Female urban wage premium in Brazil*

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

Working paper - draft version
Submitted to XVIII ENABER - 2020

June 27, 2020

Abstract

The urban wage premium (UWP), is given by the positive wage differential in dense areas when compared to less dense regions and remains even after controlling the characteristics of individuals, occupations, and firms. Few studies identify the UWP between genders or, more specifically, that analyze the female UWP. Most studies focus on men, as they have a more stable behavior in the labor market. However, it is precise because there are different behaviors between genders that the inclusion of women becomes relevant. Such differences begin with the decision to participate in the labor market, which is influenced by factors and takes into account a coordination process, affecting men and women differently. The UWP investigation seeks to understand (i) whether the agglomeration economies in dense areas benefit men and women differently, and (ii) whether the allocation or composition of the labor market generates different magnitude for the UWP between genders. With this context, this paper aims to evaluate the UWP for women in Brazil from 2012 to 2019 using the Continuous National Household Sample Survey (PNADC). These data allow the analysis of several characteristics of individuals, occupations, firms, and household positions, with national coverage and representativeness. The results show that women's UWP is double that of men. Women show a UWP of 11.3%, while the male UWP is 5.7%. Using quantile regressions we reveal that the magnitude of the UWP is different over the wage distribution with different trajectories between men and women. In general, we show that the Brazilian UWP may be underestimated when considering only the group of men, given the higher female UWP. Still, one can overestimate the UWP when not considering the wage distribution (as we show with quantile regressions). These results are relevant to the UWP explanation in Brazil, by including women and highlighting the differences between genders.
Keywords: Female labor market. Urban wage premium. Agglomeration levels.
JEL classification: R23, J16, J21, J31
General Theme: 12 - *Questões espaciais no mercado de trabalho*

Resumo

O prêmio de salário urbano (UWP) é dado pelo diferencial de salário positivo em áreas densas quando comparado a regiões menos densas e existe mesmo após o controladas as características dos indivíduos, ocupações e firmas. Poucos estudos identificam a UWP entre gêneros ou, mais especificamente, que analisam o UWP das mulheres. A maioria dos estudos se concentra nos homens, pois eles têm um comportamento mais estável no mercado de trabalho. No entanto, é exatamente porque existem comportamentos diferentes entre os gêneros que a inclusão de mulheres se torna relevante. Tais diferenças começam com a decisão de participar do mercado de trabalho, que é influenciado por fatores diversos e leva em consideração um processo de coordenação, afetando homens e mulheres de maneira

*We gratefully acknowledge the financial support for this research by the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)*.
†M.Sc. Candidate in Economics and Development at Federal University of São Paulo-Brazil (UNIFESP). eloiza.regina@unifesp.br
‡Professor of Economics and Vice coordinator, Graduate Program in Economics and Development at Federal University of São Paulo-Brazil (UNIFESP). veneziano.araujo@unifesp.br
§Professor of Economics at Federal University of São Paulo-Brazil (UNIFESP). Researcher in the study group on Family and Gender Economics (*Grupo de Estudos em Economia da Família e do Gênero* - GeFam). slgoncalves@unifesp.br

diferente. A investigação do UWP procura entender (i) se as economias de aglomeração em áreas densas beneficiam homens e mulheres de maneira diferente e (ii) se a alocação ou composição do mercado de trabalho gera magnitude diferente para a UWP entre os sexos. Nesse contexto, este artigo tem como objetivo avaliar o UWP para mulheres no Brasil de 2012 a 2019 usando a Pesquisa Nacional por Amostra de Domicílios Contínuos (PNADC). Esses dados permitem a análise de várias características de indivíduos, ocupações, firmas e posições no domicílio, com cobertura e representatividade nacionais. Os resultados mostram que o UWP das mulheres é o dobro dos homens. As mulheres mostram um UWP de 11,3%, enquanto o UWP masculino é de 5,7%. Regressões quantílicas revelam magnitude diferente para a UWP ao longo da distribuição salarial e diferentes trajetórias entre homens e mulheres. De maneira geral, mostramos que o UWP no Brasil pode estar subestimado se considerarmos apenas o grupo de homens, dada o maior UWP feminino. Ainda assim, pode-se superestimar o UWP quando não se considera a distribuição salarial (como mostramos nas regressões quantílicas). Esses resultados são relevantes para a explicação da UWP no Brasil, incluindo mulheres e destacando as diferenças entre gêneros.
Palavras-chave: Mercado de trabalho feminino. Prêmio salarial urbano. Aglomerações.

1 Introduction

The urban wage premium (UWP) is the positive wage differential in dense geographical areas as compared to less dense regions. Several national and international studies have confirmed the existence of this differential even after controlling for the characteristics of individuals, occupations, and firms[1]. Agglomeration economies and human capital are the main traditional mechanisms that can explain the existence of the UWP[2]. However, a more recent line of investigation regarding the effect of specific individual characteristics on this premium has added new elements to this discussion. Analysis based on the level of education or skill, experience, and age group, for example, is found in recent studies[3]. Additionally, studies addressing developing countries have sought to identify the UWP considering the heterogeneity of individuals in the labor market[4].

Of these, most studies focus on men, a group with a higher level of participation in the labor market, with relatively few studies analyzing and identifying the female UWP. This choice seeks to eliminate possible variations due to discrimination and for the male workforce to have more stable behavior, despite adverse conditions such as low wages, high unemployment, and poor working conditions (MENEZES-FILHO; MENDES; ALMEIDA, 2004). However, it is precisely because the level of participation, wages, and decisions of both genders are distinct in the labor market, that the inclusion of women in UWP analysis becomes relevant.

Such differences begin with the decision of whether to participate in the labor market or not which is affected by factors traditionally linked to family responsibilities and care, household chores, and family composition. The existence of infrastructure and public services, such as child care, is an example of an incentive for the participation of women in the labor market (PHIMISTER, 2005). Additionally, the decision to participate in the labor market also considers family coordination in the decision-making process, especially related to the location of individuals, where thick labor markets can be a possible solution to the location problem for married couples, as they are large enough to offer good professional matches for both partners (MORETTI, 2012).

The gender income gap and the level of female labor force participation are well-documented issues in the labor market literature. Therefore, we will direct our analysis to the UWP differential between genders, which does not reflect the regional wage gaps, but rather, reveals the possible advantages and disadvantages of being in denser regions for both men and women. The female UWP investigation seeks to understand whether (i) agglomeration economies in denser areas benefit men and women differently, and (ii) the allocation or composition of the labor market influences the existence of a wage premium of different magnitudes between genders.

First, despite having mobility restrictions in the search for jobs[5], related to time, distance, or finance, it may

[1] Some examples of studies on developed countries are Combes, Duranton e Gobillon (2008), D'Costa e Overman (2014), and Glaeser e Mare (2001). In Brazil, we have Chauvin et al. (2017), Silva, Santos e Freguglia (2016), and Barufi (2015).

[2] Duranton e Puga (2004) conceptualizes the agglomeration economies, decomposing them into sharing, matching, and learning. The studies by Moretti (2011), Moretti (2013), Behrens, Duranton e Robert-Nicoud (2014), and Roca e Puga (2017) deal with the role of human capital, including the sorting process of more qualified individuals in dense areas.

[3] Examples are Moretti (2004), Gould (2007), Bacolod, Blum e Strange (2009), Andersson e Thulin (2013), Silva, Santos e Freguglia (2016), and Neves-Jr, Azzoni e Chagas (2017).

[4] Such as studies like Duranton (2016), Matano, Obaco e Royuela (2020), and García (2019).

[5] According to Passos e Guedes (2018), economically active women have to reconcile work with domestic duties—approximately 90% of

not be a disadvantage for women in metropolitan areas (MAs), given the greater diversity and the broader labor market, which facilitates and provides better matching between firms and workers (NISIC, 2017; MADDEN; CHIU, 1990; MEEKES; HASSINK, 2018). Additionally, women may experience higher levels of career interruption, labor turnover, and job-to-job transitions, and work fewer hours than men. These facts can have positive or negative effects on their earnings throughout their career, affect their accumulation of human capital, and imply losses of earnings that may be greater in MAs due to the agglomeration effects promoted by the learning process (PHIMISTER, 2005). By contrast, frequent career changes and interruptions could mean that women benefit more from job matching in urban areas; therefore, the negative effects from the breaks itself should be lower in denser areas. Thus, the agglomeration economies of dense areas may benefit men and women differently.

Second, factors related to the composition and allocation of women in the labor market can also be related to differences in the magnitude of male and female UWP. The concentration of women in service occupations—that can enable the combination of paid work and domestic activities, if flexible enough—can generate positive effects in urban areas, as they can promote better access to promotions to positions with higher wages (YAHMED, 2018). Additionally, women in MAs are more frequently in white-collar occupations and large companies with a higher level of productivity and wages. In contrast, outside dense areas they frequently work in smaller companies and occupations linked to manufacturing or personal services that pay less (KRUG; NISIC, 2011).

The few existing empirical studies reported controversial results. Meekes e Hassink (2018) finds no relevant difference in the UWP between men and women in the Netherlands between 2006 and 2014. Jones, D'Aoust e Bernard (2017) identifies an UWP only for men in three African labor markets (Nigeria, Tanzania, and Uganda) between 2010 and 2013. On the other hand, some studies that find the UWP for both genders differ in terms of magnitude. Krug e Nisic (2011) reports a higher UWP for men, albeit at just 3.7 euros per month, in Germany between 1995 and 2007.

Three studies find a higher UWP for women. Nisic (2017) estimates an UWP of 3.0% for married women and 1.4% for married men, in Germany between 1992 and 2012, whereas Duranton (2016) reports an UWP of 6.3% for women and 4.9% for men in Colombia between 1996 and 2012[6]. Similar results are reported by Phimister (2005), for the United Kingdom from 1991-1998, with a 6.4% UWP for women and a 3.8% UWP for men. Our contribution to this fledgling literature is the female UWP estimation comparing it with the well-studied male UWP, in addition to doing it for a developing country, with a large informal sector, for which, to the best of our knowledge, no attempt has been made to study female UWP.

Considering this context and the scant literature on the subject, this paper aims to investigate the female urban wage premium in Brazil. The analysis concentrates on MAs and uses data from the Continuous Brazilian National Household Sample Survey (PNADC) for the period between 2012 and 2019. The PNADC is a longitudinal database that provides rich individual information about labor market transitions and the characteristics of the occupations and firms in which workers are employed. An essential advantage of this database is that it allows us to investigate workers in both the formal and informal labor markets[7]. All results are compared with the corresponding figures for men.

Traditionally, UWP studies adopt a panel data approach, with estimates carried out using Ordinary Least Squares (OLS) and Fixed Effects (FE) regressions. The OLS estimation relies on a linear Mincer-type equation that identifies the average remuneration of each observable characteristic, and FE estimation, while also controlling for those characteristics that are not observable and that do not vary over time. However, the application of this method requires the observation of individuals who change regions over time, thus isolating the premium associated with being in a specific area. It is not possible to follow individuals who change homes with PNADC database, which implies that it does not include migrations between areas.

In the absence of this type of data, alternative approaches can be used and tested. We also contribute to the UWP literature by conducting empirical investigations using a alternative approaches as robustness test and several heterogeneity estimations. Likewise, quantile regressions have rarely been applied in UWP studies, but considering the concentration of women in unskilled and low-wage occupations, the study of the UWP throughout the distribution of wages, helps in a better interpretation of the UWP. Matano e Naticchioni (2016) observes a higher UWP in the top quantiles for different groups of individuals across migration patterns to less or highly dense areas. For Brazil, the

economically active women in Brazil perform household activities versus 52% of men—irrespective of whether they have to do household chores, or care for children or the elderly.

[6]His results, however, are very sensitive to different specifications, and not always favorable to women.

[7]Administrative registers in Brazil only cover the formal labor market, being the primary database of the Annual Social Information Report (RAIS).

only relevant study is that of Cruz e Naticchioni (2012), who found a UWP that is 3.1 p.p. higher in the 90th quantile compared to the 10th quantile in 2002 and 2009. Although both studies make use of quantile regressions, neither analyzes the UWP of women. Examining the UWP using quantile regressions is an innovative way to explore whether the effects of agglomeration benefit both genders differently.

Additionally, this paper empirically contributes to the UWP literature by estimating and correcting the sample selection bias related to participation in the labor market, which is particularly relevant in the case of female workers. Another contribution of this paper is that it integrates the literature on gender and regional economics, by extending the discussion beyond gender gaps and labor participation, and including the UWP at different agglomeration levels across genders.

The main results of this paper are as follows: there is a female UWP of 11.3%, while the UWP for men is half of that, at 5.76%—this result is robust to several specifications and robustness tests. We also find a higher UWP for women in subgroups based on the formality status in the labor market and the agglomeration scale. The quantile regressions show that the UWP has different magnitudes throughout the distribution and that it has different trajectories between men and women, mainly in medium, large and, extra-large MAs. The female UWP is lower at the beginning of the distribution, starting close to the male UWP—at around 7%—but increases and exceeds 12% in the groups with higher salaries (90th and 95th quantiles). Meanwhile, the male UWP is close to the regression by POLS at the beginning of the wage distribution, but becomes smaller right after the 10th quantile. Additionally, the estimation of the probability of labor market participation shows that different factors are associated with male and female participation, corroborating the findings of the gender economics literature.

Therefore, the main conclusion of this paper is that the previous UWP results are underestimated since female workers are neglected in most of the traditional UWP literature. Moreover, the UWP could be over or underestimated for women and men if wage distribution is disregarded. We find strong evidence that agglomeration effects do not benefit workers equally, and a higher female UWP can also be attributed to changes in the female labor market composition. It seems that denser areas tend to be more favorable to women, as these areas have a higher share of women in high-productive occupations. This scenario is consistent across different subgroups, which indicates that agglomeration effects have overcome possible constraints, pointed out by the literature, as women low spatial mobility.

This paper has four sections in addition to this introduction. Section 2 examines the Brazilian female labor participation, while Section 3 describes the database and sample used in this paper, and also presents the econometric model and estimation methods. Section 4 shows the main estimation results and robustness tests using different specifications and methods. Finally, Section 5 summarizes the main conclusions of the paper.

2 Brazilian Female Labor Market

Women represent 51.4% of the working-age population (18-55 years) in Brazil. Men and women in this age group are very heterogeneous in terms of individual characteristics, family composition, and labor market participation (Table 2.1). It is possible to observe that working-age women typically belong to families with more members and a larger number of children. Besides, women in this age group are less likely to be the head of their household, work fewer hours per week, and have less tenure in their current job compared to men.

Women also present a lower employment and formality rate, have lower participation in the labor market, even though have a higher level of schooling (10.6 years for women versus 9.9 years for men). 41.% of women are in MAs, while the figure for men is 39%. Between 2012 and 2019, the participation of women in the Brazilian labor market reached 59%, whereas the participation of men reached 81% in the same period. For women, we observed a more significant regional disparity in terms of labor market participation, as shown in Figure 2.1, varying between 36% and 71%, with the largest share of workers in the MAs, and the Southern and Southeast Regions.

Table 2.1: Working-age individuals' profile by Gender

	Women	Men	
	Avg/Share (1)	*Avg/Share* (2)	Diff. (3)
Individual characteristics			
Age	35.8	35.3	-0.459***
Years of schooling	10.6	9.9	-0.680***
Race (whites)	45.0%	44.0%	-0.015***
Marital status (married)	57.0%	55.0%	-0.018***
Household head	30.0%	51.0%	0.206***
Household composition			
With child	42.0%	39.0%	-0.033***
Child 0-6	0.26	0.25	-0.015***
Child 7-14	0.38	0.34	-0.036***
Household labor income^{a}	2,104.2	1,503.2	-600.998***
Family size	4.1	3.6	-0.43
Head/spouse employed	63.0%	51.0%	-0.123***
Labor Market			
Employment rate	0.88	0.92	0.037***
Labor participation	0.59	0.81	0.215***
Formal worker	61.0%	63.0%	0.018***
Weekly working hours	37.6	42.9	5.294***
Tenure (months)	70.9	82.1	11.195***
In Metropolitan areas	41.0%	39.0%	-0.013***
N	871,888	822,785	1,694,673
Share of individuals	51.4%	48.6%	

Source: Elaborated by the authors based on PNADC (IBGE, 2020) from 2012 to 2019 with individual sample weights.
Notes: We only consider data from the first of five interviews conducted by the PNADC for individuals aged between 18 to 55 years. ^{a} Values in R$. The significance levels are *** p <0.01, ** p <0.05, and * p <0.1.

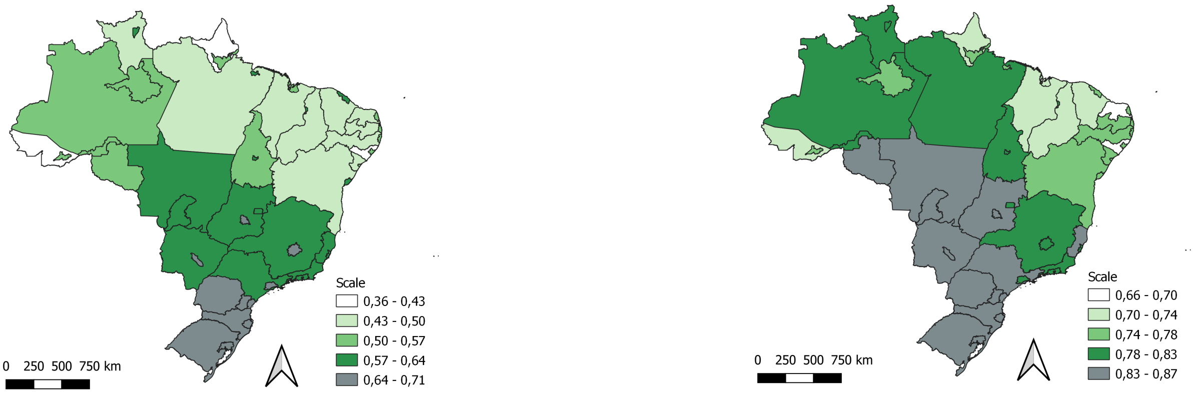
Women are in service, sales, and elementary occupations[8] (Figure 2.2), which accounts for 52% of women employed in non-MAs and 45% in MAs. Such occupations are among those that pay the lowest wages in the labor market. For men, these two occupations—services, sales, and elementary activities—also represent a large share of workers but reaches at most 33% and 30% in non-MAs and MAs respectively. The largest share of male workers is in skilled construction occupations in both areas, which represents 21% and 19%, respectively. Even though wages are slightly higher for women in elementary occupations, they are higher for men in all other occupations, independently of the area.

Despite this, non-MAs concentrate workers from both genders in occupations with lower wage levels (Figure 2.3) than in MAs, where we observe the opposite, with a larger group of workers with higher wages. Even so, women are more concentrated than men in occupations with lower wages, as their distribution curve is left-shifted. Thus, the facts highlight that the Brazilian labor market is similar to the description given by the literature, and culminates in a lower level of wages for women. Although our focus is not on the gender gap, we found elements that corroborate with the two possible explanations of the UWP magnitudes by gender, as discussed in the literature. The agglomeration effects may impact workers differently, according to their gender, and women are more present in occupations with lower wage levels.

[8] These occupations involve domestic workers in general; interior cleaning workers in buildings, offices, hotels, and other establishments; laundries for clothes and handrails; vehicle washers; window cleaners; other cleaning workers; elementary workers in agriculture, fishing and forestry, and those in mining, construction, manufacturing and transportation (IBGE, 2019b).

Figure 2.1: Labor Market Participation by Gender

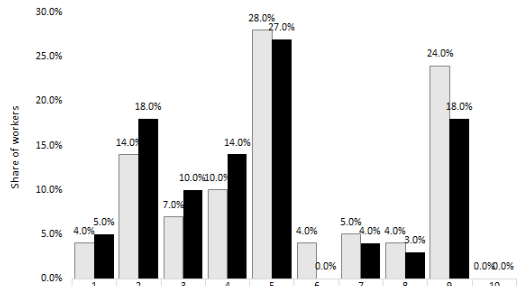
(a) Women (b) Men



Source: Created by the authors based on data from the PNADC (IBGE, 2020) from 2012 to 2019 with individual sample weights; the Territorial Brazilian Division (DTB) (IBGE, 2018b), and municipalities *shape file* in 2010 from IBGE (2019a).
Notes: The map plots the state frontiers and the respective capitals and MAs. Only workers aged between 18 and 55 years.

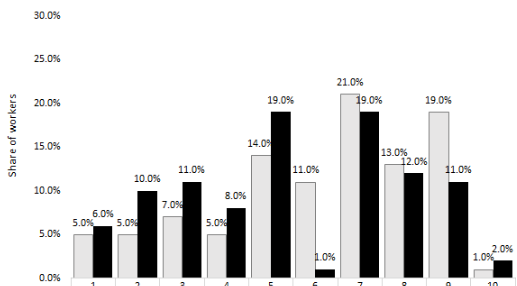
Figure 2.2: Share of workers and hourly wage by Occupation and Gender

(a) Women



	1	2	3	4	5	6	7	8	9	10
Non-MA	4.0%	14.0%	7.0%	10.0%	28.0%	4.0%	5.0%	4.0%	24.0%	0.0%
MA	5.0%	18.0%	10.0%	14.0%	27.0%	0.0%	4.0%	3.0%	18.0%	0.0%
Non-MA (hourly wage)	21.6	23.2	14.2	10.6	8.8	7.9	8.2	8.0	7.2	28.0
MA (hourly wage)	32.6	33.8	19.3	12.1	10.1	9.6	9.8	9.1	8.5	34.5

(b) Men



	1	2	3	4	5	6	7	8	9	10
Non-MA	5.0%	5.0%	7.0%	5.0%	14.0%	11.0%	21.0%	13.0%	19.0%	1.0%
MA	6.0%	10.0%	11.0%	8.0%	19.0%	1.0%	19.0%	12.0%	11.0%	2.0%
Non-MA (hourly wage)	29.2	33.3	18.8	12.2	12.1	8.7	11.1	11.0	6.6	25.3
MA (hourly wage)	43.5	44.9	22.6	13.8	13.0	9.1	12.0	12.2	7.8	29.7

Source: Elaborated by the authors based on PNADC (IBGE, 2020) from 2012 to 2019, with individual sample weights.
Notes: We only consider data from the first of five interviews for individuals aged between 18 to 55 years. Hourly wages in R$. Occupation levels: 1-Directors and managers; 2-Science and intellectuals professionals; 3-Mid-level professionals and technicians; 4-Administrative support workers; 5-Service workers and salespeople; 6-Skilled farming forestry, Hunting and fishing workers; 7-Skilled construction workers; 8-Mechanical workers, plant and machinery operators; 9-Elementary occupations; and 10-Members of the army forces, police and military firefighters.

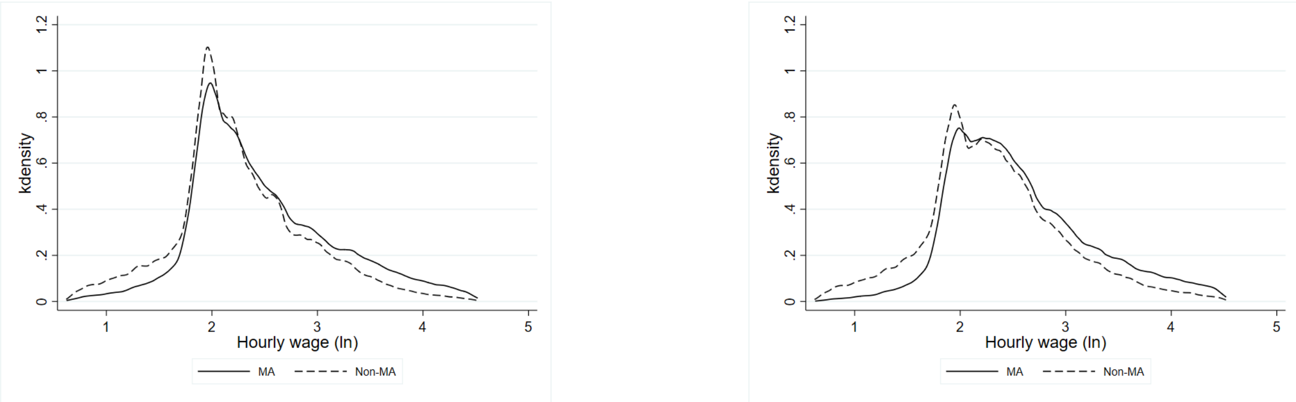
3 Empirical Strategy

3.1 Database and sample

The database used for the empirical analysis is the PNADC, collected and released by the Brazilian Institute of Geography and Statistics (IBGE). PNADC is representative of the Brazilian population, has a broad geographic scope, and covers the entire Brazilian territory, excluding some select areas. It is a quarterly longitudinal data survey based on a rotating panel in which a household is surveyed one month and then excluded from the sampling process for the following two months, before returning for the next interview. The process is repeated until five complete interviews on each household have been carried out, once every quarter. Thus, the database presents approximately 211,000 interviews by quarter. The analyzed period comprises the period of 2012 to 2019, totaling more than 17.0 million observations (IBGE, 2020).

Figure 2.3: Hourly Wage distribution by area

(a) Women (b) Men



Source: Elaborated by the authors based on PNADC (IBGE, 2020) from 2012 to 2019, with individual sample weights.
Notes: We only consider data from the first of five interviews for individuals aged between 18 to 55 years. The average natural logarithm of the hourly wage is 2.36 (R$ 13.73) for women and 2.44 (R$ 15.33) for men.

The PNADC records a variety of information on individuals, households, their occupations, and the firms in which they work, which makes it possible to investigate the socioeconomic conditions of different members of a household. It is also possible to observe the statuses and transitions of individuals in the labor market, including those of unemployed workers. PNADC microdata present a household identifier[9] that allows us to follow families through the panel, but there is no individual longitudinal identifier specified by the IBGE. Thus, following the empirical literature on Brazilian household surveys, we create an identifier using the household identifier, the gender, and the date of birth of each individual. Some data were excluded in constructing the sample for this paper. The final sample only comprises employed and in-paid-work individuals[10], aged between 18 and 55 years, except public sector workers, the military, and family workers[11]. We also exclude individuals with labor income in the 1st and 99th percentile of the wage distribution in each year, as well as those with only one interview in the panel[12].

Following the literature, as working in large firms implies wage gains, and as large firms are located disproportionately in MAs, their size could affect the UWP estimated for denser areas[13]. Therefore, we exclude three quarters, from 2015 Q4 to 2016 Q2, from the sample due to the lack of information about firm size. The individual makes known the size of the firm for which he or she works. It is reported discretely with four ranges according to the number of workers: from 1 to 5, from 6 to 10, from 11 to 50, and above 50 workers. Domestic workers are not asked about this, however, given the relevance of this employment category among the group of women (13% of women), we perform an input procedure to complete the data for these workers, determining their firm size range as from 1 to 5 employees.

The final sample is comprised of more than 3.76 million observations, of which 1.46 million are women, and 2.30 million are men. Since PNADC is a household survey, we can use general information about the household to study each individual member. This fact is particularly relevant to the objective of this paper, as these characteristics may differently influence the participation of men and women in the labor market and their respective wage premiums. Thus, we create variables for each individual based on their household composition. Marital status is created from the existence of a household head and a spouse (of the same or different gender) in the same household[14]. Information

[9]This identifier is a combination of the primary sampling unit (UPA), the household number (V1008), and the panel identifier (V1014).

[10]Before the exclusion of unemployed and inactive individuals, we performed a correction of the sample selection bias related to the probability of being employed, as will be explained ahead.

[11]Such workers do not have earnings, although they report helping a family member in some economic activity.

[12]As household survey, PNADC data shows attrition caused by two types of missing data: (i) Individuals who change homes are missing after they move, which implies that PNADC does not include migrations between areas (individuals are in the same place throughout the survey); (b) Due to the non-response of individuals at some point between the five quarters. The data loss is approximately 14% between the first and second interview, reaching 39% by the fifth interview. Correction for attrition was not applied, but we try to deal with this possible bias looking at different sample cohorts as robustness tests.

[13]As in studies like Yankow (2006), Badaoui, Strobl e Walsh (2010), Andersson e Thulin (2013), Dauth et al. (2016) and Silva (2018).

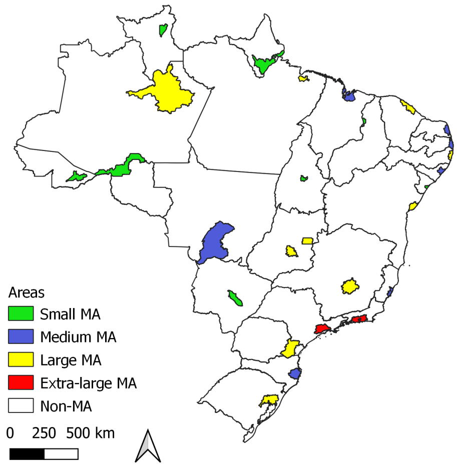
[14]We applied the same reasoning to the identification of a married son or daughter.

about the household position of each member is available, as well as their labor income, and the number of children by age group. Aggregated variables for the family are also created, such as the family size and the presence of grandparents or domestic workers.

The wages of individuals in the PNADC are self-reported[15] and, therefore, are not originally corrected for price variation across time and area. Since wages are our primary dependent variable, it is necessary to deflate it for proper identification[16]. We deflate wages to the prices of the 4rd quarter of 2019 using an index based on the Extended National Consumer Price Index (IPCA), released by IBGE (IBGE, 2018a). When considering the IPCA, we are also partially correcting regional inequalities since it covers states by quarter and year[17]. PNADC provides a representative sample for capitals, MAs, and states. Each MA corresponds to the state capital and to the respective MA, if any, or otherwise only to the state capital, totaling 27 MAs that cover 41.1% of the Brazilian population. Disregarding the MAs not linked to a capital city means the exclusion of the MAs located in the countryside, which accounts for approximately 4.6% of the total population. Regions outside MAs are called non-MAs. MAs are classified into four groups according to their population size, small, medium, large, and extra-large (as shown in Figure 3.1). This classification is justified by the heterogeneity of MAs in terms of the population size, varying between 291 thousand and 21 million.

Finally, PNADC is a complex sampling survey. It contemplates a probabilistic sample, extracted from a master sample of sectors based on the Census (IBGE), which is representative at different geographic levels. The units selected for the master sample constitute the PSU. In each quarter, the interviewed individuals receive a sample weight, with a correction for non-interview events, and with post-stratification by population projection. However, there is no longitudinal weight that considers individuals' representativeness throughout the panel. Since a constant weight is necessary to perform estimations, we use the average of the individual weights for the periods in which he or she is interviewed.

Figure 3.1: States and metropolitan areas in Brazil



Source: Created by the authors based on data from the PNADC (IBGE, 2020), the Territorial Brazilian Division (*Divisão Territorial Brasileira (DTB)* (IBGE, 2018b), and a *shape file* of municipalities in 2010 based on IBGE (2019a).

[15] There is a possible bias due to declaration errors. Although the question about wages specifies gross values, individuals can interpret the question as net values or can be unaware of the exact gross amount. This fact, however, is not exclusive to this paper and will not be subject to analysis or correction.

[16] Such correction, however, is ignored in several studies on UWP in Brazil and other countries, given the lack of a price index at a specific geographical scale. Exceptions are the works of Yankow (2006), Baum-Snow e Pavan (2011) and Chauvin et al. (2017) that correct the wages from price variation across time and area. Cruz e Naticchioni (2012), Barufi (2015), Matano, Obaco e Royuela (2020) and Silva (2018) are examples of studies that only time-corrected wages.

[17] In an attempt to account for price variation across areas, robustness tests were carried out considering regional consumption baskets (RCB) of each Brazilian state is based on the National Survey of the Basic Food Basket from the Inter-Union Department of Statistics and Socioeconomic Studies (DIEESE, 2019). This survey provides monthly average values of consumption basket considering different compositions (linked to cultural aspects) for all the state capitals. It had a national coverage since 2016, so for some places, mainly North and Northeast Regions, we consider the average value of the respective Macro-Region. Based on the RCB values in each state, we build an index considering as base level the average of the two baskets that have the lowest values. Finally, a third correction attempt considers both time and areas criteria, firstly 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.

3.2 Econometric Model

3.2.1 Benchmark framework and the sample selection

The first step of the econometric strategy is evaluating the sample selection bias related to the probability of working that can be distinct for women and men. The estimation of the UWP requires us to estimate a regression in which wages (W) are the dependent variable. However, a positive value for wages is only observed if the individual is employed, generating an issue of sample selection since employed and unemployed workers can be different in observable and unobservable characteristics[18].

The identification and correction of the sample selection bias, according to Heckman (1979), is based on using information about the individual's decision to participate in the labor market and be employed. It is possible through a two-step procedure: (i) the estimation of the probability of being employed through the selection equation, which is estimated by considering employed and non-employed individuals, and (ii) the use of this predicted probability in the estimation of the equation for wages. The first stage is given by $P(LMP = 1|X) = \Phi(\delta X)$. Thus, we estimate the probability of being employed for the entire employed and non-employed population aged between 18 and 55, excluding workers in the public sector, the military, and family workers, and we also carry out estimates separately for men and women. We chose a non-linear *Probit* model with the following specification:

$$P_{it} = \alpha + \omega X_{it} + e_{it} \tag{1}$$

In this equation 1, P_{it} is the probability of being employed and receiving a nonzero wage. The vector of independent variables X_{it} includes the individual characteristics that affect this probability, such as the age and the age squared, race, marital status, household position, schooling level, the presence of a child, region, quarter, and year. Besides, we include some household variables, such as the number of children up to six years old, the number of children between seven and fourteen years old, the family size, the employment status of the spouse or the household head, the number of sons or daughters of working-age, a dummy for at least one married son or daughter in the household, the labor income of the spouse, of the son or daughter and of other household members excluding the wages of individual i, a dummy for the presence of a grandfather or grandmother, and a dummy for the presence of a domestic worker at the household[19]. The inclusion of these variables, particularly in the analysis of women, is motivated by the theoretical discussion about the factors that differently influence the participation of women and men in the labor market. These factors are intrinsically related to the family composition and the household position of women. Finally, the variable e_{it} is the error term.

From the first step, we recover the inverse of the Mills ratio (IMR) as $\widehat{\lambda}_i = \lambda(\mathbf{x}_i\widehat{\delta})$, in which $\lambda(.) \equiv \frac{\phi(.)}{\Phi(.)}$ is the ratio between the probability density function and the cumulative distribution function. This ratio is included in the second step as an independent variable capturing the sample selection bias. The second step consists of estimating Mincer's equation using the Pooled OLS (POLS) method, which can allow parameters to change 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*)—as in our case—POLS estimations allow for aggregate time effects that have the same influence on W for all i. We estimate the following specification:

$$W_{it} = a + \beta X_{it} + \theta MA_{it} + \gamma\widehat{\lambda}_{it} + u_{it} \tag{2}$$

in which W is the logarithm of the hourly wage ($lnhwage$) of the worker i in the period t. The vector of independent variables X_{it} presents three groups of observable characteristics: i) individual characteristics: gender, age, age squared, race, marital status, household position, schooling level, year, quarter, and region; ii) occupational factors: industry, occupation, firm size, tenure, and formality status; and, iii) household characteristics: the presence and number of children under six years old or between seven and fourteen years old. The variable u_{it} is the error term, and the variable $\widehat{\lambda}$ is the predicted IMR.

The UWP is captured by the coefficient θ that indicates the wage premium associated with being in a MA.

[18]One of the advantages of the PNADC is the availability of information for both employed and unemployed individuals—including unemployed and inactive individuals—which allows us to estimate and correct the sample selection bias. Most studies on the UWP do not perform this type of correction. According to Dalberto e Cirino (2018), this question becomes irrelevant in the case of men, since most of them are in the labor market. As women experience higher unemployment and inactive rates, this fact needs to be considered.

[19]Considering a domestic worker that lives in the household, not daily workers.

Additionally, for some estimations, the MA can be categorical, and we consider four groups of MAs: small, medium, large, and extra-large, with non-MA as omitted category. The estimations separated by gender allow the identification of the UWP magnitude within each group. Robustness test are applied to verify whether the results obtained are consistent with different samples cohorts. As heterogeneity exercises we also estimate the UWP separately for formal and informal workers, and for different sub-samples. Additionally, to qualify the results considering the individuals' characteristics, we estimate Equation 2 including the interaction of the agglomeration levels with some chosen variables. By doing this, we identify the UWP by the sum of the θ coefficient and the β from the interaction term.

The estimation via POLS allows us to identify the average UWP level for each specification and sample considering the observable characteristics of individuals. The UWP literature traditionally estimates UWP first by using the OLS method and then by using the FE, then controlling for observable and non-observable characteristics that are fixed in time[20]. However, the application of this method requires following individuals even under housing and or region changes over time, allowing them to isolate the exact premium associated with being in a specific area. Even though PNADC is a longitudinal database, when housing and or region change occurs, the household leaves the sample, which makes it impossible to identify the UWP using FE.

3.2.2 Quantile regressions

POLS estimation allow us to identify the UWP for women, men, or for the full sample. In section 2, we argue that one of the characteristics of the female labor market is the concentration of workers in occupations that pay lower salaries, regardless of the area they live in. This fact raises an essential hypothesis in the UWP study: how different is the urban wage premium over the distribution of wages? That is, is the urban wage premium in an occupation that pays low wages higher or lower than the urban wage premium in an occupation with a higher salary? The estimation of quantile regressions can indicate whether the UWP has a different magnitude based on the quantile of the distribution of wages.

According to Cameron e Trivedi (2005), quantile regressions provide the interpretation of effects along with the wage distribution. They are an alternative to OLS estimations as they use an asymmetric absolute loss or check function—that assigns different weights to the negative and positive residues among the distribution—instead of minimizing squared errors. The estimation of quantile regression performed in this paper follows the specification of Koenker e Jr (1978) and Wooldridge (2010). Our objective is to identify the UWP in each quantile of the conditional distribution and the explanatory variables X, assuming linearity of the parameters:

$$Quant_\tau(Y_i|X_i) = \varphi_0(\tau) + \beta X_i(\tau) + \theta MA_i(\tau) \tag{3}$$

where the τ coefficient represents the chosen quantile and is between 0 and 1. Under these conditions, the error minimization solution is given by $\rho_\tau(u) = (\tau - 1[u < 0])$, thus the consistent estimator for each $\widehat{\beta}_\tau$ (or for $\widehat{\theta}(\tau)$) will be given by: $\widehat{\beta}_\tau = argmin_\beta \dfrac{1}{N} \sum_{i=1}^{N} (Y_i - \varphi_0 - \beta X_i - \theta MA_i)[(\tau - 1)\{Y_i - \varphi_0 - \beta X_i - \theta MA_i \geq 0\}]$

Each $\widehat{\beta}_\tau$ will capture the heterogeneity of returns obtained by individuals for the characteristics observable (X) throughout the wage distribution. The coefficient $\widehat{\theta}(\tau)$ captures the heterogeneity of the UWP conditional on the characteristics observed (X) throughout the wage distribution in each chosen quantile. The vector X includes the same group of variables as estimated by POLS. For this estimation, we use only the first interview for each individual in a cross-sectional approach[21] by gender and for the full sample, and for the 5th, 10th, 25th, 50th, 75th, 90th, and 95th quantiles[22].

[20] Examples of this are found in studies carried out by Glaeser e Mare (2001), Baum-Snow e Pavan (2011), D'Costa e Overman (2014), Barufi (2015), Chauvin et al. (2017), and Silva (2017).

[21] Although it is possible to estimate quantile regressions for panel data, their use considers FE, which is not possible with our database, as previously explained. For more details, see Machado e Silva (2019).

[22] An issue involving quantile regressions is their estimation in the context of a sample with selection bias. An alternative to correct the sample selection bias in quantile regressions is to correct it in a semi non-parametric (SNP) way. This method follows the specification of Buchinsky (2002). The compared results of the two correction methods are available under request. We compare the quantile regressions obtained by both corrections of the sample selection bias, and verify very close UWP coefficients. For all the agglomeration levels, the UWP obtained by both methods is similar in magnitude and trajectory throughout the distribution. The confidence intervals overlap, indicating that the hypothesis of error linearity does not significantly affect the results, and the results by the traditional method are consistent under a different procedure. For a case study, see Coelho, Veszteg e Soares (2010).

4 Results

4.1 Descriptive analysis

This session initially provides a comparison of the profiles of men and women in MAs and non-MAs (Table 4.1a-b). Regarding the educational profile, women have more schooling years in total compared to men, 10.5 and 11.6 versus 9.1 and 10.8. In non-MAs, both genders have a family with more members, with a higher presence and number of children, and present a lower household labor income level. Women are more often the head of the household in MAs than in non-MAs—36% versus 32%. For men, the opposite occurs (58% versus 52%). Women overcome men in household labor income, as they are more frequently in a household with another person who is employed. Precisely, at least 60% of women have a spouse who is employed, while only 52% of men present this characteristic.

Concerning occupational and firm characteristics, formality is higher in MAs, reaching 65% for women and 68% for men, while in non-MAs, 58% of the total number of workers of both genders are formal. The average female wage is lower in both areas with a higher gender gap in MA —R$1.9 per hour—, with lower differentials in tenure and working hours in these areas. As pointed out by the literature, women work fewer hours and have a lower tenure.

Analyzing agglomeration levels (columns c to f), we verify that the differential in years of schooling between genders decreases as the area becomes denser with less than one year at large and extra-large MAs. Again, we see smaller families in denser areas, with fewer married individuals in both genders. Women still have a lower wage compared to men, and the gap between them increases in denser areas. Additionally, we have an increase in the average hourly wage with the agglomeration levels for both genders. In small MAs, the wage gap between men and women is 18%, while in extra-large MAs, it reaches only 13%. The reasons for this behavior can be related to better labor market conditions for women in larger cities as the formality rate increases according to the agglomeration level, since extra-large MAs have 68% of women in formal jobs versus 57% in small MAs.

Table 4.1: Descriptive analysis: Workers' profile by Gender

	(a) Non-MA		(b) MA		(c) Small MA		(d) Medium MA		(e) Large MA		(f) Extra-large MA	
	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women
Avg/Share	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Age	36.1	35.6	36.0	36.2	35.3	35.6	35.9	35.9	35.9	36.1	36.3	36.5
Schooling level (years)	9.1	10.5	10.8	11.6	10.2	11.2	10.3	11.3	10.6	11.3	11.3	11.9
Race (white)	46.0%	51.0%	42.0%	45.0%	25.0%	27.0%	35.0%	39.0%	39.0%	41.0%	50.0%	53.0%
Marital status (married)	63.0%	57.0%	58.0%	51.0%	60.0%	51.0%	61.0%	52.0%	59.0%	51.0%	57.0%	50.0%
Household head	58.0%	32.0%	52.0%	36.0%	51.0%	40.0%	53.0%	37.0%	52.0%	37.0%	52.0%	35.0%
With child	44.0%	43.0%	39.0%	38.0%	43.0%	42.0%	41.0%	39.0%	40.0%	39.0%	38.0%	37.0%
Child 0-6	0.29	0.24	0.25	0.21	0.30	0.24	0.27	0.22	0.26	0.21	0.24	0.20
Child 7-14	0.40	0.39	0.33	0.33	0.38	0.39	0.34	0.34	0.33	0.34	0.31	0.32
Household income*	1148.52	1866.83	1673.79	2316.15	1425.8	1922.86	1317.22	1926.57	1577.75	2141.99	1878.89	2621.11
Family size	3.63	3.56	3.55	3.51	3.82	3.76	3.60	3.55	3.58	3.53	3.47	3.44
Head/spouse employed	52.0%	65.0%	52.0%	60.0%	52.0%	59.0%	50.0%	59.0%	53.0%	60.0%	52.0%	60.0%
Formal worker	58.0%	58.0%	68.0%	65.0%	59.0%	57.0%	66.0%	62.0%	67.0%	63.0%	72.0%	68.0%
Weekly worked hours	43.3	37.6	43.6	39.0	42.3	37.8	42.8	37.9	43.3	38.7	44.3	39.7
Average hourly wage*	11.1	9.4	14.0	12.0	11.8	9.7	12.3	10.6	13.1	11.2	15.5	13.5
Tenure (months)	84.1	61.9	71.3	58.3	67.0	54.2	70.6	57.7	68.9	56.6	74.3	60.7
N	294,383	162,936	150,682	120,642	20,844	15,859	26,405	20,542	68,091	55,464	35,342	28,777
Share of workers	64.4%	35.6%	55.5%	44.5%	56.8%	43.2%	56.2%	43.8%	55.1%	44.9%	55.1%	44.9%

Source: Elaborated by the authors based on PNADC (IBGE, 2020) from 2012 to 2019, with individual sample weights.
Notes: * Values in R$. We only consider data from the first interview for individuals in the sample.

4.2 Labor Market Participation

The first step of the estimation to identify female UWP is to investigate the existence of sample selection bias. For this, we apply the Heckman (1979) procedure in two stages, as detailed in section 3.2. We estimate the probability of being employed for the sample as a whole, and separately for men and women, to identify whether individual or household characteristics affect genders differently. Table 4.2 shows the results for each group. For the individual characteristics, we identify that age, household head status, and schooling level above one year of schooling (the omitted category) positively influence the participation of men and women. However, the opposite occurs for women in terms of racial and married status, with a negative effect on female labor market participation.

Although the presence of children increases the probability of being employed for both men and women, a deeper analysis of the number of children per age group reveals a different scenario between genders. Female participation is negatively influenced by the presence of children, independently of the age group, and even for a working-age son or daughter. This result can be explained by the intensity of female dedication to childcare activities. In addition, a recent piece of the literature shows that mothers and working-age children living in the same household can be substitutes in the labor market. On the other hand, having a married son or daughter at home has a positive influence on women's participation and a negative impact on men.

The size of the family negatively influences the participation of both genders, as well as having a grandparent in the household, although the last one is not statistical significant for women . This last result is particularly interesting because the direction of the coefficient for the presence of a grandparent on the participation of other members of the household in the labor market is not well investigated and can be controversial. Such influence can be positive if the grandparent supports domestic activities and is involved in the childcare needs of the grandchildren, mainly for women. However, it can have a negative influence if this same grandparent demands special care, preventing the man or woman from having a paid occupation, or if the grandparent provides additional income that adds to the family budget, therefore reducing the labor supply of other members of the household.

Still, having a domestic employee at home positively affects the participation of men and women. Additionally, male participation is positively related to having an employed spouse, but it is not significant for women. We see similar impacts between genders in relation to household wages. The wages of the spouse, son/daughter and other members positively affect individual participation for both gender, but remain positive for the full sample only for son/daughter wages. These results indicate the existence of differences in the effects of individual and household characteristics on the participation of men and women in the labor market.

Table 4.2: Labor Market Participation: Probabilities by gender

Dep.Var.: Be employed	Women (1)	Men (2)	Full Sample (3)
Age (ln)	5.402***	5.352***	4.468***
	(0.0275)	(0.0280)	(0.0186)
Age^2 (ln)	-2.205***	-2.205***	-1.798***
	(0.0116)	(0.0119)	(0.00791)
Race	-0.00385*	0.0163***	0.0518***
	(0.00222)	(0.00258)	(0.00155)
Marital status	-0.239***	0.430***	0.297***
	(0.00408)	(0.00374)	(0.00219)
Household head	0.243***	0.242***	0.527***
	(0.00266)	(0.00302)	(0.00165)
Incomplete Elementary	0.413***	0.453***	0.368***
	(0.00580)	(0.00478)	(0.00356)
Elementary school	0.581***	0.576***	0.502***
	(0.00606)	(0.00528)	(0.00378)
High school	0.762***	0.619***	0.628***
	(0.00586)	(0.00505)	(0.00366)
College or higher	1.036***	0.815***	1.035***
	(0.00652)	(0.00664)	(0.00443)
With child	0.0989***	0.153***	0.0172***
	(0.00433)	(0.00542)	(0.00269)
Child 0-6 (ln)	-0.312***	0.0803***	-0.171***
	(0.00478)	(0.00581)	(0.00274)
Child 7-14 (ln)	-0.0268***	-0.0162***	-0.0478***
	(0.00466)	(0.00563)	(0.00265)
Family size (ln)	-0.333***	-0.218***	-0.0322***
	(0.00338)	(0.00338)	(0.00185)
Head/spouse employed	-0.000486	0.113***	0.267***
	(0.00492)	(0.00404)	(0.00224)
Working-age son/daughter	-0.0179***	0.0587***	-0.104***
	(0.00374)	(0.00441)	(0.00218)
Married son/daughter	0.0442***	-0.0213**	0.0582***
	(0.00815)	(0.0107)	(0.00505)

– Continued on next page –

Table 4.2 — continued from previous page

Dep.Var.: Be employed	Women (1)	Men (2)	Full Sample (3)
Spouse wage (ln)	0.0269***	0.00337***	-0.0642***
	(0.000867)	(0.000741)	(0.000382)
Son/daughter wage (ln)	0.0306***	0.0267***	0.0138***
	(0.000419)	(0.000463)	(0.000241)
Other members wage (ln)	0.0183***	0.00407***	-0.0145***
	(0.000619)	(0.000588)	(0.000310)
Grandparents	-0.0421	-0.132***	-0.104***
	(0.0296)	(0.0381)	(0.0174)
Domestic worker	0.885***	0.0817**	0.293***
	(0.0276)	(0.0399)	(0.0169)
Year/quarter dummies	Yes	Yes	Yes
Macro-region dummies	Yes	Yes	Yes
Observations	3,275,058	3,238,093	6,513,151

Notes: Estimates for the PNADC population between 18 and 55 years old, employed and non-employed from 2012 to 2019. All models include a constant term, robust errors and individual sample weights. Omitted categories: non-MA, less than one year of schooling, the year 2012, 1st quarter, southeast region. Significance levels are: *** p <0.01, ** p <0.05, * p <0.1.

4.3 Urban Wage Premium

We describe the results of the selection equation in a previous section, analyzing the probability of being employed. This procedure allows us to estimate the IMR, which is applied to the wage equation. We test the existence of the selection bias by applying a t test to the coefficient associated with the IMR ($\widehat{\lambda}$) under the null hypothesis of no selection bias ($H_0 : \gamma = 0$). For the three groups (women, men, and full sample), this test shows that such a hypothesis should be rejected, indicating the existence of bias and the need for correction[23]. Table 4.3 shows the results after including the Heckman correction. The coefficient associated with MAs and which represents the UWP magnitude (2) suffers a reduction in the three groups, but as expected, the reduction after sample selection correction is higher for women, from 22.5% to 15.4%, while it is 21.8% to 20.2% for men.

Considering selection bias correction, we follow with the inclusion of control variables for time and macro-regions (3), firm and occupation (4), individual characteristics (5) and finally, the variables concerning children (6), thus completing the specification of Equation 2 by the POLS method, our base model. For the full sample, the UWP is 7.94% and we observe relevant differences between genders. The female UWP reaches 11.3%, while the male UWP is about half of that, reaching 5.76%. This result is in line with the theory that suggests that women benefit more from the agglomeration effects of MAs than men, given their greater diversity and the broader labor market, which facilitates and provides better matching between firms and workers, especially for women who concentrate their job search near home.

The difference in favor of the female UWP in such magnitude contrasts with the results found in the few existing empirical studies that report the closest UWP by gender, although still, a higher female UWP. This is the case in a study by Duranton (2016), which estimates that Columbia had a 6.3% UWP for women and a 4.9% UWP for men, between 1996 and 2012[24], and it is also the case in a study by Phimister (2005), which found that the United Kingdom had a UWP of 6.4% for women and a UWP of 3.8% for men, from 1991-1998.

Given the heterogeneity of Brazilian MAs, Equation 2 was also estimated, considering the agglomeration levels—the four groups of MAs—and Table 4.4 shows the results of the three sample groups. The female UWP ranges between 10.3% and 13.7%, and is higher in the two extremes, in small and extra-large MAs. Men have an UWP that varies between 4.66% and 11.1%, decreasing as the area becomes denser. The main results remain, with the female UWP surpassing that of men in all agglomeration levels.

[23] The test is applied to the coefficient associated with the IMR corresponding to column (2) of Table 4.3. For the group of women with 1,468,775 observations, the t test is equal to 3.4e+03, for the group of men with 2,296,206 observations, the $t =$ test is equal to 2.5e+03 and, for the complete sample with 3,764,982 observations, the $t =$ test is equal to 4.7e+03.

[24] The higher UWP for women in Duranton (2016) is very sensitive to different specifications, which is another contrast compared to our results.

Table 4.3: POLS Regressions: MAs

Dep.Var.: lnhwage	(1)	(2)	(3)	(4)	(5)	(6)
(a) Women						
MA	0.225***	0.154***	0.165***	0.111***	0.113***	0.113***
	(0.0155)	(0.00259)	(0.00190)	(0.00113)	(0.00109)	(0.00109)
Observations	1,468,775	1,468,775	1,468,775	1,468,775	1,468,775	1,468,775
R-squared	0.034	0.215	0.234	0.393	0.433	0.435
(b) Men						
MA	0.218***	0.202***	0.193***	0.0695***	0.0576***	0.0576***
	(0.00119)	(0.00110)	(0.00108)	(0.000980)	(0.000942)	(0.000942)
Observations	2,296,206	2,296,206	2,296,206	2,296,206	2,296,206	2,296,206
R-squared	0.028	0.157	0.219	0.418	0.459	0.459
(c) Full Sample						
MA	0.207***	0.181***	0.182***	0.0874***	0.0798***	0.0794***
	(0.000901)	(0.000817)	(0.000810)	(0.000743)	(0.000712)	(0.000712)
Observations	3,764,982	3,764,981	3,764,981	3,764,981	3,764,981	3,764,981
R-squared	0.027	0.181	0.222	0.396	0.445	0.446
Household children	No	No	No	No	No	**Yes**
Worker controls	No	No	No	No	**Yes**	**Yes**
Industry dummies (6)	No	No	No	**Yes**	**Yes**	**Yes**
Firm size dummies (4)	No	No	No	**Yes**	**Yes**	**Yes**
Occupation controls (10)	No	No	No	**Yes**	**Yes**	**Yes**
Year/quarter dummies	No	No	**Yes**	**Yes**	**Yes**	**Yes**
Macro-region dummies	No	No	**Yes**	**Yes**	**Yes**	**Yes**
Heckman's Correction	No	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**

Notes: All models include a constant term, robust errors and individual sample weights. Omitted categories: non-MA, less than one year of schooling, the year 2012, 1st quarter, southeast region, agricultural industry, firm size from 1 to 5 employees, and occupations in the service sector and salespeople. Controls and errors are omitted due to space restrictions and are available upon request. Significance levels are: *** p <0.01, ** p <0.05, * p <0.1.

Table 4.4: POLS Regression: Agglomeration levels

	Women	Men	Full Sample
	Base Model	Base Model	Base Model
Dep.Var.: lnhwage	(1)	(2)	(3)
Small MA	0.137***	0.111***	0.121***
	(0.00216)	(0.00185)	(0.00141)
Medium MA	0.111***	0.0664***	0.0838***
	(0.00184)	(0.00153)	(0.00117)
Large MA	0.103***	0.0590***	0.0767***
	(0.00125)	(0.00110)	(0.000825)
Extra-large MA	0.121***	0.0466***	0.0761***
	(0.00182)	(0.00160)	(0.00120)
Observations	1,468,775	2,296,206	3,764,981
R-squared	0.435	0.459	0.446

Notes: All models include a constant term, robust errors and individual sample weights. Omitted categories: non-MA, less than one year of schooling, the year 2012, 1st quarter, southeast region, agricultural industry, firm size from 1 to 5 employees, and occupations in the service sector and salespeople. Controls and errors are omitted due to space restrictions and are available upon request. Significance levels are: *** p <0.01, ** p <0.05, * p <0.1.

4.4 UWP Quantile Analysis

POLS estimation allow us to identify the average level of the UWP of women, men, and for the full sample. However, the estimation of quantile regressions can indicate whether the UWP has a different magnitude throughout the wage distribution, and in the agglomeration levels. Thus, we estimate quantile regressions focusing on the first interview[25], in a cross-section approach for separate groups of men and women. We estimate the coefficients of seven quantiles (5th, 10th, 25th, 50th, 75th, 90th, and 95th), considering robust errors and individual sample weights.

Quantile regressions are first estimated with a dummy for MA and then for the agglomeration levels. Figure 4.1 presents the UWP coefficients on graphs to facilitate results visualization[26], for which we have two different UWP estimates for each gender: (i) the result obtained by POLS for the base model—which remains fixed over the quantiles (BM)—and, (ii) the quantile regression result (Q).

The result of the quantile regression by gender is totally different (panel (a)). For men, the UWP at the beginning of the wage distribution is close to the regression by POLS, but it becomes lower after the 10th quantile. For women, the opposite occurs, and the UWP is lower at the beginning of the distribution, starting close to the male UWP, but it increases and exceeds 12% in the groups with higher salaries—- the 95th quantile. In both cases, the UWP obtained by POLS overcomes the quantile UWP for most of the wage distribution, the exception occurs for women around 75%, which have a lower UWP under POLS.

Although there is a non-negligible female UWP at any point of wage distribution—from 6.92% to 12.1%—the higher UWP at top salaries means that more productive occupations (which pay higher salaries) also provide higher premiums in MAs for women. Meanwhile, the male UWP varies less—from 4.52% to 6.17%, being similar for occupations at the bottom of the wage distribution. This scenario corroborates with both motivations for UWP differentials by gender: the agglomeration effects benefit women differently, and their occupation in the labor market could generate different premiums.

In turn, graphs (b) to (e) portray the other four graphs, showing the UWP for each MA level and the coefficient associated with each quantile by gender. Again, we compare the results of quantile regressions with those obtained by POLS. In small MAs (b), the UWP for men and women is similar, as well as the behavior throughout the distribution. However, men have a higher UWP at the bottom 10% wages and are surpassed by the female UWP in the rest of the distribution. Both genders present a lower quantile UWP compared to POLS from the top 50% of higher wages. Thus, it seems that in small MAs, the position in the wage distribution matters more than the gender as a determinant of a higher/lower UWP.

Denser areas, besides small MAs, show totally different results by gender, as the gap between their UWP increases with wage distribution. Medium (c) and large MAs (d) show a similar scenario: both start with an UWP at the same level—around 8% in medium and 6% in large MAs—but it increases throughout the distribution for women and decreases for men. In extra-large MAs, men show a constant UWP, with the result for each quantile practically equal to the average UWP by POLS regressions. Women, on the other hand, show higher premiums than men throughout the entire distribution.

An interesting duality of quantile results that is not observable with POLS estimations is the fact that occupations with lower productive levels benefit men more in less dense areas—small MAs—while the opposite occurs for women, with high-productive occupations providing higher premiums in denser areas—extra-large MAs. Another fact not observable with POLS estimations is that the female UWP could be smaller than male depending on the position in the wage distribution. In medium and large MAs, at least until the 50th quantile, the average UWP by POLS overcomes quantile estimation, and the same occurs in extra-large MAs below 25th quantile, indicating lower premiums than the average for less-productive occupations.

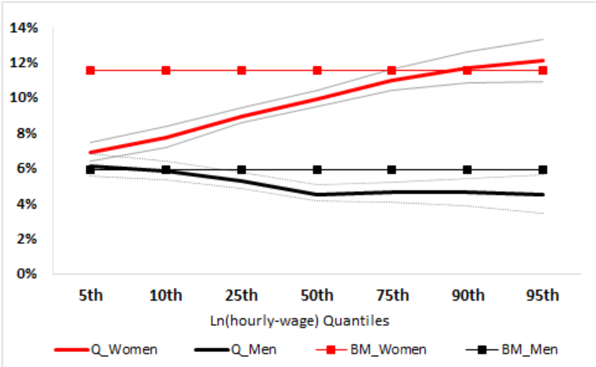
Again, we observe an UWP differential by gender, mainly in denser MAs—medium, large, and extra-large—in line with average UWP by POLS estimations, presenting higher premiums for women. Besides, we also observe in these MAs, a higher UWP at the top of the wage distribution, meaning that occupation in the labor market also matters in the female UWP magnitude. It seems that denser MAs do provide an environment where women in high-productive occupations can benefit more than when they are in low-productive ones.

[25] The option to use only the first interview of each individual for quantile regressions maximizes the number of individuals, and also allows for a higher rate of assertiveness of the responses with the engagement of research participation. In section 4.5, the first robustness test showed that POLS estimation results using only the first observation of each individual are similar to those obtained for the primary sample.

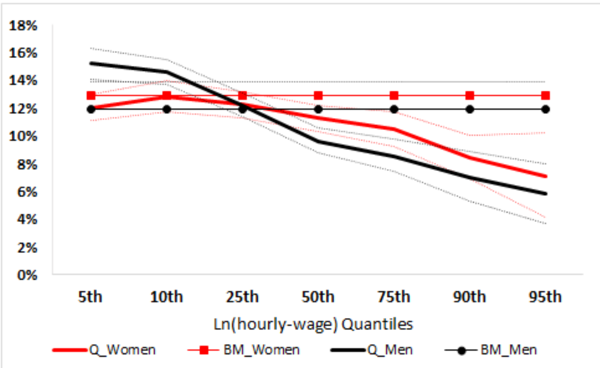
[26] The complete estimation result is available upon request.

Figure 4.1: UWP Quantile Regressions by Agglomeration Levels

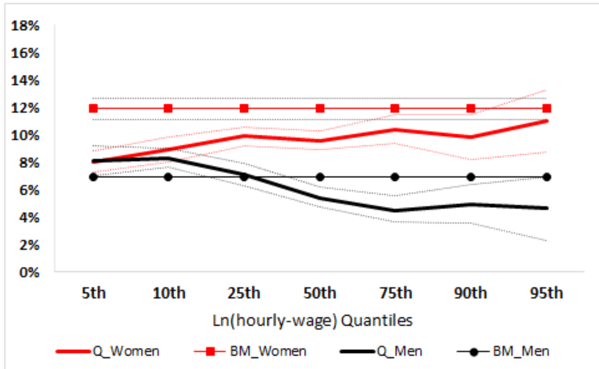
(a) MAs



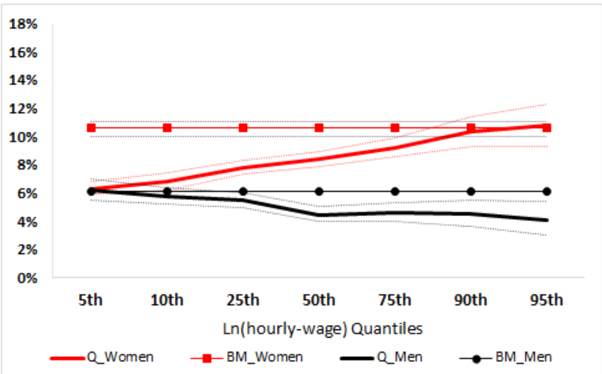
(b) Small MAs



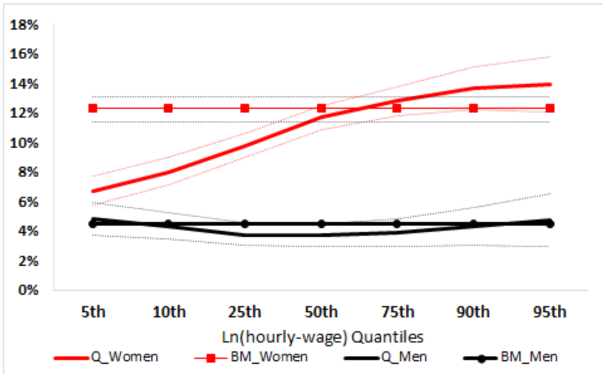
(c) Medium MAs



(d) Large MAs



(e) Extra-large MAs



Source: Elaborated by the authors based on the estimation. The complete results is available upon request.
Notes: Vertical axis denotes the UWP magnitude. Gray lines indicate a 95% confidence interval. Base Model considering only the first interview for each individual in the sample.

4.5 Robustness tests

4.5.1 Different samples

The sample construction of this paper is based on some selections from the PNADC database. Thus, we apply some tests to verify the robustness of the results in the context of a less restricted set of selections. The sample of women and men used in POLS estimates is an unbalanced panel, which includes only individuals with at least two interviews, among the five possible. The first robustness check consists of an estimation of the first observation of each individual. This test intends to analyze whether individuals who were interviewed only once—and, consequently, are out of the sample—have a different UWP in comparison to the primary sample.

The second test, in turn, considers the individuals who answered all five interviews. In this case, the tested hypothesis examines if different UWP results for men and women would be found with a balanced panel. A third robustness check is the estimation of the UWP considering individuals employed in the public sector, armed forces, and in the military—groups that we exclude from the primary sample[27]. Although ruled by specific labor laws, the motivation for this test is due to an attempt to verify whether the significant differences in the UWP between genders is related to the exclusion of those occupations that have a male predominance among its employees and a higher relevance in non-MAs. For each of the three tests, we perform the same selection bias correction procedures and estimate the UWP with the same base model specification. Table 4.5 shows the results of these tests where we observe very similar coefficients in comparison to the base model, the same pattern of agglomeration levels, and the highest values of the female UWP.

Table 4.5: Robustness: POLS with different samples

	Women				Men			
	Base Model (1)	1st interview (2)	Balanced Panel (3)	With Public Sector[a] (4)	Base Model (5)	1st interview (6)	Balanced Panel (7)	With Public Sector[a] (8)
Small MA	0.140***	0.131***	0.146***	0.181***	0.113***	0.119***	0.126***	0.150***
	(0.00316)	(0.00520)	(0.00428)	(0.00324)	(0.00284)	(0.00446)	(0.00388)	(0.00283)
Medium MA	0.113***	0.121***	0.112***	0.130***	0.0669***	0.0713***	0.0683***	0.0945***
	(0.00260)	(0.00437)	(0.00363)	(0.00269)	(0.00232)	(0.00372)	(0.00328)	(0.00238)
Large MA	0.105***	0.108***	0.106***	0.138***	0.0595***	0.0614***	0.0635***	0.0895***
	(0.00186)	(0.00298)	(0.00248)	(0.00190)	(0.00170)	(0.00260)	(0.00229)	(0.00171)
Extra-large MA	0.120***	0.123***	0.119***	0.143***	0.0465***	0.0463***	0.0487***	0.0596***
	(0.00291)	(0.00426)	(0.00384)	(0.00303)	(0.00263)	(0.00374)	(0.00345)	(0.00266)
Observations	1,404,056	269,563	888,658	1,870,589	2,199,215	423,711	1,394,276	2,494,106
R-squared	0.422	0.409	0.418	0.491	0.460	0.452	0.462	0.493

Notes: Dependent variable = $lnhwage$. All models include a constant term, robust errors and individual sample weights. All models follow the base model specification with controls for household children, worker's characteristics, industry, firm size, occupation, year or quarter, macro-region, and IMR. Omitted categories: non-MA, less than one year of schooling, the year 2012, 1st quarter, southeast region, agricultural industry, firm size from 1 to 5 employees, and occupations in the service sector and salespeople. Controls and errors are omitted due to space restrictions and are available upon request. [a] Disregarding Firm Size variable as public sector workers did not report it. Significance levels are: *** p <0.01, ** p <0.05, * p <0.1.

4.5.2 Aggregated UWP by gender

We now evaluate the UWP by interacting with the agglomeration level the gender identification for the full sample. This exercise replicate the decreasing pattern for the UWP according to the agglomeration levels, as already showed in previous results, ranging between 4.87% to 10.7%. However, when the coefficients obtained from the interactions with each MA level are taking to account the UWP pattern change, with a higher magnitude at the extreme (small and extra-large MAs) and more equal between MAs, raging from 9.9% to 14.1%[28].

Looking at the gender return on wages we see that women present a wage penalty of 18.3%. Taking to account again the coefficients obtained from the interactions with each MA level, this penalty can be lower depending on the area. In small MAs the wage penalty is reduced to -14.9% (-18.3% plus 3.37%), whereas in extra-large MAs it reaches only -12.1% (-18.3% plus 6.19%). A more deep investigation is needed, but it is possible that denser areas provide a less unequal environment for women, with a higher wage premium as a mechanism to reduce the wage gap.

4.6 Heterogeneity

4.6.1 UWP by specific groups

Deepening the analysis of the UWP, we run several heterogeneity exercises with estimates for different sub-samples of men and women based on specific characteristics related to the household position, marital status, and

[27]For this exercise we disregard firm size variable. We also include, statutory and auxiliary family workers.

[28]Given by the some of the UWP coefficient and the interaction term for each MA, as for example 14.1% for small MAs (10.7% plus 3.37%).

Table 4.6: Aggregated UWP by gender

Dep.Var. lnhwage	Coeff. (1)	SE (2)
Small MA	0.107***	(0.00182)
Medium MA	0.0664***	(0.00152)
Large MA	0.0598***	(0.00109)
Extra-large MA	0.0484***	(0.00152)
Gender (Female)	-0.183***	(0.000970)
Interactions		
Small MA x Gender	0.0337***	(0.00266)
Medium MA x Gender	0.0406***	(0.00230)
Large MA x Gender	0.0389***	(0.00161)
Extra-large MA x Gender	0.0619***	(0.00213)
Observations	3,764,981	
R-squared	0.446	

Notes: Include a constant term, robust errors and individual sample weights. Following the base model specification with controls for household children, worker's characteristics, industry, firm size, occupation, year or quarter, macro-region, and IMR. Omitted categories: non-MA, less than one year of schooling, the year 2012, 1st quarter, southeast region, agricultural industry, firm size from 1 to 5 employees, and occupations in the service sector and salespeople. Controls and errors are omitted due to space restrictions and are available upon request. Significance levels are: *** p <0.01, ** p <0.05, * p <0.1.

the presence of children. The complete results for this tests are available upon request. As Section 4.2 depicted, the influence of individual and household characteristics on labor market participation is different between genders, so the purpose of these exercises is to verify whether they may also indicate different UWPs and if the largest female UWP presented remains in different sub-samples.

The results show that the female UWP is higher than male UWP in all ten subgroups analyzed, and, with rare exceptions, is below 10%. At extra-large MAs, the female UWP far surpasses the male UWP, especially in samples related to the presence of children. The closest UWP level between genders occurs in small MAs and among married people, where the difference is only 1.7 percentage points, with a 12.7% UWP for women and an 11.0% UWP for men.

These exercises also bring other interesting facts, such as the differences within each gender group. Single women have a higher UWP than married women at all agglomeration levels, while between men, this scenario is the opposite. The female UWP is also always higher for those with older children compared to those with young children, which is not the behavior presented in the male group. Women without children have a higher UWP than those with children only at large and extra-large MAs, while the same scenario occurs for men in medium to extra-large MAs. Being married or not does not mean a big difference in the UWP magnitude for women in extra-large MAs if they already have children—10.8 versus 10.2%—but this difference is higher in lower-density areas, reaching 2 percentage points.

4.6.2 Formality status

The existence of a large informal sector is a factor to be considered when analyzing wages in the urban areas of developing countries. In Brazil, being an informal worker means not having access to benefits provided by law, such as unemployment insurance, an additional salary in December, paid leave due to illness, maternity leave, and accidents at work, in addition to not having access to public retirement benefits. Therefore, the formality status in the Brazilian labor market is not a choice for many unskilled workers and can have different impacts on men and women.

We perform a heterogeneity exercise based on the formality status of occupations. The results show that the UWP is higher for informal workers, both men, and women, except in extra-large MAs, where formal jobs have the highest UWP for men (Table 4.7). We identify different patterns for men and women. Although the overall results for women show a higher UWP at both extreme MA levels (small and extra-large MAs), the formal female UWP increases according to agglomeration levels—from 6.01% to 11.9%, whereas the informal UWP decreases according to scale—from 20.6% to 12.8%. We observe the same pattern for men by formality status, but the overall result for them follows the decreasing behavior of the informal UWP. We now observe that including informal workers influences the female UWP. Disregarding female informal workers leads to an underestimated UWP and an increasing pattern according to agglomeration levels.

Table 4.7: POLS Regressions: Formality Status

	(a) Women			(b) Men			(c) Full Sample		
	Formal	Informal	Both	Formal	Informal	Both	Formal	Informal	Both
Var.Dep.: lnhwage	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Small MA	0.0643***	0.206***	0.137***	0.0488***	0.168***	0.111***	0.0530***	0.182***	0.121***
	(0.00238)	(0.00379)	(0.00216)	(0.00224)	(0.00310)	(0.00185)	(0.00167)	(0.00243)	(0.00141)
Medium MA	0.0601***	0.167***	0.111***	0.0268***	0.114***	0.0664***	0.0400***	0.136***	0.0838***
	(0.00196)	(0.00341)	(0.00184)	(0.00177)	(0.00291)	(0.00153)	(0.00134)	(0.00222)	(0.00117)
Large MA	0.0656***	0.149***	0.103***	0.0330***	0.0873***	0.0590***	0.0460***	0.111***	0.0767***
	(0.00135)	(0.00236)	(0.00125)	(0.00126)	(0.00207)	(0.00110)	(0.000941)	(0.00157)	(0.000825)
Extra-large MA	0.119***	0.128***	0.121***	0.0543***	0.0367***	0.0466***	0.0834***	0.0776***	0.0761***
	(0.00196)	(0.00384)	(0.00182)	(0.00182)	(0.00323)	(0.00160)	(0.00135)	(0.00250)	(0.00120)
	(0.0247)	(0.0443)	(0.0227)	(0.0215)	(0.0290)	(0.0172)	(0.0164)	(0.0228)	(0.0132)
Observations	841,439	627,336	1,468,775	1,301,839	994,367	2,296,206	2,143,276	1,621,705	3,764,981
R-squared	0.469	0.324	0.435	0.435	0.371	0.459	0.438	0.334	0.446

Notes: All models include a constant term, robust errors and individual sample weights. All models follow the base model specification with controls for household children, worker's characteristics, industry, firm size, occupation, year or quarter, macro-region, and IMR. Omitted categories: non-MA, less than one year of schooling, the year 2012, 1st quarter, southeast region, agricultural industry, firm size from 1 to 5 employees, and occupations in the service sector and salespeople. Controls and errors are omitted due to space restrictions and are available upon request. Significance levels are: *** p <0.01, ** p <0.05, * p <0.1.

5 Final Remarks

This paper sought to evaluate the urban wage premium of women in Brazilian MAs for various subgroups of individuals, in terms of occupational, firm, and household characteristics, using different estimation methods. The use of the PNADC allows us to correct the sample selection bias of the probability of labor market participation, while ensuring national coverage and representativeness. The UWP study in this paper seeks to understand whether: (i) the economies of agglomeration in dense areas benefit men and women differently, and (ii) whether the allocation or composition of the labor market influences the existence of a wage premium of different magnitude between genders.

The strategy applied comprised a deep analysis separately for men and women, being able to compare them, taking advantage of the rich database, and exploring the differentials between household composition and formality status. Additionally, we also used a quantiles regressions approach to assess the heterogeneity across the wage distribution. The main result of the paper is that women have a higher UWP than men, and this result is robust to several specifications and estimation methods. For the POLS estimation, women have a 11.3% UWP, whereas men have only a 5.76% UWP. This finding holds even in some heterogeneity exercises. Quantile regressions reveal that the UWP has a different magnitude throughout the wage distribution and that it has different trajectories between men and women, mainly in medium, large, and extra-large MAs.

Additionally, the results show different determinants for the participation of men and women in the labor market. The participation of women is negatively affected by their marital status and the presence of children, irrespective of whether these children are young or of working age. Men, on the other hand, participate less in the labor market if they have a child between 7 and 14 years old, or if there is a married child at home.

The results of this paper are relevant to the study of the UWP in Brazil, including women as the interest group of analysis, and contribute to the fledgling female UWP literature by highlighting the differences between genders, using different estimation methods and tests. The main conclusion indicate that the previous UWP results are underestimated since female workers are neglected in most of the traditional UWP literature, and they could be over or underestimated for women and men if wage distribution is disregarded.

We find strong evidence that agglomeration effects do not benefit workers equally, and a higher female UWP can also be attributable to the changes in female labor market composition. It seems that denser areas tend to be more favorable to women, as they have a higher share of women in high-productive occupations. This scenario is indeed consistent across different subgroups, which indicates that the agglomeration effects have overcome possible constraints, pointed out by the literature, as women low spatial mobility.

The results also indicate a relevant perspective in the gender wage differentials discussion. First, denser areas appear to enjoy a greater awareness of the gender gap; therefore, a higher female UWP can contribute to closing it. However, the higher female UWP can increase the inequality between women from MAs and those from non-MAs, harming their group as a whole. From a policy point of view, this represents a challenge, demanding different and coordinated actions across agglomeration levels. This is particularly important for non-MAs, as their lack of agglomeration effects leads to worse opportunities in the labor market, compared to MAs, especially for women.

The mechanisms behind the UWP differentials by gender should be the focus of future research. The analysis of the effects of local infrastructure—such as the higher availability of housekeepers and childcare facilities in denser areas—on UWP, could be very important as they provide better conditions for the improvement of the careers of women. Similarly, weddings, cultural issues, intra-household bargaining, and knowledge externalities are factors that can influence not only an individual's behavior in the labor market, but also have different degrees of relevance across agglomeration levels.

The remaining issues involve exploring the features associated with women's participation in the labor market with the application of multinomial estimates, in addition to investigating whether, besides the UWP, the wage returns to worker characteristics are different throughout the wage distribution by gender. Additionally, as in the first paper, other possibilities for further research involves (i) the role Formal job benefits and how they could influence the higher Informal UWP, (ii) a proper FE estimation, and (iii) the firm size effects in denser areas, as we only account for big firms with fifty or more employees as a group.

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. 2, 7

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. 2

BADAOUI, E. E.; STROBL, E.; WALSH, F. The formal sector wage premium and firm size. *Journal of Development Economics*, Elsevier, v. 91, n. 1, p. 37–47, 2010. 7

BARUFI, A. M. B. *Agglomeration economies and labour markets in Brazil*. Tese (Doutorado) — Universidade de São Paulo, 2015. 2, 8, 10

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. 8, 10

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. 2

BUCHINSKY, M. Quantile regression with sample selection: Estimating women's return to education in the US. In: *Economic applications of quantile regression*. [S.l.]: Springer, 2002. p. 87–113. 10

CAMERON, A. C.; TRIVEDI, P. K. *Microeconometrics: methods and applications*. [S.l.]: Cambridge University Press, 2005. 10

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. 2, 8, 10

COELHO, D.; VESZTEG, R.; SOARES, F. V. Regressão quantílica com correção para a seletividade amostral: estimativa dos retornos educacionais e diferenciais raciais na distribuição de salários das mulheres no Brasil. Instituto de Pesquisa Econômica Aplicada (Ipea), 2010. 10

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. 2

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, v. 126, n. 24, p. 91–114, 2012. 4, 8

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. 9

DAUTH, W. et al. Spatial wage disparities: Workers, firms, and assortative matching. 2016. 7

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. 2, 10

DIEESE. *Pesquisa Nacional da Cesta Básica de Alimentos*. 2019. Disponível em: <https://www.dieese.org.br/cesta/>. 8

DURANTON, G. Agglomeration effects in Colombia. *Journal of Regional Science*, Wiley Online Library, v. 56, n. 2, p. 210–238, 2016. 2, 3, 13

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. 2

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. 2

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. 2, 10

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. 2

HECKMAN, J. J. Sample selection bias as a specification error. *Econometrica*, v. 47, n. 1, p. 153–161, 1979. 9, 11

IBGE. *Nota Técnica. Deflacionamento dos rendimentos do trabalho dos trimestres móveis da PNAD-Contínua*. Instituto Brasileiro de Geografia e Estatítica, 2018. Disponível em: <http://www.ibge.gov.br>. 8

IBGE. *Organização do território: Divisão Territorial Brasileira (DTB)*. IBGE Rio de Janeiro, 2018. Disponível em: <http://www.ibge.gov.br>. 6, 8

IBGE. *Malhas municipais*. IBGE Rio de Janeiro, 2019. Disponível em: <http://www.ibge.gov.br>. 6, 8

IBGE. *PNADC. Notas Técnicas.* Instituto Brasileiro de Geografia e Estatítica, 2019. Disponível em: <http://www.ibge.gov.br>. 5

IBGE. *Pesquisa Nacional por Amostra de Domicilios Contínua (PNADC) 2012-2019*. 2020. Disponível em: <http://www.ibge.gov.br>. 5, 6, 7, 8, 11

JONES, P.; D'AOUST, O.; BERNARD, L. The urban wage premium in Africa. In: *Wage Inequality in Africa*. [S.l.]: Springer, 2017. p. 33–53. 3

KOENKER, R.; JR, G. B. Regression quantiles. *Econometrica: Journal of the Econometric Society*, JSTOR, p. 33–50, 1978. 10

KRUG, G.; NISIC, N. Is there an urban wage premium for women? A difference-in-difference analysis using Propensity Score Matching. *LASER Discussion Paper - n.54*, 2011. 3

MACHADO, J. A.; SILVA, J. S. Quantiles via moments. *Journal of Econometrics*, Elsevier, v. 213, n. 1, p. 145–173, 2019. 10

MADDEN, J. F.; CHIU, L.-i. C. The wage effects of residential location and commuting constraints on employed married women. *Urban Studies*, Publications Sage UK: London, England, v. 27, n. 3, p. 353–369, 1990. 3

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. 3

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. 2, 8

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. 3

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. 2

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. 2

MORETTI, E. Local labor markets. In: *Handbook of labor economics*. [S.l.]: Elsevier, 2011. v. 4, p. 1237–1313. 2

MORETTI, E. *The new geography of jobs*. [S.l.]: Houghton Mifflin Harcourt, 2012. 2

MORETTI, E. Real wage inequality. *American Economic Journal: Applied Economics*, v. 5, n. 1, p. 65–103, 2013. 2

NEVES-JR, E. C.; AZZONI, C. R.; CHAGAS, A. Skill wage premium and city size. *Working paper series n. 19 - Department of Economics*, University of São Paulo (FEA-USP), 2017. 2

NISIC, N. Smaller differences in bigger cities? Assessing the regional dimension of the gender wage gap. *European Sociological Review*, Oxford University Press, v. 33, n. 2, p. 292–3044, 2017. 3

PASSOS, L.; GUEDES, D. R. Participação feminina no mercado de trabalho e a crise de cuidados da modernidade: conexões diversas. *Planejamento e Políticas Públicas - n.50*, 2018. 2

PHIMISTER, E. Urban effects on participation and wages: are there gender differences? *Journal of Urban Economics*, Elsevier, v. 58, n. 3, p. 513–536, 2005. 2, 3, 13

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. 2

SILVA, D. L. G. *Economias de aglomeração e heterogeneidade do trabalhador e firma na determinação de salários no Brasil*. Tese (Doutorado) — Universidade de São Paulo, 2017. 10

SILVA, D. L. G. Contribuição dos efeitos da firma e de invidivíduo para os efeitos de localização sobre os salários e para a variação salarial do trabalhador formal no Brasil. *Pesquisa e Planejamento Econômico*, v. 48, n. 2, 2018. 7, 8

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. 2

WOOLDRIDGE, J. M. *Econometric analysis of cross section and panel data*. [S.l.]: MIT press, 2010. 9, 10

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. 3

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. 7, 8