Department of Economics- FEA/USP

Urban wage premium in a labor market with informality

Eloiza Regina Ferreira de Almeida

Veneziano C. Araújo

Solange Gonçalves

Working Paper Series Nº 2024-25

DEPARTMENT OF ECONOMICS, FEA-USP
WORKING PAPER Nº 2024-25

Urban wage premium in a labor market with informality

Eloiza Regina Ferreira de Almeida (eloiza.almeida@usp.br)
Veneziano C. Araújo (veneziano.araujo@unifesp.br)
Solange Gonçalves (solange.goncalves@usp.br)

Abstract:

Few studies analyze how the urban wage premium is different for informal workers, and their results are controversial. This paper aims to clarify the reason for this mixed evidence, evaluating how workers' heterogeneity in terms of labor contract - formal or informal - and occupational position - as wage-earner or self-employed - may impact the magnitude and direction of the UWP estimates. We address this investigation by analyzing the Brazilian labor market using the PNADC (IBGE) longitudinal database for the period from 2012 to 2019. The results show that formal workers present an increasing UWP according to the urban scale, as seen in many previous studies for developed or developing countries. In its turn, informal workers UWP is double the formal ones but is reduced in denser areas. Thus, our study shows previous UWP studies that focus only on formal workers could underestimate the magnitude of this premium for the whole labor market and that disregarding the groups of workers hides the complexity inherent of their insertion in the large urban labor markets. Also, different estimations highlight some mechanisms on the UWP explanation, such as sorting and matching. These results add new insights to the UWP in Brazil, signaling the importance of analyzing the whole labor market.

Keywords: Urban wage premium. Informality. Workers' heterogeneity.

JEL Codes: R23, J46, J31

Urban wage premium in a labor market with informality

Eloiza Regina Ferreira de Almeida*
Veneziano de Castro Araújo†
Solange Ledi Gonçalves‡

Abstract

Few studies analyze how the urban wage premium is different for informal workers, and their results are controversial. This paper aims to clarify the reason for this mixed evidence, evaluating how workers' heterogeneity in terms of labor contract - formal or informal - and occupational position - as wage-earner or self-employed - may impact the magnitude and direction of the UWP estimates. We address this investigation by analyzing the Brazilian labor market using the PNADC (IBGE) longitudinal database for the period from 2012 to 2019. The results show that formal workers present an increasing UWP according to the urban scale, as seen in many previous studies for developed or developing countries. In its turn, informal workers UWP is double the formal ones but is reduced in denser areas. Thus, our study shows previous UWP studies that focus only on formal workers could underestimate the magnitude of this premium for the whole labor market and that disregarding the groups of workers hides the complexity inherent of their insertion in the large urban labor markets. Also, different estimations highlight some mechanisms on the UWP explanation, such as sorting and matching. These results add new insights to the UWP in Brazil, signaling the importance of analyzing the whole labor market.

Keywords: Urban wage premium. Agglomeration levels. Informality. Workers' heterogeneity
JEL Classification: R23, J46, J31

*University of São Paulo-Brazil (FEA-USP). Email address: eloiza.almeida@usp.br
†Department of Economics of the Federal University of São Paulo-Brazil (UNIFESP). Email address: veneziano.araujo@unifesp.br
‡Department of Economics of the University of São Paulo-Brazil (FEA-USP). Email address: solange.goncalves@usp.br

1 Introduction

Many empirical studies confirm the existence of a wage premium in denser areas. Productivity gains arising from the agglomeration effects[1] and externalities are explanations attributed to this premium. Agglomeration effects enable a better firm-employee matching (STORPER; VENABLES, 2004; BAUM-SNOW; PAVAN, 2011) and human capital accumulation, due to knowledge spillovers and gains from experience (ROSENTHAL; STRANGE, 2008; MORETTI, 2011; ANDERSSON; THULIN, 2013; MORETTI, 2013; BEHRENS; DURANTON; ROBERT-NICOUD, 2014; ROCA; PUGA, 2017). In its turn, the agglomeration externalities come from the sorting of high-skilled workers in denser areas (BEHRENS; DURANTON; ROBERT-NICOUD, 2014; COMBES *et al.*, 2012).

The urban wage premium (UWP) in a labor market with a non-negligible proportion of workers without a formal labor contract and self-employed may present different determinants and trajectories related to this additional source of heterogeneity of the labor occupation. Informal workers may benefit from a lower wage premium due to the more frequent application of labor rules (ALMEIDA; CARNEIRO, 2012); the location restrictions in urban land markets (HARRIS, 2014); the lower reserve wages in denser areas (HENLEY; ARABSHEIBANI; CARNEIRO, 2009); and the rapid cities growth (MATANO; OBACO; ROYUELA, 2020). Matano, Obaco and Royuela (2020) argue that dense areas' agglomeration effects benefit formal workers more than informal workers, especially those related to the sorting, learning, and matching processes.

On the other hand, the predominance of low-skilled workers in the informal sector may be the main factor explaining a higher UWP for this group of workers, as (i) the higher competition (MORETTI, 2011) and the constraints or lower potential to create formal jobs (GARCÍA, 2019) in denser areas may provoke a disproportional decrease in high-skilled workers wages; (ii) the existence of disamenities – such as pollution, traffic congestion, crime, and garbage excess - in larger areas may be a more critical decision factor for the high-skilled worker than for low-skilled ones, with the urban premium playing a compensatory role (GUEVARA-ROSERO; POZO *et al.*, 2020); (iii) self-employed workers may sell products locally, and their income is more directly determined by local housing and transportation costs (DURANTON, 2016).

This study investigates the urban wage premium in a labor market with a high level of informality. Our main goal is to reveal how the workers' heterogeneity in terms of their labor contract and their occupational position as a wage-earner or self-employed may impact the estimates' UWP magnitude and direction. We investigate this issue while focusing on the following questions: (i) Do the UWP results reported in the literature change whether we consider informal workers? (ii) How does the heterogeneity in workers' formality status and occupational position affect UWP by agglomeration level? (iii) Do the role of individual and occupational characteristics on the UWP explanation are distinct between workers subgroups? (iv) What are the possible mechanisms that drive the UWP in a labor market with informality?

We choose to address this issue to the greatest possible extent with Brazil's data, a middle-income developing country with 43.5% of the workforce in the informal labor market, in 2019 (IBGE, 2020a). This level of informality is close to the average found in South America (46%), and well above

[1] Duranton and Puga (2004) conceptualizes agglomeration effects in three fundamentals: sharing, matching, and learning.

developed countries (18%) (BONNET; VANEK; CHEN, 2019). Thus, nearly half of employment is taken aside in previous studies focusing only on formal workers.

We explore an additional source of occupational heterogeneity of workers that was not yet investigated in the UWP literature, as we combine information on the status of formality of the labor contract and the occupational position of workers – wage-earner or self-employed – distinguishing results of four different group of workers: a) formal wage-earner; b) informal wage-earner; c) formal self-employed; and d) informal self-employed, covering the whole Brazilian labor market. By doing this differentiation, it is possible to investigate groups of workers with similar goals and profiles that only differ in terms of formality[2]. Wage-earners comprise all employees in the private sector, while self-employed are both employers and self-employed. Although formal workers outnumber informal workers' earnings by more than 40% on average, this differential reaches 38% among wage-earners and 55% among self-employed.

We explore the data from the Continuous Brazilian Household Sample Survey (PNADC) of the Brazilian Institute of Geography and Statistics (IBGE). This rich longitudinal database that covers the whole country, at different agglomeration levels. Our analysis comprises the Brazilian Metropolitan Areas (MA), taking advantage of a highly concentrated urban hierarchy[3], by treating them at different agglomeration levels. We covered the period from 2012 to 2019-Q3 and estimated using pooled OLS several mincer-type equations according to formality status and groups of workers, testing different agglomeration levels definition, sample choices, and additional variables. Moreover, we explore different estimation procedures highlighting the role of some mechanisms behind the UWP, such as sorting and matching, as pointed out in the literature.

Our findings suggest that previous results reported in the UWP literature based on the formal sector may underestimate the premium magnitude, at least if there is a large proportion of informal workers, like the Brazilian labor market, as their premium doubles that for formal ones. In both sectors, the UWP differs by agglomeration level and exhibit different patterns. Formal workers' UWPs range between 2.70% and 5.35% and increase from medium to extra-large MAs. In contrast, informal workers UWP has a broader range of 4.19% to 18.0% and drives the overall results, as it replicates the decreasing pattern as the agglomeration level increases. In extra-large MAs, the UWPs of both groups are closer, but they distance themselves as the agglomeration level decreases.

We verify through heterogeneity analysis that the trajectory and magnitude of the urban wage premium differ for workers' subgroups. Formal wage earners UWP increases from medium to extra-large MAs, but the other three subgroups decrease with the agglomeration level. Besides, the UWP magnitude strongly differs from formal wage earners to the other three subgroups in MAs, reaching 3.4% against at least 8.0%, respectively. The UWP for formal wage earners is lower than almost all agglomeration level. Again, we find evidence that disregarding these workers hides the complexity inherent in their inclusion in large urban labor markets.

[2] By definition, informal workers do not have a formal labor contract or do not contribute to the Social Security Institute. Thus, an informal worker lacks access to several benefits provided by labor laws, such as unemployment insurance and severance payments in the case of maternity leave, death, illness, or an occupational accident.

[3] Brazil accounts for 208 million inhabitants, 41.1% living in MAs. From the 27 MAs, 19 have more than one million inhabitants, and two of which have more than ten million – São Paulo and Rio de Janeiro. This large MAs outnumber other large areas contemplated in UWP studies for Latin America.

Although urban areas traditionally present higher wage levels, the agglomeration effects may not benefit the subgroups equally. The combination of different factors may influence the existence and magnitude of a wage premium. Robustness checks confirm the UWP results even when considering different workers' samples and alternative definitions of agglomeration levels.

Our study contributes to the literature on urban wage premium by considering a not well-investigated heterogeneity source: the labor contract's formality status and the occupational position. Most of the UWP literature is concentrated in developed countries and analyzes the relationship between the agglomeration level and workers' schooling (MORETTI, 2004), job tenure (GLAESER; MARE, 2001; CARLSEN; RATTSØ; STOKKE, 2016), and skills (GOULD, 2007; ANDERSSON; THULIN, 2013; BACOLOD; BLUM; STRANGE, 2009). Even those articles concentrated on the wage premium of developing economies usually use administrative data collected from the formal labor market sector firms. For Brazil, exceptions are Chauvin *et al.* (2017) and Cruz and Naticchioni (2012) that consider the whole labor market, aggregating data on workers of the informal labor market sector does not explore possible differences by workers' formality status or occupation position.

Few studies have analyzed the heterogeneity of workers in countries with a significant share of informal workers, such as Duranton (2016), García (2019), Guevara-Rosero, Pozo *et al.* (2020), Matano, Obaco and Royuela (2020), showing controversial results. Exploring this type of heterogeneity can shed light on the explanation for these results. To the best of our knowledge, our study provides the first evidence of distinct patterns in the UWP for workers of different formality statuses and occupational positions according to agglomeration levels in Brazil.

We also contribute to studies that seek to understand the segmented labor markets and differentials between the workers in each market[4] by looking over the agglomeration role. This literature includes the specific dynamic of urban labor markets in developing countries and can be explored when we differentiate the workers by formality status and occupation position.

This paper also has a methodological contribution of taking into account information of unemployed individuals to estimate and correct the sample selection bias. This analysis is ignored by most studies, because data are not available[5]. Despite this, the sample selection bias using PNADC data is statistically significant, and its correction was necessary. As far as we know, this database has never been used to estimate the UWP.

The evidence on an urban wage premium with different magnitude, direction, and trajectory, according to the workers' heterogeneity in terms of the characteristics of the labor contract and their occupational position, can guide new policies. First, the higher informal workers UWP can have a compensatory role, highlighting the need to assess how companies in urban centers respect regulations and, more than that, may indicate that there are fiscal inefficiencies. Second, the similar UWP magnitude for wage earners in very dense areas may result from more frequent enforcement of labor regulations, reducing the premium in these areas due to the constraints of being informal, and may indicate the need to review inspection processes.

[4]Such as the recent studies of Barros and Ulyssea (2010), Dalberto and Cirino (2018), Ulyssea (2020), and the seminal study of Fields (1990).

[5]Both studies for Brazil, Chauvin *et al.* (2017), and Cruz and Naticchioni (2012) do not perform this correction, even though unemployed data were available. The other studies for Brazil use only employed formal workers' data, such as Barufi (2015).

This paper is divided into more five sections. Section 2 describes the Brazilian metropolitan labor market, followed by Section 3, that presents the data, sample, and empirical strategy. Section 4 shows the descriptive analysis, estimation results and robustness tests. Section 5 presents the final remarks.

2 Brazilian Labor Market

Brazil is a middle-income developing country with a GDP of 1.87 trillion dollars (WORLDBANK, 2019), occupying the ninth position in the world ranking, but only seventy-fifth concerning the per capita GDP (US$ 8,921), outnumbered by other Latin American countries, such as Mexico, Chile, Argentina, and Uruguay, and ahead of Peru, Colombia, Ecuador (WORLDBANK, 2019). More than 64% of its 209 million inhabitants in 2019 are of working-age (15-65 years), being the sixth most populated country globally, outnumber almost two times Mexico, the second most populated in Latin America (IBGE, 2019). Although Brazil has a demographic density of 25 inhabitants per km^2, its' continental dimensions have a highly concentrated urban hierarchy, with 86.6% of its population living in urban areas, nearly half of that, 41.1% living in MAs. At least 19 of the 27 MAs have more than one million inhabitants, two of whom have more than ten million inhabitants (i.e., São Paulo and Rio de Janeiro).

As pointed out by UWP literature, denser areas tend to pay higher salaries, and we observe this relationship in Brazilian MAs. Figure 1 shows that the average hourly wage increase with the MAs population, the first evidence of a possible wage premium in those areas.

Brazil has a non-negligible informal sector that corresponds to 43.5% of the working-age population (18-65 years) and 35.0% of the workforce. On the one hand, this informality level is close to the average in South America (46%) and lower than countries like Bolivia, Paraguay e Peru – which register 83.1%, 70.6%, and 69.2%, respectively. On the other hand, informal workers are over-represented compared to developed countries (18%) (BONNET; VANEK; CHEN, 2019), which concentrates most of UWP studies.

Within the informal group, workers are almost equally divided into wage-earners and self-employed (45% and 55%). This scenario does not change much in recent years, as shown in Figure 2. This group's differentials are due to the wage levels, where self-employed workers earn consistently 11% on average more than wage earners. Formal workers are mostly wage-earners, representing almost 80% every year. However, self-employed wages outnumber wage-earners by 50% on average between 2012 and 2019 and are the subgroup with the highest earnings between all other types. Only this group show a decrease in the wage levels in recent years, specifically between 2014-2015, remaining stable after that.

Relevant differences are also observed in the wage distribution depending on the formality status of workers' labor contract and occupational position (Figure 3). Informal workers are more concentrated at lower wage levels, whereas formal workers are more concentrated at higher ones, mainly formal self-employed workers. These differences justify our decision to analyze workers' subgroups, opposing previous studies on the Brazilian labor market, who do not take these two heterogeneity sources into account.

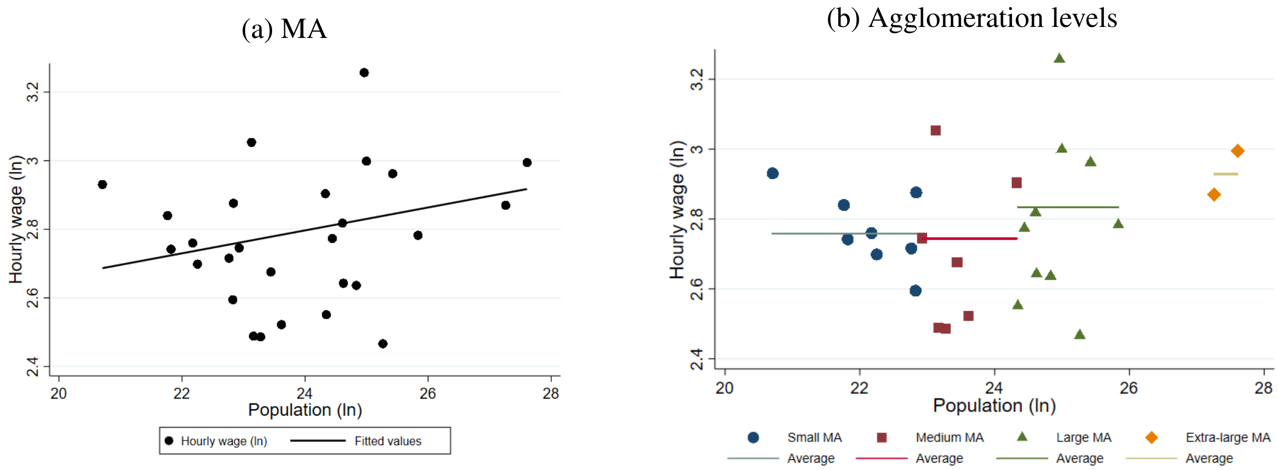

Figure 1: Average hourly wages and population

Source: Created by the authors based on PNADC from 2012 to 2019(Q3). **Notes:** We only include data from the first interview of the PNADC; employed individuals aged 18-65 years. The agglomeration levels are defined in B. The average natural logarithm of the hourly wage, time-corrected to 2019-Q3 price level, is 2.57 (R$ 18.35) for MAs and 2.29 (R$ 12.55) for non-MAs.

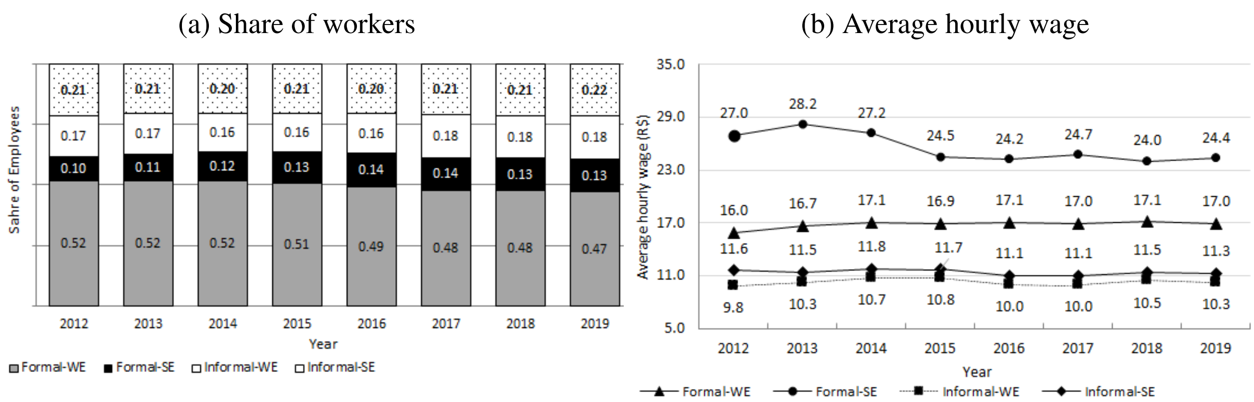

Figure 2: Workers heterogeneity

Source: Created by the authors based on PNADC from 2012 to 2019(Q3). **Notes:** Data from the first interview of the PNADC; employed men aged 18-65 years. Hourly-wages time-corrected to 2019-Q3 price level.

The regional heterogeneity of economic development and labor market dynamic also appears in the proportion of informal workers. This proportion ranges from 19% to 84% of occupied male workers across the country (Figure 4). The North and Northeast states have the highest informality levels — between 59% and 84% —, whereas the South and Southeast states have the lowest levels, with a maximum of 33%. Not coincidentally, the lower levels of informality are in the richest south states, where there is most of the country's economic activity. The map also highlights that MAs have a lower or at least equal informality rates than their counterpart in the state, as they present lighter colors, even in North and Southeast states. This scenario is consistent with the literature that underlines the difficulties of being informal in denser areas. The informality rate reaches 35.9% in MAs but is 48.3% in non-MAs.

It is essential to highlight individuals' heterogeneity in the formal and informal labor market sectors. We observe that formal workers have, on average, two more schooling years, although they

Figure 3: Average hourly wage distribution

Source: Created by the authors based on PNADC from 2012 to 2019(Q3). **Notes:** Data from the first interview of the PNADC; employed men aged 18-65 years. Excluding top and bottom 1% of wages. The average natural logarithm of the hourly wage is 2.49 (R$ 14.13) for formal wage earners, 2.65 (R$ 18.23) for formal self-employed, 2.02 (R$ 8.51) for informal wage earners, and 2.06 (R$ 9.42) for informal self-employed. The vertical line denotes the average minimum wage equal 1.85 (R$ 6.39) from 2012 to 2019, time-corrected to 2019-Q3 price level.

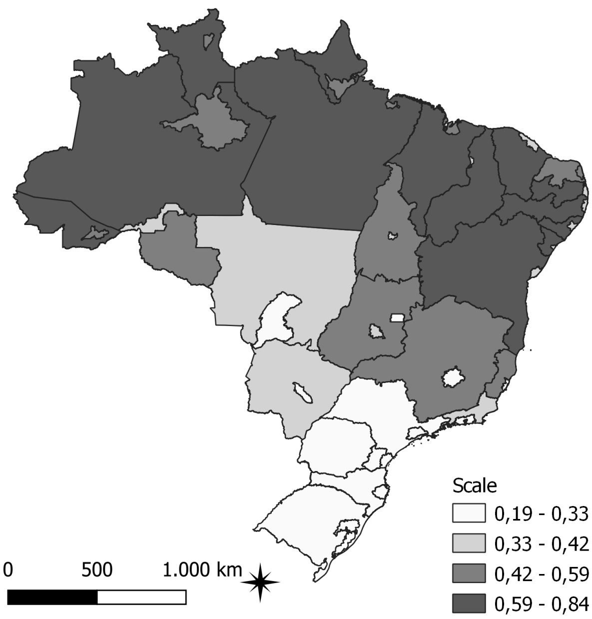

Figure 4: Share of informal workers in Brazilian states

Source: Created by the authors based on data from the PNADC for 2018, and Territorial Brazilian Division (IBGE, 2018b). **Note:** Employed men aged 18-65 years. The map plots the state frontiers and the respective capitals and MAs.

have almost the same average age[6]. Formal workers have families with a fewer number of members (3.4 versus 3.7) and are mostly married (65% versus 58%), household heads (60% versus 56%), and whites (51% versus 37%). Informal workers are over-represented in low-skilled-level occupations, with 37.5% of the workers versus only 27.4% of formal workers.

College-graduated individuals are almost three times more formal than informal workers — 18.6%

[6]Table 1 shows the sample descriptive analysis and the profile differentials.

versus 6.5% -– although both groups are concentrated in service-related activities, especially in MAs. The differentials appear in the occupation type of each group. At least 34% of formal workers are in occupations with higher wage levels such as directors and managers, science professionals, mid-level technicians, and administrative support. In comparison, informal workers represent only 14% of the same activities. Construction and elementary occupation are 48% of informal workers and only 28% within the formal ones.

More than 14% of the formal self-employed workers are directors or managers, while only 2.6% of informal self-employed have the same occupation. Although these rates are higher in MAs than in non-MAs – 16.8% and 3.4% – the occupational structure between the formal and informal self-employed shows that both groups are concentrated in service-related and construction activities. However, there are 14 percentage points more informal workers in these activities than formal workers. Wage-earners, in turn, also present relevant differentials, with an even more concentrated occupational structure. At least one-third of the informal workers are in elementary occupations. Even though they represent only 20% of informal wage earners in MAs, it is still the occupation with the highest number of workers. Within formal wage-earners, the first occupation in the number of workers is service-related activities – which pay higher wages than elementary ones – and covers 18% of total employment.

3 Empirical Strategy

3.1 Database and sample

We analyze the UWP using the PNADC from the Brazilian Institute of Geography and Statistics (IBGE), which is currently the primary survey about the Brazilian labor market. However, it is still rarely used for panel data analysis. The PNADC is a quarterly longitudinal database based on a rotating panel where a household is surveyed one month and then excluded from the sampling process for the following two months before being interviewed again. The process repeats five times, once every quarter, with approximately 211,000 interviews per quarter. The analyzed period spans the first quarter of 2012 to the third quarter of 2019, with a total of more than 17.5 million observations (IBGE, 2020a).

In addition to workers' employment statuses and wages, the PNADC allows us to access the location of families; individual characteristics, such as age, gender, race, household position, and schooling level; occupational features, such as tenure, formality status, occupational position, and skill level; and firm characteristics, such as the number of employees and industry. Following standard practice in the literature, the sample used for the analysis in this paper only includes men aged 18 to 65 years[7] with one job and at least 20 working hours. We exclude military members, public sector employees due to their specific labor laws, and family and domestic workers[8]. We also exclude

[7]The choice of men eliminates possible variation due to discrimination. Besides, the male workforce has more stable behavior in the labor market. We conduct the analysis only for men also to compare our results with those of previous studies.

[8]These groups are excluded because they do not have earnings.

workers with wages in the top or bottom 1% of the distribution. Besides, our sample does not cover the periods from 2015Q4 to 2016Q2 – three quarters – because data on firm size are not available for this period[9].

By using a household panel survey, we must consider some limitations of the database. In the PNADC, families that move homes during the survey are missing after moving, which implies that the data do not cover migration between regions. Thus, we are not modeling or considering the individuals' migration decisions. The second cause of attrition in this database is individuals that did not answer the survey at some point in the five quarters. The data loss is approximately 14% between the first and second interviews, reaching 40% by the fifth interview. We do not correct the possible bias from this attrition, but we try to address it by looking at different interviews as a robustness test.

To analyze the data by agglomeration level, we distinguish between MAs and non-MAs. MAs are the largest cities that correspond to the 27 state capitals of Brazil and its neighboring municipalities[10]. According to the population, we classify the MAs into four groups, small, medium, large, and extra-large[11], as shown in B.

We defined informal workers as those who do not have a formal labor contract or do not contribute to the Social Security Institute, following the informal employment definition from Bonnet, Vanek and Chen (2019)[12]. Based on this definition, a worker in an informal occupation lacks access to several benefits provided by labor laws, such as unemployment insurance and severance payments in the case of maternity leave, death, illness, or an occupational accident.

In turn, the occupational position includes three groups: private-sector workers, employers, and self-employed workers. We consider the private sector employees as wage earners, and the self-employed group comprises employers and self-employed workers. Following the approach adopted by Kunal *et al.* (2019), Dalberto and Cirino (2018), and Tansel and Kan (2012), we combine these dimensions to create four subgroups of workers, namely i) formal wage earners (Formal-WE), ii) formal self-employed workers (Formal-SE), iii) informal wage earners (Informal-WE), and iv) informal self-employed workers (Informal-SE).

Wages in the PNADC are self-declared. We deflate wages for the third quarter of 2019 using an index based on the Extended National Consumer Price Index (IPCA) released by IBGE (2018a). The correction based on the IPCA partially corrects regional inequalities because it covers each state by quarter and year.[13,14]

[9] A shows the sample construction and the respective losses of observations.

[10] The PNADC only identifies the state capitals MAs, which covers 41.1% of the Brazilian population. Non-capital MAs corresponds to 4.6% of the population, and we test the relevance of these areas as a robustness check in Section 4.2.4.

[11] The levels were divided as small: those with less than 1 million inhabitants (8 MAs); medium: between 1 and 2 million (7 MAs); large: between 2 and 10 million (10 MAs); and extra-large: more than 10 million (2 MAs).

[12] The labor economics literature provides different definitions of the informal sector or informal occupations. See Corseuil, Reis and Brito (2015) and Slonimczyk (2014) for more details.

[13] In an attempt to account for price variation across areas, robustness tests were carried out considering regional consumption baskets (RCB). Instead of IPCA, the use of the consumption baskets is not intended to be a correction of the total local living costs, as it only contemplates the costs of a minimum food basket, not including, for example, rent or clothing. Based on the RCB values for each state, we build an index considering as base level the average of the two lowest baskets values. A third correction attempt considers both time and area criteria, first correcting for the RCB and, subsequently, the time correction with IPCA. The details and estimation results considering both attempts, and nominal wages, are available upon request.

[14] It is worth noting that this correction is ignored by several studies on the UWP in Brazil and other countries, given the

The PNADC has a household identifier, but it does not include a longitudinal identifier for individuals. We build one based on the household identifier, date of birth, and sex of each individual[15]. Finally, PNADC is a complex sampling survey. It utilizes a probabilistic sample extracted from a master sample of sectors based on the National Demographic Census (IBGE), representing the whole population at different geographic levels.

3.2 Econometric model

The first step of our econometric strategy is to correct the sample selection bias due to the probability of participating in the labor market. Thus, we take advantage of data available on individuals' employment status and use the two-step estimation, following Heckman (1979)[16].

The second step is the estimation of the UWP using the POLS method. Studies traditionally use panel data approaches to estimate the UWP, such as pooled OLS (POLS) or fixed effects (FE), controlling for observable and non-observable and time-fixed characteristics. However, data from PNADC does not allow us to use FE models to estimate the UWP, as families that move from a city to another one during the survey are missing after migration. Thus, we estimate the urban wage premium using the POLS method, which allows for changes in parameters over time, even if some variables are not time-varying. According to Wooldridge (2010), with a large number of observations (N) and few periods (T), POLS estimation allows for aggregate time effects that have the same influence on W for all i. We estimate the following Mincer-type equation:

$$W_{it} = a + \beta X_{it} + \theta MA_i + \gamma \widehat{\lambda}_{it} + u_{it} \tag{1}$$

in which the dependent variable, W_{it}, is the natural logarithm of the hourly wage of worker i in period t. The vector of independent variables X_{it} includes i) individual and household characteristics (i.e., gender, age, age squared, race, marital status, household position, schooling level, an indicator of the presence of children, and region) including quarter and year, and ii) occupational and firm characteristics (i.e., industry, firm size, tenure, formality status, and occupational position). The variable u_{it} is the error term, and the variable $\widehat{\lambda}_{it}$ is the predicted inverse Mills ratio (IMR), recovery from sample selection correction.

Under this specification, the coefficient θ captures the UWP and indicates the wage premium associated with living in MAs. Also, for some estimates, the MA can be a categorical variable with four agglomeration levels: small, medium, large, and extra-large. We separately estimate the UWP for workers' subgroups based on the heterogeneity in terms of labor contract – formal or informal – and occupational position – wage-earner or self-employed. We also interact with the MA variable, the occupational skill and schooling level of workers to understand some individual characteristics' role.

lack of a price index at a specific geographical scale. Exceptions are the studies of Yankow (2006), Baum-Snow and Pavan (2011), and Chauvin *et al.* (2017), who correct the wages for price variations across time and areas, and those of Cruz and Naticchioni (2012), Barufi (2015), and Matano, Obaco and Royuela (2020), who only time-correct the wages.

[15] In this type of identification multiple individuals within a household may have the same identifier, as in the case of twins of the same sex. We address this possible situation by excluding duplicate observations.

[16] The results and tests for this correction are available upon request.

We proceed with some robustness checks to verify our results' sensitivity to sample selections, periods, and alternative definitions of agglomeration levels. Besides, we propose some robustness and heterogeneity exercises to test some appointments of the UWP literature. Some studies suggest that the UWP results from a high concentration of specific industries and, in turn, of certain types of jobs and workers profiles in MAs, which generate higher wages than less dense areas.

Therefore, different samples and techniques are used to investigate the sorting and matching mechanisms[17]. In this context, we investigate the firm-worker matching role by estimating the UWP for a subsample of individuals that change their job and occupational category between the first and second interviews of PNADC. Thus, we can investigate whether improvements in the matching in denser areas may positively impact wages in job-to-job transitions compared to the group of individuals who remain in the same job. Finally, we observe the sorting in dense areas through an alternative two-step procedure inspired by Combes, Duranton and Gobillon (2008)[18].

4 Results

4.1 Descriptive analysis

We first construct some descriptive statistics for our sample's individuals regarding the formality labor market status (Table 1). Using all observations from 2012 to 2019(Q3), we verify that formal and informal workers of MA and non-MA areas differ in terms of individual characteristics, household composition, and job features (columns 1-4).

Both formal and informal workers at MAs have more schooling years, fewer children, higher household income, and individual wages. However, there are relevant differences between areas within the groups. The educational level differential is higher for informal workers, with more 2.3 years for those in MAs compared to non-MAs. The same comparison for formal workers shows only 1.2 schooling years in favor of formal workers in MAs.

Wages are also higher in MAs for both formal and informal workers, but the difference reaches 29% for informal workers (an average hourly wage of R\$11.3 in MAs versus R\$8.0 in non-MAs). In contrast, it is only 13% (an average hourly-wage of R\$14.7 in MAs versus R\$12.9 in non-MAs) for formal workers. In MAs, the wage differential of formal workers compared to informal workers reaches 23.3%, while in non-MAs, the same comparison reaches 38.1%. Indeed, this is a preliminary, descriptive, and unconditional evidence that some groups of workers may experience wage differentials in urban areas.

Columns 5-12 of Table 1 present the descriptive statistics for the four subgroups of workers -

[17] Some examples of studies about the mechanisms are Combes, Duranton and Gobillon (2008), Baum-Snow and Pavan (2011), Combes *et al.* (2012), Andersson, Klaesson and Larsson (2014), Behrens, Duranton and Robert-Nicoud (2014), D'Costa and Overman (2014), Hamann, Niebuhr and Peters (2019) and Matano, Obaco and Royuela (2020).

[18] The first step is based on estimating Mincer's equation with FE, controlling for time-varying factors (Z_{it}). Then, we regress the estimated FE in a second step over time-fixed factors, including the agglomeration levels. This procedure allows us to verify how much of the FE is associated with individuals' locations. Formally, the procedure takes the following specification for the first step: $W_{it} = a + \beta Z_{it} + ci + u_{it}$ where c_i is the estimated FE for the sample, and for the second step: $\widehat{ci} = \vartheta + \theta MA_i + \varrho F_i + v_i$ where F_i represents the time-fixed control variables, that is, age, race, macro-region, and agglomeration levels.

– formal and informal wage earners, and formal and informal self-employed. Wages are higher for formal workers in MAs and non-MAs, but informal wage earners have the highest differential between areas – 51.1% -– followed by informal self-employed — with a 35.5% differential.

In MAs, informal wage earners earn 19.6% less than formal wage earners, as they work fewer hours, present less schooling years, are younger, and are more prone to be single. Informal wage earners in non-MAs earn 39.1% less than formal workers, almost double the difference in MAs. In MAs, informal self-employed earn 44.3% less than formal ones, while in non-MAs, the same comparison reaches a differential of 48.2% in favor of formal self-employed workers. Also, informal self-employed workers have bigger families, a lower household income, are non-whites over-represented, and are almost three years younger than formal workers.

The heterogeneity of workers in terms of labor contract formality and occupational position emerges from the descriptive data and justify the importance of the analysis we perform in this paper. Under competitive markets, working in the informal sector may be a choice based on personal preferences after considering all costs and benefits (TANSEL; KAN, 2012). However, descriptive analysis shows that even within informal workers, there is a high heterogeneity level with two different groups. An upper-tier group formed by high-skilled workers in specialized services, and a lower-tier group usually comprising low-skilled workers, with higher unemployment rates, and that faces entry barriers in the formal sector (HENLEY; ARABSHEIBANI; CARNEIRO, 2009; GÜNTHER; LAUNOV, 2012).

Table 1: Descriptive analysis: workers' profiles

	Formal		Informal		Formal-WE		Formal-SE		Informal-WE		Informal-SE	
	Non-MA	MA	Non-MA	MA	Non-MA	MA	Non-MA	MA	Non-MA	MA	Non-MA	MA
Avg/Share	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Individuals' characteristics												
Age	38.0	38.0	39.4	39.4	35.9	36.4	44.6	45.0	35.1	34.7	42.3	42.1
Schooling years	10.0	11.1	7.3	9.5	10.0	11.0	9.8	11.4	7.6	10.2	7.1	9.2
Race (whites)	55.0%	45.0%	37.0%	39.0%	52.0%	43.0%	66.0%	57.0%	35.0%	40.0%	38.0%	38.0%
Married	67.0%	62.0%	64.0%	58.0%	64.0%	59.0%	78.0%	74.0%	54.0%	48.0%	70.0%	64.0%
Household head	61.0%	55.0%	61.0%	54.0%	57.0%	52.0%	73.0%	69.0%	50.0%	42.0%	69.0%	61.0%
Household composition												
No Child	59.0%	63.0%	60.0%	65.0%	58.0%	62.0%	63.0%	66.0%	59.0%	66.0%	60.0%	65.0%
Children 0-6	0.26	0.24	0.27	0.22	0.28	0.24	0.22	0.20	0.29	0.22	0.26	0.23
Children 7-14	0.34	0.30	0.41	0.32	0.35	0.30	0.32	0.29	0.40	0.31	0.42	0.32
Household income^{a}	1,366.3	1,725.8	806.2	1,492.1	1,290.7	1,576.9	1,597.4	2,378.7	883.6	1,833.4	753.2	1,299.6
Family size	3.5	3.5	3.7	3.6	3.6	3.5	3.3	3.3	3.8	3.7	3.6	3.5
Head/spouse employed	55.0%	52.0%	44.0%	48.0%	53.0%	51.0%	60.0%	56.0%	42.0%	51.0%	46.0%	46.0%
Labor market												
Worked hours (week)	45.2	44.4	42.1	43.0	44.6	43.7	47.1	47.6	41.8	42.1	42.3	43.6
Avg hourly-wage^{a}	12.9	14.7	8.0	11.3	11.7	13.4	16.5	20.9	7.1	10.7	8.6	11.6
Tenure(months)	90.5	79.5	106.6	88.1	66.7	65.1	163.5	142.8	54.2	45.6	142.6	112.0
N	151,982	98,264	135,944	51,004	109,584	80,082	42,398	18,182	52,654	17,512	83,290	33,492
Share	60.7%	39.3%	72.7%	27.3%	57.8%	42.2%	70.0%	30.0%	75.0%	25.0%	71.3%	28.7%

Source: Created by the authors based PNADC sample from 2012 to 2019(Q3), with individual sample weights. **Notes:** ^{a} Values in R$ time-corrected to 2019-Q3 price level. A shows the sample construction.

4.2 Observing the UWP in agglomerations

4.2.1 MA and agglomeration levels

We estimate the UWP by adding the controls gradually, as reported in Table 2, with results without controls (1); with sample selection correction (2); with dummies for time and macro-regions (3); controls for industry, firm size, formality status and self-employment (4); and with occupation characteristics – tenure and skill level (5). Finally, the last regression is our reference model (6), including with workers' observed characteristics[19].

Panel (a) reports the model results with a single dummy for all MAs, which measures the overall UWP for metropolitan areas. MAs' effect on wages is initially 21.8% in the basic model (1), but it reduces to 5.78% with all controls (6). This result evidences a positive and significant UWP for MAs in Brazil when the whole labor market is analyzed, and it is in line with the literature.

As our focus is to deepen the analysis within the MAs, panel (b) shows the results for four agglomeration levels. The coefficients in the model (6) indicate a non-homogeneous UWP according to the agglomeration scale. Although the premium is always positive, extra-large MAs have a 4.71% premium, whereas we observe higher premia of 5.91%, 6.65%, and 11.5% for large, medium, and small MAs, respectively[20].

Several studies in the literature point to an increasing UWP as the agglomeration level increases, which is in line with the descriptive statistics and models 1-2. However, when we control for the worker, occupation, and firm characteristics in models 3 to 6, this pattern is inverted. So, we find a positive but decreasing UWP as the agglomeration level increases. This overall result may be partly related to the sorting of the most qualified workers to denser areas.

Nevertheless, our results are in line with other studies that report a lower UWP for denser areas. Guevara-Rosero, Pozo *et al.* (2020), for example, found a 3.5% UWP for the most populated city in Ecuador (Guayaquil), a lower magnitude than the 17.1% UWP of the second most populated (Quito) and lower than many other smaller cities.

4.2.2 The role of heterogeneity in the UWP

Given the high informality rates and the differences between formal and informal workers, we argue that the UWP could be distinct for these two groups of workers. Table 3 shows the POLS estimations for these groups separately. Again, we start by including a single dummy for MAs (panel a) and then considering the four agglomeration levels (panel b).

Compared to those in non-MAs, informal workers in MAs have an 8.28% UWP, double the formal workers UWP of 4.18%. Since the estimations consider wage earners as the base level, the self-employed dummy coefficient shows that these workers in the formal sector earn 22.1% more, but only 3.79% in the informal sector.

[19] The omitted categories are: non-MAs, less than one year of schooling, low occupational skill level, tenure of eight or more years, informal workers (or Informal-WE, when using workers subgroups), the year 2012, the first quarter, the southeast region, the agricultural industry, and firm size from one to five employees.

[20] The coefficients of other variables are in line with the literature and are available upon request.

Table 2: POLS regressions

	(1)	(2)	(3)	(4)	(5)	(6)
(a) MA						
MA	0.218***	0.207***	0.198***	0.0909***	0.0781***	0.0578***
	(0.00195)	(0.00186)	(0.00184)	(0.00184)	(0.00160)	(0.00151)
Observations	2,266,837	2,266,837	2,266,837	2,266,837	2,266,837	2,266,837
R-squared	0.029	0.119	0.200	0.303	0.400	0.459
(b) Agglomeration levels						
Small MA	0.0684***	0.0770***	0.282***	0.145***	0.138***	0.115***
	(0.00358)	(0.00344)	(0.00354)	(0.00336)	(0.00297)	(0.00281)
Medium MA	0.0962***	0.110***	0.208***	0.0895***	0.0841***	0.0665***
	(0.00304)	(0.00290)	(0.00289)	(0.00278)	(0.00243)	(0.00229)
Large MA	0.155***	0.149***	0.205***	0.0831***	0.0772***	0.0591***
	(0.00221)	(0.00211)	(0.00208)	(0.00204)	(0.00179)	(0.00168)
Extra-large MA	0.321***	0.298***	0.178***	0.0927***	0.0701***	0.0471***
	(0.00323)	(0.00308)	(0.00343)	(0.00326)	(0.00280)	(0.00259)
Observations	2,266,837	2,266,837	2,266,837	2,266,837	2,266,837	2,266,837
R-squared	0.037	0.126	0.201	0.304	0.400	0.459
Worker	No	No	No	No	No	**Yes**
Occupation	No	No	No	No	**Yes**	**Yes**
Industry (10)	No	No	No	**Yes**	**Yes**	**Yes**
Year/quarter	No	No	**Yes**	**Yes**	**Yes**	**Yes**
Macro-region	No	No	**Yes**	**Yes**	**Yes**	**Yes**
Heckman	No	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**

Notes: Dependent variable: *ln(hourly-wages)*. All models follow the base model specification, and include a constant term and a complex sampling design, with controls for worker characteristics, industry, firm size, year or quarter, macro-region, and IMR. The omitted category is non-MAs. Constant, controls and errors are omitted but are available upon request. The significance levels are *** p <0.01, ** p <0.05, and * p <0.1.

So, the higher UWP for informal workers could reduce the wage gap in urban areas. To illustrate this effect, we can assess it on the average wage of the sample. In non-MAs, formal workers have an average hourly wage of R$ 12.9, while informal workers earn an average of R$ 8.0 per hour, with the wage gap of 61.3% in favor of formal workers. Adding the respective UWPs in each group, formal workers have their average wage raised to R$ 14.3, and informal workers' wages to R$ 9.5 per hour, reducing the wage gap to 51.1%.

Regarding the historical perspective, one can ask if the higher UWP for informal workers is only a recent issue. To test this possibility, we estimate the same UWP specification using the PNAD database, a discontinued version of the Brazilian National Household Sample Survey (IBGE, 2020b), for years from 2001 to 2011. Figure 5 shows that informal workers UWP was always higher than the formal one at least a decade before our analysis, indicating that informal UWP superiority may be a typical result, at least for the Brazilian labor market. It also indicates that previous analyses focused on formal workers may underestimate the real UWP in developing countries.

Since little evidence of agglomeration effects for informal workers is available, the reasons for the higher informal UWP are not clear. Until now, the results show that agglomeration effects arise in both groups, but their magnitudes could be different. As our results, two studies for Colombia found a lower UWP for formal workers than informal ones. The UWP magnitude reaches a 4.7% UWP for formal workers and a 5.5% UWP for the whole labor market, meaning that informal workers overstate the formal's UWP in Duranton (2016). In turn, García (2019) found a 3.6% premium for informal workers and a penalty of 3.2% for formal workers.

Matano, Obaco and Royuela (2020) and Guevara-Rosero, Pozo *et al.* (2020) find opposite results, with a higher UWP for formal workers in Ecuador. However, both rely on a specific informal worker identification based on small firms without accounting records or taxpayer numbers. Indeed, this different concept is a confounding factor since one of these studies, Matano, Obaco and Royuela (2020), re-estimate theirs results using an identification similar to ours regarding informality (from the worker's point of view) and found results similar to ours, with a higher UWP for informal workers (4.2% versus 3.1%). Thus, the evidence of a higher UWP for informal workers seems to be a reality in developing countries, at least for Brazil, Colombia, and Ecuador.

Analyzing agglomeration levels, we see very different results between the formality status (Table 3, columns 4-6). Both sectors show a non-homogeneous UWP according to the agglomeration level. Still, for formal workers, the UWP ranges between 2.70% and 5.35% and is only increasing from medium to extra-large MA. In turn, informal workers UWP has a broader range of 4.19% to 18.0%, and it drives the results for the full sample, as it replicates the decreasing pattern as the agglomeration level increases. In extra-large MAs, the UWPs of both groups are closer, but they differ for lower agglomeration levels. This study provides, as far as we know, the first evidence of different UWP patterns of workers with different formality status according to agglomeration levels in Brazil.

Table 3: POLS regressions by formality status

	MA			Agglomeration levels		
	Informal	Formal	Full sample	Informal	Formal	Full sample
	(1)	(2)	(3)	(4)	(5)	(6)
MA	0.0828***	0.0418***	0.0578***			
	(0.00239)	(0.00180)	(0.00151)			
Small MA				0.180***	0.0498***	0.115***
				(0.00426)	(0.00338)	(0.00281)
Medium MA				0.116***	0.0270***	0.0665***
				(0.00387)	(0.00262)	(0.00229)
Large MA				0.0906***	0.0317***	0.0591***
				(0.00274)	(0.00195)	(0.00168)
Extra-large MA				0.0419***	0.0535***	0.0471***
				(0.00457)	(0.00294)	(0.00259)
Formality			0.153***			0.152***
			(0.00163)			(0.00163)
Self-employed	0.0379***	0.221***	0.117***	0.0377***	0.221***	0.117***
	(0.00223)	(0.00312)	(0.00192)	(0.00223)	(0.00312)	(0.00192)
Observations	968,372	1,298,465	2,266,837	968,372	1,298,465	2,266,837
R-squared	0.381	0.429	0.459	0.382	0.429	0.459

Notes: Dependent variable: *ln(hourly-wages)*. All models follow the base model specification, and include a constant term and a complex sampling design, with controls for worker characteristics, industry, firm size, year or quarter, macro-region, and IMR. The omitted category is non-MAs. Constant, controls and errors are omitted but are available upon request. The significance levels are *** p <0.01, ** p <0.05, and * p <0.1.

Workers' subgroups. Trying to establish how and why the previous results are different and its relationship with the urban structure, we separately estimate the UWP for four subgroups of workers. The reference model now has a dummy variable for each subgroup (Table 4), and we estimate the UWP for each subgroup (columns 1 to 4). The UWP coefficient for the full sample (5) exhibits small

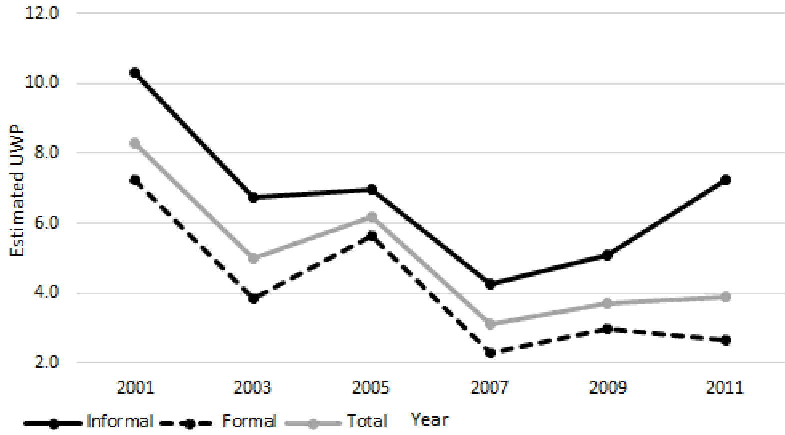

Figure 5: Historic UWP by formality status

Notes: Estimated UWP using PNAD database (IBGE, 2020b). Reported only UWP coefficients for each year in a cross-section approach for Equation 1, including a constant term, robust errors and individual sample weights. Omitted categories: non-MA, southeast region, and agricultural industry. Schooling levels not available but using a continuous variable for schooling years. Without firm size variable (not available). Considering wages time-correct at price levels of 2012. Complete results available upon request. The significance levels are: *** p <0.01, ** p <0.05, * p <0.1.

changes and the same pattern compared to previous results[21].

Results for each subgroup reveal that Formal-WE (1) has a general UWP positive but lower than other workers. Our UWP magnitude (3.44%) is lower than that found in studies for formal workers in Brazil, such as Chauvin *et al.* (2017), who find a 5.2% UWP, and Barufi (2015), who finds a 9.1% UWP. Part of this difference between our Formal-WE results and previous studies could be due to the correction of the sample selection bias that we perform. We also find an increasing UWP for Formal-WE for medium to extra-large MAs (i.e., 1.71% to 5.02% in panel (b), column 1). Higher UWP for formal workers as the agglomeration level increases are in line with previous studies; however, UWP for small MAs is higher than medium and large MAs.

The UWP of the other three groups is more than twice the Formal-WE UWP in MA's in general (panel a): 8.86% for Formal-SE, 8.03% for Informal-WE, and 8.19% for Informal-SE. Formal-SE exhibits a similar pattern by agglomeration level as Formal-WE, but with a higher UWP ranging from 7.96% to 15.7%. The subgroups of informal workers exhibit the opposite pattern; although they have similar coefficients, both Informal-WE and Informal-SE have a decreasing UWP as the agglomeration level increases. The decrease is more gradual for Informal-WE, ranging from 12.9% in small MAs to 6.32% in extra-large MAs, but it is steeper for Informal-SE, ranging from 20.0% to 2.81% from small to extra-large MAs, as shown in columns 3 and 4 of the panel (b).

This analysis indicates that disregarding intra-group heterogeneity hides the complexity inherent in these groups' inclusion in large urban labor markets, which corroborates the competitive markets approach. It also points out similar – but smaller – coefficients compared to previous studies for Formal-WE, whereas the other workers' subgroups lead to a different conclusion for the whole labor market.

[21] Similar results are found for the aggregated UWP of each subgroup when we interact with and agglomeration level. The results are available upon request.

Table 4: POLS regressions by workers' subgroups

	Formal-WE (1)	Formal-SE (2)	Informal-WE (3)	Informal-SE (4)	Full sample (5)
(a) MA					
MA	0.0344***	0.0886***	0.0803***	0.0819***	0.0597***
	(0.00185)	(0.00473)	(0.00330)	(0.00311)	(0.00150)
Observations	990,167	308,298	354,672	613,700	2,266,837
R-squared	0.471	0.333	0.439	0.368	0.464
(b) Agglomeration levels					
Small MA	0.0309***	0.157***	0.129***	0.200***	0.118***
	(0.00334)	(0.0100)	(0.00546)	(0.00546)	(0.00279)
Medium MA	0.0171***	0.0875***	0.100***	0.121***	0.0695***
	(0.00263)	(0.00735)	(0.00527)	(0.00497)	(0.00227)
Large MA	0.0215***	0.0796***	0.0822***	0.0929***	0.0603***
	(0.00198)	(0.00530)	(0.00364)	(0.00361)	(0.00167)
Extra-large MA	0.0502***	0.0916***	0.0632***	0.0281***	0.0492***
	(0.00301)	(0.00788)	(0.00653)	(0.00590)	(0.00258)
Observations	990,167	308,298	354,672	613,700	2,266,837
R-squared	0.472	0.334	0.440	0.369	0.465

Notes: All models follow the base model specification, and include a constant term and a complex sampling design, with controls for worker characteristics, industry, firm size, year or quarter, macro-region, and IMR. The omitted category is non-MAs. Constant, controls and errors are omitted but are available upon request. The significance levels are *** p <0.01, ** p <0.05, and * p <0.1.

4.2.3 Differences in the UWP by specific characteristics

Recent studies shed light on how the UWP changes according to individuals, occupations, and firms' characteristics instead of considering the UWP as fixed and homogeneous. After examining the raw UWP for workers subgroups, we re-estimate the model, including the interactions between the MA level and some selected characteristics – occupational skill and schooling level – allowing the UWPs of workers with each characteristic to vary across the urban hierarchy[22]. Table 5 reports the coefficients obtained for some chosen combination of agglomeration level, subgroup, and the selected characteristics.

Occupational skill. As expected, the interactions show that workers in high-skilled occupations have higher wage premia than low-skilled workers at all agglomeration levels (panel a). The UWP ranges from 10.6% to 16.1% for high-skilled occupations for the full sample, and it is always higher than for low-skilled workers at each agglomeration level and subgroup. Furthermore, the decreasing premium by agglomeration scale for low-skilled workers has a more stable pattern than that for high-skilled workers.

These results are in line with previous studies. Gould (2007) finds a premium in urban areas for white-collar workers and not for blue-collar workers. Bacolod, Blum and Strange (2009) report a 12.2% UWP for cognitive skills, a 2.9% UWP for people skills, and a penalty of 10.3% for motor-skill-intensive occupations. Although they use different skill specifications, both studies suggest agglomeration economies are motivated by skills and learning in denser areas, consistent with our findings.

[22]The complete estimation results for these exercises are available upon requests, such as the interaction between MA level and the firm size and tenure variables.

As pointed out in the previous section, we again evidence that only analyzing formal workers may hide the real pattern of the UWP in Brazil. High-skilled Formal-WE UWP increases with the agglomeration level, which is a typical result in the literature. Low-skilled workers have the lowest UWP among all groups and skill levels, with their UWPs near zero. Small MAs benefit low-skilled Formal-WE, whereas extra-large MAs provide higher premia for this same subgroup when they are high-skilled.

We observe different patterns and similarities across the subgroups of workers. Informal-WE overstates Formal-WE UWP in both skill levels, with the same pattern between areas. In turn, both formal and informal self-employed workers in high-skilled occupations have higher UWPs in smaller MAs, where the UWP reaches 22% for both groups. In low-skilled occupations, informal and formal self-employed workers have similar UWPs in small to large MAs, ranging between 7%-8% and 19%-18%. Extra-large MAs only provide a positive UWP for self-employed low-skilled workers in the formal sector, but not statically significant for Informal-SE. In general, the results show that workers in low-skilled occupations benefit from greater UWPs when they are at small MAs. In contrast, the benefits are greater and more stable for high-skilled workers across the urban scale.

Schooling levels. We also investigate the UWP for each schooling level and subgroup. Panel (b) shows the estimated UWP for the two schooling levels. For the full sample, a worker with a college degree or higher always has a higher UWP than workers with lower schooling levels. Only workers with a college degree or higher have an increasing UWP as the agglomeration level increases (and only from medium to extra-large MAs), a pattern similar to that previously observed for high-skilled workers.

Again, analyzing only Formal-WE, we observe a higher and increasing UWP for workers with a college degree or higher, whereas workers with only elementary barely benefit from any UWP. The results for the other three worker subgroups show different results. For Formal-SE, small MAs provide a higher UWP at any schooling level. In turn, from medium to extra-large MAs, the UWP for elementary school workers is small or does not exist.

Informal workers face a different situation regarding the UWP by schooling level. Workers with the elementary school have higher UWPs when located in small MAs, regardless of whether they are wage earners or self-employed workers. However, the results are different for workers with a college degree or higher. The UWP for wage earners varies between 14.4% and 15.9% for small to large MAs and reaches 20.6% in extra-large MAs, whereas self-employed workers have a higher UWP in small MAs (23.4%) and a lower UWP in other areas.

Other studies also report a higher UWP for high-educated workers. Glaeser and Mare (2001) find a UWP ranging from 2.3% to 13.1% for workers with more than ten schooling years in MAs in the US (equivalent to college-educated workers in Brazil). Baum-Snow and Pavan (2011) also investigate the US labor market and find a 29% UWP for college-educated workers in larger areas. For the Brazilian labor market, Silva, Santos and Freguglia (2016) find an approximately 13.2% UWP in extra-large MAs for college-educated formal workers and a non-homogeneous premium across schooling and MA levels.

Table 5: Aggregated effects by specific characteristics

		Agglomeration levels					
		Small MA	Medium MA	Large MA	Extra-large MA	Obs.	R^2
		(1)	(2)	(3)	(4)	(5)	(6)
(a) Occupational Skill							
Low	*Full sample*	0.112***	0.062***	0.037***	0.011***	2,266,837	0.465
	Formal-WE	0.04***	0.015***	0.003	0.015***	990,167	0.473
	Formal-SE	0.184***	0.099***	0.081***	0.118***	308,298	0.335
	Informal-WE	0.121***	0.097***	0.076***	0.041***	354,672	0.440
	Informal-SE	0.19***	0.11***	0.073***	-0.002	613,700	0.369
High	*Full sample*	0.161***	0.106***	0.141***	0.14***	2,266,837	0.465
	Formal-WE	0.05***	0.054***	0.128***	0.146***	990,167	0.473
	Formal-SE	0.222***	0.139***	0.135***	0.162***	308,298	0.335
	Informal-WE	0.138***	0.153***	0.138***	0.164***	354,672	0.440
	Informal-SE	0.22***	0.104***	0.13***	0.081***	613,700	0.369
(b) Schooling level							
Elementary	*Full sample*	0.098***	0.04***	0.016***	-0.003	2,266,837	0.466
	Formal-WE	0.03***	0.003	-0.013***	0.002	990,167	0.473
	Formal-SE	0.129***	0.04**	-0.002	0.02	308,298	0.335
	Informal-WE	0.116***	0.071***	0.057***	0.017*	354,672	0.441
	Informal-SE	0.162***	0.09***	0.059***	-0.002	613,700	0.370
College or higher	*Full sample*	0.144***	0.117***	0.157***	0.16***	2,266,837	0.466
	Formal-WE	0.045***	0.064***	0.13***	0.163***	990,167	0.473
	Formal-SE	0.191***	0.142***	0.171***	0.169***	308,298	0.335
	Informal-WE	0.15***	0.159***	0.144***	0.206***	354,672	0.441
	Informal-SE	0.236***	0.144***	0.15***	0.103***	613,700	0.370

Notes: Linear combination of the θ coefficient and the β related to the interaction term between the chosen variables and the Agglomeration levels. All models follow the base model specification, and include a constant term and a complex sampling design, with controls for worker characteristics, industry, firm size, year or quarter, macro-region, and IMR. The omitted category is non-MAs. Constant, controls and errors are omitted but are available upon request. The significance levels are *** p <0.01, ** p <0.05, and * p <0.1.

4.2.4 Robustness tests

This section is dedicated to robustness tests considering different workers samples and alternative definitions of agglomeration levels[23].

Attrition. The first two tests are related to the fact that the database consists of an unbalanced panel. The first test, shown in Table 6 (2), comprises only the first interview with each individual, and the second test (3) includes only individuals who remain in the survey during all five interviews. Both tests show similar results than the benchmark model (1), indicating that attrition does not significantly affect our results.

Exclusion of employers. This exercise aims to reduce the heterogeneity of our sample of workers. The exclusion of employers from the sample (4) does not modify the overall UWP results by agglomeration level. The sensitive coefficients are associated with formal and informal self-employed workers, an expected result since employers are self-employed workers in our primary definition. For formal-SE, the coefficient decreases from 28.4 to 20.3 but remains highly statistically significant and positive. We find a wage penalty for informal-SE of 2.37% – against the base level Informal-WE – compared to a positive effect of 2.01% in the base model. Still, the fact that Informal-WE faces a

[23]Results of three additional robustness tests are available upon request: i) the estimation of the UWP considering price variation across areas and time, and nominal wages; ii) the inclusion of local features as controls, such as temperature, rainfall, specialization level, and the Herfindahl index in Mincer's equation; iii) the exclusion of Brazil's capital city from the sample of cities, as this municipality has a proportionally higher share of workers in public administration. Results for the UWP are not sensitive to these sample selections.

wage penalty in MAs in this exercise indicates that employers have a wage level above the mean, which may bias the UWP results.

No restriction on working hours. We apply a common constraint on worker hours[24], focusing on workers in full-time jobs, excluding workers who report 1-19 weekly hours. Exercise (5) takes into account all workers without restrictions on working hours. Informal workers may work fewer hours compared to formal workers. Indeed, in the full PNADC database, 11% of the informal workers report 1-19 working hours, and only 1.1% of formal workers do the same. Again, the UWP in this exercise shows similar coefficients by agglomeration level compared to the base model.

Exclusion of firm size as a control variable. The firm size variable is unavailable in the PNADC in three quarters (2015Q4, 2016Q1, and 2016Q2). We perform a robustness test for the full survey period, discarding the firm size variable and re-estimating the UWP. In exercise (6), we verify that the overall results remain with slightly higher UWP levels.

Table 6: Robustness: sample selections

	Base Model (1)	Only 1st Interview (2)	Balanced panel (3)	Without employers (4)	All working hours (5)	Without firm size (6)
Small MA	0.118***	0.126***	0.132***	0.118***	0.117***	0.126***
	(0.00279)	(0.00437)	(0.00377)	(0.00272)	(0.00278)	(0.00273)
Medium MA	0.0695***	0.0744***	0.0706***	0.0717***	0.0720***	0.0814***
	(0.00227)	(0.00358)	(0.00317)	(0.00223)	(0.00229)	(0.00221)
Large MA	0.0603***	0.0621***	0.0648***	0.0603***	0.0613***	0.0719***
	(0.00167)	(0.00250)	(0.00224)	(0.00165)	(0.00168)	(0.00162)
Extra Large MA	0.0492***	0.0480***	0.0523***	0.0479***	0.0499***	0.0604***
	(0.00258)	(0.00360)	(0.00337)	(0.00255)	(0.00259)	(0.00254)
Formal-WE	0.0425***	0.0400***	0.0386***	0.0602***	0.0289***	0.124***
	(0.00191)	(0.00320)	(0.00258)	(0.00191)	(0.00192)	(0.00171)
Formal-SE	0.284***	0.284***	0.277***	0.203***	0.279***	0.233***
	(0.00273)	(0.00440)	(0.00347)	(0.00284)	(0.00273)	(0.00264)
Informal-SE	0.0201***	0.0188***	0.0133***	-0.0237***	0.0306***	-0.0422***
	(0.00205)	(0.00352)	(0.00268)	(0.00202)	(0.00203)	(0.00195)
Observations	2,266,837	437,194	1,453,949	2,144,276	2,354,554	2,533,910
R-squared	0.465	0.459	0.467	0.460	0.452	0.459

Notes: Dependent variable: *ln(hourly-wages)*. All models follow the base model specification, and include a constant term and a complex sampling design, with controls for worker characteristics, industry, firm size, year or quarter, macro-region, and IMR. The omitted category is non-MAs. The significance levels are *** p <0.01, ** p <0.05, and * p <0.1. Models 2 to 6 are the same as the base model estimations but with different samples: (2) Only observations from the first interview; (3) Only workers with all five interviews; (4) Sample excluding employers; (5) Sample including workers with 1-19 weekly working hours; (6) Sample w/o restrictions based on firm size data availability and disregarding firm size.

Exclusion of the states with non-capital MAs. Data from the PNADC only identifies state capital MAs, accounting for 41.1% of the Brazilian population. Noncapital MAs corresponds to 4.6% of the total population. We perform a robustness check excluding the São Paulo (SP) state, which concentrates almost half of the non-capitals MAs population (2.2% of the total population) to verify our results' consistency. We also rule out the other three states with non-capitals MAs, Santa Catarina,

[24]A similar method is used by Glaeser and Mare (2001), Gould (2007), Elvery (2010), Cruz and Naticchioni (2012), Carlsen, Rattsø and Stokke (2016), Berlingieri (2019), and Meekes and Hassink (2018).

Paraná, and Minas Gerais, in a second exercise. The results are in Table 7. The overall conclusions remain the same: the decreasing UWP for informal workers is even steeper when we exclude the four states, reducing from 18.3% to 2.3% with the agglomeration levels, and the trajectory of the UWP for formal workers is increasing from medium to extra-large MAs.

Table 7: Robustness: excluding non-capital MAs

	Informal			Formal		
	Base Model	Without SP state	Without four states	Base Model	Without SP state	Without four states
	(1)	(2)	(3)	(4)	(5)	(6)
Small MA	0.180***	0.184***	0.183***	0.0498***	0.0548***	0.0461***
	(0.00426)	(0.00427)	(0.00433)	(0.00338)	(0.00339)	(0.00346)
Medium MA	0.116***	0.128***	0.119***	0.0271***	0.0454***	0.0213***
	(0.00387)	(0.00388)	(0.00415)	(0.00262)	(0.00264)	(0.00298)
Large MA	0.0906***	0.101***	0.106***	0.0317***	0.0474***	0.0293***
	(0.00274)	(0.00274)	(0.00315)	(0.00195)	(0.00196)	(0.00246)
Extra-large MA	0.0419***	0.0548***	0.0237***	0.0535***	0.0534***	0.0329***
	(0.00457)	(0.00496)	(0.00580)	(0.00294)	(0.00331)	(0.00402)
Observations	968,362	909,947	761,201	1,298,475	1,139,763	787,725
R-squared	0.382	0.357	0.348	0.429	0.404	0.406

Notes: Dependent variable: *ln(hourly-wages)*. The four states excluded are São Paulo, Santa Catarina, Paraná, and Minas Gerais. All models follow the base model specification, and include a constant term and a complex sampling design, with controls for worker characteristics, industry, firm size, year or quarter, macro-region, and IMR. The omitted category is non-MAs. Constant, controls and errors are omitted but are available upon request. The significance levels are *** p <0.01, ** p <0.05, and * p <0.1.

Urban population density. We now test the use of an alternative agglomeration specification taking the MAs urban population density in natural logarithmic form. Data from the PNADC does not allow the identification of the density of municipalities. Thus, we use the average levels of urban density for non-MAs, which allows to compare the results with those of other studies[25].

Table 8 presents the results. We observe a positive effect on hourly wages for the full sample (formal and informal workers), indicating that a one percentage point increase in urban density means a 4.51% increase in wages. All four subgroups of workers present a positive and significant UWP. Again, Formal-WE has the lowest UWP (3.69%), which is also lower than those reported in previous studies that use demographic density, such as 9.1% (BARUFI, 2015) and 7.4% (SILVA, 2017)[26]. In turn, informal workers have almost the same UWP levels across subgroups, as we find UWPs of 4.91% for Informal-WE and 4.07% for Informal-SE. The highest UWP is up to the formal-SE group with 8.20%. These results are in line with the previous results of this paper and close enough to other studies' results in the literature.

[25]The same results are obtained when using the municipalities' total demographic density. We follow the literature and use historical urban density levels as instrumental variable (IV) and a 2SLS estimation, to deal with the endogeneity. Some examples of the use of this kind of IV are Combes, Duranton and Gobillon (2008), Duranton (2016), and García (2019). We find a positive and significant UWP using urban population density levels in 1950, 1980, and 1991 as IV. These results are available upon request

[26]It is worth remembering that our estimate considers the correction of the sample selection bias, which is not done by these studies. Therefore, part of the differential can be attributed to this correction.

Table 8: Robustness: urban population density as the agglomeration definition

	Formal-WE	Formal-SE	Informal-WE	Informal-SE	Full sample
	(1)	(2)	(3)	(4)	(5)
Urban density (ln)	0.0369***	0.0820***	0.0491***	0.0407***	0.0451***
	(0.00247)	(0.00647)	(0.00469)	(0.00433)	(0.00204)
Observations	990,167	308,298	354,672	613,700	2,266,837
R-squared	0.471	0.332	0.437	0.366	0.463

Notes: Dependent variable: *ln(hourly-wages)*. All models follow the base model specification, and include a constant term and a complex sampling design, with controls for worker characteristics, industry, firm size, year or quarter, macro-region, and IMR. Constant, controls and errors are omitted but are available upon request. The significance levels are *** p <0.01, ** p <0.05, and * p <0.1.

4.3 The role of firm-employee matching in agglomeration effects

Now we investigate the role of firm-employee matching in agglomeration effects. We do so by restricting our sample to workers that change jobs -– job-to-job (JTJ) transitions -– between the first and second interviews of the survey, and also change the occupational category in this job change. In this scenario, we can investigate whether the matching process in denser areas positively impacts wages compared to individuals who remain in the same job over time. Figure 6 shows the matrix of changes, and we see that JTJ transitions are more frequent in low-skilled occupations, such as elementary, mechanics, construction, and farming[27].

First interview		Second interview: A	B	C	D	E	F	G	H	I	J	Total
Directors and Managers	A	0.0%	0.8%	1.1%	0.6%	2.7%	0.5%	1.2%	0.5%	0.5%	0.0%	8.0%
Science and Intellectuals Professionals	B	0.8%	0.0%	1.1%	0.4%	0.6%	0.0%	0.4%	0.2%	0.1%	0.0%	3.6%
Mid-level Professionals and Technicians	C	1.1%	1.2%	0.0%	1.3%	1.6%	0.2%	3.2%	1.1%	0.8%	0.1%	10.4%
Administrative Support	D	0.6%	0.4%	1.3%	0.0%	1.1%	0.1%	0.8%	0.7%	1.1%	0.0%	5.9%
Service Workers and Salespeople	E	2.6%	0.6%	1.5%	1.1%	0.0%	0.6%	2.9%	1.4%	2.7%	0.0%	13.3%
Skilled Farming Forestry, Hunting and Fishing	F	0.5%	0.0%	0.1%	0.1%	0.6%	0.0%	1.0%	0.9%	6.3%	0.0%	9.5%
Skilled Construction	G	1.1%	0.4%	3.1%	0.8%	2.8%	0.9%	0.0%	3.8%	5.0%	0.0%	17.8%
Mechanical, Plant and Machinery Operators	H	0.5%	0.2%	1.0%	0.6%	1.3%	0.8%	3.8%	0.0%	3.1%	0.0%	11.1%
Elementary occupations	I	0.5%	0.1%	0.8%	1.1%	3.0%	6.2%	5.2%	3.2%	0.0%	0.0%	20.1%
Ill-defined occupations	J	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.2%
Total		7.6%	3.7%	10.1%	5.8%	13.8%	9.3%	18.3%	11.7%	19.5%	0.2%	100.0%

Figure 6: Matrix of job-to-job changes between occupations

Source: Created by the authors based on occupation changes between the first and second interview for employed individuals in the PNADC sample from 2012 to 2019(Q3).

For the sub-sample of movers, the UWP coefficient remains similar to that of the base model (Table 9, columns 1-2), although it is higher for extra-large (6.7% versus 4.9%) and large MAs (6.7% versus 6.0%). Considering all workers who responded to the first and second interviews, individuals with occupational JTJ transitions benefits from a 1.0% premia than those who stay in the same job over time (3). Forth and fifth columns present estimations with a dummy for JTJ transitions and interactions with worker subgroups and agglomeration levels. The general pattern for the UWP does

[27] Similar results are found considering a sub-sample of individuals with job-to-job transitions between industries.

not change. A job-to-job transition with a shift to another occupation increases the wages by 1.43% but can be higher or lower according to the workers' subgroup (4). An even higher premium occurs for Formal-SE (+3.20%), while Formal-WE's return is reduced by 1.39%.

Finally, we interact the dummy for the occupation transition with the agglomeration levels to obtain the aggregated UWP. In this case, a JTJ transition does not provide itself a statistically significant return on wages. Still, when it occurs in large and extra-large MAs, the workers have an additional premium of 1.30% and 2.48%, respectively. Even with this additional premium, the UWP of these areas remains lower than that of small MAs.

Two recent studies also found a positive UWP for workers with JTJ transitions between occupations. Matano, Obaco and Royuela (2020) found a higher premium in Ecuador for workers with JTJ transitions than those who remained in the same job — a 7.2% premium for formal workers and an even higher 18.4% premium for informal workers. Focusing on the German labor market, Hamann, Niebuhr and Peters (2019) found a premium of around 1.3% for workers with JTJ transitions. Despite methodological differences[28], we also evidence that JTJ transitions can influence the dynamics of UWP, which corroborates with a broader perspective on the effects of agglomeration, exploring its non-homogeneous benefits.

The higher benefits for job changes in denser areas can arise from the matching mechanism. Although our analyses focus on direct transitions between jobs, learning externalities may also be part of the UWP explanation. Human capital accumulation may benefit those workers who do not go through unemployment.

4.4 Control for individual fixed-effects

We control individual fixed-effects following a two-step procedure inspired by Combes, Duranton and Gobillon (2008). Table 10 presents the FE by formality status and agglomeration level from the first step estimation. For the full sample, the FE is higher in extra-large MAs and increases with agglomeration level. Still, non-MA workers have results similar to workers of medium MAs (43.3% versus 43.6%, respectively).

The results are different according to the formality status. Although formal workers have an increasing FE, those in non-MAs have the lowest level. In turn, informal workers have a lower FE compared to formal workers at small, large, and extra-large MAs. The exceptions are those in medium MAs, with FE of 46.5%, and workers in non-MAs, with a 44.1% FE. These first results indicate possible positive sorting in denser areas, with higher FE levels, driven by formal workers.

In the second step, we regress the estimated FE using as controls only time-fixed variables, that is, MA or agglomeration levels, macro-region, age, and race. The results (Table 11) show that being in an MA (panel a) explains 17.2% of the FE in the full sample, an average level between formal and informal workers – 12.4% and 21.0%, respectively. Again, Formal-WE presents the lower coefficients compared to the other three groups.

Looking at the agglomeration levels, we see a decreasing pattern as the MA becomes denser (panel b), varying between 15.1% and 25.1%, and distinctions between sectors. For formal workers (column

[28]Both cited studies consider demographic density as the interest variable.

Table 9: UWP for workers with job-to-job transition and occupational change

		Between occupations			
	Base Model (1)	Only movers (2)	Dummy for transitions (3)	Movers x subgroups (4)	Movers x agg. levels (5)
Small MA	0.118***	0.120***	0.120***	0.120***	0.117***
	(0.00279)	(0.00739)	(0.00352)	(0.00352)	(0.00396)
Medium MA	0.0695***	0.0694***	0.0709***	0.0708***	0.0702***
	(0.00227)	(0.00622)	(0.00281)	(0.00281)	(0.00312)
Large MA	0.0603***	0.0671***	0.0595***	0.0595***	0.0564***
	(0.00167)	(0.00429)	(0.00201)	(0.00201)	(0.00225)
Extra-large MA	0.0492***	0.0670***	0.0471***	0.0471***	0.0416***
	(0.00258)	(0.00673)	(0.00301)	(0.00301)	(0.00328)
Formal-WE	0.0425***	0.0307***	0.0430***	0.0444***	0.0430***
	(0.00191)	(0.00499)	(0.00242)	(0.00266)	(0.00242)
Formal-SE	0.284***	0.318***	0.285***	0.276***	0.285***
	(0.00273)	(0.00717)	(0.00345)	(0.00384)	(0.00345)
Informal-SE	0.0201***	0.0267***	0.0230***	0.0207***	0.0230***
	(0.00205)	(0.00556)	(0.00267)	(0.00298)	(0.00267)
Occupation movers (OC)			0.0104***	0.0143***	0.00368
			(0.00193)	(0.00429)	(0.00241)
OC x Formal-WE				-0.0139***	
				(0.00492)	
OC x Formal-SE				0.0320***	
				(0.00748)	
OC x Informal-SE				-0.00881	
				(0.00579)	
OC x Small MA					0.0114
					(0.00799)
OC x Medium MA					0.00318
					(0.00678)
OC x Large MA					0.0130***
					(0.00471)
OC x Extra-large MA					0.0248***
					(0.00651)
Observations	2,266,837	217,680	936,182	936,182	936,182
R-squared	0.465	0.456	0.461	0.461	0.461

Notes: Dependent variable: *ln(hourly-wages)*. All models follow the base model specification, and include a constant term and a complex sampling design, with controls for worker characteristics, industry, firm size, year or quarter, macro-region, and IMR. The omitted category is non-MAs. Constant, controls and errors are omitted but are available upon request. The significance levels are *** p <0.01, ** p <0.05, and * p <0.1.

3), from medium to extra-large MAs, the UWP premium increases between 9.55 and 13.8%. In turn, informal workers (column 6) showed the higher coefficients in all agglomeration levels, reaching a 30.6% premium in small areas. The decreasing pattern remains. This result confirms previous findings of the relevance of informal workers on the overall results.

The analysis for the worker's subgroups also confirms the same pattern found by POLS estimation for Formal-WE, with increasing UWP from medium to extra-large MAs, although at different magnitudes (column 1). In turn, Informal-WE shows a higher and stable UWP, varying between 20.6 to 24.3% (column 4). The overall decreasing pattern comes from the self-employed group, where both formal and informal workers present lower UWP levels in denser MAs.

Table 10: Average fixed effects

	Informal (1)	Formal (2)	Full sample (3)
Non-MA	0.441	0.411	0.433
	(0.3978)	(0.3636)	(0.3818)
Small MA	0.407	0.442	0.420
	(0.3632)	(0.383)	(0.3707)
Medium MA	0.465	0.442	0.436
	(0.4017)	(0.3794)	(0.3764)
Large MA	0.427	0.456	0.438
	(0.3782)	(0.3852)	(0.3786)
Extra-large MA	0.419	0.463	0.448
	(0.3716)	(0.3822)	(0.3794)

Notes: Dependent variable: *ln(hourly-wages)*. For formal and informal estimations, the omitted category is wage earners, and that for the full sample is wage earners in the informal sector. All models under robust standard errors (in parentheses). Controls: worker characteristics, industry, firm size, year, and quarter.

Table 11: Fixed effects explained by Agglomeration levels

	Formal			Informal			Full
	WE	SE	Total	WE	SE	Total	sample
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(a) MA							
MA	0.102***	0.225***	0.124***	0.214***	0.208***	0.210***	0.172***
	(0.00224)	(0.00511)	(0.00210)	(0.00365)	(0.00321)	(0.00251)	(0.00170)
Observations	990,167	308,298	1,298,465	354,672	613,700	968,372	2,266,837
R-squared	0.149	0.092	0.132	0.242	0.210	0.226	0.207
(b) Agglomeration levels							
Small MA	0.0828***	0.326***	0.129***	0.243***	0.336***	0.306***	0.251***
	(0.00399)	(0.0118)	(0.00408)	(0.00633)	(0.00571)	(0.00460)	(0.00326)
Medium MA	0.0665***	0.217***	0.0955***	0.219***	0.244***	0.237***	0.184***
	(0.00324)	(0.00829)	(0.00316)	(0.00618)	(0.00541)	(0.00428)	(0.00269)
Large MA	0.0833***	0.233***	0.113***	0.206***	0.217***	0.211***	0.178***
	(0.00239)	(0.00586)	(0.00229)	(0.00402)	(0.00385)	(0.00297)	(0.00194)
Extra-large MA	0.127***	0.210***	0.138***	0.219***	0.154***	0.179***	0.151***
	(0.00390)	(0.00893)	(0.00364)	(0.00771)	(0.00654)	(0.00514)	(0.00314)
Observations	990,167	308,298	1,298,465	354,672	613,700	968,372	2,266,837
R-squared	0.149	0.093	0.132	0.242	0.212	0.226	0.208

Notes: Dependent variable: *Fixed-Effects*. Second step estimation following the procedure inspired by Combes, Duranton and Gobillon (2008). All models were estimated without sample selection correction, with individual sample weights and robust standard errors (in parentheses). Constant, controls and errors are omitted but are available upon request. The significance levels are *** p <0.01, ** p <0.05, and * p <0.1.

5 Final Remarks

In this paper, we revisited the relationship between the UWP and agglomeration levels, considering an additional heterogeneity of workers in terms of the labor contract characteristics and their occupational position as a wage-earner or self-employed. We address this issue to the greatest possible extent with extensive and recent Brazilian longitudinal panel data on employment, unemployment, formality status, and workers' occupational features — the PNADC. We estimate the UWP by MA for four different workers' groups: a) formal wage-earner; b) informal wage-earner; c) formal self-employed; and d) informal self-employed, covering all Brazilian labor market. To ensure the validity of our results, we performed robustness tests with several sub-samples and agglomeration definitions.

The results add new insights to the explanation of the UWP in Brazil. First, we reveal that the results previously reported in the UWP literature based on the formal sector may underestimate the premium's magnitude, at least when there is a large proportion of informal workers, like the Brazilian labor market. The UWP of informal workers, on average, is double the formal workers UWP — 8.28% and 4.18%, respectively. Both sectors have a positive and significant UWP, but they exhibit different patterns along the agglomeration scale. The UWP of formal workers ranges between 2.70% and 5.35%, which are lower estimates than those found in studies focused on Brazil's formal workers. The UWP of informal workers varies widely, from 4.19% to 18.0%, and decreases with agglomeration levels. Thus, we find evidence that workers benefit differently from the agglomeration effects. Moreover, in an additional estimation, we evidence that the largest UWP for informal workers is not a recent Brazil scenario.

Second, the intra-group heterogeneity analysis shows that formal wage earners have a lower UWP than in previous studies, but that is positive and increasing from medium to extra-large MAs. At least part of this UWP differential can be attributable to the sample selection correction, which is possible with our data. In contrast, other subgroups exhibit a decreasing pattern by agglomeration level. Again, our results provide evidence that disregarding informal workers hides the complexity inherent in their inclusion in large urban labor markets.

Third, we observe differences in the UWP by exploring two specific characteristics: occupational skill-level and schooling levels. Both low-skilled and less-educated workers benefit from a higher premium in small-MAs, regardless of their formality status or occupational position. In contrast, throughout the agglomeration levels, the UWP trajectory depends on high-skilled or college-educated workers' occupational position. Wage-earners present a higher UWP in extra-large MAs, whereas self-employed benefits from a UWP in small-MAs. Formal-WE is outperformed by the other groups regardless of the characteristic analyzed.

Furthermore, job-to-job transitions between occupations add a wage premium for workers in MAs, meaning that workers benefit from better matching, especially in large and extra-large MAs or for Formal-SE. We also perform a two-step procedure inspired by Combes, Duranton and Gobillon (2008) to highlight the positive sorting in denser areas, with higher FE levels in general, driven by formal workers. A non-negligible part of individuals' FE (between 15.1% and 25.1%) can be explained by location in agglomerated areas in the full sample. This proportion is even higher for informal workers – between 17.9% and 30.6%. This evidence corroborates the previous results illustrating the

UWP heterogeneity.

Our main contribution relies on an in-depth analysis of the UWP by considering a heterogeneity source that was not previously investigated in this literature and by treating wage earners differently from self-employed workers, whether formal or informal. Thus, our results corroborate with recent studies for developing countries' labor markets that found different UWP magnitudes according to the worker's formality status. However, to the best of our knowledge, our study provides the first evidence of distinct patterns in the UWP for workers of different formality statuses and occupational positions according to agglomeration levels, in a developing country with a large proportion of informal workers.

By analyzing the whole labor market, we find a distinct pattern according to the agglomeration level, which is not a result elucidated in previous studies of the UWP that use employers' administrative records, thus restricting the evidence UWP to formal workers. Another contribution of this paper relies on evidence of the mechanisms that can generate the UWP in MAs (sorting and matching) using different estimations. Lastly, we take advantage of the availability of data to estimate and correct the sample selection bias.

The evidence on an urban wage premium with different magnitude, direction, and trajectory, depending on the heterogeneity of workers in terms of the labor contract's characteristics and occupational position, can guide new policies. The higher informal UWP can have a compensatory role, highlighting the need to assess how companies in urban centers respect regulations and, more than that, may indicate fiscal inefficiencies. Also, the similar magnitude between formal and informal wage-earners UWPs in very dense areas may result from more frequent enforcement of labor regulations, indicating the need to review inspection processes.

Further research agenda could involve understanding which mechanism is behind generating a higher wage premium for informal workers, including labor enforcement. Furthermore, it should involve elucidating the reasons for the UWP differences between occupational positions (subgroups of workers) and the role of JTJ transitions in the dynamics of the UWP.

A Building the sample

Table A. Building the sample

		Observations		Loss	
		Absolute	*%*	*Absolute*	*%*
		(1)	(2)	(3)	(4)
	Total Observations	17,545,779	100.0%		
Step 1	Fill errors and duplicates	15,409,648	87.8%	2,136,131	12.2%
Step 2	Household with more than on head	14,639,300	83.4%	770,348	4.4%
Step 3	Below 18 or above 65 years old	9,202,936	52.5%	5,436,364	31.0%
Step 4	Women	4,442,376	25.3%	4,760,560	27.1%
Step 5	In occupations with specific labor laws	4,040,675	23.0%	401,701	2.3%
Step 6	In temporary positions at the household	4,020,134	22.9%	20,541	0.1%
Step 7	Inactive	3,425,812	19.5%	594,322	3.4%
Step 8	Unemployed	3,015,201	17.2%	410,611	2.3%
Step 9	With more than 1 job	2,926,937	16.7%	88,264	0.5%
Step 10	Domestic workers	2,892,285	16.5%	34,652	0.2%
Step 11	With less than 20 working hours a week	2,786,172	15.9%	106,113	0.6%
Step 12	Without firm size	2,499,363	14.2%	286,809	1.6%
Step 13	With only one interview	2,313,095	13.2%	186,268	1.1%
Step 14	At top or bottom 1% wages	2,266,837	12.9%	46,258	0.3%
	Final sample	2,266,837	12.9%		

Source: Elaborated by the authors based on PNADC from 2012 to 2019(Q3).

B Agglomeration levels definition

Table B. Agglomeration levels definition

MA	State	Macro-Region	Population	Benchmark Agg. levels	MA Density	MA Urban Density	*REGIC*	Alternative Agglomeration Levels		
								MA Density	MA Urban Density	*REGIC*
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
São Paulo	São Paulo	Southeast	21,734,682	Extra-Large MA	2,714.4	7,205.4	1A	Extra-Large MA	Extra-Large MA	Extra-Large MA
Rio de Janeiro	Rio de Janeiro	Southeast	12,763,459		1,881.9	5,086.9	1B	Extra-Large MA	Extra-Large MA	Extra-Large MA
Belo Horizonte	Minas Gerais	Southeast	5,961,895	Large MA	395.0	3,290.0	1C	Medium MA	Large MA	Large MA
Porto Alegre	Rio G. do Sul	South	4,340,733		417.5	3,169.0	1C	Large MA	Medium MA	Large MA
Fortaleza	Ceará	Northeast	4,106,245		547.6	3,134.7	1C	Large MA	Medium MA	Large MA
Recife	Pernambuco	Northeast	3,999,817		1,439.5	6,357.7	1C	Extra-Large MA	Extra-Large MA	Large MA
Salvador	Bahia	Northeast	3,929,209		896.5	4,907.4	1C	Large MA	Large MA	Large MA
Curitiba	Paraná	South	3,654,960		218.1	2,896.8	1C	Medium MA	Medium MA	Large MA
Distrito Federal	Distrito Federal	Midwest	3,015,268		516.4	2,554.2	1B	Large MA	Medium MA	Extra-Large MA
Manaus	Amazonas	North	2,676,936		20.7	3,727.1	1C	Small MA	Large MA	Large MA
Goiânia	Goiás	Midwest	2,606,931		349.2	2,351.9	1C	Medium MA	Small MA	Large MA
Belém	Pará	North	2,510,274		698.3	5,106.8	1C	Large MA	Extra-Large MA	Large MA
Vitória	Espírito Santo	Southeast	1,979,337	Medium MA	839.9	3,967.1	1C	Large MA	Large MA	Large MA
São Luís	Maranhão	Northeast	1,633,117		195.2	2,560.3	2A	Medium MA	Medium MA	Medium MA
Natal	Rio G. do Norte	Northeast	1,604,067		445.6	3,065.0	2A	Large MA	Medium MA	Medium MA
Maceió	Alagoas	Northeast	1,338,756		465.3	4,993.3	2A	Large MA	Large MA	Medium MA
João Pessoa	Paraíba	Northeast	1,278,401		453.7	3,850.5	2A	Large MA	Large MA	Medium MA
Florianópolis	Santa Catarina	South	1,209,818		159.2	2,055.5	1C	Medium MA	Small MA	Large MA
Cuiabá	Mato Grosso	Midwest	1,041,307		14.1	1,835.8	2A	Small MA	Small MA	Medium MA
Aracaju	Sergipe	Northeast	961,120	Small MA	1,094.4	4,348.4	2A	Extra-Large MA	Large MA	Medium MA
Campo Grande*	Mato G. do Sul	Midwest	895,982		109.4	2,220.8	2A	Medium MA	Small MA	Medium MA
Teresina*	Piauí	Northeast	864,845		619.3	3,259.6	2A	Large MA	Large MA	Medium MA
Macapá	Amapá	North	646,323		29.6	2,745.2	2C	Small MA	Medium MA	Small MA
Porto Velho*	Rondônia	North	529,544		15.2	2,375.0	2B	Small MA	Small MA	Small MA
Rio Branco*	Acre	North	407,319		45.4	2,312.8	2C	Small MA	Small MA	Small MA
Boa Vista*	Roraima	North	399,213		66.0	1,956.9	2C	Small MA	Small MA	Small MA
Palmas*	Tocantins	North	299,127		131.5	1,187.5	2B	Medium MA	Small MA	Small MA

Source: Created by the authors. Estimated population for 2018 (IBGE, 2020a). **Notes:** *Only the state's capital. **MA density** is classified into four levels according to the number of inhabitants per area (km^2), namely, small MAs are those with less than 100, medium MAs are those between 100 and 417 (the median level), large MAs are those between 417 and 1,000, and extra-large MAs are those above 1,000. **MA urban density** is classified in four levels according to the number of inhabitants per urban area (km^2), namely, small MAs are those with less than 2,500, medium MAs are those between 2,500 and 3,200, large MAs are those between 3,200 and 5,000, and extra-large MAs are those above 5,000. The **REGIC** definition uses Cities' Influence Regions (*Região de Influência das Cidades*) from IBGE (2018c), which classifies urban centers according to the intensity of their links.

C Robustness tests: Agglomeration levels classification

Our primary agglomeration levels considering the MAs' population, but we also test three other definitions, as shown in B columns 9-11. We first reclassify the MAs according to their demographic density. Compared to the benchmark, the extra-large MAs group includes more MAs, Recife, and Aracaju, in addition to São Paulo and Rio de Janeiro, and the other groups are mixed.

The second exercise proposes another classification for MAs that considers urban population density – urban inhabitants per km^2. The motivation for this approach is that using the full demographic density may hide MAs' particularities. The final specification uses the Cities' Influence Regions (*Região de Influência das Cidades*) from IBGE (2018c), which classifies urban centers according to their links' intensity, and given their definition, they overlap the administrative boundaries of MAs. This specification shows small differences in the extra-large and medium MA groups relative to the benchmark.

We see a similar pattern compared to the base model for both MA levels (Table C), especially for Models 2 and 4. For the urban density classification (3), we observe lower UWPs for small MAs, but the other agglomeration levels show similar coefficients to those of the base model.

Table C. Robustness: alternative definitions of Agglomeration levels

	Base Model (1)	Demographic density (2)	Urban density (3)	REGIC (4)
Small MA	0.118***	0.0773***	0.0499***	0.152***
	(0.00279)	(0.00301)	(0.00267)	(0.00393)
Medium MA	0.0695***	0.0480***	0.0625***	0.0829***
	(0.00227)	(0.00216)	(0.00200)	(0.00243)
Large MA	0.0603***	0.0667***	0.0738***	0.0631***
	(0.00167)	(0.00189)	(0.00222)	(0.00164)
Extra-large MA	0.0492***	0.0579***	0.0536***	0.0468***
	(0.00258)	(0.00236)	(0.00230)	(0.00242)
Observations	2,266,837	2,266,837	2,266,837	2,266,837
R-squared	0.465	0.464	0.464	0.465

Notes: Dependent variable: *ln(hourly-wages)*. Agglomerations levels classification as shown in B. All models include a constant term and a complex sampling design with individual sample weights, and they all follow the base model specification with controls for worker characteristics, industry, firm size, year or quarter, macro-region, and IMR. The omitted categories are non-MAs, less than one year of schooling, low occupational skill level, tenure of eight or more years, informal wage earners, the year 2012, the first quarter, the southeast region, the agricultural industry, and firm size from one to five employees. Constant, controls and errors are omitted due to space restrictions and are available upon request. The significance levels are *** p <0.01, ** p <0.05, and * p <0.1.

References

ALMEIDA, R.; CARNEIRO, P. Enforcement of labor regulation and informality. *American Economic Journal: Applied Economics*, v. 4, n. 3, p. 64–89, 2012. Available at: <https://www.aeaweb.org/articles?id=10.1257/app.4.3.64>.

ANDERSSON, M.; KLAESSON, J.; LARSSON, J. P. The sources of the urban wage premium by worker skills: Spatial sorting or agglomeration economies? *Papers in Regional Science*, Wiley Online Library, v. 93, n. 4, p. 727–747, 2014. Available at: <https://doi.org/10.1111/pirs.12025>.

ANDERSSON, M.; THULIN, P. Does spatial employment density spur inter-firm job switching? *The Annals of Regional Science*, Springer, v. 51, n. 1, p. 245–272, 2013. Available at: <https://link.springer.com/content/pdf/10.1007/s00168-012-0544-y.pdf>.

BACOLOD, M.; BLUM, B. S.; STRANGE, W. C. Skills in the city. *Journal of Urban Economics*, Elsevier, v. 65, n. 2, p. 136–153, 2009. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S0094119008001083>.

BARROS, R. P. d.; ULYSSEA, G. On the empirical content of the formal-informal labor market segmentation hypothesis. *Brazilian Review of Econometrics*, Sociedade Brasileira de Econometria-SBE, v. 30, n. 2, 2010. Available at: <http://dx.doi.org/10.12660/bre.v30n22010.3648>.

BARUFI, A. M. B. *Agglomeration economies and labour markets in Brazil*. Phd Thesis (PhD Thesis) — Universidade de São Paulo, 2015. Available at: <https://www.teses.usp.br/teses/disponiveis/12/12138/tde-04022016-162856/en.php>.

BAUM-SNOW, N.; PAVAN, R. Understanding the city size wage gap. *The Review of Economic Studies*, Oxford University Press, v. 79, n. 1, p. 88–127, 2011. Available at: <https://doi.org/10.1093/restud/rdr022>.

BEHRENS, K.; DURANTON, G.; ROBERT-NICOUD, F. Productive cities: Sorting, selection, and agglomeration. *Journal of Political Economy*, University of Chicago Press Chicago, IL, v. 122, n. 3, p. 507–553, 2014. Available at: <https://doi.org/10.1086/675534>.

BERLINGIERI, F. Local labor market size and qualification mismatch. *Journal of Economic Geography*, Oxford University Press, v. 19, n. 6, p. 1261–1286, 2019. Available at: <https://doi.org/10.1093/jeg/lby045>.

BONNET, F.; VANEK, J.; CHEN, M. Women and men in the informal economy: A statistical brief. *International Labour Office, Geneva.*, 2019. Available at: <https://www.ilo.org/global/topics/employment-promotion/informal-economy/publications/WCMS_711798/lang--en/index.htm>.

CARLSEN, F.; RATTSØ, J.; STOKKE, H. E. Education, experience, and urban wage premium. *Regional Science and Urban Economics*, Elsevier, v. 60, p. 39–49, 2016. Available at: <https://doi.org/10.1016/j.regsciurbeco.2016.06.006>.

CHAUVIN, J. P. *et al.* What is different about urbanization in rich and poor countries? Cities in Brazil, China, India and the United States. *Journal of Urban Economics*, Elsevier, v. 98, p. 17–49, 2017. Available at: <https://doi.org/10.1016/j.jue.2016.05.003>.

COMBES, P.-P.; DURANTON, G.; GOBILLON, L. Spatial wage disparities: Sorting matters! *Journal of Urban Economics*, Elsevier, v. 63, n. 2, p. 723–742, 2008. Available at: <https://doi.org/10.1016/j.jue.2007.04.004>.

COMBES, P.-P. *et al.* Sorting and local wage and skill distributions in France. *Regional Science and Urban Economics*, Elsevier, v. 42, n. 6, p. 913–930, 2012. Available at: <https://doi.org/10.1016/j.regsciurbeco.2012.11.003>.

CORSEUIL, C. H.; REIS, M. C.; BRITO, A. S. Critérios de classificação para ocupação informal: consequências para a caracterização do setor informal e para a análise de bem-estar no Brasil. *Estudos Econômicos (São Paulo)*, SciELO Brasil, v. 45, n. 1, p. 5–31, 2015. Available at: <https://doi.org/10.1590/0101-4161201545151cma>.

CRUZ, B. d. O.; NATICCHIONI, P. Falling urban wage premium and inequality trends: evidence for Brazil. *Investigaciones regionales: Journal of Regional Research*, Asociación Española de Ciencia Regional, v. 126, n. 24, p. 91–114, 2012. Available at: <https://www.redalyc.org/articulo.oa?id=28924660005>.

DALBERTO, C. R.; CIRINO, J. F. Informalidade e segmentação no mercado de trabalho brasileiro: evidências quantílicas sob alocação endógena. *Nova Economia*, Universidade Federal de Minas Gerais, v. 28, n. 2, p. 417–460, 2018. Available at: <https://doi.org/10.1590/0103-6351/3191>.

D'COSTA, S.; OVERMAN, H. G. The urban wage growth premium: sorting or learning? *Regional Science and Urban Economics*, Elsevier, v. 48, p. 168–179, 2014. Available at: <https://doi.org/10.1016/j.regsciurbeco.2014.06.006>.

DURANTON, G. Agglomeration effects in Colombia. *Journal of Regional Science*, Wiley Online Library, v. 56, n. 2, p. 210–238, 2016. Available at: <https://doi.org/10.1111/jors.12239>.

DURANTON, G.; PUGA, D. Micro-foundations of urban agglomeration economies. In: *Handbook of Regional and Urban Economics*. Elsevier, 2004. v. 4, p. 2063–2117. Available at: <https://doi.org/10.1016/S1574-0080(04)80005-1>.

ELVERY, J. A. City size and skill intensity. *Regional Science and Urban Economics*, Elsevier, v. 40, n. 6, p. 367–379, 2010. Available at: <https://doi.org/10.1016/j.regsciurbeco.2010.05.006>.

FIELDS, G. S. Labour market modelling and the urban informal sector: Theory and evidence. In: *The informal sector revisited*. Organisation for Economic Co-operation and Development - OCDE, 1990. Available at: <https://digitalcommons.ilr.cornell.edu/articles/464/>.

GARCÍA, G. A. Agglomeration economies in the presence of an informal sector: the Colombian case. *Revue dEconomie Regionale Urbaine*, Armand Colin, n. 2, p. 355–388, 2019. Available at: <https://www.cairn-int.info/revue-d-economie-regionale-et-urbaine-2019-2-page-355.htm>.

GLAESER, E. L.; MARE, D. C. Cities and skills. *Journal of Labor Economics*, The University of Chicago Press, v. 19, n. 2, p. 316–342, 2001. Available at: <https://doi.org/10.1086/319563>.

GOULD, E. D. Cities, workers, and wages: a structural analysis of the urban wage premium. *The Review of Economic Studies*, Wiley-Blackwell, v. 74, n. 2, p. 477–506, 2007. Available at: <https://doi.org/10.1111/j.1467-937X.2007.00428.x>.

GUEVARA-ROSERO, G. C.; POZO, D. D. *et al.* Determination of the urban wage premium in ecuador. *Investigaciones Regionales*, Investigaciones Regionales, v. 47, n. 2, p. 57–77, 2020. Available at: <https://investigacionesregionales.org/numeros-y-articulos/consulta-dearticulos>.

GÜNTHER, I.; LAUNOV, A. Informal employment in developing countries: opportunity or last resort? *Journal of Development Economics*, Elsevier, v. 97, n. 1, p. 88–98, 2012. Available at: <https://doi.org/10.1016/j.jdeveco.2011.01.001>.

HAMANN, S.; NIEBUHR, A.; PETERS, J. C. Does the urban wage premium differ by pre-employment status? *Regional Studies*, Routledge, v. 53, n. 10, p. 1435–1446, 2019. Available at: <https://doi.org/10.1080/00343404.2019.1577553>.

HARRIS, J. The messy reality of agglomeration economies in urban informality: evidence from Nairobi's handicraft industry. *World Development*, Elsevier, v. 61, p. 102–113, 2014. Available at: <https://doi.org/10.1016/j.worlddev.2014.04.001>.

HECKMAN, J. J. Sample selection bias as a specification error. *Econometrica*, v. 47, n. 1, p. 153–161, 1979. ISSN 00129682, 14680262. Available at: <https://www.jstor.org/stable/1912352>.

HENLEY, A.; ARABSHEIBANI, G. R.; CARNEIRO, F. G. On defining and measuring the informal sector: Evidence from Brazil. *World Development*, Elsevier, v. 37, n. 5, p. 992–1003, 2009. Available at: <https://doi.org/10.1016/j.worlddev.2008.09.011>.

IBGE. *Nota Técnica. Deflacionamento dos rendimentos do trabalho dos trimestres móveis da PNAD-Contínua*. Instituto Brasileiro de Geografia e Estatística, 2018. Available at: <http://www.ibge.gov.br>.

IBGE. *Organização do território: Divisão Territorial Brasileira (DTB)*. IBGE Rio de Janeiro, 2018. Available at: <http://www.ibge.gov.br>.

IBGE. *Regiões de influência das cidades (REGIC)*. 2018. Available at: <https://www.ibge.gov.br/geociencias/cartas-e-mapas/redes-geograficas/15798-regioes-de-influencia-das-cidades.html?=&t=o-que-e>.

IBGE. *INPC. Séries históricas,*. [S.l.], 2019. Available at: <http://www.ibge.gov.br>.

IBGE. *Pesquisa Nacional por Amostra de Domicílios Contínua (PNADC) 2012-2019*. 2020. Available at: <http://www.ibge.gov.br>.

IBGE. *Pesquisa Nacional por Amostra de Domicílios (PNAD) 2001-2015*. 2020. Available at: <http://www.ibge.gov.br>.

KUNAL, S. *et al. Informal work in sub-Saharan Africa: Dead end or steppingstone?* [S.l.], 2019. Available at: <https://ideas.repec.org/p/unu/wpaper/wp-2019-107.html>.

MATANO, A.; OBACO, M.; ROYUELA, V. What drives the spatial wage premium in formal and informal labor markets? The case of Ecuador. *Journal of Regional Science*, Wiley Online Library, 2020. Available at: <https://doi.org/10.1111/jors.12486>.

MEEKES, J.; HASSINK, W. H. Endogenous local labour markets, regional aggregation and agglomeration economies. *USE Working Paper series*, USE Research Institute, v. 18, n. 03, 2018. Available at: <https://ssrn.com/abstract=3516371>.

MORETTI, E. Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data. *Journal of Econometrics*, Elsevier, v. 121, n. 1-2, p. 175–212, 2004. Available at: <https://doi.org/10.1016/j.jeconom.2003.10.015>.

MORETTI, E. Local labor markets. In: *Handbook of Labor Economics*. Elsevier, 2011. v. 4, p. 1237–1313. Available at: <http://www.nber.org/papers/w15947>.

MORETTI, E. Real wage inequality. *American Economic Journal: Applied Economics*, v. 5, n. 1, p. 65–103, 2013. Available at: <https://www.aeaweb.org/articles?id=10.1257/app.5.1.65>.

ROCA, J. D. L.; PUGA, D. Learning by working in big cities. *The Review of Economic Studies*, Oxford University Press, v. 84, n. 1, p. 106–142, 2017. Available at: <https://doi.org/10.1093/restud/rdw031>.

ROSENTHAL, S. S.; STRANGE, W. C. The attenuation of human capital spillovers. *Journal of Urban Economics*, Elsevier, v. 64, n. 2, p. 373–389, 2008. Available at: <https://doi.org/10.1016/j.jue.2008.02.006>.

SILVA, D. L. G. *Economias de aglomeração e heterogeneidade do trabalhador e firma na determinação de salários no Brasil*. Phd Thesis (PhD Thesis) — Universidade de São Paulo, 2017. Available at: <https://www.teses.usp.br/teses/disponiveis/12/12138/tde-13042017-112501/en.php>.

SILVA, D. L. G.; SANTOS, G. F.; FREGUGLIA, R. S. Distribuição espacial dos efeitos de aglomeração sobre os retornos à educação no Brasil entre 1995 e 2008. *Pesquisa e Planejamento Econômico*, v. 46, n. 2, 2016. Available at: <http://repositorio.ipea.gov.br/handle/11058/7295>.

SLONIMCZYK, F. Informal employment in emerging and transition economies. *IZA World of Labor*, n. 59, 2014. Available at: <https://wol.iza.org/articles/informal-employment-in-emerging-and-transition-economies/long>.

STORPER, M.; VENABLES, A. J. Buzz: face-to-face contact and the urban economy. *Journal of Economic Geography*, Oxford University Press, v. 4, n. 4, p. 351–370, 2004. Available at: <https://doi.org/10.1093/jnlecg/lbh027>.

TANSEL, A.; KAN, E. O. The formal/informal employment earnings gap: evidence from Turkey. *Available at SSRN 2049336*, 2012. Available at: <http://dx.doi.org/10.2139/ssrn.2049336>.

ULYSSEA, G. Informality: Causes and consequences for development. *Annual Review of Economics*, v. 12, n. 1, p. null, 2020. Available at: <https://doi.org/10.1146/annurev-economics-082119-121914>.

WOOLDRIDGE, J. M. *Econometric analysis of cross section and panel data*. [S.l.]: MIT press, 2010.

WORLDBANK. *World Bank Open Data.* 2019. Available at: <https://data.worldbank.org>.

YANKOW, J. J. Why do cities pay more? An empirical examination of some competing theories of the urban wage premium. *Journal of Urban Economics*, Elsevier, v. 60, n. 2, p. 139–161, 2006. Available at: <https://doi.org/10.1016/j.jue.2006.03.004>.