Informal jobs, sample selection and urban wage premium in Brazil: evidence from 2012 to 2018 *

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

Working paper - draft version
Submitted to 47th Encontro Nacional de Economia - São Paulo 2019

July 16, 2019

Abstract

This paper aims to empirically analyse the Urban Wage Premium (UWP) of Metropolitan Areas (MAs) in Brazil but it differs from previous studies by: (i) investigating the UWP for both sectors (formal and informal) by using a recent longitudinal database, the Continuous National Household Sample Survey (PNADC); (ii) the possibility of applying a sample selection correction, since it's possible to reconstruct the panel of individuals who become unemployed; (iii) a geographic approach that includes a very concentrated urban structure and; (iv) addressing the role of personal, firm and occupational characteristics in UWP composition at different agglomeration levels and evaluating the role of the sorting of individuals with the decomposition of wage differentials between the MAs. The results, supported by robustness checks, indicate that there is a UWP for workers in MAs, but it's decreasing as agglomeration becomes denser. The formal workers UWP is in line with previous studies while the informal workers UWP drives the overall results, showing a decreasing behaviour. The fact that there is a distinct UWP pattern between these groups, according to the agglomeration level, is new in UWP studies for Brazil since the majority of them are focused on formal workers. The analysis of the characteristics indicates that workers in high-skilled occupations or with a higher level of education are better paid when located in denser areas, regardless if in formal or informal jobs. However, informal workers in low-skilled occupations or with lower education level have a higher premium if located in Small MAs.
Keywords: urban wage premium, agglomeration scale, informality.
JEL classification: R23, J24, J31
Area: 13 - *Economia do trabalho*

Resumo

O objetivo deste artigo é analisar empiricamente o Prêmio Salarial Urbano (UWP) das Áreas Metropolitanas (MAs) no Brasil, mas difere de estudos anteriores por: (i) investigar o UWP para ambos os setores (formal e informal), que é possível ao utilizar uma recente base de dados longitudinal, a Pesquisa Nacional por Amostra de Domicílios Contínua (PNADC); (ii) possibilitar a aplicação da correção de Heckman para a seleção amostral, uma vez que é possível reconstruir o painel de indivíduos que ficam desempregados; (iii) uma abordagem geográfica que inclui uma estrutura urbana concentrada e; (iv) abordar o papel das características pessoais, da firma e da ocupação na composição da UWP para diferentes níveis de aglomeração e avaliar o papel do *sorting* dos indivíduos com a decomposição dos diferenciais salariais entre as MAs. Os resultados, apoiados por testes de robustez, indicam que há UWP para trabalhadores nas MAs, mas ele diminui à medida que a aglomeração torna-se mais densa. O UWP dos trabalhadores formais se mostra em linha com estudos anteriores, enquanto o UWP dos trabalhadores informais UWP impulsiona os resultados globais, mostrando um comportamento decrescente conforme a MA se torna mais densa. O fato de existir um padrão distinto para o UWP entre esses grupos, de acordo com o nível de aglomeração, é novo nos estudos sobre UWP para o Brasil, uma vez que a maioria deles compreende apenas trabalhadores formais. A análise das características indica que trabalhadores em ocupações altamente qualificadas ou com maior nível de escolaridade são obtêm prêmios salariais maiores quando localizados em áreas mais densas, independentemente de estarem em empregos formais ou informais. No entanto, os trabalhadores informais em

*We gratefully acknowledge the financial support for this research by the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)*. We thank for helpful comments and suggestions from Ana Maria Bonomi Barufi, Diana Lúcia Gonzaga da Silva, Diogo de Prince Mendonça, Luciana Rosa de Souza and the participants of *Núcleo de Economia Regional e Urbana - NEREUS* seminars at the University of São Paulo.
†MSc Candidate in Economics and Development at Universidade Federal de São Paulo. eloiza.regina@unifesp.br
‡Professor at Federal University of São Paulo - Department of Economics. veneziano.araujo@unifesp.br
§Professor at Federal University of São Paulo - Multidisciplinary Department. slgoncalves@unifesp.br

ocupações de baixa qualificação ou com menor nível de escolaridade têm um prêmio maior se estiverem localizados em MAs Pequenas.
Palavras-chave: prêmio salarial urbano, escala de aglomeração, informalidade.

1 Introduction

A variety of studies on local labour markets analyse the existence of a positive wage differential in urban areas, known as Urban Wage Premium (UWP). There is extensive empirical evidence of this premium, even after control for observable and unobservable workers' and firms' characteristics and in the recent period, studies have shed light on how the UWP changes according to the agglomeration scale, instead of considering the UWP as fixed and homogeneous. As UWP empirical studies rely on individual micro-data, they differ in scope, sample choice, and data availability, pointing results in different directions demanding further research. Even with the recent advances in the UWP analysis, the vast majority of UWP studies for Brazil do not comprises informal workers due to the use of administrative records. This fact is an essential gap, since this group have non-despicable participation on the workforce. Another limitation is although recognizing that the movements between employment and unemployment are distinct not only due to different workers' profiles but also due to the dynamics of the local labour markets in large agglomerations, data for the unemployed is not always available, which is why the authors, with few exceptions, do not work out the potential sample selection bias.

In this context, the main aim of this paper is to analyse the UWP according the agglomeration scale in Brazil. Our contribution relies on the investigation of the UWP for both sectors (formal and informal) and the application of Heckman's correction. With this focus, we evidence how the different workers' characteristics are related to agglomeration levels for UWP explanation and the possible sorting effect, decomposing the wage differentials by Oaxaca-Blinder method. This was only possible due to the use of a recent longitudinal database on the Brazilian labour market, the Continuous National Household Sample Survey (PNADC) from Institute of Geography and Statistics (IBGE), with more than 577,000 workers from 2012 to 2018, a broad geographic coverage, considering workers of all sectors and formality status.

After correcting the sample selection bias and controlling for workers' observable characteristics, we estimated the Metropolitan Area (MA) UWP, grouping the 27 Brazilian MAs based on their population levels. The results pointed to a non-homogeneous and decreasing UWP as agglomeration becomes larger: 8.15% for Extra-Large MAs and 13.8% for Small MAs. By exploring the workers' formality status, it was possible to verify that formal workers UWP is in line with previous studies, while informal workers UWP drives the overall results, showing a decreasing behaviour. The fact that there is a distinct UWP pattern between these groups, according to the agglomeration level, was never elucidated in UWP studies for Brazil.

Through additional regressions, including variables interaction, it was possible to understand how workers' and firms' characteristics explain the UWP according to agglomeration scale. Robustness checks confirm the UWP results even when including different samples and periods, alternative agglomeration scale definitions or additional variables. Even with data limitations, a non-negligible part of individuals' fixed effects can be explained by being located in MAs and investigating the components of wage differentials between MAs and Non-MAs or between Formal and Informal workers, it was possible to identify that the level of individuals' attributes can explain the highest part.

This paper is composed of 4 sections, besides this introduction: section 2 presents the background and motivation, and reviews the UWP literature; section 3 presents data, descriptive analysis and empirical strategy; section 4 shows the estimation results based on different econometric models and methods, as well as robustness tests; and finally, section 5 presents the final remarks.

2 Background

The literature presents extensive evidence of a wage differential favourable to workers in Metropolitan Areas (MAs), even after controlling for personal, occupational and firm characteristics, known as the urban wage premium (UWP). Researches Duranton and Puga (2004) consolidate the theoretical micro-foundations of agglomerations economies - a milestone in this research. They identify how agglomerations can foster wage gains as they provide: i) a reduction of costs associated with the search for new jobs by individuals and the search for candidates by firms (sharing); ii) greater possibilities to find employment and promote better matches between firms and workers (matching); and iii), spillover of ideas and new skills development (learning). Therefore, human capital (learning) can be a channel to analyse UWP, in which populated areas have a higher concentration of more educated people (sorting), and foster a more extensive and faster accumulation of human capital, with knowledge spillovers and gains from experience (ANDERSSON; THULIN, 2013; MORETTI, 2011; MORETTI, 2013; BEHRENS; DURANTON; ROBERT-NICOUD, 2014; ROCA; PUGA, 2017).

Studies report that wage differentials are not homogeneous among individuals but may vary according to the composition of workers' abilities and gains from experience, as well as the combined effect of them (as in Rosenthal and Strange (2008)).

Similarly, the UWP can also be influenced by the specific features of the labour market in denser areas due to the matching process between individuals and firms and through the transitions between jobs and occupations. In denser labour markets, firms gain access to a large pool of skilled workers, who in turn gain access to a more significant number of potential employers and suitable job positions, reducing the period of unemployment and risk of being unemployed (BAUM-SNOW; PAVAN, 2011). However, job transitions, which occur in the higher degree in populated areas, are also a way to obtain wage gains and to promote career advancements (NEVES-JR; AZZONI; CHAGAS, 2018; MATANO; NATICCHIONI, 2016).

In this context, according to Heuermann, Halfdanarson and Suedekum (2010), UWP can be empirically analysed by highlighting when the worker benefits from it. Concerning the UWP size at the wage level, one of the most important empirical studies is Glaeser and Mare (2001), which found evidence of a UWP in the United States between 1970 and 1990 of approximately 25% for large agglomerations, which, when controlling for individual observable and unobservable characteristics reduces to 4.5% remaining significant. Other studies also reveal the existence of a UWP different countries, and the results vary from 2.6% to 14.0% in benchmark estimations, reducing to 0.3% to 5,1% after controlling for workers' and firms' characteristics[1]. For Brazilian labour market, a UWP of 5.2% was identified by Chauvin et al. (2017) in benchmark estimations for Micro-regions, and of 3.4% by Silva, Santos and Freguglia (2016) in MAs between 1995 and 2008, after controlling for workers' characteristics[2].

Recent studies tend to identify how specific workers' and firms' characteristics can explain the UWP. Most of these studies are focused on analysing the relationship between the agglomeration level and the workers' schooling level (MORETTI, 2004; SILVA; SANTOS; FREGUGLIA, 2016) and job tenure (GLAESER; MARE, 2001; ROCHA; NETO; GOMES, 2011; CARLSEN; RATTSØ; STOKKE, 2016). Another group of studies evaluates the relevance of the workers' ability to determine UWP, with different concepts for skill level (GOULD, 2007; BACOLOD; BLUM; STRANGE, 2009; ANDERSSON; THULIN, 2013; NEVES-JR; AZZONI; CHAGAS, 2017).

From the methodological point of view, it is important to evaluate the existence of sample selection bias, since wages are only observe for employed individuals. However, the data for unemployed are not always available, so the vast majority of UWP studies disregard the propensity to be employed. Exceptions are the studies of Baum-Snow and Pavan (2011) and Gould (2007) that take into account the fact of being unemployed as a counter-factual analysis.Still concerning the applied methods, although much used in the investigation of the wage differential among groups of workers, few studies use the decomposition as an alternative to investigate wage differentials between regions[3]. The use of this method is appropriate because it makes possible to separate the wages differentials between the part corresponding to the workers' characteristics (which may indicate the presence and the magnitude of sorting in denser areas) and the part related to the local labour market dynamics.

Another limitation, quite frequent in the Brazilian literature, is to identify the UWP using administrative registers filled by firms, then restricting the analysis to formal workers[4]. In Brazil, this is relevant because informal workers have non-despicable participation in the workforce, from 2012 to 2018 approximately 25.4% of the working-age population according to IBGE (2018a).

3 Data and empirical strategy

3.1 Data and sample

The chosen database is the Continuous National Household Sample Survey (PNADC) from Brazilian Institute of Geography and Statistics (IBGE). It's a household survey with national coverage and a quarterly frequency for approximately 211,000 interviews in a rotating scheme, five consecutive quarters per household (IBGE, 2018b). PNADC allows workers analysis according to their location, type of contract, individuals', occupations' and firms' characteristics. For the period from 2012 to 2018 PNADC presents more than 15.8 million observations. After inconsistencies exclusions, we withdrew public sector workers and those from the Army, due to their specific labour laws. The sample consists of workers with a single job and with at least 20 hours a week and contemplates only employed men aged 18-65[5] totalising almost 557,000 workers and 2.8 million observations.

Workers that move during the survey are missing after the change, which implies that it does not cover migration between

[1] Such studies report a UWP in order of 5.1% in Norway (CARLSEN; RATTSØ; STOKKE, 2016), 3.0% in France (COMBES; DURANTON; GOBILLON, 2008), between 1.4% and 7.1% in England according to the MA size (D'COSTA; OVERMAN, 2014), 0.2% in Italy (BELLOC; NATICCHIONI; VITTORI, 2018), between 2.0% and 4.9% in Germany (BERLINGIERI, 2017; DAUTH et al., 2016), between 0.3% e 1.4% in Netherlands depending of the local employment density (MEEKES; HASSINK, 2018).

[2] Other studies also evidence UWP in Brazil using different specifications as Rocha, Neto and Gomes (2011) who find an additional of 0.45 minimum wage in benchmark estimations and 0.12 minimum wage after controlling for the individuals' characteristics.

[3] Cirino and Lima (2012) applied the Oaxaca-Blinder decomposition to understand the wage differentials, between two specific MAs in Brazil for 2006

[4] We also observe this fact in studies across the globe, mainly for the group of self-employed. In Brazil, the informal workers are often not included, as in the studies of Silva, Santos and Freguglia (2016), Rocha, Neto and Gomes (2011) and Neves-Jr, Azzoni and Chagas (2018), due to the use of RAIS-MIGRA, the Annual Social Information Relation database, an administrative record of firms and workers collected by the Ministry of Economy. The only study that included informal workers in UWP analysis for Brazil is Cruz and Naticchioni (2012), which found evidence of it for MAs (17%) and formal workers (23%), but without sample selection correction and not controlling for individuals' unobservable characteristics, given database limitations.

[5] The choice of men eliminates possible variations susceptible to discrimination, and the male workforce has a more stable behaviour despite adverse conditions such as low wages, high unemployment and poor working conditions, according to Menezes-Filho, Mendes and Almeida (2004).

regions; all the workers are in the same place throughout the research. The second case of attrition in this database emerges from workers that didn't answer the survey at any moment in the five quarters. Correction for attrition was not applied in this paper, but we try to deal with this possible bias looking at different sample cohorts as robustness check (see section 4.3). Besides, the values reported as wages are self-declared, which may be biased for declaration errors[6]. The PNADC shows a household identifier[7], but does not present a longitudinal identifier for individuals at the household. To be able to evaluate individuals over time, we construct an identifier based on the date of the birth of each member at the household. Besides that, all descriptive analysis and estimations make use of individual sample weight available on PNADC (variable V1028).

In order to analyse the agglomerations, we made a distinction between MAs (that correspond to State Capital and respective MAs or only State Capital totalling 27 MAs) and Non-MA. To evaluate the agglomeration level we subdivided the MAs according to their population into four groups. Table 1 shows the composition of each group. These approach condenses two applications found in the literature (by MA and by the number of inhabitants) and is the most suitable approach possible with this database. [Table 1 about here.]

3.2 Descriptive analysis

A non-negligible part of the working-age individuals are informal in Brazil. As can be seen in Table 2, although formal workers still represent a greater part of the working-age population, informal workers and non-employed represents 58.1% of them. The definition for informal workers comprises the cases where employed individuals or self-employed reported not have a formal contract (those in compliance with labour regulations) or do not contribute to the Social Security Institute. Being an informal worker means they do not have access to several benefits provided, by labour laws, such as unemployment benefits, job severance payments (for maternity leaves, death, illnesses, occupational accidents, among others) and public retirement. According to Yahmed (2018), workers with identical productivity, in a frictionless market, should earn higher wages to compensate for the absence of those benefits. However, by the dualistic view of the labour markets, formal jobs offers, on average, higher wages than informal jobs. The heterogeneity on workers' observable and unobservable attributes could explains this situation.

Descriptive analysis show differences between the individuals' characteristics for the working age population in Brazil (Table 2). Formal workers have, on average, almost three schooling years above informal workers, although they are slightly younger. They also have a lower household structure with fewer children and members, and are mostly composed of workers in high or medium-skilled occupations and with at least High School. While informal workers, on average, work fewer hours per week and have a lower job tenure. More in-depth analyses shown that, in general, MAs have a bigger share of workers in high-skilled occupations and with high-schooling levels compared to Non-MAs and in both sectors (formal/informal). Concerning the formal workers, those college-educated represent between 27.6% and 30.7% of total employees in MAs and only 20% in Non-MAs. Workers in high-skilled occupations reach 27.3% in Extra-large MAs but are only 18.5% in Non-MAs. Informal workers are more concentrated in Non-MAs (42.9% vs 31.8%) and have smaller participation in the total number of workers as the agglomeration increases, reaching 28.4% in Extra-Large MAs. College-educated informal workers only reach a maximum of 14.8%, being, on average, only 8.5% of total employees. They are also mostly in low-skilled occupations (easily surpassing 46% of total informal workers on each agglomeration level).

The spatial wage differentials in Brazil is easily observed in a wages comparison between locations. Metropolitan Areas (MAs) show higher mean wages than Non-MAs with a non-homogeneous trend growth according to the size of the MAs. The descriptive analysis for the sample shows that the wages difference between MAs and Non-MAs is relevant (about 30% in favour of MA, 15.9 vs 11.1) and within the MAs group, the difference is also meaningful (about 36% in favour of Extra-large MAs compared to Small MAs, 18.8 vs 12.0). Table 4 shows the mean hourly wage in different locations for workers in different categories. Following the literature, wages are higher among more educated workers and with higher job tenure. The different skill levels reveal relative wage differentials in favour of workers in high-skilled occupations in Extra-large MAs compared to Small MAs (47.2 vs 27.8). Among workers in occupations with low skill level, the difference is only 12% in favour of Extra-large MAs (10.6 vs 9.3). Wages are also higher in Extra-Large MAs for both formality status (formal/informal). [Table 2, 3 and 4 about here.]

3.3 Empirical strategy

The methodology traditionally used in UWP studies, and adopted here for different agglomeration sizes, is a panel data approach with Ordinary Least Squares Pooled (Pooled OLS) and Fixed-Effects estimates, controlling for individuals', occupations', and firms' characteristics. We first investigate the existence of sample selection bias adopting the correction procedure in two steps following Heckman (1979). The first one is an estimation for the probability of being employed where we consider a data set with

[6]Although the survey question specifies gross values is expected that the interviewees could interpret it as net taxes amounts. In this sense, Corseuil and Santos (2002) estimate that PNADC wages are 10% lower than the existing administrative records, comparing only formal workers.

[7]The household identifier comprises the variables: Primary Sampling Unit (UPA), household number (V1008) and panel identification(V1014)

all men aged 18-65 employed, non-employed or inactive, in the period of analysis, with the following specification:

$$L_i = \alpha_1 + \alpha_2 MA_i + \alpha_3 Age_i + \alpha_4 Age_i^2 + \alpha_5 Race_i + \alpha_6 Maritalstatus_i + \alpha_7 SchLevel_i$$
$$+\alpha_8 Unemployment + \alpha_9 HHhead_i + \alpha_{10} NoChild_i + \alpha_{11} Child6_i + \alpha_{12} Child14_i$$
$$+\alpha_{13} HHwage_i + \alpha_{14} HHpeople_i + \alpha_{15} PosHS_i + MacroRegion_i + T_i + \nu_i \quad (1)$$

In equation (1) L_i represents the probability of being employed. Besides the variables that represent the individual characteristics (see Appendix A for the variables definition) six household variables denote explicitly the probability of being employed: $NoChild$, $Child6$, $Child14$, $HHwage$, $HHpeople$, and $PosHS$. The T_i variable represents the set of dummies for year and quarters, while ν_{it} is the error term. In this first step, the Inverse Mills' Ratio (IMR) will indicate the existence of sample selection bias and then the second step consists in estimate the equation for wages considering the IMR as a control variable, obtaining consistent parameters. This estimation presents the following mincerian form, applied to Pooled OLS:

$$lnhwage_{it} = \beta_1 + \beta_2 MA_{it} + \beta_3 Age_{it} + \beta_4 Age_{it}^2 + \beta_5 Race_{it} + \beta_6 Maritalstatus_{it}$$
$$+\beta_7 HHhead_{it} + \beta_8 OccSkill_{it} + \beta_9 Formality_{it} + \beta_{10} Tenure_{it} + \beta_{11} SchLevel_{it} +$$
$$\beta_{12} Unemployment + \beta_{13}\lambda + Industry_{it} + MacroRegion_{it} + T_t + \epsilon_{it} \quad (2)$$

where the dependent variable $lnhwage$ is the logarithm form of the hourly wage of worker i at time t temporally deflated using the INPC deflator, λ is the Inverse Mills' Ratio and ϵ_{it} is the error term. The analysis will focus on the coefficient β_2 that will show how much the agglomeration is determinant for the UWP. Appendix B shows the estimation details of these two steps. Following the literature, to capture the demand side effects considering the state of the local labour markets we include the variable $Unemployment$ that measure the level of unemployed according to the year, quarter, Macro-Region, MA (or Non-MA) and Schooling level.

The results were obtained for all MAs together and with a separation by agglomeration scale. Considering the differences between formal and informal workers, we perform separately Pooled OLS estimations. Besides that, to deepen the analysis about the role of individuals' characteristics, interactions were made between some variables to identify their relation to the UWP of each agglomeration level. The econometric specification of the equation (2) then includes the interaction of the variable MA_{it} with Occupational skill ($Occskill_{it}$), Schooling level ($SchLevel_{it}$), and Tenure ($Tenure_{it}$). As robustness checks, several exercises were applied, as described in section 4.3.

For a precise UWP identification it's necessary to control the individuals' unobservable characteristics using Fixed-Effects method. However, the available data do not contemplates the workers that migrate across regions, making it impossible to estimate the coefficients for the MAs dummies. So, we chosen to perform a procedure inspired by Combes, Duranton and Gobillon (2008) and Meekes and Hassink (2018) that consists of an estimation of fixed effects and its' regression by the time fixed characteristics, obtaining an indication of how much of the fixed-effects is associated with characteristics of the individuals' location. The intention of doing this is to isolate the individuals' effects and evaluate the role of the agglomerations.

Finally, we applied the Oaxaca-Blinder method[8] to understand how the wages differentials between MAs can be explained by the individuals' attributes or by the remuneration dynamics in the local labour markets. This method offered additional evidence for the role of sorting and was performed comparing MAs versus Non-MAs and formal versus informal workers. All estimations make use of individual sample weight provided by IBGE (2018a) and clustered standard errors by individuals.

4 Results

4.1 Observing the UWP in agglomerations

Since, the existence of sample selection bias was verified (Appendix B report the results for each stage of the process), the results reported below always refer to estimates corrected by the Heckman's procedure. Table 5 presents estimates of different specifications for the equation (2). The first two regressions, only with the MA dummy, show the raw effect of sample selection correction, then controls are added gradually: dummies for time, industries, formality, and Macro-Region (3), workers' observable characteristics and unemployment rates (4) and, finally, occupations' characteristics (tenure and skill levels)(5). However, as the focus of this paper is to analyse the effect of the agglomeration scale on UWP, the most interesting specification is regression 10 with dummies for four MA agglomeration levels, being this model our reference. Estimates 6 to 9 are reduced forms following the addition of the same controls described earlier. The regressions for the agglomeration levels show relevant changes with sample selection correction (models 6 and 7) especially for Small and Medium MAs, increasing their coefficient in more than three percentage points. In general, other variables coefficients are in line with the literature. All models confirm the existence of a UWP in Brazil. The effect of MAs on wages is initially 26.9% in the most basic model (1) and reduces to 9.4% with all controls (5). The coefficients in the model 10 indicate a non-homogeneous UWP according to agglomeration scale. Although always positive, while

[8]For more details of the method see Oaxaca (1973), Blinder (1973) and Fortin, Lemieux and Firpo (2011).

Extra-large MAs have a differential of 8.1%, the other MAs showed a higher level of 13.8% and 10.0% respectively for Large, Medium and Small MAs.

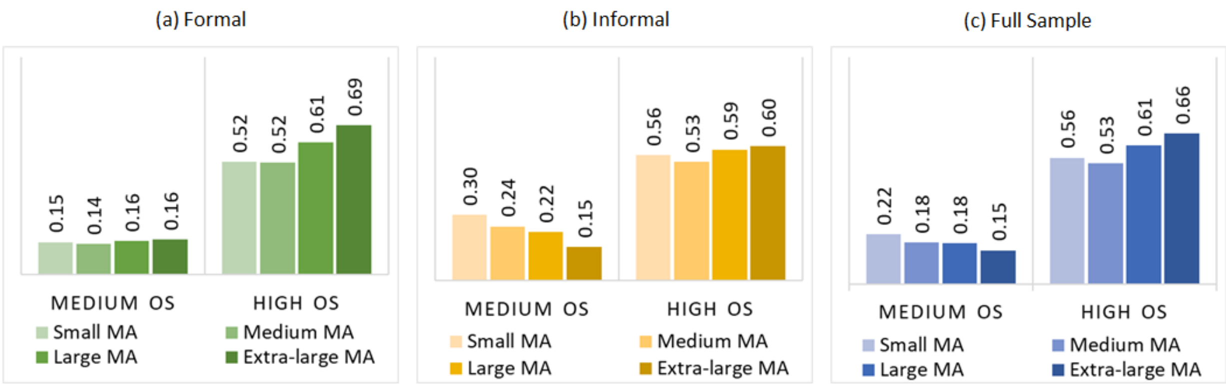
This decreasing behaviour is not in line with the majority of studies in the literature, that point an increasing UWP with agglomeration scale. One possible approach is to investigate, considering the differentials between formal and informal workers, whether the UWP is different. Table 6 shows the Pooled OLS estimations separately for the two groups and the reference model. Comparing against the Non-MA, MAs' formal workers have an 8.2% UWP, a lower result than informal workers, with 11.9%. The UWP results are very different analysing by Agglomeration level: formal workers have a higher UWP in Extra-large MAs (9.5%) while informal workers in Small MAs (20.3%). Both groups show a non-homogeneous UWP according to agglomeration scale, but for formal workers, they vary between 5.6% and 9.5%, while informal workers have more elasticity, between 6.8% and 20.3%. Informal workers UWP drives the overall results, that replicates the decreasing behaviour as the agglomeration level rise. It was never elucidated in UWP studies the fact that there is a distinct pattern between these groups according to the agglomeration level. Workers' characteristics of each group could motivate these results, the reason why they will be tested in the next section, where the UWP analysis is deepened using variables interactions. [Table 5 and 6 about here.]

4.2 UWP deep analysis

Our first results point to a non-homogeneous UWP in agglomeration scale and distinct behaviour between formal and informal workers. In this context, we analyse in more detail how the UWP can be explained by different characteristics of local labour market composition. Results in Table 7 include three interaction terms with agglomeration scale dummies: occupational skills, schooling levels and tenure. Replicated estimations were made for formal and informal workers separately (see the Appendix C), but besides the coefficients for the interaction terms, we summarised the aggregate effects of each interaction on each group, to ease results interpretation, and reported below the most important findings. [Table 7 about here.]

Occupational Skill. Workers in high and middle-skilled occupations in all MA levels present a wage premium compared to low-skilled ones. The net effect range from 15% to 17% for medium-skilled and from 56% and 66% to high-skilled occupations (Figure 1c). However, the decreasing premium according to agglomeration scale for Medium-OS workers turns into an increasing pattern for those with High-OS. So, workers in high-skilled occupations are the most benefited in larger MAs, fact observed in others studies even with different skills specification, like Gould (2007) and Bacolod, Blum and Strange (2009). Besides that, as can be seen in Figure 1(a-b), informal workers in medium-skilled occupations present a higher UWP in all MAs, comparing with formal workers, and it is double in Small MAs compared to Extra-large MAs. These results indicate that workers in medium-skilled occupations benefits of a greater UWP when allocated at Small MAs and in informal jobs.

Figure 1: Interactions with Agglomeration Scale: **Aggregated effects for Occupational Skill**

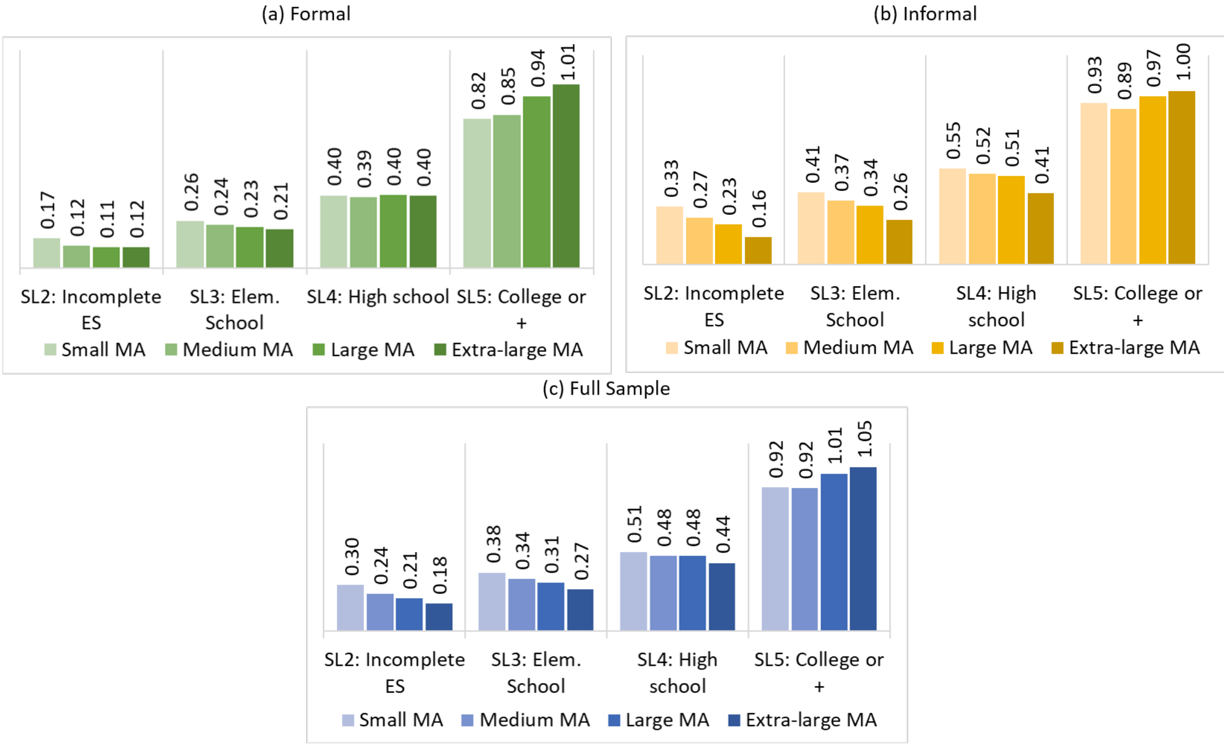


Source: Organized by the authors based on Table 7 and Appendix C results.

Schooling levels. Educated workers have higher premiums than similar ones in Non-MAs locations. The aggregate effect shows that workers with lower schooling levels (until high school) present a wage premium in MAs, but it is lower in denser areas (Figure 2c). For college-educated workers, that is a reversed pattern as the higher the MA, the greater the UWP. Both groups, formal and informal workers, present similar results, although with distinct magnitudes. Again, the UWP for informal worker surpasses the formal workers at any agglomeration and schooling level (Figure 2a-b), and very educated workers earn more in very dense regions on both groups.

Tenure. There is a wage penalty that reaches up to 10.0% for new employees (up to one year) in Extra-Large cities, but it reduces as experience consolidates (Figure 3c). After 4 years in the same job, the penalty dissipates, and a positive UWP are found, being higher in less dense areas. The aggregate effect report different results for formal and informal workers (Figure 3a-b). While

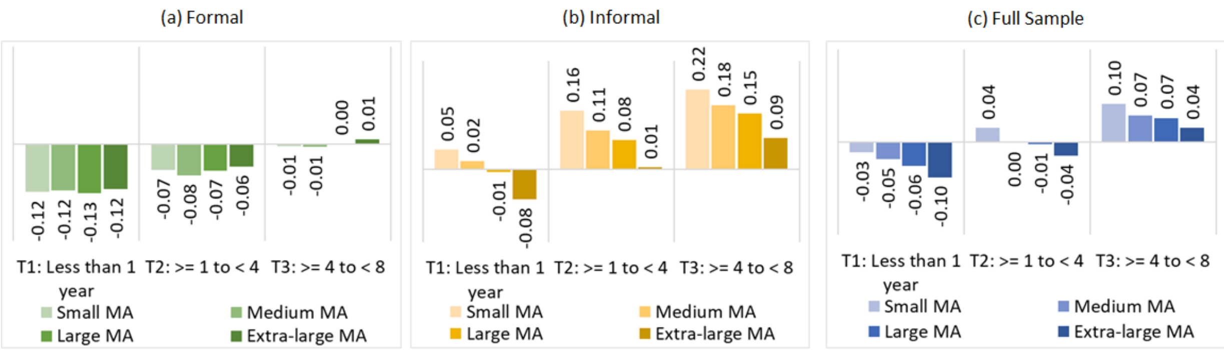
Figure 2: Interactions with Agglomeration Scale: **Aggregated effects for Schooling levels**



Source: Organized by the authors based on Table 7 and Appendix C results.

formal workers with a shorter job tenure have a wage penalty (T1 and T2) and slightly close effects for workers with 4 to 8 years tenure, informal workers have a positive UWP from 1 year regardless the agglomeration level (or before that if in a Small or Medium MA). Even though our analyses do not evaluate time trends, this evidence suggests an acceptance of wage penalties on formal jobs, at the beginning of employment, maybe because workers foresee more significant gains over time. Same prospects are not seeing in informal jobs.

Figure 3: Interactions with Agglomeration Scale: **Aggregated effects for Tenure**



Source: Organized by the authors based on Table 7 and Appendix C results.

4.3 Robustness checks

This section is dedicated to robustness exercises that include different samples, periods, alternative agglomeration scale definitions, and additional variables. In general, the UWP primarily results for agglomerations levels hold on each exercise, and they are explained below and reported in Table 8 and 9.

Attrition and panel tests. The two first tests are related to the fact that the database consists of an unbalanced panel. The first test (A) comprises only the first observation (first interview answer of each worker) followed by test (B) that comprises only individuals with at least two answers (the first to interviews or any combination between the 5 quarters). Both tests show coefficient for MAs similar to the benchmark model.

Firm size. Following the literature, we evidence the role of firm size in the UWP performing an exercise with the same specification for equation 2 including a new variable, $Firmsize$. Exercise C reports, as expected, that firm size has a positive and increasing effect on wages, and MAs UWP levels are quite similar to benchmark model results. Since new results indicate, an even lower wage gap Extra-large MAs, as pointed in the literature, not control for the presence of large firms in denser areas could overestimate UWP in these locations. Exercises D and E deepen the analysis of firm size, with separated regression according to the periods when this variable is available. On both, firm size remains significant.

Weekly worked hours. The primary sample comprises only workers with a minimum of 20 weekly worked hours than Exercise F took into account all workers without worked hours restriction. The UWP in this exercise showed similar coefficients for Agglomeration levels compared to the benchmark model.

Alternative Agglomeration Scale definitions. Concerning the agglomeration specification and the possibilities of our database, we performed three final robustness checks. The former (exercise G) reclassifies the MAs according to their demographic density (inhabitants per km^2) in four new groups being Small MAs the ones' with less than 100, Medium MAs between 100 and 417 (median level), Large MAs between 417 and 1.000 and Extra-large MAs above 1.000. The following specification (exercise H) takes the Cities' Influence Regions (*Região de Influência das Cidades*) from IBGE (2008), that classifies the urban centres according to the intensity of the links between them. Both of these exercises showed similar coefficients for Agglomeration levels compared to the base model. Finally, the last exercise takes the MA population logarithm as the area identifier. The effect of population logarithm on hourly wages is small (1.1%) but positive and statistically significant. Between the two groups analysed (formal/informal), different behaviour was found, although both significant the population logarithm affect hourly wages positively for formal workers (+2.2%) and negatively for informal workers (-1.3%), consistent with the previous results on Pooled OLS estimations. [Table 8 and 9 about here.]

4.4 Controlling the unobservable characteristics

In an attempt to control the unobservable characteristics of individuals, even with the limitations of the database (described in section 3.3), we made two estimations[9]. The first one is the estimation by Fixed Effects, shown in Table 10(a) and comparing the coefficients with Pooled OLS. As expected, the coefficients are smaller but continue to demonstrate wage differentials according to specific characteristics. The second estimation corresponds to regress the individuals' fixed effects, previously estimated, as the dependent variable. In this case, the control variables are only those that do not vary on time: individuals' location (agglomeration Scale and Macro-Region), age and race. The results (Table 10(b)) show that being in an MA explains between 20.6% and 28.5% of the individuals' fixed effects. [Table 10 about here.]

4.5 Oaxaca-Blinder Decomposition

The last set of estimates comprises the application of the decomposition method. As shown in Table 11, the higher wage levels in favour of MAs versus Non-MAs are mainly due to the individuals' attributes (65.4%). The lower part is due to the difference in coefficients, 32.5%, that is, explained by the difference in the remuneration dynamics of the local labour market. To deepen the analysis, we applied the decomposition method for formal workers against informal workers and identified that the wage differentials are also explained mostly by the individuals' attributes (50.1%) than by the local labour market dynamics (34.8%). [Tables 11 about here.]

5 Final remarks

In this paper, we analyse empirically the relationship between UWP and agglomeration levels deepening the analysis of workers' characteristics, formality status and understanding the wage differentials explanation (by the individual's attributes or the local labour markets dynamics). We use a large and recent longitudinal panel with information for Brazilian workers, based on a household survey (PNADC), which also allows us to investigate whether there is a different evidences of UWP between formal and informal workers, and correct the sample selection bias with Heckman's procedure. The results add new insights for UWP explanation in Brazil, especially related to informal workers.

The UWP was estimated for MAs (versus Non-MAs), and between MAs subdivided according to agglomeration size. The results pointed to a non-homogeneous and decreasing UWP as agglomeration becomes larger: 8.15% for Extra-Large MAs and 13.8% for Small MAs. By exploring the workers' formality status, it was possible to verify that for formal workers UWP is in

[9]In this section, we do not perform the Heckman's correction.

line with previous studies while informal workers UWP drives the overall results, repeating the decreasing behaviour. It was never elucidated in UWP studies the fact that there is a distinct pattern between these groups according to the agglomeration level.

Through additional regressions, including variables interaction, it was possible to understand how they explain the UWP according to agglomeration scale. Important differences in these interactions show up in the separated analyses according to the formality status. Workers in high-skilled occupations are the most benefited in larger MAs while workers in medium-skilled occupations benefits of a greater UWP when allocated at Small MAs and in informal jobs. Workers with low-educated levels (up to High School) earn more in MAs, but this premium decreases as the agglomeration size increase. Reversed results were found for college-educated workers, the higher the MA, the higher the UWP. Whereas for formal workers with a shorter job tenure a wage penalty shows up (T1 and T2) and slightly close premium for workers with 4 to 8 years tenure versus older tenure workers. Informal workers have a positive UWP from 1 year regardless the agglomeration level (or before that if allocated in a Small or Medium MA). Robustness checks confirm the UWP results even when including different samples, periods, alternative agglomeration scale definitions, additional variables, or using the population logarithm as the area specification. Even with data limitations, a non-negligible part of individuals' fixed effects can be explained by being located in agglomerated areas. Finally, investigating the components of wage differentials it was possible to identify that the level of individuals' attributes can explain the greatest part.

References

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.

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.

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.

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.

BELLOC, M.; NATICCHIONI, P.; VITTORI, C. *Urban Wage Premia, Cost of Living, and Collective Bargaining*. [S.l.], 2018.

BERLINGIERI, F. Local labor market size and qualification mismatch. *ZEW-Centre for European Economic Research Discussion Paper*, n. 17-055, 2017.

BLINDER, A. S. Wage discrimination: reduced form and structural estimates. *Journal of Human resources*, JSTOR, p. 436–455, 1973.

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.

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.

CIRINO, J. F.; LIMA, J. E. de. Diferenças de rendimento entre as Regiões metropolitanas de Belo Horizonte e Salvador: uma discussão a partir da decomposição de Oaxaca-Blinder. *Revista Econômica do Nordeste*, v. 43, n. 2, p. 371–390, 2012.

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.

CORSEUIL, C. H.; SANTOS, D. D. Determinantes da renda do trabalho no setor formal da economia brasileira. Instituto de Pesquisa Econômica Aplicada - (IPEA), n. 885, 2002.

CRUZ, B. de O.; NATICCHIONI, P. Falling urban wage premium and inequality trends: evidence for Brazil. *Investigaciones regionales: Journal of Regional Research*, Asociación Española de Ciencia Regional, n. 24, p. 91–114, 2012.

DAUTH, W. et al. Spatial wage disparities: Workers, firms, and assortative matching. Mimeo, Heinrich Heine University Düsseldorf, 2016.

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.

DURANTON, G.; PUGA, D. Micro-foundations of urban agglomeration economies. In: *Handbook of regional and urban economics*. [S.l.]: Elsevier, 2004. v. 4, p. 2063–2117.

FORTIN, N.; LEMIEUX, T.; FIRPO, S. Decomposition methods in economics. In: *Handbook of labor economics*. [S.l.]: Elsevier, 2011. v. 4, p. 1–102.

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.

GONÇALVES, S. L.; MENEZES-FILHO, N. A. Minimum wage and the labor supply of poor families: a collective approach using data from the continuous PNAD (2012-2015). *Working paper series - Department of Economics*, n. 40, 2015.

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.

HECKMAN, J. J. Sample selection bias as a specification error. *Econometrica*, v. 47, n. 1, p. 153–161, 1979.

HEUERMANN, D.; HALFDANARSON, B.; SUEDEKUM, J. Human capital externalities and the urban wage premium: two literatures and their interrelations. *Urban Studies*, SAGE Publications Sage UK: London, England, v. 47, n. 4, p. 749–767, 2010.

IBGE. *Regiões de influência das cidades 2007-REGIC 2007*. [S.l.]: IBGE Rio de Janeiro, 2008.

IBGE. Pesquisa Nacional por Amostra de Domicilios Contínua (PNADC) 2012-2018. Instituto Brasileiro de Geografia e Estatítica, 2018. Available at: <http://www.ibge.gov.br>.

IBGE. PNADC. Notas Técnicas. Instituto Brasileiro de Geografia e Estatítica, 2018. Available at: <http://www.ibge.gov.br>.

MATANO, A.; NATICCHIONI, P. What drives the urban wage premium? Evidence along the wage distribution. *Journal of Regional Science*, Wiley Online Library, v. 56, n. 2, p. 191–209, 2016.

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.

MENEZES-FILHO, N. A.; MENDES, M.; ALMEIDA, E. S. O diferencial de salários formal-informal no Brasil: segmentação ou viés de seleção? *Revista Brasileira de Economia*, SciELO Brasil, v. 58, n. 2, p. 235–248, 2004.

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.

MORETTI, E. Local labor markets. In: *Handbook of labor economics*. [S.l.]: Elsevier, 2011. v. 4, p. 1237–1313.

MORETTI, E. Real wage inequality. *American Economic Journal: Applied Economics*, v. 5, n. 1, p. 65–103, 2013.

NEVES-JR, E. C.; AZZONI, C. R.; CHAGAS, A. Skill wage premium and city size. *Working paper series - Department of Economics*, University of São Paulo (FEA-USP), n. 19, 2017.

NEVES-JR, E. C.; AZZONI, C. R.; CHAGAS, A. Mobilidade interfirma, salários e tamanho urbano: evidências para o Brasil. 2018.

OAXACA, R. Male-female wage differentials in urban labor markets. *International economic review*, JSTOR, p. 693–709, 1973.

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.

ROCHA, R. de M.; NETO, R. d. M. S.; GOMES, S. M. F. P. O. Maiores cidades, maiores habilidades produtivas: ganhos de aglomeração ou atração de habilidosos? Uma análise para as cidades brasileiras. *Revista Econômica do Nordeste*, v. 42, n. 4, p. 675–696, 2011.

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.

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.

YAHMED, S. B. Formal but less equal. gender wage gaps in formal and informal jobs in urban Brazil. *World Development*, Elsevier, v. 101, p. 73–87, 2018.

Tables

Table 1: Agglomeration Scale definition

MA	State	Macro-Region	Population	Agglomeration Scale
São Paulo	São Paulo	Southeast	21,571,281	Extra-Large MA
Rio de Janeiro	Rio de Janeiro	Southeast	12,699,743	
Belo Horizonte	Minas Gerais	Southeast	5,916,189	Large MA
Porto Alegre	Rio Grande do Sul	South	4,317,508	
Fortaleza	Ceará	Northeast	4,074,730	
Recife	Pernambuco	Northeast	3,975,411	
Salvador	Bahia	Northeast	3,899,533	
Curitiba	Paraná	South	3,615,027	
Distrito Federal	Distrito Federal	Midwest	2,974,703	
Manaus	Amazonas	North	2,631,239	
Goiânia	Goiás	Midwest	2,564,755	
Belém	Pará	North	2,491,052	
Grande Vitória	Espírito Santo	Southeast	1,951,673	Medium MA
Grande São Luís	Maranhão	Northeast	1,621,102	
Natal	Rio Grande do Norte	Northeast	1,587,055	
Maceió	Alagoas	Northeast	1,330,291	
João Pessoa	Paraíba	Northeast	1,266,463	
Florianópolis	Santa Catarina	South	1,189,947	
Vale do Rio Cuiabá	Mato Grosso	Midwest	1,032,714	
Aracaju	Sergipe	Northeast	949,342	Small MA
Campo Grande*	Mato Grosso do Sul	Midwest	885,711	
Teresina*	Piauí	Northeast	861,442	
Macapá	Amapá	North	634,450	
Porto Velho*	Rondônia	North	519,531	
Rio Branco*	Acre	North	401,155	
Boa Vista*	Roraima	North	375,374	
Palmas*	Tocantins	North	291,855	

Source: Organized by the authors. Estimated population for 2018 (IBGE, 2018a). *Only State's Capital.

Table 2: Descriptive analysis for the working-age population

	Formal	Informal	Non-employed	Total
	Avg/Share	Avg/Share	Avg/Share	Avg/Share
Individuals' characteristics				
Age	38.2	38.7	40.3	39.1
Years of Schooling	10.4	7.8	7.7	8.8
Race (whites)	53.1%	38.7%	42.1%	45.7%
Marital status (married)	61.2%	58.0%	53.5%	57.8%
Home household head	49.1%	47.6%	33.2%	43.3%
Household Composition				
No Child	62.5%	61.9%	68.8%	64.5%
Children under 7	0.23	0.24	0.21	0.22
Children between 7 and 14	0.30	0.37	0.29	0.31
Total Household wage (R$ month)	2,088.2	1,447.7	1,762.9	1,815.3
Total Household members	3.4	3.7	3.7	3.6
Household head/spouse occupied	56.9%	53.1%	50.8%	53.9%
Job-related characteristics				
Tenure (months)	90.3	85.5	-	88.4
Weekly worked hours	43.0	36.5	-	40.5
Weekly Hourly wage	16.6	10.0	-	14.2
High School and above	66.9%	38.9%	40.3%	50.7%
Medium or High Occupational Skill	67.4%	50.6%	-	40.2%
Tenure below 8 years	65.7%	68.5%	-	44.1%
In Metropolitan Areas	44.8%	33.6%	38.3%	39.8%
Share of individuals'	45.9%	25.4%	32.7%	100.0%

Source: Elaborated by the authors based on PNADC (IBGE, 2018a) from 2012 to 2018. Working-age population (18-65 years), referring to the first interview.
Notes: Appendix A shows the variables definition.

Table 3: Composition of metropolitan labour markets

	Formal							Informal						
				MA							MA			
	Total	MA	Non-MA	Small	Medium	Large	Extra-large	Total	MA	Non-MA	Small	Medium	Large	Extra-large
Skill														
High	21.4%	25.0%	18.5%	23.6%	22.4%	23.3%	27.3%	9.1%	12.7%	7.3%	11.5%	12.5%	12.0%	13.8%
Medium	45.9%	43.2%	48.2%	41.9%	42.6%	44.0%	42.6%	41.5%	39.3%	42.7%	39.7%	37.4%	39.4%	39.8%
Low	32.7%	31.9%	33.3%	34.6%	35.0%	32.6%	30.1%	49.4%	48.0%	50.1%	48.8%	50.2%	48.6%	46.5%
Schooling level														
Less than 1 year	2.0%	1.5%	2.5%	1.9%	2.3%	1.6%	1.2%	7.9%	4.4%	9.7%	5.7%	6.1%	4.6%	3.2%
Incomplete elementary school	17.1%	12.5%	20.8%	12.6%	13.5%	13.8%	11.0%	35.4%	26.4%	40.0%	27.8%	27.2%	28.0%	23.9%
Elementary school	14.0%	12.9%	14.9%	12.3%	12.9%	13.6%	12.2%	17.9%	18.9%	17.3%	19.1%	18.1%	18.7%	19.4%
High school	43.3%	45.1%	41.8%	45.6%	45.4%	45.0%	45.0%	30.4%	37.5%	26.8%	36.1%	36.3%	37.0%	38.7%
College or more	23.6%	28.1%	20.0%	27.6%	26.0%	26.0%	30.7%	8.5%	12.9%	6.3%	11.4%	12.3%	11.6%	14.8%
Tenure														
Less than 1 year	16.9%	16.8%	17.0%	18.3%	17.9%	18.0%	15.2%	28.3%	28.0%	28.5%	29.7%	29.6%	29.6%	25.2%
From 1 to 3 years and 11 months	29.1%	30.5%	28.0%	29.8%	29.4%	30.7%	30.6%	25.0%	27.8%	23.6%	27.5%	26.9%	27.6%	28.3%
From 4 to 7 years and 11 months	19.6%	19.9%	19.5%	19.5%	19.0%	19.6%	20.4%	15.2%	15.7%	14.9%	15.7%	15.8%	15.3%	16.2%
8 years or more	34.3%	32.8%	35.6%	32.5%	33.6%	31.7%	33.8%	31.5%	28.6%	33.0%	27.1%	27.8%	27.6%	30.3%
Share of individuals'	61.6%	68.2%	57.1%	62.6%	65.8%	66.3%	71.6%	38.4%	31.8%	42.9%	37.4%	34.2%	33.7%	28.4%

Source: Elaborated by the authors based on PNADC (IBGE, 2018a) from 2012 to 2018. Only employed individuals, between 18 and 65 years, referring to the first interview. Agglomeration levels definition on Table 1. **Notes:** Appendix A shows the variables definition.

Table 4: Sample descriptive analysis for mean hourly wages

	Total		MA		Non-MA		MA							
							Small		Medium		Large		Extra-Large	
	Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD
Hourly wage	**13.0**	25.8	**15.9**	35.1	**11.1**	16.4	**12.0**	19.1	**12.5**	16.1	**14.1**	21.6	**18.8**	47.8
Idade														
18 a 24	7.7	6.5	8.5	6.9	7.2	6.1	7.5	5.2	7.6	5.1	8.1	6.9	9.3	7.5
25 a 34	11.7	13.2	13.8	16.3	10.3	10.3	11.0	18.9	11.5	11.8	12.7	13.5	15.9	19.0
35 a 44	14.1	19.0	17.1	23.6	12.0	14.6	12.9	20.4	13.4	15.5	15.3	21.1	20.3	27.4
45 a 54	15.2	27.4	18.6	30.5	12.9	24.9	14.5	22.3	14.4	20.3	16.6	29.3	21.7	33.5
55 a 65	16.4	53.9	21.7	81.2	12.8	20.3	14.8	22.1	16.4	24.7	18.6	31.6	25.7	112.6
Schooling level														
Less than 1 year	6.1	6.5	8.0	9.0	5.5	5.4	7.8	14.6	7.0	6.0	7.6	9.5	9.1	7.3
Incomplete Elementary	8.3	8.1	9.0	8.1	8.1	8.0	8.5	7.0	8.2	6.8	8.8	9.0	9.7	7.2
Elementary	9.7	10.7	9.9	8.3	9.6	12.0	9.2	8.4	9.4	9.7	9.6	8.5	10.3	7.7
High school	12.6	14.0	13.2	15.4	12.0	12.5	11.8	14.4	12.2	13.6	12.9	16.2	13.9	14.9
College or more	36.7	67.1	41.3	80.1	30.2	42.4	28.9	46.1	30.6	30.9	37.0	44.2	46.1	100.3
Tenure														
Less than 1 year	8.7	10.2	10.1	14.0	7.8	6.7	8.2	6.6	8.6	7.9	9.3	13.7	11.9	16.4
From 1 to 3 years and 11 months	11.4	14.2	13.2	18.0	9.9	9.9	10.7	12.2	10.9	12.7	12.0	13.4	15.2	22.6
From 4 to 7 years and 11 months	13.9	18.4	16.5	23.1	11.9	13.5	13.1	23.9	13.5	15.3	15.0	22.9	18.8	24.3
8 years or more	16.7	39.1	21.7	55.8	13.6	23.7	15.9	26.7	16.4	22.0	19.3	29.5	25.5	76.4
Informal	**9.2**	13.7	**12.1**	18.8	**7.8**	9.9	**10.1**	17.3	**10.1**	12.8	**11.2**	18.6	**13.9**	20.3
Formal	**15.3**	30.5	**17.6**	40.2	**13.4**	19.5	**13.2**	20.2	**13.7**	17.4	**15.6**	22.8	**20.6**	54.5
Occupational Skill														
High	35.1	66.9	40.9	81.5	28.3	42.6	27.8	32.3	30.0	33.5	35.6	46.4	47.2	104.5
Medium	11.2	11.5	12.4	12.8	10.4	10.5	10.5	10.4	11.1	10.3	11.8	12.0	13.6	14.1
Low	8.7	10.0	9.9	11.8	7.9	8.5	9.3	21.7	9.0	9.7	9.4	10.6	10.6	11.4

Source: Elaborated by the authors based on PNADC (IBGE, 2018a). Appendix A shows the variables definition.

Table 5: Pooled OLS Ln(Hourly Wage) Regressions

Dep.Var.=lnhwage	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Agglomeration scale										
MA	0.269***	0.273***	0.157***	0.100***	0.0944***					
	(0.00218)	(0.00208)	(0.00212)	(0.00222)	(0.00207)					
Small MA						0.0980***	0.121***	0.0411***	0.139***	0.138***
						(0.00345)	(0.00331)	(0.00336)	(0.00318)	(0.00299)
Medium MA						0.124***	0.152***	0.0487***	0.102***	0.100***
						(0.00292)	(0.00281)	(0.00282)	(0.00258)	(0.00242)
Large MA						0.201***	0.211***	0.103***	0.103***	0.100***
						(0.00228)	(0.00218)	(0.00222)	(0.00224)	(0.00210)
Extra-large MA						0.389***	0.379***	0.249***	0.0923***	0.0815***
						(0.00387)	(0.00369)	(0.00368)	(0.00348)	(0.00323)
Formality			0.318***	0.195***	0.189***			0.315***	0.194***	0.188***
			(0.00174)	(0.00146)	(0.00140)			(0.00172)	(0.00145)	(0.00140)
SL2: Incomplete elementary				0.139***	0.130***				0.139***	0.130***
				(0.00304)	(0.00293)				(0.00304)	(0.00293)
SL3: Elementary				0.278***	0.256***				0.278***	0.256***
				(0.00412)	(0.00393)				(0.00413)	(0.00393)
SL4: High				0.431***	0.385***				0.432***	0.385***
				(0.00406)	(0.00388)				(0.00407)	(0.00388)
SL5: College or more				1.104***	0.839***				1.104***	0.839***
				(0.00556)	(0.00534)				(0.00557)	(0.00534)
Age (ln)				0.0243	-0.00489				0.0296	0.00128
				(0.0434)	(0.0409)				(0.0435)	(0.0409)
Age^2 (ln)				0.128***	0.0937***				0.125***	0.0910***
				(0.0181)	(0.0171)				(0.0181)	(0.0171)
Race				0.128***	0.106***				0.128***	0.106***
				(0.00152)	(0.00142)				(0.00152)	(0.00142)
Maritalstatus				0.0495***	0.0446***				0.0499***	0.0450***
				(0.00297)	(0.00280)				(0.00297)	(0.00280)
HHhead				0.0623***	0.0556***				0.0626***	0.0560***
				(0.00223)	(0.00209)				(0.00223)	(0.00209)
Unemployment rate				-0.430***	-0.350***				-0.441***	-0.363***
				(0.0459)	(0.0429)				(0.0460)	(0.0430)
High OS					0.491***					0.491***
					(0.00340)					(0.00341)
Medium OS					0.103***					0.103***
					(0.00130)					(0.00130)
T1: Less than 1 year					-0.155***					-0.156***
					(0.00182)					(0.00182)
T2: >= 1 to < 4					-0.120***					-0.120***
					(0.00188)					(0.00188)
T3: >= 4 to < 8					-0.0627***					-0.0628***
					(0.00206)					(0.00206)
Industry dummies	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Year/Quarter dummies	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Macro-Region dummies	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Heckman's Correction	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Constant	2.216***	2.563***	1.954***	1.331***	1.438***	2.216***	2.559***	1.955***	1.331***	1.441***
	(0.00115)	(0.00206)	(0.00311)	(0.0320)	(0.0302)	(0.00115)	(0.00203)	(0.00309)	(0.0320)	(0.0302)
Observations	2,842,954	2,842,954	2,842,954	2,842,954	2,842,954	2,842,954	2,842,954	2,842,954	2,842,954	2,842,954
R-squared	0.035	0.120	0.217	0.435	0.470	0.044	0.127	0.222	0.435	0.470

Notes: Base levels: Non-MA, Schooling Level = less than 1 year, Low occupation skill level, tenure of 8 or more years, agriculture, Southeast Region. All models were estimated with individuals' sample weight available on PNADC (IBGE, 2018a). Robust standard errors in parentheses clustered by individuals. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Pooled OLS Ln(Hourly Wage) Regressions by Formality status

	(a)			(b)		
Dep.Var.=lnhwage	*Formal*	*Informal*	*Full Sample*	*Formal*	*Informal*	*Full Sample*
Agglomeration scale						
MA	0.0823***	0.119***	0.0944***			
	(0.00254)	(0.00300)	(0.00207)			
Small MA				0.0717***	0.203***	0.138***
				(0.00363)	(0.00446)	(0.00299)
Medium MA				0.0569***	0.154***	0.100***
				(0.00288)	(0.00384)	(0.00242)
Large MA				0.0755***	0.133***	0.100***
				(0.00250)	(0.00326)	(0.00210)
Extra-large MA				0.0951***	0.0683***	0.0815***
				(0.00377)	(0.00518)	(0.00323)
Formality			0.189***			0.188***
			(0.00140)			(0.00140)
SL2: Incomplete elementary	0.0705***	0.125***	0.130***	0.0711***	0.122***	0.130***
	(0.00404)	(0.00386)	(0.00293)	(0.00404)	(0.00386)	(0.00293)
SL3: Elementary	0.206***	0.246***	0.256***	0.206***	0.244***	0.256***
	(0.00510)	(0.00554)	(0.00393)	(0.00510)	(0.00555)	(0.00393)
SL4: High	0.344***	0.365***	0.385***	0.345***	0.363***	0.385***
	(0.00502)	(0.00557)	(0.00388)	(0.00502)	(0.00558)	(0.00388)
SL5: College or more	0.795***	0.799***	0.839***	0.795***	0.800***	0.839***
	(0.00654)	(0.00920)	(0.00534)	(0.00654)	(0.00920)	(0.00534)
Age (ln)	0.153***	0.0354	-0.00489	0.153***	0.0445	0.00128
	(0.0524)	(0.0571)	(0.0409)	(0.0524)	(0.0571)	(0.0409)
Age^2 (ln)	0.0388*	0.0570**	0.0937***	0.0390*	0.0527**	0.0910***
	(0.0219)	(0.0237)	(0.0171)	(0.0219)	(0.0237)	(0.0171)
Race	0.107***	0.102***	0.106***	0.106***	0.102***	0.106***
	(0.00173)	(0.00217)	(0.00142)	(0.00173)	(0.00217)	(0.00142)
Maritalstatus	0.0607***	0.0320***	0.0446***	0.0606***	0.0331***	0.0450***
	(0.00366)	(0.00384)	(0.00280)	(0.00366)	(0.00384)	(0.00280)
HHhead	0.0639***	0.0493***	0.0556***	0.0637***	0.0504***	0.0560***
	(0.00265)	(0.00300)	(0.00209)	(0.00265)	(0.00300)	(0.00209)
Unemployment rate	-0.601***	-0.392***	-0.350***	-0.600***	-0.374***	-0.363***
	(0.0544)	(0.0619)	(0.0429)	(0.0543)	(0.0621)	(0.0430)
High OS	0.507***	0.434***	0.491***	0.507***	0.434***	0.491***
	(0.00386)	(0.00619)	(0.00340)	(0.00385)	(0.00619)	(0.00341)
Medium OS	0.112***	0.0952***	0.103***	0.112***	0.0957***	0.103***
	(0.00167)	(0.00189)	(0.00130)	(0.00167)	(0.00189)	(0.00130)
T1: Less than 1 year	-0.197***	-0.113***	-0.155***	-0.196***	-0.115***	-0.156***
	(0.00247)	(0.00241)	(0.00182)	(0.00247)	(0.00241)	(0.00182)
T2: >= 1 to < 4	-0.158***	-0.0597***	-0.120***	-0.158***	-0.0602***	-0.120***
	(0.00243)	(0.00269)	(0.00188)	(0.00242)	(0.00269)	(0.00188)
T3: >= 4 to < 8	-0.0992***	-0.00667**	-0.0627***	-0.0992***	-0.00708**	-0.0628***
	(0.00263)	(0.00298)	(0.00206)	(0.00263)	(0.00297)	(0.00206)
Industry dummies	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Year/Quarter dummies	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Macro-Region dummies	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Heckman's Correction	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Constant	1.663***	1.430***	1.438***	1.658***	1.441***	1.441***
	(0.0390)	(0.0421)	(0.0302)	(0.0390)	(0.0421)	(0.0302)
Observations	1,624,244	1,218,710	2,842,954	1,624,244	1,218,710	2,842,954
R-squared	0.442	0.384	0.470	0.442	0.385	0.470

Notes: Panel (a) considering a dummy for MA and Panel (b) four agglomeration levels. Base levels: Non-MA, Schooling Level = less than 1 year, Low occupation skill level, tenure of 8 or more years, agriculture, Southeast Region. All models were estimated with individuals' sample weight available on PNADC (IBGE, 2018a). Robust standard errors in parentheses clustered by individuals. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 7: Interactions with Agglomeration Scale

	Benchmark Model		MA*Occup. Skill		MA*Schooling Level		MA*Tenure	
Dep. Variable	lnhwage	SE	lnhwage	SE	lnhwage	SE	lnhwage	SE
Agglomeration scale								
Small MA	0.138***	(0.00299)	0.130***	(0.00414)	0.230***	(0.00936)	0.206***	(0.00569)
Medium MA	0.100***	(0.00242)	0.0895***	(0.00320)	0.185***	(0.00684)	0.148***	(0.00436)
Large MA	0.100***	(0.00210)	0.0645***	(0.00271)	0.169***	(0.00617)	0.170***	(0.00360)
Extra-large MA	0.0815***	(0.00323)	0.00961**	(0.00391)	0.160***	(0.0104)	0.164***	(0.00570)
Occupational Skill								
High OS	0.491***	(0.00341)	0.403***	(0.00438)	0.483***	(0.00338)	0.491***	(0.00340)
Medium OS	0.103***	(0.00130)	0.0912***	(0.00154)	0.102***	(0.00130)	0.105***	(0.00129)
Schooling level								
SL2: Incomplete elementary school	0.130***	(0.00293)	0.128***	(0.00292)	0.151***	(0.00314)	0.127***	(0.00292)
SL3: Elementary school	0.256***	(0.00393)	0.249***	(0.00390)	0.280***	(0.00413)	0.252***	(0.00392)
SL4: High school	0.385***	(0.00388)	0.381***	(0.00386)	0.394***	(0.00415)	0.382***	(0.00387)
SL5: College or more	0.839***	(0.00534)	0.836***	(0.00533)	0.780***	(0.00652)	0.835***	(0.00534)
Tenure								
T1: Less than 1 year	-0.156***	(0.00182)	-0.157***	(0.00180)	-0.157***	(0.00180)	-0.105***	(0.00206)
T2: >= 1 to < 4	-0.120***	(0.00188)	-0.121***	(0.00187)	-0.121***	(0.00187)	-0.0779***	(0.00217)
T3: >= 4 to < 8	-0.0628***	(0.00206)	-0.0635***	(0.00205)	-0.0638***	(0.00205)	-0.0330***	(0.00245)
Formality	0.188***	(0.00140)	0.189***	(0.00139)	0.188***	(0.00139)	0.188***	(0.00140)
Interactions								
Agglomeration scale x Occupational skill								
Small MA x High OS			0.0249**	(0.0107)				
Small MA x Medium OS			-0.000873	(0.00476)				
Medium MA x High OS			0.0408***	(0.00920)				
Medium MA x Medium OS			0.00262	(0.00375)				
Large MA x High OS			0.144***	(0.00705)				
Large MA x Medium OS			0.0235***	(0.00292)				
Extra-large MA x High OS			0.251***	(0.01000)				
Extra-large MA x Medium OS			0.0484***	(0.00437)				
Agglomeration scale x Schooling Level								
Small MA x SL2					-0.0851***	(0.00974)		
Small MA x SL3					-0.135***	(0.0102)		
Small MA x SL4					-0.116***	(0.0101)		
Small MA x SL5					-0.0905***	(0.0149)		
Medium MA x SL2					-0.0967***	(0.00731)		
Medium MA x SL3					-0.130***	(0.00769)		
Medium MA x SL4					-0.0986***	(0.00754)		
Medium MA x SL5					-0.0488***	(0.0116)		
Large MA x SL2					-0.109***	(0.00631)		
Large MA x SL3					-0.137***	(0.00658)		
Large MA x SL4					-0.0795***	(0.00654)		
Large MA x SL5					0.0564***	(0.00976)		
Extra-large MA x SL2					-0.135***	(0.0107)		
Extra-large MA x SL3					-0.173***	(0.0108)		
Extra-large MA x SL4					-0.117***	(0.0108)		
Extra-large MA x SL5					0.110***	(0.0140)		
Agglomeration scale x Tenure								
Small MA x T1							-0.129***	(0.00647)
Small MA x T2							-0.0894***	(0.00666)
Small MA x T3							-0.0710***	(0.00748)
Medium MA x T1							-0.0880***	(0.00506)
Medium MA x T2							-0.0726***	(0.00521)
Medium MA x T3							-0.0428***	(0.00609)
Large MA x T1							-0.129***	(0.00402)
Large MA x T2							-0.0984***	(0.00410)
Large MA x T3							-0.0720***	(0.00458)
Extra-large MA x T1							-0.155***	(0.00674)
Extra-large MA x T2							-0.123***	(0.00640)
Extra-large MA x T3							-0.0931***	(0.00719)
Worker controls	Yes		Yes		Yes		Yes	
Occupation controls	Yes		Yes		Yes		Yes	
Industry dummies	Yes		Yes		Yes		Yes	
Year/Quarter dummies	Yes		Yes		Yes		Yes	
Macro-Region dummies	Yes		Yes		Yes		Yes	
Heckman's Correction	Yes		Yes		Yes		Yes	
Constant	1.441***	(0.0302)	1.436***	(0.0300)	1.405***	(0.0300)	1.425***	(0.0301)
Observations	2,842,954		2,842,954		2,842,954		2,842,954	
R-squared	0.470		0.472		0.473		0.472	

Notes: Base levels: Non-MA, Schooling Level = less than 1 year, Low occupation skill level, tenure of 8 or more years, agriculture, Southeast Region. All models were estimated with individuals' sample weight available on PNADC (IBGE, 2018a). Robust standard errors in parentheses clustered by individuals. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 8: Robustness checks

Dep.Var.=lnhwage	Benchmark Model	A	B	C	D	E	F	G	H
Agglomeration scale									
Small MA	0.138***	0.144***	0.140***	0.132***	0.134***	0.183***	0.136***	0.108***	0.169***
	(0.00299)	(0.00447)	(0.00319)	(0.00312)	(0.00410)	(0.00553)	(0.00300)	(0.00333)	(0.00415)
Medium MA	0.100***	0.104***	0.0996***	0.0896***	0.0896***	0.137***	0.103***	0.0816***	0.105***
	(0.00242)	(0.00371)	(0.00258)	(0.00252)	(0.00322)	(0.00472)	(0.00246)	(0.00245)	(0.00224)
Large MA	0.100***	0.105***	0.101***	0.0925***	0.0979***	0.140***	0.101***	0.103***	0.103***
	(0.00210)	(0.00309)	(0.00223)	(0.00223)	(0.00288)	(0.00447)	(0.00212)	(0.00223)	(0.00216)
Extra Large MA	0.0815***	0.0885***	0.0833***	0.0691***	0.0447***	0.174***	0.0809***	0.0933***	0.0816***
	(0.00323)	(0.00442)	(0.00340)	(0.00333)	(0.00396)	(0.00696)	(0.00323)	(0.00296)	(0.00303)
Firm size									
F2: 6 to 10				0.0515***	0.0384***	0.0697***			
				(0.00228)	(0.00279)	(0.00385)			
F3: 11 to 50				0.0650***	0.0494***	0.0871***			
				(0.00235)	(0.00284)	(0.00401)			
F4: Above 50				0.0880***	0.0720***	0.110***			
				(0.00208)	(0.00253)	(0.00353)			
Formality	0.188***	0.186***	0.190***	0.148***	0.141***	0.156***	0.165***	0.189***	0.188***
	(0.00140)	(0.00227)	(0.00148)	(0.00171)	(0.00210)	(0.00286)	(0.00142)	(0.00140)	(0.00140)
Worker controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupation controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year/Quarter dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Macro-Region dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Heckman's Correction	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	1.441***	1.458***	1.399***	1.495***	1.561***	1.609***	1.475***	1.439***	1.442***
	(0.0302)	(0.0447)	(0.0319)	(0.0316)	(0.0396)	(0.0519)	(0.0302)	(0.0302)	(0.0302)
Observations	2,842,954	557,742	2,648,714	2,510,067	1,583,061	927,006	2,949,642	2,842,954	2,842,954
R-squared	0.470	0.466	0.472	0.470	0.467	0.480	0.452	0.470	0.470

Notes: (A) Only the first interview of each worker; (B) Only workers with at least two interviews in the period; (C) With an additional variable of firm size, available between 2012Q1 and 2015Q3 and 2016Q3 to 2018Q4, in four categories according the number of workers (1 to 5 as a base level); (D) Version with firm size only for the period from 2012Q1 to 2015Q3; (E) Version with firm size only for the period from 2016Q3 to 2018Q4; (F) Considering also workers who report weekly hours between 1 and 19; (G) With agglomeration level specification according to demographic density; and, (H) With agglomeration level specification according to Cities' Influence Regions (*Região de Influência das Cidades*) from IBGE (2008)). Base levels: Non-MA, Schooling Level = less than 1 year, Low occupation skill level, tenure of 8 or more years, agriculture, Southeast Region. All models were estimated with individuals' sample weight available on PNADC (IBGE, 2018a). Robust standard errors in parentheses clustered by individuals. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 9: Robustness checks - MAs Population Log

	(a)		
Dep.Var.=lnhwage	*Formal*	*Informal*	*Full Sample*
logpoprm	0.0224***	-0.0130***	0.0113***
	(0.00184)	(0.00255)	(0.00158)
Formality			0.130***
			(0.00238)
Unemployment rate	0.0789	-0.0218	0.0811
	(0.0869)	(0.112)	(0.0724)
Worker controls	Yes	Yes	Yes
Occupation controls	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes
Year/Quarter dummies	Yes	Yes	Yes
Macro-Region dummies	Yes	Yes	Yes
Heckman's Correction	Yes	Yes	Yes
Constant	1.701***	1.895***	1.702***
	(0.0843)	(0.110)	(0.0705)
Observations	666,879	339,743	1,006,622
R-squared	0.495	0.355	0.472

Notes: MAs' Population log as area' identifier. Estimated with individuals' sample weight available on PNADC (IBGE, 2018a). Robust standard errors in parentheses clustered by individuals. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 10: Controlling the unobservable characteristics

(a) Fixed-Effects

	Benchmark w/o Heck		Fixed Effects	
Dep.Var	lnhwage	SE	lnhwage	SE
High OS	0.492***	(0.00342)	0.0703***	(0.00259)
Medium OS	0.103***	(0.00130)	0.0190***	(0.00125)
Formality	0.189***	(0.00140)	0.0704***	(0.00142)
SL2: Incom. elementary school	0.164***	(0.00224)	0.0112***	(0.00220)
SL3: Elementary school	0.304***	(0.00293)	0.0360***	(0.00273)
SL4: High school	0.434***	(0.00281)	0.0600***	(0.00303)
SL5: College or more	0.892***	(0.00448)	0.125***	(0.00527)
T1: Less than 1 year	-0.156***	(0.00182)	-0.0385***	(0.00185)
T2: >= 1 to < 4	-0.120***	(0.00188)	-0.0222***	(0.00172)
T3: >= 4 to < 8	-0.0630***	(0.00206)	-0.00263	(0.00171)
Maritalstatus	0.0823***	(0.00171)	-0.00300	(0.00345)
HHhead	0.0770***	(0.00163)	0.0300***	(0.00271)
Unemployment rate	-0.537***	(0.0420)	-0.248***	(0.0297)
Aglomeration Scale	Yes		No	
Industry (dummies/FE)	Yes		Yes	
Year/Quarter dummies	Yes		No	
Macro-Region dummies	Yes		No	
Heckman's Correction	No		No	
Constant	0.952***	(0.00935)	2.141***	(0.00492)
Observations	2,842,954		2,842,954	
R-squared	0.470		0.010	

(b) Fixed Effects explained by time fixed variables

	POLS	
Dep.Var	Fixed Effects	SE
Small MA	0.285***	(0.00315)
Medium MA	0.226***	(0.00257)
Large MA	0.232***	(0.00201)
Extra-large MA	0.206***	(0.00372)
Age	0.00613***	(7.33e-05)
Race	0.247***	(0.00189)
Year/Quarter dummies	Yes	
Heckman's Correction	No	
Macro-Region dummies	Yes	
Constant	-0.375***	(0.00370)
Observations	2,842,954	
R-squared	0.202	

Notes: Panel (a) reportes only common variables between methods.
All models were estimated with individuals' sample weight available on PNADC (IBGE, 2018a). Robust standard errors in parentheses clustered by individuals.
Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 11: Oaxaca-Blinder Decomposition

(a) MA versus Non-MA

Dep.Var.=lnhwage	**Coeff.**	**Rob. Std.Err.**	**Hourly Wage**
Non-MA	2.216	0.00113	9.173
MA	2.486	0.00172	12.010
Difference	-0.269	0.00205	0.764
Decomposition			% Difference
Endowments	-0.176	0.00224	65.4%
Coefficients	-0.087	0.00192	32.5%
Interactions	-0.006	0.00197	2.2%

(b) Formal versus Informal

Dep.Var.=lnhwage	**Coeff.**	**Rob. Std.Err.**	**Hourly Wage**
Informal	2.030	0.00128	7.615
Formal	2.499	0.00125	12.170
Difference	-0.469	0.00173	0.626
Decomposition			% Difference
Endowments	-0.235	0.0014	50.1%
Coefficients	-0.163	0.00189	34.8%
Interactions	-0.071	0.00139	15.0%

Notes: Estimated with sample selection correction. All models were estimated with individuals' sample weight available on PNADC (IBGE, 2018a). Robust standard errors in parentheses clustered by individuals.

Appendix

A. List of explanatory variables

Variable name	Definition
MA	1 if Metropolitan Area or 4 dummies according to the number of inhabitants. Base level is Non-MA below 1 million=**Small MA**, from 1 to 2 million=**Medium MA**, from 2 to 10 million=**Large MA**, above 10 million=**Extra-large MA**
Age	The worker age measured in decades
$Race$	1 if white
$Maritalstatus$	1 if married
$HHhead$	Home household head. 1 if the head of the family
$SchLevel$	Years of schooling or 5 levels dummies as SL1: Less than 1 year (base level), SL2: Incomplete elementary school, SL3: Elementary school, SL4: High school, SL5: College or more
$OccSkill$	Occupational skill in 3 levels according to groups available on PNADC (VD4011 - *Grupamentos ocupacionais do trabalho principal da semana de referência para pessoas de 14 anos ou mais de idade*, following Gonçalves and Menezes-Filho (2015): **Low OS** as base level (variable items 5, 9 and 11), **Medium OS** (variable items 3, 4, 6, 7 and 8) **High OS** (variable items 1 and 2)
$Tenure$	Number of years the worker has been employed at her current workplace or 4 dummies as follow: **T1**=less than 1 year, **T2**=from 1 to 3 years and 11 months, **T3**=from 4 to 7 years and 11 months and **T4**= 8 years or more (base level)
$Unemployment$	Unemployment rate according to Macro-Region, MA, Year, Quarter and Schooling level
$Formality$	1 if formal, corresponding to workers legally contracted or self-employed which contributes to Social Security Institute
$Industry$	8 dummies according to the firm main activity: agriculture (base level), manufacture, construction, services, transportation, and public administration
$MacroRegion$	5 dummies: North, Northeast, Midwest, Southeast (base level), and South
	Selection Equation variables
$NoChild$	1 if there is at least one child at the household
$Child6$	Number of children up to 6 years old at the household
$Child14$	Number of children between 7 and 14 years old at household
$HHwage$	Total household wages, not including worker' i wage
$HHpeople$	Number of household members
$PosHS$	1 if home household head or spouse are occupied (if worker i is the household head this variable report the spouse position - if married - and if worker i is the spouse or other positions at the household is reported the household head position).

B. Heckman's correction detail

	(a)			(b)		
	POLS	Probit	POLS-Heck	POLS	Probit	POLS-Heck
Dep. Var.	*lnhwage*	*employed*	*lnhwage*	*lnhwage*	*employed*	*lnhwage*
Agglomeration scale						
MA	0.129***		0.128***			
	(0.00205)		(0.00205)			
Small MA				0.181***		0.179***
				(0.00302)		(0.00302)
Medium MA				0.135***		0.134***
				(0.00246)		(0.00246)
Large MA				0.140***		0.139***
				(0.00213)		(0.00212)
Extra-large MA				0.109***		0.108***
				(0.00324)		(0.00324)
Age (ln)	0.620***	6.501***	-0.0696*	0.618***	6.501***	-0.0613
	(0.0158)	(0.0339)	(0.0408)	(0.0158)	(0.0339)	(0.0408)
Age2 (ln)	-0.158***	-2.709***	0.129***	-0.158***	-2.709***	0.125***
	(0.00650)	(0.0135)	(0.0170)	(0.00651)	(0.0135)	(0.0170)
Race	0.105***	0.0115***	0.103***	0.105***	0.0115***	0.103***
	(0.00143)	(0.00343)	(0.00144)	(0.00143)	(0.00343)	(0.00143)
Maritalstatus	0.0801***	0.357***	0.0396***	0.0802***	0.357***	0.0403***
	(0.00172)	(0.00443)	(0.00280)	(0.00172)	(0.00443)	(0.00280)
HHhead	0.0686***	0.266***	0.0457***	0.0688***	0.266***	0.0463***
	(0.00165)	(0.00414)	(0.00210)	(0.00165)	(0.00414)	(0.00210)
SL2: Incomplete elementary school	0.191***	0.316***	0.153***	0.189***	0.316***	0.152***
	(0.00229)	(0.00504)	(0.00298)	(0.00229)	(0.00504)	(0.00298)
SL3: Elementary school	0.354***	0.466***	0.300***	0.352***	0.466***	0.300***
	(0.00297)	(0.00667)	(0.00398)	(0.00298)	(0.00667)	(0.00398)
SL4: High school	0.492***	0.470***	0.438***	0.490***	0.470***	0.437***
	(0.00283)	(0.00629)	(0.00391)	(0.00283)	(0.00629)	(0.00392)
SL5: College or more	0.958***	0.536***	0.900***	0.957***	0.536***	0.900***
	(0.00439)	(0.00827)	(0.00528)	(0.00439)	(0.00827)	(0.00528)
Unemployment rate	-0.330***	-1.823***	-0.139***	-0.344***	-1.822***	-0.156***
	(0.0421)	(0.0733)	(0.0431)	(0.0421)	(0.0733)	(0.0431)
High OS	0.527***		0.526***	0.527***		0.526***
	(0.00336)		(0.00336)	(0.00336)		(0.00336)
Medium OS	0.116***		0.116***	0.116***		0.116***
	(0.00124)		(0.00124)	(0.00124)		(0.00124)
Formality	0.171***		0.171***	0.170***		0.169***
	(0.00137)		(0.00137)	(0.00137)		(0.00137)
T1: Less than 1 year	-0.136***		-0.136***	-0.137***		-0.136***
	(0.00181)		(0.00181)	(0.00181)		(0.00181)
T2: >= 1 to < 4	-0.101***		-0.101***	-0.101***		-0.101***
	(0.00186)		(0.00186)	(0.00186)		(0.00187)
T3: >= 4 to < 8	-0.0488***		-0.0485***	-0.0489***		-0.0487***
	(0.00205)		(0.00205)	(0.00205)		(0.00205)
NoChild		-0.154***			-0.154***	
		(0.00734)			(0.00734)	
Child6 (ln)		0.00840			0.00846	
		(0.00776)			(0.00776)	
Child14 (ln)		-0.0531***			-0.0530***	
		(0.00741)			(0.00741)	
HHpeople (ln)		-0.110***			-0.110***	
		(0.00417)			(0.00417)	
HHwage (ln)		0.00696***			0.00692***	
		(0.000670)			(0.000670)	
PosHS		0.153***			0.153***	
		(0.00429)			(0.00429)	
Inverse Mill's Ratio			-0.233***			-0.230***
			(0.0123)			(0.0123)
Industry dummies	No	No	No	No	No	No
Year/Quarter dummies	Yes	Yes	Yes	Yes	Yes	Yes
State/Urban dummies	Yes	Yes	Yes	Yes	Yes	Yes
Heckman's Correction	**No**	**-**	**Yes**	**No**	**-**	**Yes**
Constant	1.142***	-3.157***	1.675***	1.153***	-3.157***	1.678***
	(0.00939)	(0.0226)	(0.0301)	(0.00939)	(0.0226)	(0.0301)
Observations	3,034,668	4,045,651	3,034,668	3,034,668	4,045,651	3,034,668
R-squared	0.442		0.442	0.442		0.442

Notes: Results based on estimations for the full PNADC data set with men aged 18-65, employed and unemployed from 2012 to 2018. Panel (a) considering a dummy for MA and Panel (b) four agglomeration levels. All models were estimated with individuals' sample weight available on PNADC (IBGE, 2018a). Robust standard errors in parentheses clustered by individuals. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

C. Interactions with Agglomeration Scale

	(a) Formal workers								(b) Informal workers							
	Benchmark Model		MA*Occup. Skill		MA*Schooling Level		MA*Tenure		Benchmark Model		MA*Occup. Skill		MA*Schooling Level		MA*Tenure	
Dep. Variable	lnhwage	SE	lnhwage	SE	lnhwage	SE	lnhwage	SE	lnhwage	SE	lnhwage	SE	lnhwage	SE	lnhwage	SE
Agglomeration scale																
Small MA	0.0717***	(0.00366)	0.0716***	(0.00507)	0.0882***	(0.0111)	0.119***	(0.00810)	0.203***	(0.00451)	0.177***	(0.00635)	0.269***	(0.0131)	0.278***	(0.00701)
Medium MA	0.0569***	(0.00289)	0.0443***	(0.00379)	0.0735***	(0.00824)	0.0757***	(0.00556)	0.154***	(0.00387)	0.135***	(0.00544)	0.200***	(0.0101)	0.218***	(0.00626)
Large MA	0.0755***	(0.00252)	0.0349***	(0.00318)	0.105***	(0.00824)	0.125***	(0.00455)	0.133***	(0.00329)	0.0983***	(0.00451)	0.158***	(0.00845)	0.211***	(0.00512)
Extra-large MA	0.0951***	(0.00382)	0.0239***	(0.00452)	0.118***	(0.0121)	0.152***	(0.00682)	0.0683***	(0.00524)	0.000146	(0.00685)	0.110***	(0.0170)	0.157***	(0.00878)
Occupational Skill																
High OS	0.507***	(0.00386)	0.411***	(0.00498)	0.498***	(0.00384)	0.507***	(0.00386)	0.434***	(0.00620)	0.359***	(0.00782)	0.430***	(0.00616)	0.433***	(0.00619)
Medium OS	0.112***	(0.00167)	0.102***	(0.00209)	0.110***	(0.00166)	0.113***	(0.00166)	0.0957***	(0.00189)	0.0816***	(0.00212)	0.0957***	(0.00189)	0.0987***	(0.00189)
Schooling level																
SL2: Incomplete elementary school	0.0711***	(0.00404)	0.0708***	(0.00403)	0.0933***	(0.00453)	0.0712***	(0.00403)	0.122***	(0.00386)	0.121***	(0.00385)	0.132***	(0.00400)	0.119***	(0.00385)
SL3: Elementary school	0.206***	(0.00511)	0.199***	(0.00507)	0.225***	(0.00553)	0.205***	(0.00510)	0.244***	(0.00555)	0.240***	(0.00553)	0.255***	(0.00572)	0.238***	(0.00554)
SL4: High school	0.345***	(0.00502)	0.342***	(0.00500)	0.349***	(0.00551)	0.345***	(0.00502)	0.363***	(0.00559)	0.360***	(0.00557)	0.353***	(0.00589)	0.359***	(0.00557)
SL5: College or more	0.795***	(0.00654)	0.798***	(0.00652)	0.735***	(0.00780)	0.794***	(0.00654)	0.800***	(0.00922)	0.791***	(0.00919)	0.723***	(0.0118)	0.795***	(0.00922)
Tenure																
T1: Less than 1 year	-0.196***	(0.00247)	-0.198***	(0.00245)	-0.198***	(0.00245)	-0.154***	(0.00293)	-0.115***	(0.00242)	-0.116***	(0.00241)	-0.116***	(0.00241)	-0.0654***	(0.00263)
T2: >= 1 to < 4	-0.158***	(0.00243)	-0.159***	(0.00240)	-0.159***	(0.00240)	-0.126***	(0.00290)	-0.0602***	(0.00270)	-0.0601***	(0.00269)	-0.0604***	(0.00270)	-0.0234***	(0.00301)
T3: >= 4 to < 8	-0.0992***	(0.00263)	-0.100***	(0.00261)	-0.101***	(0.00261)	-0.0783***	(0.00325)	-0.00708**	(0.00298)	-0.00706**	(0.00297)	-0.00730**	(0.00297)	0.0161***	(0.00339)
Interactions																
Small MA x High OS			0.0378***	(0.0123)							0.0258	(0.0178)				
Small MA x Medium OS			-0.0243***	(0.00580)							0.0375***	(0.00735)				
Medium MA x High OS			0.0606***	(0.0101)							0.0388**	(0.0175)				
Medium MA x Medium OS			-0.00459	(0.00442)							0.0268***	(0.00644)				
Large MA x High OS			0.162***	(0.00783)							0.130***	(0.0132)				
Large MA x Medium OS			0.0185***	(0.00346)							0.0394***	(0.00487)				
Extra-large MA x High OS			0.253***	(0.0109)							0.244***	(0.0198)				
Extra-large MA x Medium OS			0.0351***	(0.00500)							0.0695***	(0.00790)				
Small MA x SL2					-0.0162	(0.0120)							-0.0665***	(0.0134)		
Small MA x SL3					-0.0520***	(0.0123)							-0.110***	(0.0144)		
Small MA x SL4					-0.0365***	(0.0118)							-0.0696***	(0.0148)		
Small MA x SL5					-0.000512	(0.0169)							-0.0642***	(0.0242)		
Medium MA x SL2					-0.0434***	(0.00885)							-0.0609***	(0.0109)		
Medium MA x SL3					-0.0588***	(0.00924)							-0.0874***	(0.0117)		
Medium MA x SL4					-0.0313***	(0.00891)							-0.0311***	(0.0119)		
Medium MA x SL5					0.0365***	(0.0129)							-0.0315	(0.0225)		
Large MA x SL2					-0.0839***	(0.00846)							-0.0588***	(0.00869)		
Large MA x SL3					-0.103***	(0.00868)							-0.0748***	(0.00936)		
Large MA x SL4					-0.0498***	(0.00856)							-0.00102	(0.00958)		
Large MA x SL5					0.104***	(0.0116)							0.0844***	(0.0177)		
Extra-large MA x SL2					-0.0959***	(0.0126)							-0.0833***	(0.0176)		
Extra-large MA x SL3					-0.130***	(0.0126)							-0.106***	(0.0180)		
Extra-large MA x SL4					-0.0684***	(0.0124)							-0.0520***	(0.0182)		
Extra-large MA x SL5					0.160***	(0.0157)							0.164***	(0.0264)		
Small MA x T1							-0.0890***	(0.00880)							-0.158***	(0.00877)
Small MA x T2							-0.0596***	(0.00888)							-0.0916***	(0.00966)
Small MA x T3							-0.0463***	(0.00986)							-0.0737***	(0.0109)
Medium MA x T1							-0.0418***	(0.00631)							-0.129***	(0.00772)
Medium MA x T2							-0.0319***	(0.00636)							-0.0863***	(0.00864)
Medium MA x T3							-0.00577	(0.00742)							-0.0562***	(0.00998)
Large MA x T1							-0.0977***	(0.00503)							-0.153***	(0.00585)
Large MA x T2							-0.0680***	(0.00504)							-0.106***	(0.00639)
Large MA x T3							-0.0458***	(0.00561)							-0.0728***	(0.00733)
Extra-large MA x T1							-0.115***	(0.00795)							-0.174***	(0.0106)
Extra-large MA x T2							-0.0853***	(0.00750)							-0.127***	(0.0108)
Extra-large MA x T3							-0.0616***	(0.00839)							-0.0854***	(0.0123)
Worker controls	Yes		Yes		Yes		Yes		Yes		Yes		Yes		Yes	
Occupation controls	Yes		Yes		Yes		Yes		Yes		Yes		Yes		Yes	
Industry dummies	Yes		Yes		Yes		Yes		Yes		Yes		Yes		Yes	
Year/Quarter dummies	Yes		Yes		Yes		Yes		Yes		Yes		Yes		Yes	
Macro-Region dummies	Yes		Yes		Yes		Yes		Yes		Yes		Yes		Yes	
Heckman's Correction	Yes		Yes		Yes		Yes		Yes		Yes		Yes		Yes	
Constant	1.658***	(0.0390)	1.645***	(0.0387)	1.612***	(0.0387)	1.638***	(0.0389)	1.441***	(0.0421)	1.449***	(0.0420)	1.431***	(0.0419)	1.442***	(0.0420)
Observations	1,624,244		1,624,244		1,624,244		1,624,244		1,218,710		1,218,710		1,218,710		1,218,710	
R-squared	0.442		0.445		0.445		0.443		0.385		0.386		0.387		0.387	

Notes: Base levels: Non-MA, Schooling Level = less than 1 year, Low occupation skill level, tenure of 8 or more years, agriculture, Southeast Region. All models were estimated with individuals' sample weigth available on PNADC (IBGE, 2018a). Robust standard errors in parentheses clustered by individuals. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.