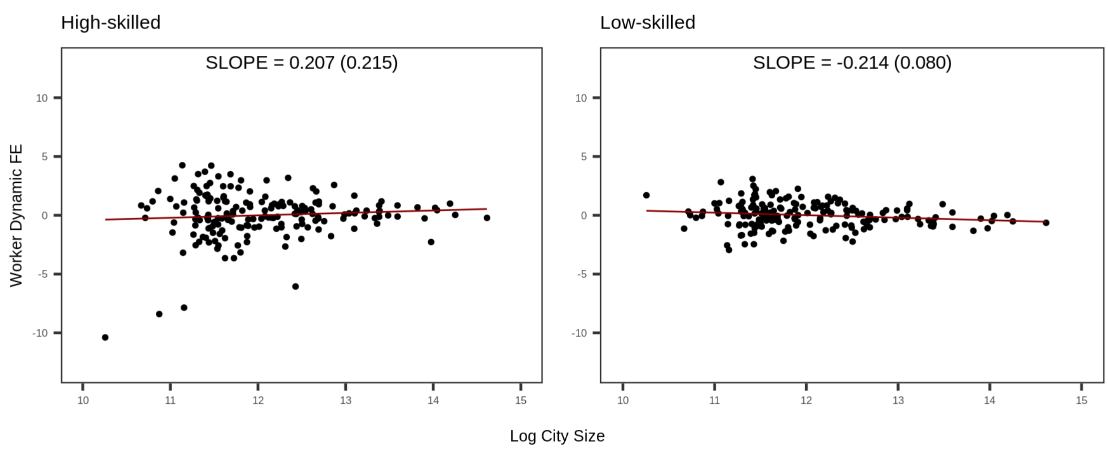
Notes: Dependent variable is the average number of employment observations between ages 23 and 30 for non-migrants. Coefficients are reported with robust standard errors in parenthesis. *, ** and *** indicate statistical significance at the 1, 5 and 10% levels.

Figure A1: Worker fixed-effects and city size - static estimation



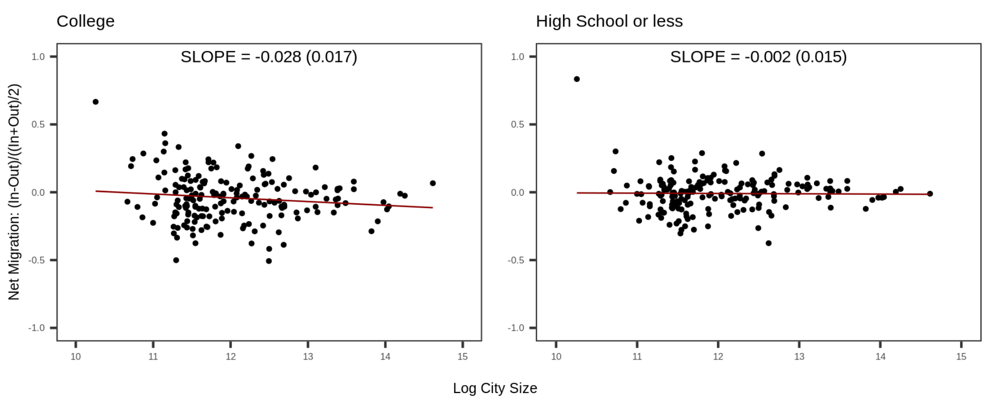
Notes: Each dot represents a city-average worker fixed-effect subtracted by the sample average. Estimates refers to Column (3) of Table 2. Coefficients are reported with robust standard errors in parenthesis.

Figure A2: Worker fixed-effects and city size - dynamic estimation



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

Figure A3: Net migration and city size



Notes: Each dot represents an average value at the city level. Coefficients are reported with robust standard errors in parenthesis. The data used refers to the sub-sample used in Section 6

B - 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

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

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 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 to city-group variables (B_c^s) and other associated to 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

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

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