

**UNIVERSIDADE FEDERAL DE ALAGOAS**  
**FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE – FEAC**  
**CURSO DE MESTRADO EM ECONOMIA APLICADA – CMEA**

**JEFFERSON DOGLAS DA SILVA PEREIRA**

**DIFERENCIAIS SALARIAIS NO BRASIL: A IMPORTÂNCIA DAS ECONOMIAS  
DE AGLOMERAÇÃO E DO *SORTING* ESPACIAL DOS TRABALHADORES E  
FIRMAS**

**MACEIÓ**

**2021**

**JEFFERSON DOGLAS DA SILVA PEREIRA**

**DIFERENCIAIS SALARIAIS NO BRASIL: A IMPORTÂNCIA DAS ECONOMIAS  
DE AGLOMERAÇÃO E DO *SORTING* ESPACIAL DOS TRABALHADORES E  
FIRMAS**

Dissertação apresentada ao Curso de Mestrado em Economia Aplicada da Faculdade de Economia, Administração e Contabilidade, da Universidade Federal de Alagoas, como requisito parcial para obtenção do título de Mestre em Economia Aplicada.

**Orientador: Prof. Dr. Keuler Hissa Teixeira**

**MACEIÓ**

**2021**

**Catálogo na fonte**  
**Universidade Federal de Alagoas**  
**Biblioteca Central**  
**Divisão de Tratamento Técnico**

Bibliotecário: Marcelino de Carvalho Freitas Neto – CRB-4 – 1767

P436d Pereira, Jefferson Douglas da Silva.

Diferenciais salariais no Brasil : a importância das economias de aglomeração e do *sorting* espacial dos trabalhadores e firmas / Jefferson Douglas da Silva Pereira. – 2021.

109 f. : il.

Orientador: Keuler Hissa Teixeira.

Dissertação (Mestrado em Economia Aplicada) – Universidade Federal de Alagoas. Faculdade de Economia, Administração e Contabilidade. Programa de Pós-Graduação em Economia. Maceió, 2021.

Bibliografia: f. 84-89.

Apêndices: f. 90-109.

1. Economias de aglomeração. 2. Salários. 3. *Sorting*. 4. Disparidades regionais - Brasil. I. Título.

CDU: 332.142.2(81)

## FOLHA DE APROVAÇÃO

JEFFERSON DOGLAS DA SILVA PEREIRA

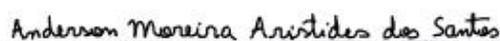
### DIFERENCIAIS SALARIAIS NO BRASIL: A IMPORTÂNCIA DAS ECONOMIAS DE AGLOMERAÇÃO E DO *SORTING* ESPACIAL DOS TRABALHADORES E FIRMAS

Dissertação aprovada como requisito parcial para obtenção do título de Mestre em Economia Aplicada no Curso de Mestrado em Economia Aplicada da Universidade Federal de Alagoas, pela seguinte banca examinadora:



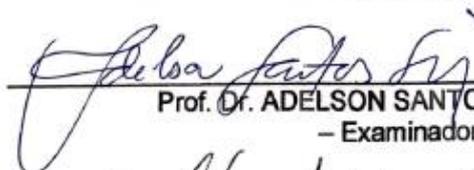
---

Prof. Dr. KEULER HISSA TEIXEIRA – (FEAC/UFAL)  
– Orientador –



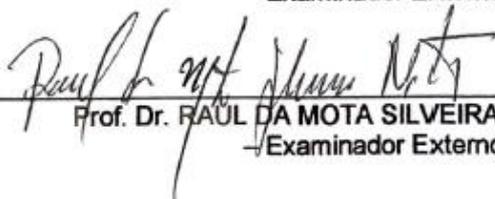
---

Prof. Dr. ANDERSON MOREIRA ARISTIDES DOS SANTOS – (FEAC/UFAL)  
– Examinador Interno –



---

Prof. Dr. ADELSON SANTOS DA SILVA – (UFRPE)  
– Examinador Externo –



---

Prof. Dr. RAUL DA MOTA SILVEIRA NETO – (UFPE)  
– Examinador Externo –

Maceió, 17 de fevereiro de 2021.

## AGRADECIMENTOS

Primeiramente, eu agradeço a Deus por todas as conquistas que tenho alcançado em minha vida. Agradeço a Deus pela minha boa saúde física e, principalmente, mental, e por ter sido o meu refúgio de todas as horas. Por tudo isso, toda honra e toda glória a ti, Senhor.

A toda a minha família por todo o apoio e confiança depositados em mim: meu pai, meus irmãos e, em especial, minha mãe, Josefa, e minhas irmãs, Josélia e Liandra. Elas foram pessoas guiadas e iluminadas por Deus que me ajudaram a cursar este mestrado, sempre ao meu lado em todos os momentos de dificuldades. Sem esse apoio, certamente a caminhada teria sido ainda mais árdua. Agradeço também ao meu cunhado por todo o seu suporte.

Ao meu orientador Keuler Teixeira por todo o apoio e suporte (emocional, inclusive) dispensados a mim e por todo o seu empenho em conseguir a base de dados.

Aos professores Adelson Santos, Anderson Santos e Raul Silveira Neto por terem aceitado o convite em compor a banca examinadora desta dissertação e por todas as sugestões e orientações dadas para o aperfeiçoamento deste trabalho.

Aos professores do Mestrado em Economia da UFAL por todo o conhecimento compartilhado, especialmente, à professora Ana Milani por ter sempre acreditado no meu potencial e ao professor Anderson Santos por toda vivência e conhecimento compartilhados.

Aos meus companheiros de mestrado por terem feito esta caminhada mais leve e também por todo suporte dado no decorrer do curso: Alexandre, Andriano, Jailson, Lucas, Moisés e Pablo. Quero registrar o quão importante foram os momentos de risos e sofrimentos compartilhados de maneira mais direta com Moisés e Jailson.

As minhas amigas Clédia, Luiza e Rosana por sempre acreditarem no meu potencial e por sempre vibrarem junto comigo. A amizade de vocês foi, por muitas vezes, a injeção de ânimo para que eu seguisse adiante. Luiza, muito obrigado por toda a ajuda com os dados.

A Bárbara por todo o apoio, por cada palavra de incentivo, por todas as alegrias compartilhadas, por todos os conselhos, por todas as conversas, enfim, por tudo. A minha querida amiga Lívia por todo o apoio e confiança em mim depositados, por todas as nossas conversas e por toda nossa cumplicidade. A Aldiana e a Mirabel por toda a convivência.

A todos os meus professores desde o ensino básico até a graduação, em especial, a Adelson Santos, meu orientador e professor de graduação e a Everlândia por toda a ajuda no tratamento da base de dados.

Por fim, eu agradeço às minhas amigas e aos meus amigos aqui não citados, mas que sei o quanto torcem por mim e o quanto acreditam no meu potencial. A todos vocês, o meu muito obrigado!

## RESUMO

O objetivo deste estudo é mensurar os efeitos das economias de aglomeração sobre os salários individuais no mercado de trabalho formal brasileiro no período de 2010 a 2018, considerando os efeitos do *sorting* espacial dos trabalhadores, das firmas e do *match* trabalhador-firma sobre os salários. A interação espacial entre os agentes econômicos no mercado de trabalho brasileiro pode resultar em aumentos de produtividade tanto das firmas quanto dos trabalhadores, elevando seus rendimentos. Dada a maior disponibilidade de microdados pareados trabalhadores-firmas, como é o caso dos dados aqui utilizados, oriundos da RAIS (2010-2018), é possível avançar no controle das heterogeneidades não observadas tanto dos trabalhadores quanto das firmas. É consenso na literatura que o *sorting* espacial tanto dos trabalhadores quanto das firmas é uma importante fonte de explicação para os diferenciais salariais. Neste sentido, esta dissertação busca controlar essas fontes de possíveis vieses na estimação dos efeitos das economias de aglomeração sobre os salários. Além disso, busca-se controlar o viés decorrente do *match* trabalhador-firma nas áreas de mercado de trabalho local. Tendo em vista as grandes disparidades regionais no Brasil, este trabalho busca analisar também como os efeitos das economias de aglomeração sobre os salários se comporta entre as regiões e os diferentes tamanhos das áreas de mercado de trabalho local, fornecendo assim evidências que podem auxiliar no planejamento estratégico de políticas públicas voltadas para o mercado de trabalho. A partir de uma estrutura de dados em painel que acompanha o trabalhador ao longo do período que vai de 2010 a 2018, o estudo utiliza modelos de dados em painel que permitem estimações por múltiplos efeitos fixos. O estudo também utiliza o método de estimação por variáveis instrumentais para controlar a possível endogeneidade entre a densidade do emprego e os salários. Os resultados encontrados no modelo principal deste estudo apontam para um efeito líquido das economias sobre os salários na ordem de 1,69%. Ademais, os resultados fornecem evidências de que há diferenças substanciais no efeito das economias de aglomeração sobre os salários de cada uma das regiões brasileiras e de cada tamanho das áreas de mercado de trabalho local. De modo geral, os resultados encontrados nesta dissertação apontam para relevância do *sorting* espacial dos trabalhadores e firmas e do *match* trabalhador-firma na determinação dos salários das áreas de mercado de trabalho local do Brasil.

Palavras-chave: Economias de aglomeração; Salários; *Sorting*; *Match* trabalhador-firma; Disparidades regionais.

## ABSTRACT

The objective of this study is to measure the effects of agglomeration economies on individual wages in the formal Brazilian labor market from 2010 to 2018, controlling the effects of the spatial sorting of workers and firms, and the match worker-firm on wages. The spatial interaction between economic agents in the Brazilian labor market can result in increases in productivity for both firms and workers, increasing their revenues. Given the greater availability of paired worker-firm microdata, as the data used in this study, extracted from the RAIS (2010-2018), it is possible to advance in the control of unobserved heterogeneities of both workers and firms. There is a consensus in the literature that the spatial sorting of workers and firms is an important source of explanation for wage differentials. In this sense, this dissertation aims to control these possible sources of bias in estimating agglomeration effects on wages. In addition, this dissertation seeks to control the bias resulting from the match worker-firm in the local labor market areas. In view of the great regional disparities in Brazil, this study also aims to analyze how the effects of agglomeration economies on wages behave among Brazilian regions and among different sizes of Labor Market Areas (LMA), thus providing evidence that can assist in the strategic planning of public policies directed to the labor market. Based on a panel data structure that follows the worker over the period from 2010 to 2018, this study uses panel data models that allow estimates for multiple fixed effects. Besides, this study used the method of estimation by instrumental variables to control the possible endogeneity between employment density and wages. The results found in the main model of this study point to a net effect of agglomeration economies on wages of around 1.69%. Furthermore, the results provide evidence that there are substantial differences in the effect of agglomeration economies on wages in each of the Brazilian regions and in each size of the local labor market areas. In general, the results found in this dissertation point to the relevance of the spatial sorting of workers and firms and of the match worker-firm in determining LMA wages in Brazil.

Keywords: Agglomeration economies; Wages; Sorting; Match worker-firm; Regional disparities.

## **LISTA DE FIGURAS**

Figura 4.1: Salário-hora-real médio conforme mobilidade (2010-2018) .....	67
Figura 4.2: Salário-hora-real e a densidade do emprego nas LMA (2010-2018).....	68
Figura 4.3: Salário-hora-real e a densidade do emprego nas LMA (2010-2018).....	69

## LISTA DE TABELAS

Tabela 3.1: Total de vínculos ativos e inativos registrados na RAIS: 2010-2018.....	34
Tabela 3.2: Descrição das etapas de tratamento da base identificada da RAI (2010-2018).....	39
Tabela 3.3: Características das Áreas de Mercado de Trabalho Local (2018).....	43
Tabela 4.1: Estatísticas descritivas das variáveis contínuas .....	61
Tabela 4.2: Rendimento médio e distribuição dos trabalhadores conforme suas características .....	62
Tabela 4.3: Rendimento médio e distribuição dos trabalhadores conforme características das firmas .....	64
Tabela 4.4: Rendimento médio e distribuição dos trabalhadores conforme características das LMA.....	65
Tabela 4.5: Média do salário-hora-real das LMA conforme seu tamanho e macrolocalização (2010-2018).....	66
Tabela 4.6: Mobilidade das firmas e dos trabalhadores entre 2010 e 2018.....	66
Tabela 4.7: Efeitos das economias de aglomeração sobre os salários individuais nas LMA (2010-2018).....	72
Tabela 4.8: Efeitos das economias de aglomeração sobre os salários individuais nas LMA (2010-2018) – VI-MQ2E.....	75
Tabela 4.9: Efeitos das economias de aglomeração sobre os salários individuais conforme tamanho das LMA (2010-2018).....	77
Tabela 4.10: Efeitos das economias de aglomeração sobre os salários individuais conforme macrolocalização das LMA (2010-2018).....	79

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>11</b>
<b>2 REVISÃO DA LITERATURA</b> .....	<b>17</b>
2.1 FUNDAMENTAÇÃO TEÓRICA DAS ECONOMIAS DE AGLOMERAÇÃO .....	17
2.2 ECONOMIAS DE AGLOMERAÇÃO E SALÁRIOS: EVIDÊNCIAS EMPÍRICAS.....	22
<b>2.2.1 Economias de aglomeração e seus efeitos sobre os salários: evidências para o Brasil</b> .....	<b>25</b>
<b>3 METODOLOGIA</b> .....	<b>27</b>
3.1 MODELAGEM ECONOMETRICA: EMBASAMENTO TEÓRICO .....	27
3.2 DADOS E VARIÁVEIS .....	32
<b>3.2.1 Base de Dados</b> .....	<b>33</b>
<b>3.2.2 Tratamento dos microdados da RAIS Identificada</b> .....	<b>35</b>
<b>3.2.3 Áreas de Mercado de Trabalho Local (<i>Labour Market Areas</i>)</b> .....	<b>40</b>
<b>3.2.4 Descrição das Variáveis</b> .....	<b>44</b>
3.3 ESTRATÉGIA EMPÍRICA .....	46
<b>3.3.1 Modelo Econométrico</b> .....	<b>46</b>
<b>3.3.2 Estratégias de Identificação</b> .....	<b>50</b>
<b>3.3.3 Métodos de Estimação</b> .....	<b>52</b>
<b>3.3.3.1 Estimador de Múltiplos Efeitos Fixos</b> .....	<b>54</b>
<b>3.3.3.2 Variável Instrumental</b> .....	<b>57</b>
<b>4 RESULTADOS</b> .....	<b>60</b>
4.1 ESTATÍSTICAS DESCRITIVAS .....	60
4.2 RESULTADOS DAS ESTIMAÇÕES EM PAINEL .....	70
<b>4.2.1 Economias de aglomeração e prêmio salarial: resultado da estimação por efeitos fixos</b> .....	<b>70</b>
<b>4.2.2 Economias de aglomeração e prêmio salarial: resultado da estimação por efeitos fixos e variáveis instrumentais</b> .....	<b>73</b>
<b>4.2.3 Efeitos das economias de aglomeração sobre os salários individuais: Tamanho e Macrolocalização das LMA importam?</b> .....	<b>76</b>
<b>5 CONSIDERAÇÕES FINAIS</b> .....	<b>80</b>
<b>REFERÊNCIAS</b> .....	<b>84</b>
<b>APÊNDICES</b> .....	<b>90</b>

## 1 INTRODUÇÃO

Os estudos sobre os diferenciais salariais têm sido abordados com certa frequência na literatura econômica, em especial na Economia do Trabalho, a partir das equações salariais de Mincer (1974). Uma regularidade desses estudos revela que, mesmo controlando características observáveis e não observáveis dos trabalhadores, diferenças salariais persistem em favor dos grandes centros urbanos, indicando a importância de um componente espacial na determinação salarial. A questão espacial tem sido foco da Economia Regional e Urbana, que tem direcionado esforços em compreender a relação entre a concentração espacial das atividades econômicas e os ganhos de produtividade por ela gerados. Neste sentido, uma nova vertente de pesquisa, unificando os principais elementos de ambas as áreas de estudo, tem buscado investigar os diferenciais salariais espaciais por meio das economias de aglomeração.

Os estudos dessa nova vertente de pesquisa teve que lidar, durante muito tempo, com grandes dificuldades no que diz respeito à microfundamentação dos mecanismos das economias de aglomeração, que foram propostos inicialmente por Marshall (1890). Essa dificuldade só começa a ser superada com o surgimento da Nova Geografia Econômica (NGE), a partir do modelo centro-periferia desenvolvido por Krugman (1991a). A partir de então, foram desenvolvidos outros trabalhos que também buscaram fundamentar esses mecanismos (KRUGMAN, 1991b; FUJITA; KRUGMAN; VENABLES, 2001).

As pesquisas sobre a existência de economias de aglomeração têm por base as grandes disparidades regionais que são observadas em diferentes territórios ou, muitas vezes, dentro de um mesmo território. Por exemplo, em 1990, as regiões centrais japonesas, com apenas 0,18% da área total, representavam 29% do PIB do leste asiático. Na Coreia do Sul, a região da capital (província de Seul e Kyungki), que possui uma área correspondente a 11,8% do país e inclui 45,3% da população, produz 46,2% do Produto Interno Bruto (PIB). Na França, o contraste é ainda maior: a Ilha de França (região metropolitana de Paris), que representa 2,2% da área do país e 18,9% da população, produz 30% do seu PIB (FUJITA; THISSE, 2002). No Brasil, conforme o Instituto Brasileiro de Geografia e Estatística (IBGE), em 2017, o município de São Paulo tinha 5,8% da população e 0,018% do território, produzindo 10,6% do PIB nacional.

Essa concentração de população e de atividade econômica nos grandes centros urbanos pode ser decorrente da atratividade que esses locais possuem para produzir, tanto pelo grande mercado local quanto pela disponibilidade dos bens e serviços produzidos neles (KRUGMAN, 1991a). Neste sentido, em locais mais concentrados há uma maior interação

entre os agentes, o que pode resultar em aumentos de produtividade tanto das firmas quanto dos trabalhadores (ROSENTHAL; STRANGE, 2004). É possível, portanto, que esse aumento de produtividade decorrente de uma maior interação espacial entre os agentes provoque elevações na renda dos trabalhadores e das firmas.

A propósito, uma evidência documentada frequentemente na literatura sobre economias de aglomeração diz respeito à existência de salários mais elevados em regiões onde a atividade econômica é mais densa e produtiva. Glaeser e Maré (2001) são os primeiros a mensurar o efeito geral das economias de aglomeração sobre a determinação dos salários. Para capturar esse efeito geral, os autores utilizaram a densidade do emprego como medida de aglomeração. Os resultados identificaram a existência de um prêmio salarial a favor das áreas metropolitanas densas dos EUA, na ordem de 33%. Na Inglaterra, Fingleton (2003) encontrou evidências de que, conforme a densidade do emprego aumenta, o nível de salários aumenta mais do que proporcionalmente, chegando a 18% na cidade de Londres. Para o Brasil, Rocha; Silveira Neto e Gomes (2011) constataram um diferencial positivo de 9,4% para os trabalhadores das regiões metropolitanas. Mais recentemente, Silva (2017) encontrou um efeito positivo da densidade do emprego sobre os salários locais no valor de 4,9%.

Esses estudos, assim como outros que buscam mensurar os efeitos das economias de aglomeração sobre os salários, usam em suas análises a metodologia desenvolvida pioneiramente por Mincer (1974), que busca capturar as diferenças de rendimentos entre os trabalhadores. Contudo, tais estudos incluem além dos determinantes de capital humano- escolaridade e experiência-, a dimensão espacial como determinante salarial. Essa inclusão é ponto de integração entre a Economia do Trabalho e a Economia Regional e Urbana.

Aliado a esse incremento, as pesquisas têm avançado metodologicamente para evitar a superestimação dos efeitos das economias de aglomeração sobre os salários (GLAESER; MARÉ, 2001; COMBES; DURANTON; GOBILLON, 2008; ROSENTHAL; STRANGE, 2008). Isso pode ocorrer se trabalhadores mais habilidosos e firmas mais produtivas se aglomerarem em mercados de trabalhos mais densos e produtivos. Neste sentido, conforme Combes e Gobillon (2015), as habilidades desempenham dois papéis específicos na estimação desses efeitos. Primeiro, as próprias habilidades podem ser um dos determinantes das economias de aglomeração. Segundo, pode haver algum *sorting* de habilidades entre os locais, que deve ser controlado para evitar vieses ao medir o impacto das características locais relacionadas às economias de aglomeração. Dito de outra maneira, o controle do *sorting* de habilidades, isto é, da autosseleção entre mão de obra qualificada e mercados de trabalhos

densos, é fundamental para determinação do efeito líquido das economias de aglomeração sobre os salários.

Para lidar adequadamente com o *sorting* espacial de habilidades, deve-se controlar a heterogeneidade não observada dos trabalhadores, o que exige um painel de trabalhadores (COMBES; DURANTON; GOBILLON, 2008). Na literatura internacional, os estudos tem reconhecido a importância desse controle, uma vez que o prêmio salarial urbano poder ser fruto, ainda que parcialmente, do *sorting* de trabalhadores qualificados nas cidades (GLAESER; MARÉ, 2001; COMBES; DURANTON; GOBILLON, 2008; MION; NATICCHIONI, 2009; MATANO; NATICCHIONI, 2016). Combes *et al* (2012) encontraram evidências sobre a distribuição de salários e habilidades em áreas de emprego mais densas e menos densas na França. De acordo com os autores, em média, os trabalhadores em áreas mais densas são mais qualificados. Há também uma forte super-representação de trabalhadores com habilidades particularmente altas e baixas em áreas mais densas. Os autores destacam que esses resultados são consistentes com os padrões de migração, que incluem uma seleção negativa de migrantes para áreas menos densas e uma seleção positiva para áreas mais densas.

No Brasil, os estudos têm considerado em suas análises a presença de habilidades não observadas dos trabalhadores na determinação salarial (FREGUGLIA, 2007; ROCHA; SILVEIRA NETO; GOMES, 2011; BARUFI; HADDAD; NIJKAMP, 2016). Silva, Santos e Freguglia (2016) encontraram que existe um prêmio salarial urbano de 3% para os trabalhadores formais das áreas metropolitanas do Brasil, após controlar a heterogeneidade individual não observada e as características observadas dos trabalhadores e do emprego. No entanto, o prêmio salarial não foi identificado em todas as 24 regiões metropolitanas do Brasil. De acordo com os autores, isso pode sugerir que o aparente ganho de aglomeração seja um retorno às habilidades não observadas do trabalhador formal ou fatores não observados de firmas.

Como se vê, não apenas o *sorting* espacial dos trabalhadores podem confundir os ganhos reais das economias de aglomeração, como também a heterogeneidade não observada das firmas. De acordo com Combes e Gobillon (2015), as empresas podem diferir em suas equipes de gestão, sendo algumas mais eficientes que outras, e isso cria certa heterogeneidade na produtividade. Além disso, os autores consideram a possibilidade de que empresas com gerenciamento mais eficiente sejam estabelecidas em locais maiores. Portanto, a maior produtividade nestas regiões pode decorrer da concentração de empresas mais eficientes nesses locais e não do tamanho da localidade em si. Essa autoseleção das firmas

para locais em que possam otimizar sua produção, é o que caracteriza o *sorting* espacial das firmas.

Os estudos brasileiros têm buscado avançar no controle não só do *sorting* espacial dos trabalhadores, mas também no controle por *sorting* espacial das firmas (SILVA, 2017; CAMPOS, 2018; NEVES JÚNIOR, 2018). Silva (2017) realiza o primeiro grande estudo para o Brasil que leva em conta esses dois controles na determinação salarial e na identificação dos efeitos das economias de aglomeração. A autora encontrou que o efeito trabalhador é mais relevante do que o efeito firma na variação individual dos salários. Além disso, os resultados mostraram que 4,9% da diferença salarial individual observada são decorrentes dos efeitos puros de aglomeração. Por fim, autora ressalta a importância desses controles para uma mensuração adequada dos efeitos de aglomeração sobre os salários.

Outro ponto que pode levar a aumentos de produtividade em uma região é a interação simultânea entre as características dos trabalhadores e das firmas. Conforme Mion e Naticchione (2009) essa interação entre trabalhadores e firmas, que é denominada *assortative matching*, deve ser incluída nos estudos sobre os efeitos das economias de aglomeração. Analisando o mercado de trabalho italiano, os autores encontraram que o mecanismo de *matching* assortivo é negativamente correlacionado com o tamanho do mercado de trabalho nas províncias. Dauth *et al.* (2016) atribui a esse mecanismo a principal explicação para existência do prêmio salarial urbano no mercado de trabalho alemão. Woodcock (2008) destaca a importância de se controlar o *match* trabalhador-firma para que não se obtenha estimativas enviesadas.

Em um país, como o Brasil, que possui um padrão aglomerativo diferente entre as regiões, a aglomeração pode afetar a produtividade delas de maneira distinta. De acordo com os dados da Relação Anual de Informações Sociais (RAIS), no período de 2010 a 2018, a região Sudeste apresentou a maior média de aglomeração do mercado de trabalho formal do país (56,7 trabalhadores por Km<sup>2</sup>), sendo este número 3,6 vezes maior que a média de aglomeração da região Nordeste (15,8 trabalhadores por Km<sup>2</sup>) e 5 vezes maior do que a média da região Sul (11,4 trabalhadores por Km<sup>2</sup>), regiões estas que possuem a segunda e a terceira maiores médias de aglomeração, respectivamente. As regiões Centro-Oeste (3,2 trabalhadores por Km<sup>2</sup>) e Norte (5 trabalhadores por Km<sup>2</sup>) apresentam, nesta ordem, as menores médias de aglomeração do país. Se essas diferenças no padrão de aglomeração brasileira refletir na produtividade local e, conseqüentemente, nos salários, tem-se uma importante evidência empírica que pode auxiliar na construção de políticas voltadas para o mercado de trabalho, de modo a atenuar as disparidades salariais do Brasil. Neste sentido, para que as evidências

sejam robustas e possam embasar essas políticas, o controle dos vieses do *sorting* espacial dos trabalhadores e firmas e do *match* trabalhador-firma é fundamental na relação entre aglomeração e produtividade.

É importante destacar que o Brasil vem passando por importantes mudanças econômicas. De acordo com o IBGE, entre 2010 e 2018, o PIB anual brasileiro cresceu, em média, 1,67%, tendo experimentado neste período, um crescimento de 7,5%, em 2010, e quedas de 3,5% e 3,3% nos anos de 2015 e 2016, respectivamente. Essa oscilação de crescimento econômico pode influenciar a composição do mercado de trabalho formal nas localidades brasileiras e impactar o padrão de aglomeração dessas localidades se, em decorrência dessa variação de crescimento, trabalhadores e firmas realizarem algum tipo de mobilidade, seja entre firmas e áreas de trabalho, no primeiro caso, seja entre localidades, no último caso. O fato é que impactos sobre o mercado de trabalho e sobre o padrão de aglomeração de uma localidade podem afetar a produtividade e, conseqüentemente, os salários locais.

Diante deste contexto, esta dissertação busca mensurar os efeitos das economias de aglomeração sobre os salários individuais no mercado de trabalho formal brasileiro no período de 2010 a 2018, considerando os efeitos do *sorting* espacial dos trabalhadores, das firmas e do *match* trabalhador-firma sobre os salários. Além desse objetivo, são traçados dois objetivos específicos. O primeiro busca identificar a contribuição das economias de aglomeração sobre os diferenciais salariais para cada região do país. O segundo busca mensurar como a densidade do emprego influencia a determinação salarial a partir do tamanho das áreas de mercado de trabalho do Brasil. Dessa forma, será possível compreender como a aglomeração impacta os salários e se há diferenças substanciais no seu efeito a depender da particularidade da unidade espacial analisada.

As análises dos efeitos da densidade do emprego (*proxy* para as economias de aglomeração) sobre os salários do mercado de trabalho brasileiro, realizadas até então, não têm considerado o efeito do *match* trabalhador-firma como um determinante dos salários locais. Neste sentido, esta dissertação busca preencher esta lacuna ao controlar, simultaneamente, o efeito desse *match* e do *sorting* espacial dos trabalhadores e firmas sobre as estimativas salariais. Além disso, diferentemente dos estudos brasileiros sobre economias de aglomeração, que optam por usar como unidade espacial de análise os arranjos populacionais e concentrações urbanas ou as regiões de influência das cidades ou os municípios isolados, esta dissertação implementa a metodologia adotada por Neves Júnior (2018) que combina essas três unidades espaciais para construção das áreas de mercado de

trabalho local, que são as unidades espaciais de análise desta pesquisa. Em suma, esta dissertação, a partir dos seus objetivos traçados, contribui com a literatura de economias de aglomeração ao avançar no controle dos vieses de estimação e ao adotar uma nova medida de unidade espacial para realizar as análises aqui propostas. Ademais, ao mostrar como a aglomeração impacta os salários de acordo com as regiões brasileiras e com o tamanho das unidades espaciais consideradas, este estudo pode auxiliar na tomada de decisão dos agentes públicos e privados no que diz respeito a medidas que sejam direcionadas ao mercado de trabalho formal brasileiro, com vistas a reduzir as disparidades produtivas nele existentes.

Além desta introdução, esta dissertação está dividida em mais quatro capítulos. O capítulo 2 traz a revisão da literatura, que está dividida em duas seções. A primeira aborda a fundamentação teórica das economias de aglomeração. A segunda seção refere-se às evidências empíricas das economias de aglomeração e seus efeitos sobre os salários. O capítulo 3 diz respeito à metodologia e está dividido em três seções. A primeira descreve o modelo teórico em que se baseia este estudo. A segunda aborda os dados e as variáveis consideradas para realização desta pesquisa. E, a terceira descreve a estratégia econométrica adotada neste trabalho. O capítulo 4 apresenta os resultados e está dividido em 2 seções. A primeira reporta as estatísticas descritivas da amostra e a segunda apresenta os resultados das estimações realizadas nesta dissertação. O capítulo 5 apresenta as considerações finais deste estudo. Na sequência, são apresentadas as referências usadas neste trabalho e, em seguida, está disposto o apêndice.

## 2 REVISÃO DA LITERATURA

Este capítulo tem dois objetivos principais. O primeiro deles é apresentar o desenvolvimento da teoria sobre as economias de aglomeração, partindo das contribuições pioneiras até as mais recentes, que dizem respeito à microfundamentação dos mecanismos de aglomeração. O segundo objetivo é descrever as principais evidências empíricas das economias de aglomeração e seus efeitos sobre os salários, no contexto internacional e também nacional.

### 2.1 FUNDAMENTAÇÃO TEÓRICA DAS ECONOMIAS DE AGLOMERAÇÃO

As primeiras abordagens acerca das economias de aglomeração estiveram presentes nas obras de Smith (1776) e Marshall (1890), sendo este último o mais conhecido e mais influente dos primeiros analistas das economias de aglomeração. A partir da publicação de *Principles of Economics*, uma série de estudos sobre os efeitos da aglomeração na produtividade tem sido desencadeada. Mas, apesar do grande destaque de Marshall (1890), o primeiro autor a abordar os benefícios da aglomeração sobre a produtividade foi Smith (1776), ainda que de maneira limitada, já que seus argumentos baseavam-se apenas na divisão do trabalho. Na sua obra *An Inquiry into the Nature and Causes of the Wealth of Nations*, Smith (1776) argumenta que, com o aumento da especialização, isto é, quando os trabalhadores dedicam-se mais tempo a uma mesma tarefa, há aumento de produtividade. Para Smith (1776), ao se especializar em uma única tarefa, o trabalhador aperfeiçoa-se nela, reduzindo os custos fixos associados à troca de ferramentas e de posição dentro da fábrica.

Os autores da geografia econômica, em especial, os autores da Teoria da Localização, também forneceram *insights* importantes para os estudos das aglomerações econômicas, ao tentarem explicar a decisão locacional das atividades econômicas. Von Thünen (1826), ao analisar como as atividades agrícolas se distribuía em torno dos centros urbanos, afirmou que a localização ideal seria aquela em que os custos de transportes fossem minimizados. Para Weber (1909), a localização da indústria se daria onde o custo de produção fosse o menor possível, sendo este influenciado principalmente pelo custo de transporte. Lösch (1954), por sua vez, afirmava que a localização ótima da empresa fica onde o lucro líquido é maior. Contribuições teóricas desses autores foram reunidas na análise de Isard (1956), que considerou além da localização da indústria, o conjunto das atividades econômicas. Os resultados da análise é que a decisão locacional das firmas orienta-se pela minimização dos custos de transporte.

Mesmo com todos esses esforços, o espaço e as aglomerações só ganharam relevância na teoria econômica, com o surgimento da Nova Geografia Econômica (NGE), a partir do modelo centro-periferia de Krugman (1991a). Os modelos da NGE superam a limitação da geografia econômica em modelar o espaço na teoria econômica e, diferentemente desta, não toma como dada a existência das cidades. Fujita, Krugman e Venables (2001), em trabalho seminal, buscaram explicar como se dá o surgimento das novas cidades e a formação das aglomerações. A grande contribuição da nova geografia econômica está em microfundamentar os argumentos teóricos outrora expostos por Marshall (KRUGMAN, 1991; FUJITA; KRUGMAN; VENABLES, 2001).

Os trabalhos seminais de Krugman (1991a), Krugman (1991b) Fujita, Krugman e Venables (2001) e Fujita e Thisse (2002) apresentam as bases para os estudos sobre economias de aglomeração. A aglomeração pode ser entendida como o agrupamento de atividades, criado e sustentado por algum tipo de lógica circular que ocorre em muitos níveis, desde os distritos, que atendem a demanda local, até regiões mais especializadas, que atendem o mercado mundial como um todo (FUJITA; KRUGMAN; VENABLES, 2001). Em um estudo mais recente, Combes e Gobillon (2015) consideram economias de aglomeração como sendo qualquer efeito que aumente a renda das empresas e dos trabalhadores quando o tamanho da economia local crescer. Portanto, a aglomeração consiste em um agrupamento de atividades econômicas, sustentado por uma lógica circular que aumenta a renda de firmas e de trabalhadores quando a economia cresce.

A literatura propõe várias classificações para os diferentes mecanismos por trás das economias de aglomeração. Marshall (1890) divide os efeitos de aglomeração em três: transbordamento de conhecimentos (*knowledge spillovers*), agrupamentos de mão-de-obra (*labor pooling*) e encadeamento intersetorial (*intermediate input linkages*) (COMBES; GOBILLON, 2015). Os *spillovers* de conhecimentos ocorrem quando uma indústria é localizada, permitindo que os trabalhadores aprendam uns com os outros. O *labor pooling* é o mecanismo de aglomeração que permite uma melhor correspondência entre as necessidades de um empregador e as habilidades de um trabalhador, reduzindo o risco para ambos. O *sharing* de insumos ocorre quando a produção envolve retornos internos crescentes de escala (ROSENTHAL; STRANGE, 2004).

Na perspectiva *Marshalliana*, conforme Duranton e Puga (2004), os *spillovers* de conhecimento são discutidos em relação à aquisição de habilidades pelos trabalhadores e seu aprendizado sobre novas tecnologias. Para Marshall (1890) se um trabalhador, em uma indústria localizada, inicia uma nova ideia que traz invenções e melhorias em máquinas, em

processos e na organização geral dos negócios, ela pode ser adotada e modificada por outros trabalhadores. Essa modificação torna-se, então, a fonte de novas ideias. Os *spillovers* de conhecimento podem, portanto, dar a esta indústria localizada certa vantagem competitiva.

Ao falar sobre *spillovers*, Krugman (1991b) afirma que os *spillovers* de conhecimento são invisíveis e não podem ser medidos e rastreados. Van Der Panne (2004) reforça que os *spillovers* são variantes à distância, já que se referem ao conhecimento tácito, que só pode ser adquirido através do processo de interação social. As repercussões do conhecimento podem, portanto, surgir entre empresas do mesmo setor e só podem ser suportadas por concentrações regionais de um setor específico (VAN DER PANNE, 2004). Isto indica, portanto, que em localidades onde a atividade econômica é especializada, os *spillovers* são facilitados.

Na discussão sobre os *linkages*, Marshall (1890) menciona explicitamente os benefícios do *sharing* de fornecedores intermediários que produzem sob retornos crescentes (DURANTON; PUGA, 2004). Marshall (1890) afirma que, ao fornecerem suplementos e materiais, as indústrias subsidiárias permitem a indústria localizada de uma região reduzir os seus custos devido a diversos fatores, dentre eles, a redução dos custos de transportes. Ainda de acordo com o autor, as indústrias subsidiárias também se beneficiam, uma vez que ao se dedicarem a um pequeno ramo do processo de produção e fornecerem material para muitos vizinhos podem manter em uso constante máquinas altamente especializadas, pagando, assim, suas despesas. Desse modo, o encadeamento intersetorial torna o processo produtivo menos oneroso e mais eficiente para ambas às indústrias.

Uma indústria localizada pode oferecer suporte a fornecedores locais mais especializados, o que, por sua vez, torna essa indústria mais eficiente e reforça a localização (KRUGMAN, 1991b). Esse suporte só é possível se considerado que o *sharing* de insumos depende crucialmente da existência de economias de escala na produção dos insumos. Como Rosenthal e Strange (2004) descrevem, na presença de economias de escala, uma empresa isolada estará em desvantagem produtiva em relação àquelas concentradas, pois seus custos de produção são mais elevados. Os autores deixam claro que, somente, empresas à jusante localizadas onde a indústria está concentrada podem terceirizar suas demandas de insumos para produtores que possuem uma escala eficiente de produção. Portanto, os ganhos de produtividade se dão, justamente, devido a essa capacidade de terceirização.

Em se tratando do *labor pooling*, Marshall (1890) enfatiza que, em todas as etapas, exceto nos primeiros estágios de desenvolvimento econômico, uma indústria localizada obtém uma grande vantagem por ofertar um mercado constante de habilidade. Os empregadores tendem a se localizarem onde possam encontrar trabalhadores com as habilidades de que

necessitam; os trabalhadores, por sua vez, vão a lugares onde há empregadores que precisam de suas habilidades. Esse argumento sobre o *labor pooling* aponta para um mecanismo de *matching* no mercado de trabalho (DURANTON; PUGA, 2004).

O modelo teórico mais conhecido que busca demonstrar como o *labor pooling* funciona, é o proposto por Krugman (1991b). Nele, a indústria, especializada em uma atividade, é constituída por um número limitado de firmas e elas produzem utilizando uma mão de obra homogênea. Ao se localizarem em um mesmo espaço, tanto trabalhadores como firmas poderão lidar melhor com os choques adversos de produtividade. Krugman (1991b) descreve esta situação: se duas firmas estão em um mesmo local, pelo menos ocasionalmente, os bons tempos de uma coincidirão com os maus tempos de outra; os trabalhadores adicionais disponíveis da firma que está em má situação, serão absorvidos pela firma que estiver em boas condições no momento. Essa situação deixa claros os benefícios do *labor pooling* do mercado de trabalho, apontando para uma aglomeração induzida por a especialização na produção, que promove ganhos de produtividade tanto para as firmas quanto para os trabalhadores.

Os estudos mais recentes sugerem outros mecanismos para o surgimento das economias aglomerações para além dos já discutidos nesta seção. Baseada em Marshall (1980), a classificação mais usada atualmente foi proposta por Duranton e Puga (2004), que consideram os efeitos *sharing*, *matching* e *learning* como sendo os três principais mecanismos para existência das aglomerações. O primeiro considera que um maior mercado permite uma maior eficiência no compartilhamento (*sharing*) das infraestruturas e das instalações locais, uma variedade de fornecedores de insumos intermediários e um grupo de trabalhadores cujas habilidades são similares. O segundo garante que um maior mercado permite uma melhor correspondência (*matching*) entre empregadores e funcionários, compradores e fornecedores ou parceiros de negócios. Essa melhor correspondência pode garantir uma junção mais proveitosa entre esses agentes. Por fim, o terceiro considera que um maior mercado facilita a aprendizagem (*learning*), por exemplo, promovendo o desenvolvimento e a adoção generalizada de novas tecnologias e práticas de negócios (PUGA, 2010).

Os efeitos do *sharing* incluem os ganhos de uma variedade maior de insumos e especialização industrial, do uso comum de bens e instalações indivisíveis locais e da associação de riscos (COMBES; GOBILLON, 2015). De acordo com Duranton e Puga (2004), o compartilhamento de instalações e infraestrutura permite uma redução de custos aos usuários, assim como partilha de mão de obra especializada aumenta a produção local devido

ao aumento da produtividade individual. No que diz respeito ao *sharing* de fornecedores, ele só terá alguma relevância na determinação das economias de aglomeração se a indústria fornecedora for espacialmente concentrada, contudo, possui grande relevância na aglomeração setorial (OVERMAN; PUGA, 2010).

Duranton e Puga (2004) consideram que o compartilhamento de insumos intermediários variados deve gerar retornos crescentes de escala, já que um aumento na produção final em virtude desse compartilhamento requer um aumento menor que proporcional nos fatores primários. Os efeitos dessas partilhas aliados aos benefícios de se compartilhar um mercado constante de habilidade (*labor pooling*), que reduzem os riscos de flutuações no emprego, são as razões pelas quais Duranton e Puga (2004) consideram o *sharing* como uma importante fonte de aglomeração.

Em relação à importância do *matching* como mecanismo de aglomeração, Combes e Gobillon (2015) destacam que o seu efeito está ligado à melhoria da qualidade ou quantidade de correspondências (*matches*) entre empresas e trabalhadores. Para Puga (2010), em mercados de trabalho densos, essa melhor correspondência pode assumir a forma de chances melhores de encontrar uma correspondência adequada, uma qualidade de correspondência mais alta ou uma combinação de ambos. Portanto, em mercados com muitas oportunidades de emprego a serem exploradas ao mesmo tempo, é pouco provável que não haja um bom *match* entre empregadores e funcionários.

A fundamentação microeconômica para esse mecanismo foi elaborada por Duranton e Puga (2004), que apresentam um modelo de *matching* baseado em Salop (1979), introduzindo, contudo, uma estrutura urbana ao modelo. Nesse contexto, os autores destacam que há duas fontes de economias de aglomeração. A primeira é uma externalidade de *matching*, destacada pela primeira vez por Helsley e Strange (1990), em que um aumento no número de agentes que tentam corresponder melhora a qualidade esperada de cada correspondência. A segunda é decorrente da ampliação do modelo de Salop (1979) para permitir a concorrência no mercado de trabalho, em que uma concorrência mais forte ajuda a economizar os custos fixos, fazendo com que o número de empresas aumente menos que proporcionalmente à força de trabalho.

Os efeitos do *learning* envolvem a geração, difusão e acumulação de conhecimento. Uma cidade diversificada oferece muitas alternativas e pode ser o berço de novas empresas, que ao atingirem sua maturidade deslocam-se para cidades mais especializadas, criando aglomeração e, assim, reduzindo os seus custos. Em equilíbrio, permanecerão nas cidades apenas os trabalhadores jovens para tentar adquirir habilidades e os trabalhadores velhos

qualificados para transmitir suas habilidades, todos os trabalhadores velhos não qualificados mudam-se para o interior. Por meio do acúmulo de capital humano e aumento da produção, o preço dos bens diminui. Isso se torna, portanto, incentivo para o crescimento populacional das cidades e reforça a aglomeração (DURANTON; PUGA, 2004).

Todas as teorias de aglomeração que enfatizam a transmissão de ideias sugerem que cidades e capital humano são complementos, porque níveis mais altos de habilidades significam mais conhecimento a ser transmitido. No caso da inovação, é mais provável que trabalhadores e empresas mais qualificados sejam inovadores. No caso da acumulação urbana de capital humano, uma força de trabalho mais qualificada significa que há mais professores em potencial para jovens trabalhadores que aprendem com seus pares urbanos (GLAESER; GOTTLIEB, 2009).

Como se viu nesta seção, as economias de aglomeração, enquanto sendo um agrupamento de atividades econômicas apontam para dois efeitos principais. Um deles é a redução dos custos envolvidos no processo produtivo. O outro é o aumento da produtividade dos agentes econômicos decorrentes desse agrupamento. Esta dissertação ao analisar as diferenças salariais a partir da contribuição das economias de aglomeração foca neste último efeito.

## 2.2 ECONOMIAS DE AGLOMERAÇÃO E SALÁRIOS: EVIDÊNCIAS EMPÍRICAS

Esta seção apresenta alguns dos principais estudos empíricos sobre as economias de aglomeração. Contudo, esta análise empírica não é tão simples, como pontua Puga (2010), dado que todas as fontes de aglomeração compartilham a previsão de que a produtividade aumenta com a escala de atividade em um local. Isso torna difícil mensurar o impacto individual de cada mecanismo sobre a produtividade. A fim de contornar essa dificuldade, muitos trabalhos têm buscado quantificar os ganhos de produtividade decorrentes das economias de aglomeração por meio da análise de diferentes escopos<sup>1</sup>. Neste sentido, a estratégia adotada consiste em utilizar uma medida que capture o efeito geral das economias de aglomeração sobre a produtividade.

Ciccone e Hall (1996) introduziram a densidade o emprego como medida de aglomeração para capturar os efeitos das economias de aglomeração sobre a produtividade. Utilizando dois modelos, sendo um baseado nas externalidades geográficas locais e outro na diversidade dos serviços intermediários locais, os autores estimaram funções de produção por meio de Mínimos Quadrados Ordinários (MQO) e Variáveis Instrumentais (VI), com dados

---

<sup>1</sup> Rosenthal e Strange (2004) fornece um rico detalhamento sobre os escopos das economias de aglomeração.

da economia dos Estados Unidos da América (EUA), a fim de verificar os efeitos da densidade do emprego sobre os diferenciais de produtividade. Os resultados do estudo indicam que a duplicação da densidade do emprego eleva a produtividade média do trabalho em 6%. Para Ciccone e Hall (1996), a densidade da atividade econômica explica mais de 50% das diferenças de produtividade observada nos estados americanos. Por fim, os autores consideram que a aglomeração compensa o efeito de congestionamento nas áreas mais densas e pode aumentar o crescimento econômico.

A fim de estimar os efeitos de aglomeração para Alemanha, Espanha, França, Itália e Reino Unido, Ciccone (2002) combinou dados espaciais sobre o valor agregado com dados sobre emprego e educação desses países. Os resultados empíricos do artigo demonstram efeitos substanciais de aglomeração nos cinco países europeus da amostra e os efeitos de aglomeração não parecem ser significativamente diferentes entre os países. As estimativas dos mínimos quadrados sugerem que uma duplicação da densidade do emprego aumenta a produtividade média do trabalho em aproximadamente 5%, enquanto que as estimativas com variáveis instrumentais sugerem um aumento de 4,5%. Essas estimativas são muito semelhantes às obtidas para os EUA (CICCONE; HALL, 1996).

Também utilizando a densidade do emprego como medida de aglomeração, e os salários como forma de mensurar a produtividade<sup>2</sup>, Glaeser e Maré (2001), por meio das estimações por Mínimos Quadrados Ordinários (MQO) e Variáveis Instrumentais (VI), relatam evidências de que os residentes mais antigos nas cidades maiores ganham um prêmio salarial em relação aos trabalhadores recém-chegados à mesma cidade. Os autores também identificaram que os salários dos trabalhadores mais antigos ao deixarem suas cidades são mais altos em seu novo local de trabalho quanto maior o tamanho da cidade anterior de residência. Glaeser e Maré (2001) identificaram que há um diferencial salarial de 33% a favor dos trabalhadores das áreas urbanas dos EUA quando comparados àqueles das áreas não urbanas. Os autores consideram que uma fração significativa desse prêmio salarial urbano é acumulada para os trabalhadores ao longo do tempo e permanece com eles quando saem das cidades. Esses resultados corroboram com a existência de um componente dinâmico associado às economias de aglomeração.

A partir desse estudo de Glaeser e Maré (2001), outros estudos buscaram investigar os efeitos das economias de aglomeração sobre os salários. Wheaton e Lewis (2002) reforçam

---

<sup>2</sup> Essa abordagem baseia-se no pressuposto de que, em mercados competitivos, o trabalho recebe o valor de seu produto marginal. Mesmo sem concorrência perfeita, em locais mais produtivos, os salários serão, portanto, mais altos (Rosenthal e Strange, 2004).

que o salário reflete os ganhos de produtividade dos trabalhadores. Neste sentido, os autores, usando dados do censo americano, em 1990, testaram os efeitos das economias de aglomeração sobre os salários, tendo como medida de aglomeração a especialização e a concentração do emprego por ocupação e por setor. Os resultados mostram que as especializações ocupacional e setorial variam de 0 a 7% e de 0 a 68%, elevando os salários em 23% e 30%, respectivamente. Além disso, as concentrações ocupacional e setorial variam de 0 a 46% e de 0 a 41%, o que gera ganhos salariais de 12% e 16%, respectivamente. Por fim, Wheaton e Lewis (2002) verificaram que os trabalhadores do setor manufatureiro ganham salários mais altos quando estão nos mercados de trabalho urbanos, com elasticidades entre 1,2 e 3,6 para esses efeitos.

Fingleton (2003) busca investigar a relação entre a densidade do emprego, usada como medida de aglomeração, e os salários no Reino Unido para os anos de 1999 e 2000, por meio da estimação de Mínimos Quadrados em Dois Estágios (MQ2E). O autor encontrou evidências que sustentam a hipótese de retornos crescentes para densidade do emprego em áreas urbanas. Esses retornos são mais elevados nos grandes centros urbanos, chegando a 18% para cidade de Londres.

Utilizando a acessibilidade ao produto regional bruto na Suécia como *proxy* para as economias de aglomeração, Klaesson e Larsson (2009) buscaram investigar sua influência nos salariais regionais. Para isso, os autores utilizaram dados regionais de 2004 estimaram por MQO e por VI. Além disso, o nível salarial médio regional foi decomposto em um índice de produtividade regional e um índice de composição industrial regional. O resultado geral da análise é que as economias de aglomeração são uma característica predominante nas regiões da Suécia e influencia a produtividade e, conseqüentemente, os salários. As elasticidades estimadas mostram que uma região duas vezes maior deve ter salários aproximadamente 1,7% mais altos. Por fim, o índice de produtividade é 1,8% maior e o de composição industrial 0,9% maior.

Groot, Groot e Smit (2014) buscaram identificar a natureza e as causas das diferenças salariais na Holanda. Com base em microdados de trabalhadores individuais para o período de 2000-2005 e através de equações salariais *mincerianas*, os autores mostraram que há um prêmio salarial em favor dos trabalhadores das áreas urbanas mais densas da Holanda. Conforme, os resultados dos autores, o efeito geral das economias de aglomeração, medido pela densidade do emprego, apresentou elasticidade de 4,8% sobre os salários na Holanda. Os autores também encontraram evidências de que a participação local de um setor aumenta a produtividade, se essa participação for dobrada, a produtividade aumenta em 2,9%.

De modo geral, apesar de ser difícil a mensuração dos efeitos particulares dos mecanismos de aglomeração, diversos estudos têm mensurado o efeito geral das economias de aglomeração. Para isso, os autores têm adotado *proxies* para a aglomeração, como densidade populacional, tamanho populacional e também a densidade do emprego, como demonstrado nesta seção. O consenso é que um dos efeitos das economias de aglomeração é o aumento da produtividade, que em muitas pesquisas têm sido representada pelo salário. Neste sentido, muitos estudos internacionais têm buscado capturar a relação entre as economias de aglomeração e os salários.

### **2.2.1 Economias de aglomeração e seus efeitos sobre os salários: evidências para o Brasil**

No Brasil, a literatura empírica sobre as economias de aglomeração ainda é muito limitada e recente, contudo, o interesse por essa área de estudo vem se ampliando de maneira considerável, em especial, no que diz respeito aos efeitos sobre os salários. Os estudos existentes apontam para uma relação positiva entre o prêmio salarial e às aglomerações urbanas brasileiras. Azzoni e Servo (2002) e Campos e Silveira Neto (2009) verificaram uma redução nos diferenciais salariais ao controlarem em suas análises o custo de vida, as habilidades observadas e não observadas e as características ocupacionais dos trabalhadores. Contudo, mesmo após o controle de tais variáveis, persistiu um diferencial salarial a favor das regiões metropolitanas brasileiras. Isso sugere, portanto, que devem existir ganhos salariais decorrentes do tamanho da aglomeração urbana.

Rocha, Silveira Neto e Gomes (2011) chamam a atenção para a dificuldade em identificar a existência dos efeitos de aglomeração sobre os salários em áreas urbanas grandes, dado que quanto maior é a cidade ou o centro urbano, maior é a pluralidade de observações. Por isso, segundo os autores, os retornos para as características não observáveis e observáveis dos trabalhadores são maiores assim como para os custos de vida. Deste modo, para identificar a influência das economias de aglomeração sobre os salários, é necessário controlar as características individuais dos trabalhadores e levar em conta as características e atributos locais. Mesmo após controlar tais variáveis, Rocha, Silveira Neto e Gomes (2011) encontraram um diferencial salarial positivo de 9,4% em favor dos trabalhadores das regiões metropolitanas brasileiras, no período de 2000 a 2008.

Silva, Santos e Freguglia (2016) encontraram um prêmio salarial urbano de 3% para os trabalhadores das regiões metropolitanas do Brasil, no período de 1995 a 2008, isso após o controle das características observáveis e não observáveis do trabalhador e do emprego. O

resultado encontrado pelos autores é similar aos encontrados por Heuermann, Halfdanarson e Suedekum (2010), onde o prêmio salarial à aglomeração variou de 5% a 10%.

Barurufi, Haddad e Nijkamp (2016) investigam o retorno salarial às economias de aglomeração para os diferentes setores da indústria. Nas estimações sem efeitos fixos o prêmio salarial fruto da densidade do emprego variou entre 5% e 9,5%, dependendo do setor analisado. Ao se considerar os efeitos fixos, os autores encontraram variação salarial entre 3% e 19% a depender do setor analisado. Em alguns setores, a elasticidade da densidade em relação os salários aumentou e em outros diminuiu quando comparada aos modelos sem efeitos fixos. Ao estimar por variáveis instrumentais e efeitos fixos, Barurufi, Haddad e Nijkamp (2016) encontraram um retorno salarial à densidade do emprego que varia entre 3,5% e 18,5%, apresentando pouca variabilidade em relação ao modelo estimado apenas por efeitos fixos. Esses resultados mostram que a endogeneidade, tratada no modelo estimado por VI, pode subestimar os efeitos da densidade sobre os salários.

Silva (2017) apresenta importantes evidências sobre os diferenciais salariais do Brasil, ao controlar, simultaneamente, os efeitos do *sorting* espacial dos trabalhadores e das firmas. As estimações que controlam esses efeitos indicam que há um retorno salarial à densidade do emprego entre 2,6% e 4,2%. No modelo estimado principal, estimado por VI e efeitos fixos, a autora encontra um prêmio salarial urbano de 4,9% decorrente da densidade do emprego. Esses resultados mostram que o *sorting* espacial dos trabalhadores e das firmas podem enviesar as estimativas do efeito da densidade sobre os salários.

Em análise feita para a região metropolitana de São Paulo, Campos (2018) encontrou um efeito positivo das aglomerações sobre os salários. As evidências mostram que ao considerar múltiplos efeitos fixos (trabalhador, firma, célula), simultaneamente, o efeito da densidade do emprego é na apropriada área é de 0,039%. Ao tratar a endogeneidade entre a densidade do emprego e os salários, esse efeito sobe para 1,785. Isto indica, portanto, que a endogeneidade pode subestimar os efeitos da densidade do emprego e os salários.

Em linhas gerais os estudos para o Brasil vêm avançando no controle de viés nas estimações dos efeitos das economias de aglomeração sobre os salários. Esses estudos mostram que os efeitos fixos explicam parte das diferenças salariais atribuídas à densidade do emprego e que a endogeneidade entre a densidade do emprego e a produtividade age subestimando os efeitos das aglomerações sobre os salários. Neste sentido, para obter-se estimativas mais robustas, além de tratar do *sorting* espacial dos trabalhadores e das firmas, é preciso tratar a possível endogeneidade entre a densidade do emprego e os salários. A estratégia empírica utilizada para se obter essas estimativas é descrita a seguir no capítulo 3.

### 3 METODOLOGIA

Este capítulo tem por objetivo apresentar todos os aspectos metodológicos que foram adotados nesta dissertação. Ele está dividido em três seções. A primeira seção apresenta o embasamento teórico em que este estudo se ancora. A segunda seção descreve a base de dados e os procedimentos de tratamento aplicados a ela, bem como apresenta as variáveis utilizadas nesta pesquisa. Por fim, a terceira seção detalha a estratégia econométrica de estimação utilizada para que se possa alcançar os objetivos traçados para esta dissertação.

#### 3.1 MODELAGEM ECONOMÉTRICA: EMBASAMENTO TEÓRICO

Este estudo utiliza como referência o modelo teórico de *sorting* por habilidades proposto por Combes, Duranton e Gobillon (2008), desenvolvido no âmbito da Economia Regional e Urbana. Uma especificação econométrica deste modelo será utilizada nesta dissertação para identificar os efeitos das economias de aglomeração sobre os salários individuais, controlando o *sorting* espacial dos trabalhadores, das firmas e de suas interações. Ademais, serão incluídas mais variáveis explicativas do que as consideradas no modelo original, conforme a disposição dos dados.

Na economia urbana, a associação entre maior produtividade nos mercados de trabalho mais densos e trabalhadores mais habilidosos são conhecidos como *sorting* espacial por habilidades (COMBES; DURANTON; GOBILLON, 2008). A ocorrência desse fenômeno é frequentemente associada às diversas configurações espaciais na distribuição de habilidades por regiões. Sendo assim, é um elemento fundamental para a compreensão das disparidades regionais de produtividade.

Combes, Duranton e Gobillon (2008) consideram, pelo menos, três fatores principais para explicar a existência das grandes disparidades espaciais de salários: i) diferenças espaciais na composição das habilidades da força de trabalho; ii) diferenças nas dotações espaciais não humanas, tais como fatores geográficos e produtivos, e iii) interações entre trabalhadores ou entre empresas que ocorrem localmente e podem resultar em ganhos de produtividade. Em resumo, as disparidades salariais espaciais resultam das diferenças regionais na composição de atributos dos trabalhadores, nas dotações não humanas e nas interações produtivas locais.

A partir desses fatores, Combes, Duranton e Gobillon (2008) desenvolveram um modelo teórico que é referência na análise do *sorting* espacial salarial e que sistematiza a ideia de *sorting* preconizada por Marshall (1890). No modelo proposto, a remuneração do trabalhador  $i$ ,  $w_i$ , é determinada por:

$$w_i = A s_i, \quad (3.1)$$

onde  $s_i$  representa as habilidades do trabalhador e  $A$  a produtividade. Assim, o salário médio da região  $a$ ,  $w_a$ , é dado pelo produto da média das habilidades na região  $a$ ,  $\bar{s}_a$ , pela produtividade do trabalho,  $A$ , como definido na equação seguinte:

$$w_a = A \bar{s}_a \quad (3.2)$$

No modelo Combes, Duranton e Gobillon (2008), a produtividade pode ser influenciada por dotações não humanas, isto é, tanto por características geográficas (clima, recursos naturais, etc) quanto por dotações locais não naturais (capital público e privado, tecnologia, instituições). Neste sentido, a produtividade do trabalho da região  $a$  pode ser afeta pelas dotações não humanas,  $E_a$ , de modo que o salário nesta região,  $w_a$ , será dado por:

$$w_a = A(E_a) \quad (3.3)$$

Além disso, Combes, Duranton e Gobillon (2008) também definem o salário a partir das interações locais entre trabalhadores ou entre firmas. Na verdade, os benefícios gerados por tais interações refletem economias de urbanização e efeitos de localização e ainda a concentração setorial de firmas, as quais respaldam importantes fontes de economias de aglomeração, como o *matching* de trabalhadores ou o *sharing* regional. Assim, o salário médio na região  $a$ , no setor  $k$  é dado por:

$$w_{a,k} = A(I_a, I_{a,k}), \quad (3.4)$$

onde  $I_a$  e  $I_{a,k}$  são dois vetores de variáveis de interação que capturam as economias de urbanização e de localização, respectivamente. Desse modo, a partir da equação (3.4), os efeitos das economias de aglomeração são capturados de maneira agregada.

O modelo inova, ao considerar em uma única estrutura de determinação salarial, os três fatores que consideram explicar a existência das grandes disparidades espaciais de salários. Desse modo, o salário de equilíbrio de um trabalhador  $i$ , na área  $a(i)$  e no setor  $k(i)$  será dado por:

$$w_i = A(E_{a(i)}, I_{a(i)}, I_{a(i),k(i)}) s_i. \quad (3.5)$$

O modelo microeconômico desenvolvido por Combes, Duranton e Gobillon (2008) assumem uma firma competitiva representativa, com função Cobb-Douglas, operando em uma área  $a$ , no setor  $k$  e no período  $t$ , cuja função de lucro é:

$$\pi_{a,k,t} = p_{a,k,t} y_{a,k,t} - \sum_{i \in (a,k,t)} w_{i,t} l_{i,t} - r_{a,k,t} z_{a,k,t}, \quad (3.6)$$

onde,  $p_{a,k,t}$  é preço da produção  $y_{a,k,t}$ . Para qualquer trabalhador  $i$  empregado nesta firma no ano  $t$ ,  $w_{i,t}$  é o salário por dia e  $l_{i,t}$  é o número de dias trabalhado. Os outros fatores de produção são representados por  $z_{a,k,t}$ , e  $r_{a,k,t}$  é o preço desses fatores. A função de produção

assumida no modelo de Combes, Duranton e Gobillon (2008) é uma função Cobb-Douglas em trabalho efetivo e os outros fatores de produção, como a seguinte:

$$y_{a,k,t} = A_{a,k,t} \left( \sum_{i \in (a,k,t)} s_{i,t} l_{i,t} \right)^b (z_{a,k,t})^{1-b} \quad (3.7)$$

A função de produção (3.7) assume uma firma representativa competitiva com retornos constantes de escala, onde  $0 < b \leq 1$ ;  $s_{i,t}$  são as habilidades do trabalhador  $i$  no período  $t$  e  $A_{a,k,t}$  é a produtividade total dos fatores em  $(a, k, t)$ . No equilíbrio competitivo, o trabalhador  $i$  empregado na área de emprego  $a(i, t)$  e no setor  $k(i, t)$  no ano  $t$  recebe um salário igual ao seu produto marginal:

$$w_{i,t} = b p_{a(i,t),k(i,t),t} A_{a(i,t),k(i,t),t} \left( \frac{z_{a(i,t),k(i,t),t}}{\sum_{i \in (a,k,t)} s_{i,t} l_{i,t}} \right)^{1-b} s_{i,t} \quad (3.8)$$

Combes, Duranton e Gobillon (2008) fazem uso da condição de primeira ordem para maximizar o lucro em relação aos outros fatores ( $z_{a,k,t}$ ) e o substitui na equação (3.8):

$$w_{i,t} = b(1-b)^{\frac{(1-b)}{b}} \left( p_{a(i,t),k(i,t),t} \frac{A_{a(i,t),k(i,t),t}}{(r_{a(i,t),k(i,t),t})^{1-b}} \right)^{\frac{1}{b}} s_{i,t} = B_{a(i,t),k(i,t),t} s_{i,t} \quad (3.9)$$

Onde,  $B_{a(i,t),k(i,t),t} \equiv b(1-b)^{\frac{(1-b)}{b}} \left( p_{a(i,t),k(i,t),t} \frac{A_{a(i,t),k(i,t),t}}{(r_{a(i,t),k(i,t),t})^{1-b}} \right)^{\frac{1}{b}}$ .

Por um lado, através da expressão para  $w_{i,t}$ , as diferenças salariais entre as áreas podem refletir diferenças nas habilidades individuais,  $s_{i,t}$ , ou, alternativamente, também podem refletir as verdadeiras diferenças de produtividade causadas por dotações e interações locais,  $B_{a(i,t),k(i,t),t}$ . Por outro lado, o diferencial de produtividade entre regiões pode estar associado ou à produtividade dos fatores ( $A_{a,k,t}$ ), ou à relação de preços dos produtos ( $p_{a,k,t}$ ), ou mesmo ao preço dos insumos não trabalhistas ( $r_{a,k,t}$ ). Contudo, não é possível identificar os efeitos de preço e de tecnologia separadamente sobre os salários. (COMBES; DURANTON; GOBILLON, 2008).

Neste contexto, Combes, Duranton e Gobillon (2008) sugerem que é possível que características locais, como a densidade do emprego, tenham efeitos sobre a produtividade ( $B_{a(i,t),k(i,t),t}$ ), tanto positivos (economias de aglomeração) quanto negativos (congestionamento), não sendo possível identifica-los separadamente. Neste caso, a estimação capta apenas o efeito geral de uma variável. Assim, apesar de não conseguir identificar como os efeitos tecnológicos e o preço afetam isoladamente os salários, e esta é uma limitação nos estudos sobre economias de aglomeração como um todo, o modelo Combes, Duranton e

Gobillon (2008) é uma sofisticada ferramenta para a identificação dos diferenciais salariais por habilidades, especialmente, em uma perspectiva espacial.

A fim de poder testar empiricamente a equação (3.9), Combes, Duranton e Gobillon (2008) adotam uma especificação em dois estágios, tanto para o termo de habilidades dos trabalhadores,  $s_{i,t}$ , quanto para o termo de produtividade causada por dotações e interações locais,  $B_{a(i,t),k(i,t),t}$ . Assumindo primeiramente que as habilidades do trabalhador  $t$  são dadas por:

$$\log s_{i,t} = X_{i,t}\varphi + \delta_i + \epsilon_{i,t}, \quad (3.10)$$

onde  $X_{i,t}$  é um vetor de características do trabalhador variáveis no tempo;  $\delta_i$  são os efeitos fixos do trabalhador e  $\epsilon_{i,t}$  é o termo de erro. Nesta especificação, os autores assumem que os erros são independentes e identicamente distribuídos entre períodos e trabalhadores. Voltando-se para  $B_{a(i,t),k(i,t),t}$ , que reflete as reais diferenças de produtividade da equação (3.9), Combes, Duranton e Gobillon (2008) assumem que ela é dada por:

$$\log B_{a,k,t} = \beta_{a,t} + \mu_{k,t} + I_{a,k,t}\gamma_k, \quad (3.11)$$

em que  $\beta_{a,t}$  representa o efeito fixo área-ano (efeito puro de área);  $\mu_{k,t}$  denota o efeito fixo setor-ano (efeito puro do setor);  $\gamma_k$  é o vetor dos coeficientes associados com o vetor de variáveis de interações intra-setoriais para cada área-setor-ano,  $I_{a,k,t}$ .

Combes, Duranton e Gobillon (2008) combinaram as equações (3.9), (3.10) e (3.11), de modo que:

$$\log w_{i,t} = \beta_{a(i,t),t} + \mu_{k(i,t),t} + I_{a(i,t),k(i,t),t}\gamma_{k(i,t)} + X_{i,t}\varphi + \delta_i + \epsilon_{i,t} \quad (3.12)$$

Os autores destacam que na equação (3.12) as interpretações de  $I_{a,k,t}\gamma_k$  e de  $X_{i,t}\varphi$  são problemáticas. Neste sentido, para que se possa decompor o efeito total do setor em efeito fixo do setor e um efeito de tempo é preciso considerar uma mesma tendência temporal para todas as indústrias, reduzindo o número de coeficientes a serem estimados. A especificação econométrica final do primeiro estágio corresponde à demanda inversa por trabalho:

$$\log w_{i,t} = \beta_{a(i,t),t} + \mu_{k(i,t),t} + \tilde{I}_{a(i,t),k(i,t),t}\gamma_{k(i,t)} + \tilde{X}_{i,t}\varphi + \delta_i + \epsilon_{i,t}, \quad (3.13)$$

onde  $\tilde{I}_{a(i,t),k(i,t),t}$  é o vetor centralizado das variáveis de interações intra-setorial e  $\tilde{X}_{i,t}$  é o vetor centralizado das características do trabalhador variáveis no tempo. Essa estimativa do primeiro estágio permite identificar separadamente o poder explicativo dos efeitos de trabalhador, baseados nas suas habilidades ( $\tilde{X}_{i,t}\varphi + \delta_i$ ); das interações intra-setoriais ( $\tilde{I}_{a,k,t}\gamma_k$ ) e do conjunto de dotações e interações entre os setores ( $\beta_{a,t}$ ), sendo considerados os efeitos fixo de indivíduo, setor e área.

Até então, a partir da equação (3.13), foi possível avaliar a importância relativa de pessoas *versus* espaço, para explicar as disparidades espaciais dos salários (COMBES;

DURANTON; GOBILLON, 2008). A fim de avaliar os determinantes das diferenças salariais entre áreas, os autores estimaram a importância relativa de dotações e interações inter-setoriais na explicação dos efeitos fixos de área-ano. Assim, a especificação econométrica do segundo estágio é dada da seguinte maneira:

$$\beta_{a,t} = \omega_0 + \theta_t + I_{a,t}\gamma + E_{a,t}\alpha + v_{a,t}, \quad (3.14)$$

em que  $\theta_t$  representa os efeitos fixos de tempo;  $\alpha$  é um vetor de coeficientes associados com as variáveis das dotações do local,  $E_{a,t}$ ;  $\gamma$  é um vetor de coeficientes ligados com as interações locais inter-setoriais,  $I_{a,t}$  e  $v_{a,t}$  é o termo de erro que reflete os choques locais de tecnologia distribuídos entre áreas e períodos, assumidos como sendo independentes e identicamente distribuídos. Por fim, os autores fixaram  $\omega_0$  em zero, que é o coeficiente do período, isto é, o ano de referência.

As equações (3.13) e (3.14) constituem a especificação econométrica completa do modelo Combes, Duranton e Gobillon (2008). A estimação é feita em dois estágios, pois na equação do segundo estágio (3.14), os autores utilizam como variável dependente, os efeitos fixos de área estimados na equação (3.13), do primeiro estágio. Os autores apresentam dois motivos para realizarem a estimação em dois estágios. O primeiro diz respeito ao fato de que, em estágio único, não é possível calcular a variação dos choques locais,  $v_{a,t}$ , já que o modelo é projetado na dimensão *within* da equação salarial, e os trabalhadores podem se mover entre áreas. Se não estimado em dois estágios, o modelo terá vieses nos erros padrão para os coeficientes estimados de variáveis explicativas agregadas. A segunda razão é que, a estimação em estágio único, não permite a distinção entre os choques locais e os choques puramente idiossincráticos no nível do trabalhador, o que é importante tendo em vista a ausência das variáveis de dotação local.

Combes, Duranton e Gobillon (2008) ressaltam que para identificar os efeitos fixos do setor na equação (3.13), é necessário que haja mobilidade suficiente entre os setores para que todos os setores estejam conectados uns aos outros, ainda que indiretamente, por meio de fluxos de trabalhadores. A identificação dos efeitos fixos de área é um pouco mais sutil. Para os autores, quando os trabalhadores se deslocam pelas áreas, fornecem a identificação das diferenças entre as áreas ao longo do tempo. Assim, os trabalhadores que permanecem identificam mudanças ao longo do tempo em sua área. Portanto, para identificar os efeitos fixos de área, Combes, Duranton e Gobillon (2008) destacam a necessidade de que alguns trabalhadores permaneçam em cada uma das áreas de emprego entre duas datas consecutivas e que existam áreas ou grupo de áreas com fluxo de trabalhadores para o resto do país. Os

autores ressaltam que tendo uma grande base de dados, todas essas condições são facilmente satisfeitas.

O modelo CDG (2008) adota algumas restrições de identificação, tendo em vista que os efeitos fixos por área são identificados apenas um em relação ao outro, assim como os efeitos fixos do setor. Neste sentido, é preciso uma categoria de comparação tanto para área quanto para setor. Em sua análise, Combes, Duranton e Gobillon (2008) escolhem uma área e um setor, como categorias de comparação, cujos coeficientes sejam zero.

Embora útil para identificação, Combes, Duranton e Gobillon (2008) ponderam que uma base de dados muito grande, restringe a estimação da equação (3.13) do primeiro estágio a um procedimento mais simples. Primeiramente, os autores estimam a equação (3.13) usando um estimador de dimensão *within*, em que todas as variáveis estão centralizadas em relação à sua média para cada indivíduo, o que resulta nos coeficientes de todas as variáveis, exceto os efeitos fixos do trabalhador. Em seguida, é possível recuperar um estimador de efeito fixo de cada trabalhador, calculando seu erro médio de previsão. Pelo teorema de Frish-Waugh, este é o estimador de fluxo para o efeito fixo individual. Os autores ressaltam que apenas os trabalhadores que aparecem pelo menos duas vezes no painel contribuem para a estimativa.

A especificação econométrica do modelo Combes, Duranton e Gobillon (2008), supõe que as escolhas da área e do setor são estritamente exógenas. Contudo, essa suposição não deve ser muito restritiva dado que a especificação contém efeitos fixos de área e de setor. De acordo com Combes, Duranton e Gobillon (2008), os resultados da equação salarial podem ser enviesados se houver uma *sorting* espacial ou de setor com base nos erros. No entanto, eles não serão enviesados se o *sorting* for baseado nas variáveis explicativas, incluindo os efeitos fixos de indivíduos, de área e de setor. De maneira mais clara, existe um viés quando a decisão de localização por parte do trabalhador é determinada pelo salário exato que ele pode obter nos locais em um determinado ano, mas não existe um viés quando os trabalhadores baseiam sua decisão locacional com base no salário esperado.

### 3.2 DADOS E VARIÁVEIS

Esta seção tem por objetivo apresentar e detalhar a construção da base de dados final utilizada nas estimações realizadas nesta dissertação. Para isto, ela está dividida em 4 subseções. A primeira delas descreve a principal base de dados utilizada neste estudo. A segunda detalha o tratamento dos dados. A terceira subseção descreve a unidade espacial de análise considerada para fins desta dissertação. Por fim, na quarta subseção são apresentadas as variáveis utilizadas nos modelos de regressão deste estudo.

### 3.2.1 Base de Dados

Os principais dados utilizados neste estudo são provenientes da Relação Anual de Informações Sociais Identificada (RAIS identificada) do Ministério da Economia (ME)<sup>3</sup>. A RAIS<sup>4</sup> é um banco de dados, gerido pelo governo federal, no âmbito do ME, que recolhe informações, de forma compulsória e anual, de todos os estabelecimentos do mercado de trabalho formal brasileiro. O seu objetivo é fornecer informações sobre o mercado de trabalho, subsidiar o controle da arrecadação de benefícios trabalhistas, viabilizar a concessão de benefícios sociais e trabalhistas e gerar estatísticas sobre o mercado formal do Brasil.

Segundo nota técnica divulgada pelo então Ministério do Trabalho<sup>5</sup> (2015) sobre a base de dados da RAIS 2014, a RAIS cobre cerca de 97% do setor formal da economia, sendo, portanto, uma espécie de censo do mercado de trabalho formal brasileiro. Ainda de acordo com a nota, o tratamento aplicado aos dados dos estabelecimentos e vínculos empregatícios permite que as informações divulgadas, a partir da RAIS, tenham desagregação até em nível de município, de subatividades e classes econômicas e de ocupações.

A principal vantagem da RAIS é que o seu formato longitudinal permite construir dados em painel para os trabalhadores formais e acompanhar as suas trajetórias geográfica, ocupacional e setorial ao longo do tempo, bem como acompanhar características selecionadas do estabelecimento empregador, através de códigos de identificação. Para o trabalhador, a identificação pode ser feita através de seu número de inscrição no Programa de Integração Social (PIS)<sup>6</sup> ou por meio do seu número de inscrição no Cadastro de Pessoas Físicas (CPF)<sup>7</sup>. Já o estabelecimento pode ser identificado pelo seu número no Cadastro Nacional de Pessoas Jurídicas (CNPJ)<sup>8</sup> ou pelo número no Cadastro Específico do Instituto Nacional de Seguro Social (INSS) – CEI<sup>9</sup>.

A RAIS é organizada de modo que cada relação de trabalho formal declarada entre um trabalhador e uma firma é registrada. Sendo assim, um registro na base é composto por

---

<sup>3</sup> Esta base de dados foi adquirida mediante solicitação feita ao Ministério da Economia pela coordenação do Curso de Mestrado em Economia Aplicada (CMEA) da UFAL. Reitera-se que todos os protocolos de sigilo acordado com o ME estão sendo cumpridos por este autor.

<sup>4</sup> A RAIS foi instituída pelo decreto nº 76.900, de 23/12/75.

<sup>5</sup> Desde 01 de janeiro de 2019 o Ministério deixou de existir e suas funções foram redistribuídas a outros ministério, como o da Economia.

<sup>6</sup> O PIS possui 11 dígitos e foi estabelecido pela lei complementar nº 7, de 7 de setembro de 1970 (<https://is.gd/nLAn45>).

<sup>7</sup> O CPF possui 11 dígitos e foi estabelecido pela decreto-lei no 401, de 30 de dezembro de 1968 (<https://is.gd/54CcMz>).

<sup>8</sup> O CNPJ possui 14 algarismos e foi instituído pela lei no 4.503, de 30 de novembro de 1964 (<https://is.gd/h2JZEv>).

<sup>9</sup> O CEI possui 12 dígitos e foi instituído pela lei nº 8.213, de 24 de julho de 1991 (<https://is.gd/tZRKAJ>).

informações do trabalhador (gênero, escolaridade, raça/cor, idade, nacionalidade), do estabelecimento (atividade econômica, localização geográfica, tamanho) e do próprio vínculo (data de admissão, data e causa de afastamento, ocupação, remuneração, horas contratadas, tempo de emprego, tipo de vínculo). A base de dados é organizada por unidade da federação e divulgada com periodicidade anual.

O banco de dados da RAIS registra os vínculos empregatícios estabelecidos entre 01 de janeiro e 31 de dezembro de cada ano e, como já mencionado, cada registro corresponde a uma relação de trabalho formal, regida pela Consolidação das Leis do Trabalho (CLT) ou pelo regime estatutário<sup>10</sup> (IBGE, 2021). Como a RAIS registra cada vínculo formal estabelecido no decorrer do ano-base, é possível que um trabalhador seja registrado mais de uma vez na base, caso tenha tido mais de um vínculo formal no ano-base. Por este motivo, o banco de dados disponibiliza uma variável que informa a quantidade de vínculos ativos e inativos em 31 de dezembro de cada ano para que se conheça a situação empregatícia dos trabalhadores ao fim de cada ano-base.

No período de 2010 a 2018, o registro total de vínculos ativos e inativos em 31 de dezembro de cada ano cresceu entre os anos de 2010 (66.747.302) e de 2014 (76.107.279), mas a partir deste último até 2018 (66.214.692), último ano da base, este número recuou<sup>11</sup>, conforme pode ser visto na **Tabela 3.1**.

**Tabela 3.1: Total de vínculos ativos e inativos registrados na RAIS: 2010-2018**

Ano	Total de vínculos	Vínculos inativos em 31/12	Vínculos ativos em 31/12
2010	66.747.302	22.678.947	44.068.355
2011	70.971.125	24.660.494	46.310.631
2012	73.326.485	25.867.773	47.458.712
2013	75.400.510	26.452.077	48.948.433
2014	76.107.279	26.535.769	49.571.510
2015	73.629.639	24.644.090	48.985.549
2016	67.144.598	21.084.400	46.060.198
2017	65.655.882	19.374.292	46.281.590
2018	66.214.692	19.583.577	46.631.115
<b>Total</b>	<b>635.197.512</b>	<b>210.881.419</b>	<b>424.316.093</b>
<b>Média</b>	<b>70.577.501</b>	<b>23.431.269</b>	<b>47.146.233</b>

Fonte: Elaboração própria a partir dos microdados da RAIS identificada (2010-2018).

É a partir dos dados da **Tabela 3.1** que escopo empírico desta pesquisa começa a ser desenhado. Contudo, para que a base de dados da RAIS se adeque ao modelo utilizado como base nesta dissertação, o modelo de Combes, Duranton e Gobillon (2008), alguns ajustes são necessários. Sendo assim, a fim de assegurar que a formatação em painel dos dados e os

<sup>10</sup> Os trabalhadores estatutários são ocupam cargos públicos regidos por leis próprias de cada esfera de governo.

<sup>11</sup> Este recuo pode ser reflexo da recessão econômica enfrentada pelo Brasil entre meados de 2014 e fim de 2016.

resultados encontrados sejam consistentes e confiáveis, a base de dados passou por alguns tratamentos que serão descritos a seguir na subseção 3.2.2.

### **3.2.2 Tratamento dos microdados da RAIS Identificada**

A finalidade desta subseção é descrever em detalhes o processo de tratamento dos microdados da RAIS identificada que deu origem ao painel de dados balanceado utilizados nesta pesquisa, que contém características relacionadas aos indivíduos e às firmas, além de aspectos regionais e do vínculo empregatício. Os filtros aplicados aos microdados da RAIS têm por objetivo final obter um painel balanceado para os anos de 2010 a 2018, de modo que apenas o vínculo empregatício principal do trabalhador permaneça na base, o que permitirá o controle do *sorting* espacial de trabalhadores e firmas, como propõe a metodologia desta pesquisa. Neste sentido, será apresentado, com riqueza de detalhes, o processo de tratabilidade dos dados até que se chegue a esse objetivo final.

Antes de tudo, é importante ressaltar que os procedimentos adotados neste trabalho seguem outros estudos similares a este (BARUFI, 2015; SILVA, 2017; CAMPOS, 2018; NEVES JÚNIOR, 2018) que, com algumas adaptações, adotaram a abordagem proposta pelo então Ministério do Trabalho para o arranjo da RAIS-MIGRA<sup>12</sup>, que permite o acompanhamento do trabalhador ao longo do tempo.

O tratamento da base de dados para construir o painel de dados requeridos nesta pesquisa está dividido em seis blocos. O bloco 1 conta com seis etapas que buscam eliminar inconsistências inerentes à própria base e delimitar o escopo empírico do estudo. Neste bloco, os procedimentos são realizados ano a ano em decorrência de limitações computacionais<sup>13</sup>. Após concluir as etapas de tratamento do bloco 1, as bases anuais são agregadas. O bloco 2 conta com apenas uma etapa que gera o painel inicial. O bloco 3 de tratamento tem por objetivo tratar o identificador de estabelecimento, necessário na análise aqui proposta, e possui três etapas. O bloco 4 está dividido em quatro etapas e visa tratar inconsistências remanescentes referentes ao indivíduo. O bloco 5 conta com 2 etapas que refinam o escopo empírico deste estudo. Por fim, o bloco 6, último bloco de tratamento, é composto por uma única etapa que delimita de maneira definitiva o escopo empírico desta dissertação.

---

<sup>12</sup> A RAIS-MIGRA já não é mais disponibilizada desde 2012.

<sup>13</sup> Os microdados da RAIS são disponibilizados para cada ano. Em virtude do tamanho dos arquivos anuais, a agregação desses dados exigiria um esforço computacional muito grande que implicaria em perda significativa de eficiência no processamento dos dados. Sendo assim, a agregação que gera o painel só será feita após a realização de alguns filtros nas bases anuais.

Na Etapa I do bloco 1, manteve-se na base apenas os vínculos ativos em 31 de dezembro de cada ano. Na Etapa II foram excluídos os registros com remuneração nula ou *missing*, com tempo de emprego igual a zero ou *missing* e com carga horária semanal contratada inferior a 20 horas<sup>14</sup>. Ainda na Etapa II, foram excluídos os registros sem inscrição no PIS (identificador do trabalhador), ou com erro de preenchimento<sup>15</sup>. A escolha pelo PIS como identificador do trabalhador e, não o CPF, dá-se em decorrência deste último apresentar mais erros e valores *missings* do que o primeiro, razão pela qual Campos (2018), Silva (2017) e Silva (2019) também fizeram essa escolha. Essas duas etapas trataram, principalmente, de inconsistências da própria base e iniciaram a delimitação do escopo empírico deste trabalho. As etapas seguintes do bloco 1 buscaram delimitar ainda mais esse escopo.

A Etapa III excluiu os trabalhadores da Administração Pública, tendo em vista que a determinação salarial para estes difere da lógica de mercado, e os salários não devem variar entre localidades (SILVA, 2017), sendo assim, apenas os trabalhadores do setor privado e de entidades empresariais<sup>16</sup>, excluindo também desse grupo as entidades empresariais de economia mista, foram mantidos na base. Na Etapa IV, foram excluídos os trabalhadores sem vínculo CLT e cujo tipo de admissão é registrado como proveniente de transferências, de contratos temporários ou por tempo determinado, bem como os registrados como aprendiz<sup>17</sup>. A etapa V manteve na base apenas os trabalhadores dentro da faixa etária produtiva, isto é, trabalhadores entre 18 e 65 anos<sup>18</sup>. Ao fim de mais estas 3 etapas o escopo do estudo foi sendo melhor definido. A última etapa do bloco 1 buscou definir o vínculo de trabalho principal de modo a permitir a construção do painel.

A Etapa VI, última do bloco 1, definiu um único vínculo trabalhador-firma que caracteriza a relação de trabalho principal do indivíduo. Para isso uma série de filtros foi adotada. Em um primeiro momento, excluiu-se da base aqueles trabalhadores com mais de 5 vínculos empregatícios. Em seguida, foram excluídas inconsistências de gênero e de idade, ou seja, informações divergentes sobre estas variáveis para um mesmo PIS foram excluídas. Adiante, foram mantidos os vínculos com maior carga horária semanal contratada. Além

---

<sup>14</sup> Conforme Campos (2018) e Neves Júnior (2018), contratos inferiores a 20 horas semanais são norteados por mecanismos distintos dos que possuem regime contratual superior a 20 horas semanais.

<sup>15</sup> O PIS que foi utilizado como identificador do trabalhador possui 11 dígitos, então registros com menos de 11 dígitos foram excluídos.

<sup>16</sup> A definição de entidades empresariais é baseada na natureza jurídica das empresas, conforme a Comissão Nacional de Classificação de 2018 (IBGE, 2021). Essa mesma definição foi utilizada no estudo de Amarante (2017).

<sup>17</sup> Esta etapa baseia-se, principalmente, em Amarante (2017) e Silva (2017).

<sup>18</sup> Este padrão é utilizado na literatura e foi adotado em estudos, tais como, o de Arbach (1999), Glaeser e Maré (2001), Campos (2018), Silva (2017) e Silva (2019).

disso, permaneceram na base os contratos cuja data de admissão é a mais antiga para um mesmo PIS, com base nos critérios da RAIS-MIGRA, também adotados por Silva (2017) e Neves Júnior (2018) que define o vínculo principal como sendo aquele cuja data de admissão é a mais antiga. Por fim, para aqueles trabalhadores que permaneceram com mais de um vínculo, isto é, um mesmo PIS com mais de um registro de emprego, adotou-se o procedimento de escolha aleatória de modo a selecionar apenas um vínculo. Desse modo, gerou-se uma base composta por trabalhadores associados a apenas um vínculo empregatício. O Apêndice A traz as informações sobre a quantidade de observações excluídas a cada etapa de tratamento do bloco 1.

Após os procedimentos de limpeza da base de dados realizadas no bloco 1, os microdados anuais foram agrupados. Como já mencionado, o PIS é o identificador do trabalhador, por isso a vinculação intertemporal dos trabalhadores deu-se a partir dele. Sendo assim, a Etapa I, única do bloco 2 de tratamento da base de dados, manteve na base apenas os trabalhadores que estão presentes ao longo de todos os anos, isto é, de 2010 a 2018. Ao fim dessa etapa, das 250.960.042 de observações, resultado da agregação das bases anuais, formou-se um painel inicial com 47.432.799 observações, correspondentes a 5.270.311 trabalhadores e 1.314.351 estabelecimentos. Um adendo importante é que, a cada fim de bloco de tratamento serão mantidos na base apenas os trabalhadores presentes em todos os anos da base. Com isso, a cada fim de bloco de tratamento, o painel permanece balanceado.

Com o painel inicial gerado, manteve-se na base apenas o vínculo principal trabalhador-firma. Como a análise aqui proposta faz uso do identificador de firma, além do identificador de indivíduo, é necessária a limpeza de algumas inconsistências no identificador de firma. Na RAIS, os estabelecimentos são identificados pelo CNPJ ou pelo CEI. Cada estabelecimento, filial ou sede, possui um CNPJ ou CEI específico, de modo que estabelecimentos dispensados do CNPJ são identificados pelo CEI<sup>19</sup>. De acordo com a RAIS, os estabelecimentos inscritos tanto no CNPJ quanto no CEI devem apresentar a declaração da RAIS de acordo com o contrato de trabalho dos empregados (CEI ou CNPJ)<sup>20</sup>.

Silva (2017) aponta dois problemas principais relacionados ao identificador do estabelecimento. O primeiro é que alguns apresentavam o mesmo CNPJ, no mesmo ano-base, em locais distintos, inclusive em unidades da federação distintas; o segundo problema ocorre quando estabelecimentos são registrados em setores de atividades distintas em um mesmo

---

<sup>19</sup> Neste caso, são incluídas obras, empregadores pessoas físicas, urbanas e rurais que mantiveram empregados no ano-base.

<sup>20</sup> No caso de canteiros de obras, independentemente do contrato, o CEI vinculado deve ser informado.

ano. Neste sentido o bloco 3 de tratamento dos dados busca tratar e limpar as inconsistências decorrentes desses problemas.

A Etapa I do bloco 3 buscou identificar os casos decorrentes do primeiro problema (942 estabelecimentos), em seguida, substituir o CNPJ pelo CEI dos estabelecimentos identificados e, por fim, excluir os casos remanescentes (23 estabelecimentos, ou, 40.766 observações). A Etapa II fez o mesmo procedimento para o segundo problema (aqui 275 estabelecimentos foram identificados e 39 foram excluídos, o que corresponde a 27.592). A Etapa III, última do bloco 3 de tratamento, excluiu os estabelecimentos que não estavam inscritos no CNPJ ou no CEI (935 observações deletadas). Após essas etapas, 278.370 observações foram deletadas, de modo que a base apresentou 47.154.429 de observações, sendo 5.239.381 indivíduos e 1.317.349 estabelecimentos. Como é possível notar, o número de estabelecimentos foi ampliado em decorrência da substituição do CNPJ pelo CEI, o que indica que havia estabelecimentos distintos inscritos no mesmo CNPJ. Por fim, é importante dizer que, este estudo utilizou como identificador de firma o CNPJ, sendo este substituído pelo CEI nos casos em que apresentou inconsistências (especificamente, 4.131 casos).

Após o tratamento do identificador do estabelecimento, outras inconsistências foram identificadas na base e devidamente tratadas no bloco 4 que conta com quatro etapas. Na Etapa I, foram excluídas 7.866.387 observações referentes a indivíduos que reportaram nível de escolaridade no ano anterior maior que no ano posterior. Na Etapa II, outras 279.909 observações referentes a trabalhadores, que apresentaram divergências ao longo do tempo nas variáveis ano de nascimento e idade, foram excluídas da base. Na Etapa III, foram excluídas 621.630 observações de indivíduos que reportaram diferentes gêneros entre os anos de 2010 e 2018. Por fim, a Etapa IV do bloco 4 de tratamento, excluiu 35.010 observações referentes às inconsistências no grupo ocupacional (CBO 2002) dos trabalhadores. Nesta etapa, excluíram-se os trabalhadores do grupo das forças armadas e os membros do poder público, além dos *missings* e dos casos com erro de preenchimento<sup>21</sup>. Ao fim do bloco 4, foram mantidas no painel 38.351.421 observações, correspondentes a 4.261.269 trabalhadores e 1.088.284 estabelecimentos.

O bloco 5 de tratamento é composto por duas etapas. Na Etapa I, foram excluídas<sup>22</sup>. 284.631 observações de trabalhadores que receberam, em dezembro de cada ano, salário-hora menor do que o estabelecido por lei para o ano. Na Etapa II, excluiu-se 45.836 observações

---

<sup>21</sup> Neste caso, o erro de preenchimento ocorre quando o grupo ocupacional definido pela Classificação Brasileira de Ocupações 2002 (CBO 2002) foi preenchido com código inferior a seis dígitos.

<sup>22</sup> O governo federal publica anualmente o valor mínimo da hora de trabalho com base na CLT.

cujos salários reais, em dezembro, fossem superiores a R\$ 50.000, visando, assim, eliminar observações discrepantes, ou *outliers*<sup>23</sup>. Ao fim do bloco 5, foram excluídas 1.371.216 observações, restando na base 36.695.574 observações, correspondentes a 4.077.286 trabalhadores e a 1.036.111 estabelecimentos.

Por fim, o último bloco de tratamento, o bloco 6, é composto de uma única etapa que delimita de vez o escopo empírico deste trabalho, que consiste em manter na base apenas os trabalhadores presentes nas áreas de mercado de trabalho local. Após o fim do bloco 6 de tratamento, restaram na base 31.308.255 observações, relativas a 3.478.695 indivíduos e a 858.953 estabelecimentos.

Tendo em vista as limitações computacionais enfrentadas pelo autor, optou-se, assim como fez Silva (2017), por extrair uma amostra aleatória de 5% das observações a partir dos dados obtidos no bloco 6. A seleção foi feita mantendo a representatividade das variáveis e a estrutura do painel. Assim, a amostra final contém 11.585.637 de observações, correspondentes a 520.455 firmas e 1.287.293 indivíduos e 2.232.883 combinações de indivíduos e firmas.

A **Tabela 3.2** resume todas as etapas de tratamento conforme o bloco ao qual pertencem.

**Tabela 3.2: Descrição das etapas de tratamento da base identificada da RAI (2010-2018)**

Bloco de tratamento	Etapas de tratamento
Bloco 1	<b>Etapa I:</b> Manter vínculos ativos <b>Etapa II:</b> Excluir inconsistências de renda, de tempo de emprego, do PIS e com menos de 20hrs semanais <b>Etapa III:</b> Excluir o setor público e manter apenas entidades empresariais <b>Etapa IV:</b> Excluir trabalhadores sem vínculo CLT e admitidos por transferências, tempo determinado e aprendizes <b>Etapa V:</b> Manter indivíduos entre 18 e 65 anos <b>Etapa VI:</b> Manter apenas o vínculo principal
Bloco 2	<b>Etapa I:</b> Manter apenas os trabalhadores presentes na base entre 2010 e 2018
Bloco 3	<b>Etapa I:</b> Excluir CNPJs com inconsistências geográficas <b>Etapa II:</b> Excluir CNPs com divergências de setor <b>Etapa III:</b> Excluir vínculos sem CNPJ ou CEI
Bloco 4	<b>Etapa I:</b> Excluir inconsistências intertemporais de escolaridade <b>Etapa II:</b> Excluir inconsistências intertemporais de ano de nascimento e idade <b>Etapa III:</b> Excluir divergências intertemporais de gênero <b>Etapa IV:</b> Excluir inconsistências intertemporais de ocupação
Bloco 5	<b>Etapa I:</b> Excluir trabalhadores com salário-hora menor do que a lei determina para o ano-base <b>Etapa II:</b> Excluir salários reais superiores a R\$ 50.000,00
Bloco 6	<b>Etapa I:</b> Manter apenas os trabalhadores das áreas de mercado de trabalho local

Fonte: Elaboração própria

<sup>23</sup> Este procedimento foi o mesmo adotado por Amarante (2017).

A subseção 3.2.3, detalha como as áreas de mercado de trabalho local, unidade de análise espacial adota nesta pesquisa, foram construídas.

### **3.2.3 Áreas de Mercado de Trabalho Local (*Labour Market Areas*)**

Em conformidade com a literatura de economias de aglomeração (GLAESER; MARÉ, 2001; GROOT; GROOT; SMIT, 2014; BARUFI, 2015; NEVES JÚNIOR, 2018), este estudo adota como unidade espacial de análise as áreas de mercado de trabalho local, muito conhecidas por sua grafia em inglês *Labour Markets Areas* (LMA). Esta subseção tem por objetivo descrever as três etapas que determinaram a construção dessas áreas. Foram utilizadas as versões mais atuais dos estudos mais recentes sobre arranjos populacionais e concentrações urbanas (IBGE, 2016) e sobre as regiões de influência das cidades – REGIC (IBGE, 2018). A partir desses dois estudos, as LMA foram delimitadas, seguindo metodologia adota por Neves Júnior (2018). Antes de prosseguir com essa descrição, é importante ressaltar que os dados referentes a cinco municípios criados entre 2010 e 2018 foram compatibilizados para o ano de 2010, isto é, as informações sobre esses municípios foram computadas para os municípios que os originaram. Desse modo, a definição das áreas de mercado de trabalho local não apresentará inconsistências em decorrência destes desmembramentos.

Na primeira etapa, foram considerados como áreas de mercado de trabalho local os arranjos populacionais. Para o IBGE (2016) um arranjo populacional é definido como um agrupamento de dois ou mais municípios onde existe forte integração populacional decorrente dos movimentos pendulares para trabalho ou estudo, ou devido à contiguidade entre as manchas urbanizadas principais. Seguindo os critérios<sup>24</sup> adotados pelo IBGE (2016), foram delimitados 294 arranjos populacionais com classificação a partir de faixas populacionais. Dos 294 arranjos populacionais, 189 são classificados como pequenos (população total de até 100.000 habitantes), 81 como médios (população superior a 100.000 habitantes e inferior ou igual a 750.000 habitantes) e 24 são classificados como grandes arranjos populacionais (população total superior a 750.000 habitantes).

---

<sup>24</sup> Para mensurar e identificar os arranjos populacionais, o IBGE adota três critérios de integração. O primeiro deles exige que haja forte intensidade relativa dos movimentos pendulares de trabalho e estudo, de modo que essa intensidade seja igual ou superior a 0,17 do índice de integração entre os municípios, conforme fórmula desenvolvida pelo instituto. O segundo critério exige que haja forte intensidade absoluta dos movimentos pendulares para trabalho e estudo, de tal forma que o deslocamento entre um município A e um município B seja em um volume igual ou superior a 10.000 pessoas. O terceiro e último critério diz respeito à contiguidade das manchas urbanizadas principais de dois municípios, que não deve ultrapassar 3 km de distância.

Esses arranjos populacionais eram formados por 953 municípios em 2010, que totalizavam 106.246.994 pessoas<sup>25</sup> ou 55,7% da população residente no Brasil naquele ano (IBGE, 2016). Considerando os dados populacionais (IBGE, 2018) do último ano de análise deste trabalho, esses arranjos somavam 117.301.930 pessoas ou 56,3% dos residentes do país em 2018. A grande maioria dos arranjos populacionais (89,5%) é formada por cinco municípios ou menos, sugerindo que o processo de aglomeração está se intensificando em centros de menor porte populacional, devido ao aumento da mobilidade (IBGE, 2016).

A segunda etapa considera como áreas de mercado de trabalho local as concentrações urbanas, que incluem os arranjos populacionais descritos acima e os municípios isolados<sup>26</sup>. As concentrações urbanas é uma definição utilizada pelo IBGE (2016) para nomear os arranjos populacionais e os municípios (municípios isolados) com população acima de 100.000 habitantes, que estão diretamente relacionados ao fenômeno urbano e suas dinâmicas. Seguindo os mesmos critérios populacionais descritos no parágrafo anterior e incluindo os pequenos arranjos populacionais como uma categoria das concentrações urbanas<sup>27</sup>, são definidas 371 concentrações urbanas. Ao todo são 159 médias concentrações, formadas por 80 municípios isolados, 78 arranjos populacionais e 1 arranjo populacional de 1ª integração<sup>28</sup>; 26 grandes concentrações urbanas, compostas por 20 arranjos populacionais, 1 arranjo populacional de 1ª integração<sup>29</sup> (São José dos Campos-SP), 1 arranjo populacional de 2ª integração<sup>30</sup> e 2 municípios isolados<sup>31</sup> (IBGE, 2016). Por fim, depois de alguns arranjos populacionais terem sido integrados a outros, foram classificados 186 arranjos populacionais como pequenas concentrações urbanas.

---

<sup>25</sup> Considerando que existem os arranjos populacionais fronteirços, formados não somente por municípios brasileiros, mas também por unidades político-administrativas equivalentes nos países vizinhos, esse contingente sobe para 107.167.901 pessoas (IBGE, 2016).

<sup>26</sup> São considerados municípios isolados aqueles que não atendem aos critérios para compor um arranjo populacional, mas tem uma população superior a 100.000 habitantes.

<sup>27</sup> O IBGE (2016) não faz essa inclusão, o instituto apenas define médias e grandes concentrações urbanas, porém para trabalhar com uma mesma categoria de análise, objetivando uma melhor compreensão do processo que originou as áreas de mercado de trabalho local, os pequenos arranjos foram considerados como uma categoria de concentração urbana.

<sup>28</sup> A 1ª integração de um arranjo populacional é gerada quando há intensidade muito alta de movimento pendular para trabalho e estudo entre dois ou mais arranjos populacionais. A 1ª Integração do arranjo de "Presidente Prudente-SP" é formada por 2 arranjos populacionais mais um município isolado.

<sup>29</sup> A 1ª integração do arranjo de "São José dos Campos-SP" é formada por 2 arranjos populacionais.

<sup>30</sup> A 2ª integração do arranjo populacional é gerada quando há intensidade muito alta de movimento pendular para trabalho e estudo entre duas 1ªs integrações do arranjo populacional. Neste caso, a 2ª integração do arranjo de "Porto Alegre-RS" é formada por 4 arranjos populacionais e 3 municípios isolados.

<sup>31</sup> Os municípios isolados são Campo Grande-MS e Manaus-AM.

As concentrações urbanas aqui definidas são formadas por 1042 municípios, que compreendia 64,4% da população residente no Brasil (122.846.014 de pessoas) em 2010<sup>32</sup>. Considerando a população de 2018, essas concentrações urbanas eram compostas por 65,2% da população residente no país (135.893.203 de pessoas). Como se vê, ao considerar as concentrações urbanas como unidade espacial de análise, centros urbanos importantes, como Campo Grande-MS e Manaus-AM, que não formam arranjos populacionais, são incluídos na definição de áreas de mercado de trabalho local. Isto permite, portanto, que municípios relevantes para compreensão do mercado de trabalho formal brasileiro não fiquem de fora da análise proposta nesta dissertação.

Apesar de contornar o problema da não inclusão de municípios isolados importantes para dinâmica do mercado de trabalho brasileiro, a noção de áreas de mercado de trabalho local pode ser aprimorada. Para tanto, segue-se metodologia adotada por NEVES JÚNIOR (2018), e considera-se na definição de áreas de mercado de trabalho local, os municípios que estão no campo de influência dos municípios isolados. Isto é feito a partir dos dados da pesquisa de Regiões de Influência das Cidades – REGIC – (IBGE, 2018). Sendo assim, a terceira e última etapa de delimitação das LMA que serão utilizadas neste estudo considera, além das concentrações urbanas descritas anteriormente, os municípios que compõem as regiões de influência imediata (doravante: municípios influenciados) dos municípios isolados. O Apêndice B mostra como foram determinados os municípios influenciados utilizados nesta dissertação.

A definição de áreas de mercado de trabalho local realizada nesta terceira etapa é mais abrangente do que as definidas nas duas etapas anteriores e é consistente com a abordagem inicial realizada na primeira etapa. Diante disso, este estudo utilizará como unidade de análise espacial a definição de áreas de mercado de trabalho local construída nesta terceira etapa em que engloba os arranjos populacionais, os municípios isolados e os municípios influenciados por estes últimos. Ao todo foram definidas 371 áreas de mercado de trabalho local, compostas por 1991 municípios, classificadas conforme sua população em: pequenas (população com até 100.000 habitantes), médias (população superior a 100.000 habitantes e inferior ou igual a 750.000 habitantes) e grandes (população superior a 750.000 habitantes).

As áreas de mercado de trabalho local definidas para uso nesta pesquisa compreendem 35,8% dos municípios brasileiros. Contudo, a despeito desse número pequeno de municípios, essas áreas somam 153.961.441 de pessoas, o que corresponde a 73,8% da população

---

<sup>32</sup> E 123.766.921 de pessoas ao considerar a população dos arranjos fronteiriços.

brasileiro no ano de 2018. A **Tabela 3.3** mostra como estão compostas as áreas de mercado de trabalho local. Nela é possível observar que as regiões Sudeste (145) e Sul (91) são as macrorregiões que apresentam uma maior quantidade de LMA, seguidas das regiões Nordeste (73) e Centro – Oeste (32). A região Norte é a que apresenta a menor quantidade de LMA (30), pois possui um processo de urbanização mais lento onde os processos de urbanização e desmembramento são menos intensos que nas demais regiões (IBGE, 2016). Em termos de contingente populacional, as macrorregiões Sudeste e Nordeste, juntas, possuem 70,1% (107.896.443 pessoas) da população brasileira que reside em áreas de mercado de trabalho local e 67,6 (25 de 37) das LMA de tamanho grande.

**Tabela 3.3: Características das Áreas de Mercado de Trabalho Local (2018)**

Região	Tamanho das Áreas de Mercado de Trabalho Local	População no Brasil	Quantidade de Áreas de Mercado de Trabalho Local	Quantidade de arranjos, municípios isolados e municípios influenciados
Norte	Pequena	505.620	13	23
	Média	4.193.222	12	80
	Grande	8.271.078	5	118
<b>Subtotal</b>		<b>12.969.920</b>	<b>30</b>	<b>221</b>
Nordeste	Pequena	1.655.588	36	76
	Média	8.074.966	24	203
	Grande	23.485.818	13	202
<b>Subtotal</b>		<b>33.216.372</b>	<b>73</b>	<b>481</b>
Sudeste	Pequena	2.986.003	61	133
	Média	20.072.909	72	448
	Grande	51.621.159	12	196
<b>Subtotal</b>		<b>74.680.071</b>	<b>145</b>	<b>777</b>
Sul	Pequena	2.141.961	54	103
	Média	10.235.644	34	180
	Grande	8.422.241	3	57
<b>Subtotal</b>		<b>20.799.846</b>	<b>91</b>	<b>340</b>
Centro - Oeste	Pequena	731.087	18	36
	Média	2.977.171	10	86
	Grande	8.586.974	4	50
<b>Subtotal</b>		<b>12.295.232</b>	<b>32</b>	<b>172</b>
<b>Total</b>		<b>153.961.441</b>	<b>371</b>	<b>1991</b>

Fonte: Elaboração própria a partir dos dados das Estimativas da População e do REGIC (IBGE, 2018).

Após definida a noção de áreas de mercado de trabalho local a ser utilizada nesta dissertação, combinou-se os dados das 371 LMA com os dados da RAIS, através do código dos municípios, o que resultou na exclusão de 1 pequeno arranjo formado por 2 municípios e 63 municípios de outras LMA. Portanto, a base final, definida após realização do bloco 6 de tratamento da subseção 3.2.2, conta com 370 áreas de mercado de trabalho local, compostas por 1.926 municípios. Uma vez definidas essas áreas de mercado de trabalho local e encerrados todos os blocos de tratamento, ou seja, uma vez definida a base de dados final

desta dissertação, foram construídas as variáveis nela utilizadas. A subseção 3.2.4 descreve as essas variáveis.

### 3.2.4 Descrição das Variáveis

Esta subseção descreve o processo de construção das variáveis utilizadas nesta pesquisa, após os tratamentos aplicados à base de dados da RAIS identificada e da definição das áreas de mercado de trabalho local. Elas se referem a três aos trabalhadores, às firmas e à localização do emprego. No nível do trabalhador, encontra-se a variável dependente utilizada nos modelos estimados nesta dissertação, o logaritmo natural do salário-hora real do trabalhador em dezembro de cada ano, cuja construção será apresentada nos próximos dois parágrafos.

Para construção dessa variável, inicialmente, deflacionou-se a remuneração<sup>33</sup> nominal recebida pelo trabalhador em dezembro de cada ano, utilizando o Índice de Preços Amplo ao Consumidor- IPCA, calculado pelo IBGE, tendo 2018 como ano base. Atualmente, o IPCA é calculado para 11 regiões metropolitanas e 5 municípios<sup>34</sup>. Seguindo critério adotado por Freguglia (2007)<sup>35</sup>, o IPCA das regiões metropolitanas e dos municípios foram extrapolados para seus respectivos estados, e para os estados não contemplados com o IPCA, adotou-se o critério de aproximação pelo índice da região metropolitana/município mais próximo. Neste caso, o IPCA de Belém foi utilizado para todos os estados da região Norte; o de Fortaleza para Ceará, Piauí, Maranhão e Rio Grande do Norte; o de Recife para Pernambuco, Paraíba, Alagoas e Sergipe; o de Campo Grande para Mato Grosso do Sul e Mato Grosso<sup>36</sup>. A **Tabela C.1** do Apêndice C mostra o índice de correção do IPCA adotado para deflacionar os salários.

Uma vez calculado o salário real, prosseguiu-se a construção da variável salário-hora real para construção da variável dependente. O critério utilizado para construir essa variável é baseado no artigo nº 64 da CLT, que também é o mesmo utilizado pelo governo federal para definir o valor mínimo da hora de trabalho (BRASIL, 2017). Disto isto, o salário real foi dividido por cinco vezes a quantidade de horas semanais contratadas como pode ser visto na **equação 3.15**:

$$\text{Salário} - \text{hora real} = \frac{\text{Salário real}_{\text{dezembro}}}{(5 \times \text{Quantidade de horas semanais contratadas})} \quad (3.15)$$

<sup>33</sup> A remuneração mensal informada na RAIS exclui o valor pago referente às horas extras, às férias e ao 13º.

<sup>34</sup> As regiões metropolitanas são: Belém, Fortaleza, Recife, Salvador, Belo Horizonte, Vitória (desde dezembro de 2013), Rio de Janeiro, São Paulo, Curitiba, Porto Alegre e Brasília. Os municípios são: Goiânia, Campo Grande (desde dezembro de 2013), Rio Branco, São Luís e Aracaju (estes três últimos desde março de 2018).

<sup>35</sup> Este procedimento também foi adotado por Neves Júnior (2018), Silva (2019) e Silva (2017)

Após este cálculo, apenas aplicou-se o logaritmo natural no salário-hora real.

Ainda no nível do trabalhador, extraiu-se da base da RAIS, a variável idade (18-65 anos), idade ao quadrado, experiência (tempo de emprego, em meses, no mesmo vínculo), experiência ao quadrado, gênero, raça (raça/cor), e educação. A variável raça é composta das seguintes categorias: indígena, branca, preta, amarela, parda, não identificada, multirracial. A categoria multirracial diz respeito aos indivíduos que mudaram seu pertencimento racial no decorrer do período. A variável educação está agrupada em 4 ciclos, a partir dos 11 níveis disponíveis na base da RAIS, são eles: Básico 1 (analfabeto, até o 5º ano incompleto do ensino fundamental e 5º ano completo do ensino fundamental); Básico 2 (do 6º ao 9º ano incompleto do ensino fundamental, ensino fundamental completo e ensino médio incompleto); Médio (ensino médio completo e ensino superior incompleto) e Superior (educação superior completa, mestrado completo e doutorado completo).

Por fim, extraiu-se da RAIS a variável grupo ocupacional. A ocupação foi definida de acordo com a Classificação Brasileira de Ocupações 2002 (CBO 2002), que define 10 grandes grupos ocupacionais. Nesta pesquisa utilizou-se 9 deles<sup>37</sup>, de modo que foram excluídos o Grupo 0 referente às forças armadas, policiais e bombeiros, e as ocupações referentes aos membros superiores do poder público do Grupo 1, com vistas a excluir da análise o setor público.

Em relação à firma, as variáveis consideradas no estudo são duas: setor de atividade econômica da firma e tamanho da firma. O setor da firma foi definido a partir da Classificação Nacional de Atividades Econômicas 2.0 (CNAE 2.0) do IBGE. Em decorrência da Etapa III do bloco 1 de tratamento, serão considerados nesta pesquisa 14 setores de atividade econômica dos 21 definidos pela CNAE 2.0<sup>38</sup>. Esta estratégia metodológica baseia-se em Amarante (2017), que considera em sua análise apenas o setor privado e as entidades empresariais, excluindo as entidades de economia mista, com vistas a garantir uma análise

---

<sup>37</sup> Grupo 1 (Dirigentes de empresas, gerentes, aqui se excluiu as categorias do setor público); Grupo 2 (Profissionais das ciências e das artes); Grupo 3 (Técnicos de nível médio); Grupo 4 (Trabalhadores de serviços administrativos); Grupo 5 (Trabalhadores dos serviços, vendedores do comércio em lojas e mercado); Grupo 6 (Trabalhadores agropecuários, florestais e da pesca); Grupo 7 (Trabalhadores da produção de bens e serviços industriais 1); Grupo 8 (Trabalhadores da produção de bens e serviços industriais 2); Grupo 9 (Trabalhadores em serviços de reparação e manutenção).

<sup>38</sup> Setor 1 (Agricultura, pecuária, produção florestal, pesca e aquicultura); Setor 2 (Indústrias extrativistas); Setor 3 (Indústrias de Transformação); Setor 4 (Eletricidade e Gás); Setor 5 (Água, esgoto, atividades de gestão de resíduos e descontaminação); Setor 6 (Construção); Setor 7 (Comércio; reparação de veículos automotores e motocicletas); Setor 8 (Transporte, armazenagem e correio); Setor 9 (Alojamento e alimentação); Setor 10 (Informação e Comunicação); Setor 11 (Atividades financeiras, de seguros e serviços relacionados); Setor 12 (Atividades imobiliárias); Setor 13 (Atividades profissionais, científicas e técnicas) e Setor 14 (Atividades administrativas e serviços complementares).

mais robusta sobre os diferenciais salariais, já que as regras de fixação de salários adotadas pelas entidades empresariais diferem consideravelmente dos outros tipos de entidades jurídicas e parece adequa-se melhor ao modelo base deste estudo, o modelo de Combes, Gobillon e Duranton (2008).

O tamanho da firma é determinado a partir da quantidade de trabalhadores ativos que o estabelecimento possui e do seu setor de atuação<sup>39</sup>: Microempresa (se comércio e serviços: até 9 empregados, se indústria e construção: até 19); Empresa Pequena (se comércio e serviços: de 10 a 49 empregados, se indústria e construção: de 20 a 99); Empresa Média (se comércio e serviços: de 50 a 99 empregados, se indústria e construção: de 100 a 499) e Empresa Grande (se comércio e serviços: 100 ou mais empregados, se indústria e construção: 500 ou mais).

Em relação à localização do emprego, definiram-se quatro variáveis. A primeira delas é a variável de interesse desta dissertação, a logaritmo natural densidade do emprego, que é calculada a partir da razão entre o número total de empregos e a área em quilômetros quadrados ( $\text{Km}^2$ ) da área de mercado de trabalho local em questão e, em seguida, logaritmizada. A segunda variável é o logaritmo natural da área, em  $\text{km}^2$ , da LMA. A terceira variável é o tamanho da área de mercado de trabalho local que está classificado em: pequeno, médio e grande, conforme critério populacional definido na subseção 3.2.3. E, por fim, a quarta variável diz respeito à macrorregião em que a LMA está localizada. A **Tabela C.2** do Apêndice C mostra de maneira mais detalhada as variáveis descritas nesta subseção.

### 3.3 ESTRATÉGIA EMPÍRICA

Esta seção tem por finalidade apresentar as principais estratégias empíricas utilizadas para se chegar aos resultados finais desta pesquisa. Ela divide-se em três subseções. A primeira apresenta o modelo econométrico e suas principais nuances. A segunda aborda o método de estimação adotado. E, por fim, a terceira subseção apresenta a estratégia de identificação utilizada para estimar os resultados desta dissertação.

#### 3.3.1 Modelo Econométrico

A fim de capturar os efeitos das economias de aglomeração sobre os salários dos trabalhadores do mercado de trabalho formal brasileiro, no período de 2010 a 2018, este estudo utiliza como referência o modelo de Combes, Duranton e Gobillon (2008). Para controlar os efeitos do *sorting* espacial de trabalhadores e de firmas e a *match* desses efeitos,

---

<sup>39</sup> Essa classificação para os dados da RAIS foi determinada no estudo Anuário do Trabalho na micro e Pequena Empresa organizado pelo Serviço de Apoio às micro e Pequenas Empresas (SEBRAE), que foi elaborado pelo Departamento Intersindical de Estatísticas e Estudos Socioeconômicos (DIEESE).

isto é, o *matching* assortativo, é utilizado um modelo de dados em painel. A especificação 3.19 é a principal especificação desta dissertação por controlar esses efeitos.

Contudo, antes de estimar a especificação principal, são estimadas equações que não controlam os efeitos do *sorting* espacial dos trabalhadores e das firmas e o *matching* assortativo, simultaneamente, a fim de acompanhar o que acontece com a magnitude do coeficiente do  $\ln$  da densidade do emprego à medida que se incluem esses efeitos fixos.

O primeiro modelo a ser estimado considera apenas os efeitos fixos de tempo e especificado da seguinte maneira:

$$\ln(W_{it}) = \beta_1 \ln(DensEmp)_{at} + \beta_2 X_{it} + \beta_3 F_{jt} + \beta_4 LMA_{at} + \beta_5 T_t + \varepsilon_{it} \quad (3.16)$$

Em que  $\ln(W_{it})$  é a variável dependente, o logaritmo natural do salário-hora real do indivíduo  $i$ , no tempo  $t = \{2010, \dots, 2018\}$ ;  $\ln(DensEmp)_{at}$  representa o logaritmo natural da densidade do emprego da LMA  $a = \{1, \dots, 370\}$ , no tempo  $t$ , que serve como *proxy* para capturar o efeito geral das economias sobre os salários;  $X_{it}$  é um vetor de variáveis referentes ao trabalhador  $i$  (idade, idade<sup>2</sup>, experiência, experiência<sup>2</sup>, gênero, raça, educação e ocupação), no tempo  $t$ ;  $F_{jt}$  é um vetor de características relacionadas à firma  $j$  (tamanho da firma e setor da firma) no tempo  $t$ ;  $LMA_{at}$  é um vetor de variáveis que diz respeito à LMA  $a$  (logaritmo da área da LMA, em km<sup>2</sup>, tamanho da LMA e macrorregião) no tempo  $t$  e  $\beta_5 T_t$  representam *dummies* de tempo.

O segundo modelo controla o *sorting* espacial de trabalhadores, isto é, inclui os efeitos fixos de indivíduo na estimação, controlando as heterogeneidades não observadas. Com essa inclusão, os efeitos da autoseleção entre a mão de obra qualificada e os mercados de trabalho locais, que podem enviesar as estimativas dos efeitos da aglomeração, são eliminados (COMBES; GOBILLON; DURANTON, 2008). O modelo é estimado a partir da seguinte equação:

$$\ln(W_{it}) = \beta_1 \ln(DensEmp)_{at} + \beta_2 X_{it} + \beta_3 F_{jt} + \beta_4 LMA_{at} + \beta_5 T_t + \delta_i + \varepsilon_{it} \quad (3.17)$$

O parâmetro  $\delta_i$  representa o efeito fixo de indivíduo.

Assim como os indivíduos, as firmas também podem se autosselecionar, buscando se localizar em locais que otimizem sua produção. Dessa forma, uma maior produtividade nas LMA pode ocorrer em decorrência dessa autoseleção ou de heterogeneidades da firma, tais como, gestão (COMBES; GOBILLON, 2015). Se isso ocorre, então, os efeitos das economias de aglomeração sobre os salários podem estar capturando também este *sorting* espacial de firmas e, portanto, enviesando os resultados das estimativas dos diferenciais salariais.

O controle do *sorting* espacial de firmas tem estado presente em poucos estudos. Isto ocorre, conforme cita Neves Júnior (2018), principalmente, porque os dados que permitem tal controle até pouco tempo não estavam ao alcance dos pesquisadores. A importância do efeito de firma nos estudos dos diferenciais salariais ganha destaque com Abowd, Kramarz e Margolis (1999) que desenvolveram um modelo seminal de decomposição salarial<sup>40</sup>, a fim de verificar o diferencial setorial de salários na França. Essa decomposição decorre dos resultados heterogêneos do mercado de trabalho percebidos nas diferenças de remunerações e de histórico de emprego que indivíduos notavelmente equivalentes possuem. Essa heterogeneidade pode estar relacionada às diferenças permanentes e não observáveis entre os indivíduos, isto é, o efeito de indivíduo e às diferenças permanentes entre os empregadores, ou seja, o efeito de firma.

Dessa forma, será adicionado ao modelo base desta pesquisa, cuja referência é o modelo de Combes, Duranton e Gobillon (2008), o controle do *sorting* espacial de firmas. Esse controle se dará por meio do controle das heterogeneidades não observadas das firmas (SILVA, 2017; CAMPOS, 2018; NEVES JÚNIOR, 2018), de modo que a estimação a ser feita é a seguinte:

$$\ln(W_{it}) = \beta_1 \ln(DensEmp)_{at} + \beta_2 X_{it} + \beta_3 F_{jt} + \beta_4 LMA_{at} + \beta_5 T_t + \delta_i + \varphi_j + \varepsilon_{it} \quad (3.18)$$

Em que  $\varphi_j$  captura o *sorting* espacial das firmas. A estimação 3.18 controla, simultaneamente, o *sorting* espacial dos indivíduos e das firmas.

Como já discutido na revisão da literatura e também no modelo de Combes, Duranton e Gobillon (2008), outro ponto que pode levar a aumentos de produtividade em uma determinada área é a interação simultânea entre as características dos trabalhadores e das firmas. Para Mion e Naticchione (2009) essa interação - *matching* assortivo (*assortative matching*), deve ser considerada nos estudos dos efeitos das economias de aglomeração sobre os salários. A omissão do *match* trabalhador-firma pode enviesar as estimativas (WOODCOCK; 2008; 2015). Diante disto, será adicionada ao modelo uma variável que capture o efeito desta combinação ( $\vartheta_{ij}$ ), de acordo com a seguinte especificação:

$$\ln(W_{it}) = \beta_1 \ln(DensEmp)_{at} + \beta_2 X_{it} + \beta_3 F_{jt} + \beta_4 LMA_{at} + \beta_5 T_t + \delta_i + \varphi_j + \vartheta_{ij} + \varepsilon_{it} \quad (3.19)$$

O modelo 3.19 é o modelo de interesse desta dissertação. E, assim como todas as demais especificações descritas nesta subseção, será estimado em único estágio como feito em Rosenthal e Strange (2008), Campos (2018) e Pereira (2018).

---

<sup>40</sup> Para mais detalhes sobre o modelo ver Abowd, Kramarz e Margolis (1999).

Além dos modelos já definidos anteriormente nesta seção, o modelo principal desta dissertação será estimado para cada região do Brasil e também para cada tamanho de LMA. Com isto, busca-se atingir os objetivos específicos deste estudo, demonstrando como a densidade do emprego se comporta nos diferentes recortes regionais e dimensionais país afora.

É importante salientar que, ao se controlar esses efeitos fixos, é possível que não apenas os elementos associados ao *sorting* espacial de trabalhadores e de firmas e ao *matching* assortativo sejam eliminados da análise de regressão, mas também parte dos efeitos da própria aglomeração, já que o efeito do *sorting* é considerado por algumas abordagens teóricas como sendo uma parcela importante do efeito de aglomeração (SILVA, 2017; NEVES JÚNIOR, 2018). Dessa forma, o efeito mensurado pela densidade do emprego (*proxy* utilizada para as economias de aglomeração) estaria refletindo o efeito líquido das economias de aglomeração sobre os salários, desconsiderando os componentes não observáveis que constituem as economias de aglomeração, como o *sorting* espacial os trabalhadores e firmas e o *match* trabalhador-firma.

Mesmo controlando esses efeitos fixos, as estimativas dos modelos podem continuar viesadas. Isso decorre da possível endogeneidade entre a densidade do emprego e os salários. Sendo assim, para contornar esse problema, serão feitas estimações por VI. Assim, espera-se encontrar resultados mais robustos.

Antes de concluir esta subseção, é preciso destacar dois aspectos importantes sobre a estimação dos modelos apresentados. O primeiro diz respeito às condições necessárias para se estimar as equações por meio dos efeitos fixos e a segunda se refere à factibilidade de se estimar equações lineares com múltiplos efeitos fixos. A identificação dos efeitos fixos, conforme Combes, Duranton e Gobillon (2008; 2011), requer suficiente mobilidade na amostra. Sendo assim, para capturar os efeitos fixos dos trabalhadores, das firmas e o *match* trabalhador-firma é preciso que haja mobilidade dos indivíduos entre firmas e LMA. De maneira análoga, para capturar os efeitos das firmas é preciso que haja alguma mobilidade das firmas entre as áreas de mercado de trabalho local. Sobre as estimações com múltiplos efeitos fixos, elas são realizadas a partir do algoritmo desenvolvido por Correia (2014) para o *STATA*, com base na abordagem iterativa proposta por Guimarães e Portugal (2010) para estimar modelos lineares com grande dimensão de efeitos fixos. Estes aspectos são melhores detalhados nas subseções seguintes.

### 3.3.2 Estratégias de Identificação

As literaturas internacional e nacional têm avançado na identificação dos efeitos das economias de aglomeração sobre os salários. Contudo, como destacado por Combes, Duranton e Gobillon (2008; 2011), dois problemas principais podem prejudicar essa identificação. O primeiro diz respeito à autosseleção (*sorting*) de indivíduos e de firmas para mercados de trabalho mais produtivos, o que pode afetar os salários. O segundo problema é referente à possível endogeneidade da variável densidade do emprego, a causalidade reversa entre a densidade do emprego e os salários podem enviesar as estimativas das economias de aglomeração sobre os salários. Em face disso, a estratégia de identificação adotada nesta dissertação visa encontrar as estimativas menos viesadas possíveis dos efeitos das economias de aglomeração sobre os salários no mercado de trabalho formal brasileiro.

Com a maior disponibilidade de dados sobre o mercado de trabalho se tornou possível controlar o efeito de variáveis não observadas de trabalhadores e de firmas, por meio da inclusão de efeitos fixos nos modelos. Com essa inclusão, os estudos têm buscado controlar a heterogeneidade não observada dos indivíduos (GLAESER; MARÉ, 2001; COMBES; DURANTON; GOBILLON, 2008; MION; NATICCHIONI, 2009; MATANO; NATICCHIONI, 2016; SILVA, 2019) e, mais recente, das firmas (GROOT; GROOT; SMIT, 2014; SILVA, 2017; CAMPOS, 2018; NEVES JÚNIOR, 2018). A inclusão dos efeitos fixos de indivíduos e de firmas também permite controlar a autosseleção dos indivíduos para centros urbanos mais densos e para as firmas mais produtivas e das firmas para locais mais produtivos. Dados mais desagregados do mercado de trabalho, como é o caso dos microdados da RAIS, utilizados nesta pesquisa, possibilitam ainda uma compreensão melhor do papel do espaço na determinação salarial.

No que diz respeito ao *sorting* espacial, a estratégia adotada consiste na inclusão dos efeitos fixos dos trabalhadores, das firmas e do *match* trabalhador-firma nos modelos aqui estimados. A estrutura de dados em painel que combina informações de trabalhadores e de firmas permite que esses efeitos fixos sejam controlados. Contudo, para que ocorra esse controle, algumas hipóteses fundamentais precisam ser atendidas. Para identificar os efeitos fixos é preciso que haja variação interna (*within*) dos indivíduos ou ao longo do tempo, pois com pouca variação *within* o coeficiente de um regressor será estimado de forma imprecisa, e não será identificado se essa variação for zero (CAMERON; TRIVEDI, 2009). Isso faz todo sentido, pois devido a uma transformação *within*, na estimação por Efeitos Fixos (EF), as

variáveis explicativas constantes no tempo apresentam variação *within* iguala a zero e são excluídas (WOOLDRIDGE, 2010).

De acordo com o que foi apresentado no parágrafo anterior, a identificação dos efeitos fixos dos trabalhadores requer mobilidade destes entre firmas e entre áreas de mercado de trabalho local. A identificação dos efeitos fixos das firmas requer mobilidade destas entre as LMA. Além da identificação desses efeitos fixos, é preciso considerar também que o *match* entre eles pode enviesar as estimativas dos efeitos das economias de aglomeração sobre os salários. Neste sentido, será feito o controle dos efeitos fixos do *match* trabalhador-firma, assim como fizeram Campos (2018) e Neves Júnior (2018).

A identificação de todos esses efeitos fixos, por si só, não garante que as estimativas encontradas sejam as menos enviesadas possíveis. Isto, porque, como bem documentado na literatura, a possibilidade de causalidade reversa entre a densidade do emprego e os salários é muito provável. Neste sentido, resultados mais robustos só serão encontrados se esta endogeneidade for tratada.

A simultaneidade entre a densidade do emprego e os salários locais ocorre sempre que os trabalhadores fazem sua escolha de emprego com base nos salários ofertados por determinada área de mercado de trabalho local. Para Ciccone e Hall (1996), essa escolha não exógena pode ser fruto de características naturais inerentes à LMA (clima, geografia) que aumentam o salário local, atraindo mais trabalhadores, elevando a densidade do emprego na LMA. Ao mesmo tempo, essa maior densidade do emprego pode ampliar a produtividade e, conseqüentemente, os salários. Em resumo, os salários locais podem determinar a amplitude da densidade do emprego que, por sua vez, pode influenciar na determinação dos salários locais.

A fim de contornar o viés de endogeneidade da densidade do emprego, a técnica recorrente na literatura tem sido a adotar o método de estimação por Variáveis Instrumentais (VI). Este método, que será descrito na subseção 3.3.4, fornece uma solução geral para o problema de uma variável explicativa endógena (WOOLDRIDGE, 2010), como é o caso da densidade do emprego. O método de VI consiste em adotar um instrumento ou uma *proxy* para a variável endógena (densidade do emprego) que não seja correlacionada com a variável dependente (salário), mas seja correlacionada com a variável endógena. Os estudos têm usado diferentes instrumentos para lidar com a endogeneidade presente nas análises sobre economias de aglomeração e a produtividade (salários, neste estudo).

Os instrumentos mais comuns na literatura são o *log* defasado das variáveis endógenas, tais como, população/densidade do emprego (CICCONE; HALL, 1996;

COMBES; DURANTON; GOBILLON, 2008; MION; NATTICHIONI, 2009; GROOT; GROOT; SMIT, 2014; BARUFI, 2015; PEREIRA, 2018; LEKNES; RATTSSØ; STOKKE, 2020), o uso de variáveis geográficas (CICCONE, 2002; ROSENTHAL; STRANGE, 2008; COMBES *et al*, 2012; BARUFI, 2015) e do instrumento de *shift-share* de Bartik (SILVA, 2017; SILVA, 2019). De modo geral, os resultados encontrados por esses estudos, após o uso de VI, corroboram com Combes, Duranton e Gobillon (2011) e revelam que o impacto da variável de aglomeração nos salários é, apenas, ligeiramente afetado pelo viés de endogeneidade.

Esta dissertação utiliza como instrumento para densidade do emprego a densidade da população defasada das áreas de mercado de trabalho local, em logaritmo. Para isto, utiliza-se os dados populacionais e de território, fornecidos pelo IBGE, dos anos 1920, 1940 e 1950. O pressuposto para que a densidade defasada seja um bom instrumento é que ela esteja correlacionada com a densidade do emprego atual, *proxy* para as economias de aglomeração, mas não com o termo de erro estocástico das equações estimadas. A subseção 3.3.3.2 traz mais detalhes sobre essa questão.

De maneira geral, esta subseção mostrou as principais estratégias de identificação adotadas para estimação adequada dos efeitos das economias de aglomeração sobre os salários. Em relação às duas principais fontes de viés aqui discutidas, Combes, Duranton e Gobillon (2011) destacam que o *sorting* espacial de firmas e trabalhadores afeta fortemente as estimativas entre economias de aglomeração e salários, enquanto a endogeneidade afeta apenas de maneira leve os resultados. Mesmo assim, este trabalho busca tratar essas duas fontes de viés para obtenção de estimativas mais robustas.

A próxima subseção apresenta os métodos de estimação para lidar com os problemas de identificação aqui relatados. Inicialmente, apresentam-se os métodos para controle dos efeitos fixos e, em seguida, o método de VI para lidar com a endogeneidade da densidade do emprego.

### **3.3.3 Métodos de Estimação**

Para mensurar o efeito das economias de aglomeração sobre os salários individuais no mercado de trabalho formal brasileiro, de modo a controlar o efeito do *sorting* espacial dos trabalhadores e das firmas e o do *matching* desses efeitos, é necessário adotar um método que controle os efeitos fixos. A base de dados em painel construída para esta dissertação permite a utilização de diferentes métodos de estimação, tais como Mínimos Quadrados Ordinários

Empilhados (*Pooled Ordinary Least Squares* – POLS), Efeitos Aleatórios (EA) e Efeitos Fixos (EF).

Inicialmente, os efeitos das economias de aglomeração serão mensurados sem a inclusão de efeitos fixos, por meio do método POLS, obtendo, assim, o seu efeito bruto sobre os salários. Para que este método reporte estimativas consistentes, duas hipóteses principais devem ser atendidas (WOOLDRIDGE, 2010):

$$E(\mathbf{x}'_t \varepsilon_t) = 0, t = 1, 2, \dots, T. \quad (3.20)$$

$$\text{posto} \left[ \sum_{t=1}^T E(\mathbf{x}'_t \mathbf{x}_t) \right] = K. \quad (3.21)$$

A suposição 3.20 garante a exogeneidade, isto é, o termo de erro não está correlacionado com as variáveis explicativas do modelo, contidas na matriz  $\mathbf{x}$ . A suposição 3.21 garante a condição de posto pleno, de modo que dependências lineares perfeitas entre as variáveis explicativas não existem. O estimador POLS será consistente e assintoticamente normal se essas duas suposições forem satisfeitas.

O modelo proposto nesta dissertação considera a existência de efeitos não observados dos indivíduos, das firmas e do *match* indivíduo-firma. Em decorrência disso, as estimativas POLS serão enviesadas, já que, neste caso, esses efeitos não observados estariam contidos no termo de erro, o que violaria as duas suposições descritas no parágrafo anterior.

Para lidar com a presença desses efeitos não observados ( $c_i$ ), o estimador de Efeitos Aleatórios (EA) pode ser adotado. Assim como, o POLS, o método de EA considera o efeito das variáveis não observadas como estando presente no termo de erro, porém com mais suposições (WOOLDRIDGE, 2010) que o tornam um estimador eficiente, tais como:

$$E(\varepsilon_{it} | \mathbf{x}_i, c_i) = 0, t = 1, 2, \dots, T. \quad (3.22)$$

$$E(c_i | \mathbf{x}_i) = E(c_i) = 0, t = 1, 2, \dots, T. \quad (3.23)$$

Em que  $\mathbf{x}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{iT})$ .

A suposição 3.22 contempla a hipótese de exogeneidade estrita, isto é, as variáveis explicativas condicionadas aos efeitos não observados não são correlacionados com o termo de erro. A hipótese presente em 3.23 diz respeito à condição de ortogonalidade entre  $\mathbf{x}_i$  e  $c_i$ , ou seja, os efeitos não observados são independentes das variáveis explicativas do modelo. Nas especificações adotadas por este trabalho, os efeitos não observados, isto é, o *sorting* espacial dos trabalhadores e das firmas, são correlacionados com a densidade do emprego, *proxy* para as economias de aglomeração, o que viola a hipótese 3.23. Sendo assim, o estimador de EA não produzirá estimativas consistentes para os modelos propostos.

Diante da violação da hipótese 3.23, o estimador de Efeitos Fixos (EF) passa a ser o mais adequado para estimação dos modelos. Assim como no modelo de EA, a hipótese de

exogeneidade estrita entre o termo de erro e as variáveis explicativas condicionadas aos efeitos não observados ( $E(\varepsilon_{it}|\mathbf{x}_i, c_i) = 0, t = 1, 2, \dots, T$ ) é assegurada (WOOLDRIDGE, 2010). Na análise de EF, os efeitos não observados constantes no tempo podem estar correlacionados com as variáveis explicativas observadas, sem que isso comprometa a consistência das estimativas encontradas (WOOLDRIDGE, 2010). Por este motivo, a análise de EF, conforme Wooldridge (2010), é mais robusta do que a análise de EA. Contudo, o autor pontua que essa robustez tem um preço: a não inclusão de variáveis explicativas constantes no tempo (gênero, por exemplo). Isso ocorre para que não se confundam os efeitos das variáveis explicativas constantes no tempo com os efeitos das variáveis omitidas constantes no tempo, tendo em vista a possibilidade de correlação entre essas variáveis.

Ainda que o uso dos estimadores de EF seja consenso na literatura em que se insere esta dissertação, a escolha do estimador a ser utilizado nas análises deste estudo será feita pelo teste de especificação de Hausman, que compara os estimadores de EF e EA. A hipótese nula subjacente ao teste de Hausman é que os estimadores de EF e EA não diferem substancialmente; se a hipótese nula for rejeitada, então os efeitos aleatórios (não observados) provavelmente estão correlacionados com uma ou mais variáveis observadas do modelo, o que torna o estimador de EF preferível ao de EA (GUJARATI; PORTER, 2011). Em outras palavras, se a hipótese nula é rejeitada, então a hipótese 3.23 foi violada, o que torna as estimativas de EA inconsistentes. Após realização deste teste, será escolhido o melhor estimador, que também será o único cujos resultados serão apresentados.

Como já apresentado na subseção 3.3.2, as especificações dos modelos requer a estimação, simultânea, de mais de um efeito fixo, em alguns casos, considera-se também, a combinação entre os efeitos fixos dos trabalhadores e das firmas. Em decorrência disso, é necessário realizar as estimações a partir de uma metodologia que capture esses múltiplos efeitos fixos. Sendo assim, as estimações desta dissertação são realizadas por meio do algoritmo desenvolvido no *STATA* por Correia (2014) com base nas abordagens propostas por Guimarães e Portugal (2010) e por Gaure (2013). No *STATA* este algoritmo é denominado *Linear Regression Absorbing Multiple Levels of Fixed Effects (reghdfe)*. A próxima subseção apresenta de maneira mais detalhada esse algoritmo.

### **3.3.3.1 Estimador de Múltiplos Efeitos Fixos**

Recentemente, as pesquisas que envolvem diferenciais salariais têm contado com uma grande disponibilidade de dados, cada vez mais detalhados, como é o caso dos microdados da RAIS, que permitem a utilização de dados em painel com acompanhamento ao longo dos anos

de dados trabalhador-firma, como no modelo proposto por Abowd, Kramarz e Margolis (1999). Essa maior disponibilidade tem estimulado o interesse em métodos de estimação com efeitos fixos de alta dimensão (GUIMARÃES: PORTUGAL, 2010).

O primeiro trabalho que busca lidar com a inclusão de mais de um efeito fixo é desenvolvido por Abowd, Kramarz, e Margolis (1999), que buscam, a partir de uma base de dados trabalhador-firma, controlar, simultaneamente, duas fontes de heterogeneidade não observada - das firmas e dos trabalhadores. Os autores propuseram uma série de métodos a fim de se aproximar o máximo possível da solução de Mínimos Quadrados Ordinários (MQO). Pouco tempo depois, Abowd, Creecy, e Kramarz (2002) apresentaram um algoritmo iterativo capaz de controlar simultaneamente os efeitos fixos dos trabalhadores e das firmas nas estimações via MQO.

Para lidar com modelos lineares de dois ou mais efeitos fixos, Wooldridge (2010) propõe aplicar uma transformação *within* dos grupos em relação ao efeito fixo com mais categorias e adicionar uma variável *dummy* para cada categoria fictícia de todos os efeitos fixos subsequentes, o que se mostra inviável para grandes bases de dados ou para um conjunto de efeitos fixos com muitas categorias (CORREIA, 2016). Tendo em vista que esta dissertação possui uma grande base de dados e busca realizar estimações que consideram os múltiplos efeitos fixos, a estratégia proposta por Wooldridge (2010) torna-se inviável de ser aplicada. Neste sentido, é preciso de um método capaz de lidar com esses dois fatores.

Recentemente, avanços têm sido empreendidos na literatura a fim de construir um estimador que lide com a abordagem de múltiplos efeitos fixos de forma factível e computacionalmente eficiente. Neste sentido, objetivando atender a estas condições, Guimarães e Portugal (2010) e Gaure (2013) construíram um elegante estimador combinando o Método de Projeções Alternadas (MPA) com o teorema de Frisch – Waugh – Lovell (FWL).

A ideia do teorema de FWL consiste em ajustar o modelo em duas etapas. Na primeira etapa, eliminam-se os efeitos fixos de todas as variáveis do modelo. Na segunda etapa, executa-se a regressão de interesse usando os resíduos armazenados das variáveis obtidas na primeira etapa, em vez das variáveis originais. Guimarães e Portugal (2010) destacam que algumas vantagens no uso desse teorema: i) as regressões simples da primeira etapa convergem em um ritmo mais rápido; ii) é possível testar diferentes especificações do modelo usando apenas as variáveis transformadas sem a necessidade de lidar com os efeitos fixos de alta dimensão; e iii) os requisitos computacionais para lidar com conjuntos de dados muito grandes são reduzidos.

Correia (2014) desenvolveu um algoritmo que implementa a abordagem iterativa proposta por Guimarães e Portugal (2010). O autor parte do seguinte modelo:

$$y = X\beta + D_1\alpha_1 + D_2\alpha_2 + D_3\alpha_3 + \epsilon \quad (3.24)$$

Em que  $D_i$  representa um conjunto de indicadores e  $\alpha_i$  é um conjunto de efeitos fixos associados a esses indicadores, o objetivo é obter a estimativa de MQO para  $\beta$ . O algoritmo segue Guimarães e Portugal (2010) e aplica uma interação de ponto fixo sobre as equações normais, isto é, define um valor e realiza a iteração das estimativas até que as equações normais sejam resolvidas. O autor, em vez de iterar as estimativas dos  $\hat{\alpha}_i$ , itera  $Z_i := D_i\alpha_i$ , do seguinte modo:

1- Calcula:  $P_1y$  e  $\tilde{y} = M_1y$ ;

2- Inicia com  $Z_2^{(0)} = Z_3^{(0)} = 0$ ;

3- Até  $Z_2$  e  $Z_3$  convergirem, define-se:

$$(a) Z_2^{(n)} = P_2 \left[ \tilde{y} + P_1 \left( Z_2^{(n-1)} + Z_3^{(n-1)} \right) - Z_3^{(n-1)} \right]$$

$$(b) Z_3^{(n)} = P_3 \left[ \tilde{y} + P_1 \left( Z_2^{(n)} + Z_3^{(n-1)} \right) - Z_2^{(n)} \right]$$

4- Então calcula  $Z_1 = P_1(y - Z_2 - Z_3)$ , e com ele calcula a regressão transformada  $y^* = y - Z_1 - Z_2 - Z_3$ , que o que se deseja

Depois de repetir este processo para cada variável, regride-se as variáveis transformadas ( $y^* = X^*\hat{\beta} + e$ ) para se obter as estimativas corretas de  $\hat{\beta}$ . Os resíduos da regressão FWL são os mesmos da regressão completa.

Tendo em vista que  $\hat{\beta}$  e  $e$  são os mesmos na regressão e na regressão original, para obter os efeitos fixos da regressão original, Correia (2014) calcula  $\tilde{e} := y - X\hat{\beta} = e + Z_1 + Z_2 + Z_3$  e, então, utiliza  $y = X\hat{\beta} + Z_1 + Z_2 + Z_3 + e$  para obter os efeitos fixos  $Z_1, Z_2$  e  $Z_3$  das variáveis  $D_1, D_2$  e  $D_3$ . O algoritmo proposto por Correia (2014) e aqui demonstrado pode ser estendido para um número maior de efeitos fixos e pode ser usado para modelos de variáveis instrumentais<sup>41</sup>, como é o caso deste estudo.

Com base nisto e considerando a forma matricial do modelo completo a ser estimado, especificado na equação 3.19, a solução de MQO para o modelo completo desta dissertação é a resolução do sistema de equações normais a seguir:

$$\beta = (X'X)^{-1}X'(W - \delta A - \varphi B - \vartheta C) \quad (3.25)$$

$$\delta = (A'A)^{-1}A'(W - \beta X - \varphi B - \vartheta C) \quad (3.26)$$

$$\varphi = (B'B)^{-1}B'(W - \beta X - \delta A - \vartheta C) \quad (3.27)$$

---

<sup>41</sup> Para mais detalhes, ver Correia (2014).

$$\vartheta = (C'C)^{-1}C'(W - \beta X - \delta A - \varphi B) \quad (3.28)$$

Em que  $W$  é um vetor  $N \times 1$  que representa a variável dependente, o logaritmo natural do salário-hora real;  $X$  é uma matriz  $N \times k$  que representa todas as variáveis explicativas do modelo (densidade do emprego, características dos trabalhadores, das firmas, das LMA e *dummies* de tempo) e  $A$ ,  $B$  e  $C$  representam matrizes com variáveis indicadoras de indivíduo, de firma e de indivíduo-firma, respectivamente. Já  $\beta$  é um vetor  $k \times 1$  de parâmetros a ser estimado para as variáveis explicativas, incluindo a variável densidade do emprego,  $\delta$ ,  $\varphi$  e  $\vartheta$  são vetores de efeitos fixos dos trabalhadores, das firmas e do *match* trabalhador-firma, respectivamente. A iteração entre esses conjuntos de equações fornece a solução exata do estimador de MQO (GUIMARÃES; PORTUGAL, 2010).

O algoritmo desenvolvido por Correia (2014) também pode ser utilizado para estimar pelo método de Variáveis Instrumentais. O método de VI, que será utilizado, como já discutido na subseção 3.3.2 para tratar a o viés de endogeneidade, é apresentado na subseção a seguir.

### 3.3.3.2 Variável Instrumental

Sabe-se que a inclusão dos efeitos fixos na estimação não garante a total correção dos vieses, principalmente, o viés de simultaneidade entre a densidade do emprego, *proxy* para as economias de aglomeração, e os salários. Se essa simultaneidade existe, a hipótese de exogeneidade ( $Cov(DensEmp_{at}, \varepsilon_{at})=0$ ), que garante estimativas não enviesadas, não se sustenta e, com isto, o estimador de MQO torna-se inconsistente (WOOLDRIDGE, 2010). Em situações como essa, Wooldridge (2010) propõe o uso do método de Variáveis Instrumentais (VI). Para o caso em questão, deve-se encontrar um instrumento ( $Z_{at}$ ) para a densidade do emprego de modo que ele não seja correlacionado com os salários, a não ser pelo efeito da densidade sobre os salários, mas seja correlacionado com a densidade do emprego.

O método de VI fornece estimadores consistentes se atender as seguintes condições:

$$Cov(Z_{at}, \varepsilon_{at}) = 0 \quad (3.29)$$

$$Cov(Z_{at}, DensEmp_{at}) \neq 0 \quad (3.30)$$

A condição 3.29 exige que o instrumento, a densidade da população defasada, não seja correlacionado com o termo de erro, ou seja, é preciso garantir a exogeneidade do instrumento, já que ela não pode ser testada (CAMERON; TRIVEDI, 2009), tendo em vista que o erro é não observado. A condição 3.30, que pode ser testada, requer que a variável

instrumental seja correlacionada com a variável endógena do modelo, neste caso, a densidade populacional defasada deve ser fortemente correlacionada com a densidade do emprego.

O logaritmo da densidade da população defasada é um instrumento amplamente aceito e utilizado na literatura sobre diferenciais salariais (CICCONE; Hall, 1996; BARUFI, 2015; DAUTH *et al*, 2018; PEREIRA, 2018; LEKNES; RATTSSØ; STOKKE, 2020). Argumenta-se que a densidade do emprego atual deve estar correlacionada com o padrão de concentração populacional do passado, contudo, os níveis de produtividade atuais não guardam relações com a distribuição passada da densidade da população. Desse modo, as condições requeridas para que uma variável seja considerada um bom instrumento são satisfeitas.

Uma vez definido o logaritmo natural da densidade populacional defasada no tempo como variável instrumental, procedeu-se a estimação por VI, por meio do método de Mínimos Quadrados em Dois Estágios (MQ2E). Como o próprio nome já diz, este procedimento é feito em dois estágios, o MQ2E consiste em duas aplicações sucessivas de MQO (GUJARATI; PORTER, 2011). No primeiro estágio, regride-se a variável explicativa endógena sobre todas as variáveis predeterminadas no sistema, como mostrado a seguir:

$$DensEmp_{a,t} = \beta_i \ln DensPop_{a,t} + \phi_i X_{a,t} + \epsilon_{a,t} \quad (3.31)$$

Em que  $\ln DensPop_{a,t}$  é o logaritmo da densidade populacional defasada na LMA  $a$  no tempo  $t$  que explica parte da densidade do emprego contemporâneo nesta mesma LMA,  $DensEmp_{a,t}$ ;  $X$  é uma matriz que contém todas as variáveis explicativas exógenas do modelo, e  $\epsilon_{a,t}$  é o resíduo do primeiro estágio. Ao fim do primeiro estágio, conforme Gujarati e Porter (2011), a correlação existente entre a densidade do emprego e o termo de erro estocástico, decorrente da endogeneidade entre a densidade do emprego e os salários, deixa de existir.

No segundo estágio do procedimento, o valor predito da densidade do emprego  $\widehat{DensEmp}_{a,t}$ , é incluído na regressão em substituição à variável endógena do modelo,  $DensEmp_{a,t}$ , de modo que:

$$\ln(W_{it}) = \beta_1 \widehat{DensEmp}_{a,t} + \tau_i X_{a,t} + v_{it} \quad (3.32)$$

Feito esse procedimento, obtêm-se estimativas consistentes do efeito da densidade do emprego sobre o diferencial salarial (GUJARATI e PORTER, 2011).

Segundo Wooldridge (2010), a estimação por MQ2E, realizada no *software STATA*, deve ser feita em um único comando, tendo em vista que estimar cada estágio explicitamente pode levar a erros, além disso os erros-padrão do MQO da regressão do segundo estágio são relatados de maneira incorreta. Ademais, o autor ressalta que, na presença de apenas um

instrumento, os estimadores de VI e MQ2E são idênticos, de modo que sempre que se falar de VI, a menos que se especifique o contrário, a estimação será feita por MQ2E.

É possível resumir os métodos de estimação adotados nesta dissertação em três grupos principais. O primeiro grupo estima por meio dos métodos tradicionais para a estrutura de dados em painel (POLS, EF, EA). O segundo grupo utiliza de um método específico para estimar equações com múltiplos efeitos fixos. E, por fim, o terceiro grupo utiliza os estimadores de VI para tratar os problemas de endogeneidade. As estimativas encontradas a partir da aplicação desses métodos, bem como as principais estatísticas descritivas da amostra, são apresentadas a seguir, no capítulo 4.

## 4 RESULTADOS

Neste capítulo são apresentados os resultados desta pesquisa. Ele está dividido em duas seções. Na primeira seção apresentam-se as principais estatísticas descritivas da amostra. Na segunda seção são apresentados os resultados das estimações econométricas realizadas neste estudo.

### 4.1 ESTATÍSTICAS DESCRITIVAS

As estatísticas descritivas apresentadas nesta seção são referentes à amostra de dados extraída dos microdados identificados da RAIS (2010-2018), que foram utilizados para gerar as estimativas desta dissertação. As análises são feitas a partir de um painel de dados longitudinal balanceado com 11.585.637 de observações referentes a 1.287.293 indivíduos e 520.455 firmas para o período de 2010 a 2018. É importante ressaltar que as estimativas presentes na seção 4.2 focam na relação entre as variáveis referentes aos salários e à densidade do emprego (atual e defasada no tempo), variável de interesse do presente estudo. Portanto, nesta seção o objetivo em descrever todas as variáveis é apresentar como se comportam as informações que estão sendo manipuladas nesta dissertação.

A **Tabela 4.1** apresenta os valores das estatísticas para as variáveis contínuas da base de dados, incluindo para o instrumento da densidade do emprego, o logaritmo natural da densidade populacional defasada. A fim de melhorar a compreensão, abaixo das estatísticas para os logaritmos naturais do salário, da densidade do emprego e da área das LMA foram colocados os valores dessas variáveis na sua forma não logaritmizada. Como já discutido na seção da metodologia, a variável dependente dos modelos, o salário-hora-real, foi deflacionado pelo IPCA, tendo como ano base, o ano de 2018. A média do salário-hora-real dos trabalhadores presentes na amostra é de R\$ 19,054. Como pode ser visto na **Tabela 4.1**, o desvio em relação à média é elevado, 20,836, indicando grandes disparidades salariais no mercado de trabalho formal brasileiro.

Em relação à variável de interesse, a média da densidade do emprego nas LMA é de 30,223 hab/km<sup>2</sup>. Ainda sobre a densidade do emprego, é importante destacar que a diferença entre a menor (0,0001 hab/km<sup>2</sup>) e a maior (121,3225 hab/km<sup>2</sup>) densidade do emprego é de 1.213.225 vezes, esta diferença é reforçada pelo alto valor do desvio padrão (39,174 hab/km<sup>2</sup>). Essas disparidades podem refletir na composição da mão de obra nas LMA, reforçando ou dispersando as forças aglomerativas, o que pode influenciar na determinação salarial local.

A **Tabela 4.1** também traz outras informações sobre o perfil dos trabalhadores formais. A idade média é de 38 anos, considerando o filtro feito para este trabalho, que considera apenas os trabalhadores em idade produtiva (18-65 anos). A experiência média, isto é, o tempo médio de emprego na mesma firma, é de 83 meses, com um desvio padrão elevado no valor de 72 meses. Ademais a tabela traz as estatísticas sobre a densidade populacional defasada e área das LMA.

**Tabela 4.1: Estatísticas descritivas das variáveis contínuas**

Variáveis	Média	Desvio Padrão	Mín.	Max.
<i>Ln</i> (salário-hora-real)	2,611	0,752	1,285	6,108
Salário-hora-real	19,054	20,836	3,616	449,469
Salário real	4042,148	4305,611	369,555	50000
<i>Ln</i> (densidade do emprego)	2,086	2,091	-9,468	4,798
Densidade do emprego	30,223	39,174	0,0001	121,3225
<i>Ln</i> (densidade populacional defasada)	3,971	1,593	-3,575	6,053
Idade	38,434	9,444	18	65
Idade <sup>2</sup>	1566,327	761,429	324	4225
Experiência	83,886	72,731	0,1	597,9
Experiência <sup>2</sup>	12326,671	21592,149	0,01	357484,41
<i>Ln</i> (Área da LMA, em km <sup>2</sup> )	17,848	2,181	5,545	20,773
Área da LMA (em km <sup>2</sup> )	10566,1	52259,15	7,224	564667,5

Fonte: Elaboração própria a partir dos microdados identificados da RAIS (2010-2018).

Nota: As estatísticas apresentadas são oriundas da amostra total, que possui 11.585.637.

Na **Tabela 4.2** apresenta-se a estrutura salarial conforme as categorias das variáveis ligadas diretamente ao indivíduo, apresentando também as frequências dessas variáveis. Em relação ao gênero, observa-se que a amostra é composta majoritariamente por trabalhadores do gênero masculino, aproximadamente 70%. Também são estes trabalhadores que recebem os maiores salários quando comparados ao gênero feminino (R\$ 20,09 dos homens contra R\$ 16,69 das mulheres). Trabalhadores brancos, com média de salário-hora de R\$ 22,36, representam a maioria das observações na amostra, 52,3%. Enquanto isso, os trabalhadores pretos, com a menor média salarial (R\$ 14,27), representam 3,2% da amostra e os trabalhadores pardos, com a segunda menor média salarial (R\$ 14,37), somam 18,5% das observações. Esses dados revelam que o mercado de trabalho brasileiro formal é composto majoritariamente por trabalhadores brancos e do gênero masculino e que os menores salários são auferidos por pretos e pardos, respectivamente, e pelas mulheres quando comparados às suas contrapartes.

A **Tabela 4.2** também mostra que 53,3%, isto é, a maioria do mercado de trabalho formal brasileiro é composta por trabalhadores com nível médio de instrução; em seguida estão os trabalhadores com nível básico 2 (23,2%), nível superior (18,6%) e nível básico 1 (4,8%). Em se tratando dos maiores salários, são os trabalhadores com nível superior que os

detêm, recebendo R\$ 42,30. Esse valor é quase 3 vezes superior a média do segundo maior salário (R\$ 15,07), recebidos por trabalhadores com nível médio, e mais de 4 vezes superior aos menores salários (R\$ 10,42), auferidos por indivíduos de nível básico 1. Com isso, é possível dizer que o mercado de trabalho formal do Brasil oferta maiores salários conforme o nível de educação, sendo os maiores salários ofertados àqueles trabalhadores com maiores níveis educacionais.

**Tabela 4.2: Rendimento médio e distribuição dos trabalhadores conforme suas características**

Variável	Média	Desvio Padrão	Observações	Frequência (%)
<b>Gênero</b>				
Feminino	16,686	18,692	3.527.982	30,50%
Masculino	20,092	21,627	8.057.655	69,50%
<b>Raça</b>				
Branca	22,359	23,703	6.064.812	52,30%
Indígena	17,580	20,214	7.839	0,10%
Preta	14,269	12,969	372.132	3,20%
Amarela	32,628	32,806	51.417	0,40%
Parda	14,371	14,688	2.147.139	18,50%
Não identificada	18,253	19,217	149.256	1,30%
Multirracial	15,914	17,177	2.793.042	24,10%
<b>Educação</b>				
Básico 1	10,418	67,185	561.54	4,80%
Básico 2	11,398	80,780	2.692.346	23,20%
Médio	15,072	13,192	6.178.358	53,30%
Superior	42,304	32,624	2.153.393	18,60%
<b>Ocupação</b>				
Trab.da prod. de bens e serv. Industriais 1	13,820	96,689	2.751.977	23,80%
Dir. de empresas e gerentes	43,249	39,604	833.193	7,20%
Prof. das ciên. e das artes	41,193	29,840	927.179	8%
Téc. nível médio	25,121	19,926	1.188.434	10,30%
Trab. serviços admin.	14,214	12,798	2.102.842	18,20%
Trab. serv e vend. do comérc. em lojas e mercad.	10,962	10,527	2.482.553	21,40%
Trab. agropec., florest. e da pesca	10,617	73,889	121.763	1,10%
Trab.da prod. de bens e serv. Industriais 2	15,903	14,480	706.904	6,10%
Trab. em serv. de reparação e manutenção	19,125	14,939	470.792	4,10%

Fonte: Elaborada pelo próprio autor a partir dos microdados identificados da RAIS (2010-2018).

Nota: A totalidade de cada variável categórica soma 11.585.637, total de observações da amostra.

Por fim, a **Tabela 4.2** mostra que em relação à ocupação, 23,8% são trabalhadores da produção de bens e serviços industriais 1; 21,4% trabalhadores dos serviços e vendedores do comércio em lojas e mercado e 18,2% são trabalhadores dos serviços administrativos. Em relação aos salários são os dirigentes de empresas e gerentes que recebem a maior média, R\$ 43,25, seguidos dos profissionais das ciências e das artes, com R\$ 41,19 de média salarial. Já os trabalhadores agropecuários, florestais e da pesca recebem os menores salários comparados aos trabalhadores das demais ocupações, em torno de R\$ 10,62, seguidos dos trabalhadores

dos serviços e vendedores do comércio em lojas e mercado, que recebem R\$ 10,96. Esses resultados mostram que os trabalhadores das ocupações mais complexas recebem os maiores salários, enquanto os trabalhadores das ocupações menos complexas recebem os menores salários.

Os resultados da **Tabela 4.2** mostram que os salários são distribuídos de maneira desigual conforme certas características observáveis dos indivíduos. Então, se uma LMA concentra uma força de trabalho mais escolarizada, de cor branca, do gênero masculino e de ocupações mais complexas, é possível que os salários dessa LMA sejam maiores do que em outras áreas de mercado de trabalho local. Esse maior salário pode levar indivíduos a buscar empregos nessas LMA (*sorting* espacial dos trabalhadores), impactando na determinação dos salários locais. Sendo assim, as diferenças salariais entre as LMA podem ser explicadas, ainda que parcialmente, pelo *sorting* espacial de trabalhadores.

A **Tabela 4.3** faz uma análise da estrutura salarial com base nas categorias das variáveis relacionadas de maneira mais direta com as firmas, apresentando também as a proporção das observações distribuídas entre as categorias dessas variáveis. A maioria dos trabalhadores está locada em firmas que atuam no setor da indústria da transformação (32,4%), em seguida aparece o setor do comércio; reparação de veículos automotores e motocicletas, com 26,4% das observações da amostra. Em termos de salário-hora os trabalhadores das firmas que atuam nos setores de eletricidade e gás (R\$ 36,19); atividades financeiras, de seguros e serviços relacionados (R\$ 35,70) e da indústria extrativista (R\$ 34,99) são os melhores remunerados em relação aos outros setores, respectivamente. Já os trabalhadores das firmas que atuam nos setores de alojamento e alimentação (R\$ 9,67); de água, esgoto, atividades de gestão de resíduos e descontaminação (R\$ 13,20) e de atividades administrativas e serviços complementares (R\$ 14,29) são os que auferem os menores salários, respectivamente.

A **Tabela 4.3** ainda mostra que a amostra é formada por trabalhadores que, em sua maioria, são empregados em firmas de grande porte, praticamente, 41,9% deles, logo adiante, estão os trabalhadores das pequenas firmas (25,1%), das firmas médias (17,9%) e das microempresas (15,1%). Em se tratando dos salários, os trabalhadores das firmas maiores recebem os maiores salários médios. As médias de salário-hora ofertadas pelas grandes firmas é de R\$ 22,69, pelas firmas médias é de R\$ 20,81, pelas firmas pequenas é de R\$ 16,21 e pelas microempresas é de R\$ 11,63.

**Tabela 4.3: Rendimento médio e distribuição dos trabalhadores conforme características das firmas**

Variável	Média	Desvio Padrão	Observações	Frequência (%)
<b>Setor</b>				
Indústria de Transformação	21,544	21,125	3.759.070	32,40%
Indústria extrativista	34,988	34,412	88.84	0,80%
Agricultura, pecuária, prod. florestal, pesca e aquíc.	15,137	17,209	93.784	0,80%
Eletricidade e gás	36,186	27,240	86.188	0,70%
Água, esgoto, ativ. de gestão de resíd. e descont.	13,200	13,373	68.192	0,60%
Construção	16,975	16,278	419.738	3,60%
Comércio; Reparação de veículos auto. e motocic.	14,345	16,273	3.057.088	26,40%
Transporte, armazenagem e correio	17,257	19,243	1.094.664	9,40%
Alojamento e alimentação	9,671	7,804	417.342	3,60%
Informação e comunicação	32,987	29,350	410.987	3,50%
Atividades financ., de seguros e serv. relacion.	35,696	28,460	507.203	4,40%
Atividades imobiliárias	16,023	17,851	52.515	0,50%
Atividades profissionais, científicas e técnicas	22,969	24,436	339.398	2,90%
Atividades administrativas e serv. complementares	14,294	17,624	1.190.628	10,30%
<b>Tamanho da firma</b>				
Microempresa	11,627	11,614	1.752.862	15,10%
Pequena	16,208	16,564	2.906.567	25,10%
Média	20,814	21,235	2.077.515	17,90%
Grande	22,695	24,323	4.848.693	41,90%

Fonte: Elaborada pelo próprio autor a partir dos microdados identificados da RAIS (2010-2018).

Nota: A totalidade de cada variável categórica soma 11.585.637, total de observações da amostra.

Se as firmas remuneram os trabalhadores de acordo com a produtividade deles, então é possível dizer que as maiores firmas e as que atuam nos setores industriais são mais produtivas do que as suas contrapartes, já que ofertam os maiores salários. Neste sentido, as firmas buscarão se localizar em áreas de mercado de trabalho local que tenha trabalhadores mais produtivos (*sorting* espacial das firmas), o que pode impactar os salários nessas áreas. Desse modo, o *sorting* espacial de firmas pode ser um elemento importante na formação dos salários de uma determinada LMA. Isto, contudo, será verificado nos resultados das estimações propostas nesta dissertação.

Por fim, a **Tabela 4.4** faz uma análise da distribuição salarial conforme as variáveis ligadas diretamente às áreas de mercado de trabalho local, apresentando também as frequências dessas variáveis. As LMA foram divididas em três, de acordo com o seu tamanho, seguindo os critérios apresentados na metodologia. A maioria dos trabalhadores está empregada em LMA de tamanho grande, em torno de 69% deles. As LMA médias contam com, praticamente, 30% dos trabalhadores da amostra, enquanto com 4%, as LMA pequenas são as que possuem menor contingente de trabalhadores. Os indivíduos locados nas LMA

grandes auferem as maiores médias salariais (R\$ 20, 53), em seguida estão os trabalhadores das médias LMA, com R\$ 15,96, valor bem próximo dos R\$ 14,4 recebidos pelos trabalhadores das pequenas LMA.

A **Tabela 4.4** mostra que 61% dos trabalhadores estão empregados em LMA localizadas na região Sudeste, as LMA das regiões Sul e Nordeste aparecem logo em seguida com 18,4% e 12,7% dos trabalhadores, respectivamente. As regiões Norte e Centro-oeste são as LMA com a menor parcela de trabalhadores da amostra, 2,7% e 5,2%, nesta ordem. Analisando os salários, as LMA localizadas nas regiões Sudeste e Sul apresentam as maiores médias de rendimento, R\$ 21,42 e R\$ 17,42, assim ordenadas. Já a região Nordeste apresenta a menor média (R\$ 12,36), seguida de perto pela região Norte (R\$ 14,99) e centro-oeste (R\$15,64).

**Tabela 4.4: Rendimento médio e distribuição dos trabalhadores conforme características das LMA**

Variável	Média	Desvio Padrão	Observações	Frequência (%)
<b>Tamanho da LMA</b>				
Pequena	14,395	13,761	463.818	4%
Média	15,955	15,413	3.113.342	26,90%
Grande	20,529	22,751	8.008.477	69,10%
<b>Região</b>				
Sudeste	21,418	21,525	7.065.328	61%
Nordeste	12,359	14,485	1.476.682	12,70%
Norte	14,989	15,957	313.926	2,70%
Sul	17,420	16,447	2.127.456	18,40%
Centro-Oeste	15,635	15,982	602.245	5,20%

Fonte: Elaborada pelo próprio autor a partir dos microdados identificados da RAIS (2010-2018).

Nota: A totalidade de cada variável categórica soma 11.585.637, total de observações da amostra.

Os resultados da **Tabela 4.4** indicam que há elementos inerentes às próprias LMA que influenciam na determinação salarial. A depender do tamanho e da região em que LMA está localizada, o trabalhador pode auferir diferentes salários. Sendo assim, a variação salarial entre os trabalhadores pode ser fruto não apenas dos próprios atributos deles ou das firmas, mas também das características das LMA.

A **Tabela 4.5** apresenta a média do salário-hora-real ofertado nas LMA conforme o seu tamanho e sua macrolocalização. Os trabalhadores das LMA grandes localizadas na região Sudeste são os que auferem os maiores salários (R\$ 23,20), seguidos dos trabalhadores das LMA grandes da região Sul. Enquanto isso, os trabalhadores das pequenas, médias e grandes LMA da região Nordeste são os que possuem as menores médias salariais, R\$ 8,98, R\$ 10,6 e R\$ 12,73, respectivamente. Esses resultados mostram que existem grandes disparidades salariais entre as LMA, dependendo do tamanho delas e da região brasileira em que estão localizadas.

**Tabela 4.5: Média do salário-hora-real das LMA conforme seu tamanho e macrolocalização (2010-2018)**

Região	Sudeste	Nordeste	Norte	Sul	Centro-oeste	Total
<b>Tamanho da LMA</b>						
Pequena	14,74	8,98	17,15	14,70	14,49	14,40
Média	16,64	10,60	14,11	16,03	16,29	15,95
Grande	23,20	12,73	15,20	19,68	15,49	20,53
<b>Total</b>	<b>21,42</b>	<b>12,36</b>	<b>14,99</b>	<b>17,42</b>	<b>15,64</b>	<b>19,05</b>

Fonte: Elaborada pelo próprio autor a partir dos microdados identificados da RAIS (2010-2018).

Nota: A amostra total contém 11.585.637 de observações.

De modo geral, a **Tabela 4.5** mostra que os salários das LMA aumentam conforme aumenta o seu tamanho. Essa tendência é observada em todas as regiões, com exceção das regiões Norte, em que o salário das LMA pequenas (R\$ 17,15) são menores do que os das LMA médias (R\$ 14,11), e Centro-oeste, onde os trabalhadores das LMA médias recebem maiores salários (R\$ 16,29) do que os trabalhadores das LMA grandes (R\$ 15,49). Esses resultados, pela ótica das economias de aglomeração, podem indicar que nas regiões Norte e Centro-oeste há uma maior aglomeração nas menores LMA, o que explicaria os seus maiores salários. Além disso, é possível que características das próprias LMA possam explicar essa diferença de salário. Isto será verificado nas estimativas dispostas na seção subsequente.

Uma das estratégias de identificação deste estudo busca lidar com os múltiplos efeitos fixos dos trabalhadores e das firmas. Para tanto, exige-se que haja mobilidade entre firmas e LMA, indivíduos e firmas, e indivíduos e LMA, ao longo do painel. Neste sentido, a **Tabela 4.6** mostra que as exigências presentes na estratégia de identificação deste estudo são satisfeitas.

**Tabela 4.6: Mobilidade das firmas e dos trabalhadores entre 2010 e 2018**

Mobilidade	Quantidade única	Frequência (%)
Firmas entre LMA	5.999	1,15%
<b>Total de firmas na amostra</b>	520.455	100%
Trabalhadores entre firmas	569.958	44,28%
Trabalhadores entre LMA	142.353	10,67%
<b>Total de trabalhadores na amostra</b>	1.287.293	100%

Fonte: Elaborada pelo próprio autor a partir dos microdados identificados da RAIS (2010-2018).

Nota: A amostra total contém 11.585.637 de observações.

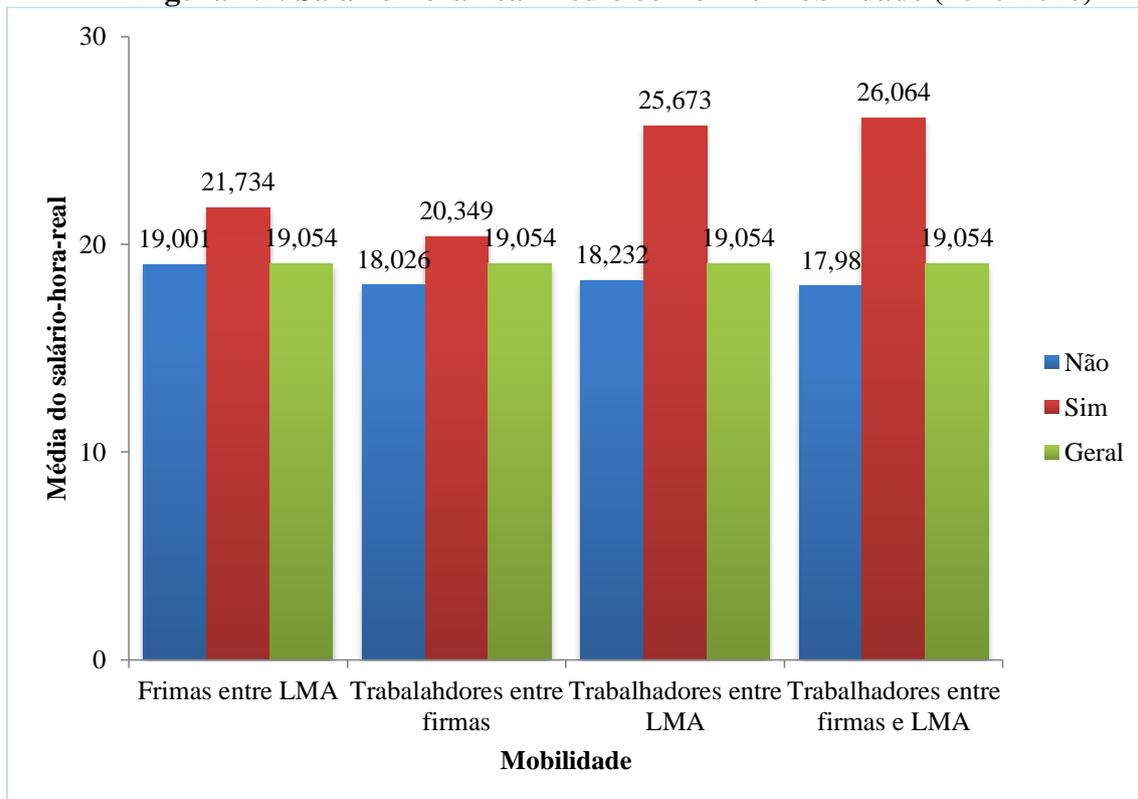
De acordo com os dados dispostos na **Tabela 4.6**, verifica-se que 1,15% das firmas presentes na amostra realizaram mobilidade entre as LMA, isto é, entraram ou saíram de alguma área de mercado de trabalho local no período de 2010 a 2018. A mobilidade de trabalhadores entre firmas representa 44,28% e entre LMA 10,67%. Esses percentuais mostram que durante o período analisado houve mobilidade de trabalhadores entre firmas e

LMA e de firmas entre LMA. Com isso, é possível identificar os efeitos fixos dos trabalhadores e das firmas, bem como sua interação, isto é, seu *match*.

Tendo em vista que a mobilidade pode ocorrer em função da busca por maiores salários (trabalhadores) e maior produtividade (firmas), é provável que haja diferenças salariais entre os indivíduos que fizeram algum tipo mobilidade. Neste sentido, a **Figura 4.1** apresenta as diferenças nas médias do salário-hora-real para os trabalhadores que realizaram mobilidade e para aqueles que não realizaram comparando-os com a média geral de todos os trabalhadores.

Na **Figura 4.1**, verifica-se que a média salarial é maior para aqueles trabalhadores que realizaram alguma mobilidade no decorrer do período de análise (2010-2018), seja entre áreas de mercado de trabalho local e/ou entre firmas, quando comparados aos que não realizaram mobilidade ou quando comparados à média salarial geral dos trabalhadores. O mesmo ocorre para aqueles trabalhadores das firmas que se moveram entre LMA.

**Figura 4.1: Salário-hora-real médio conforme mobilidade (2010-2018)**



Fonte: Elaborada pelo próprio autor a partir dos microdados identificados da RAIS (2010-2018).

Nota: A amostra total contém 11.585.637 de observações.

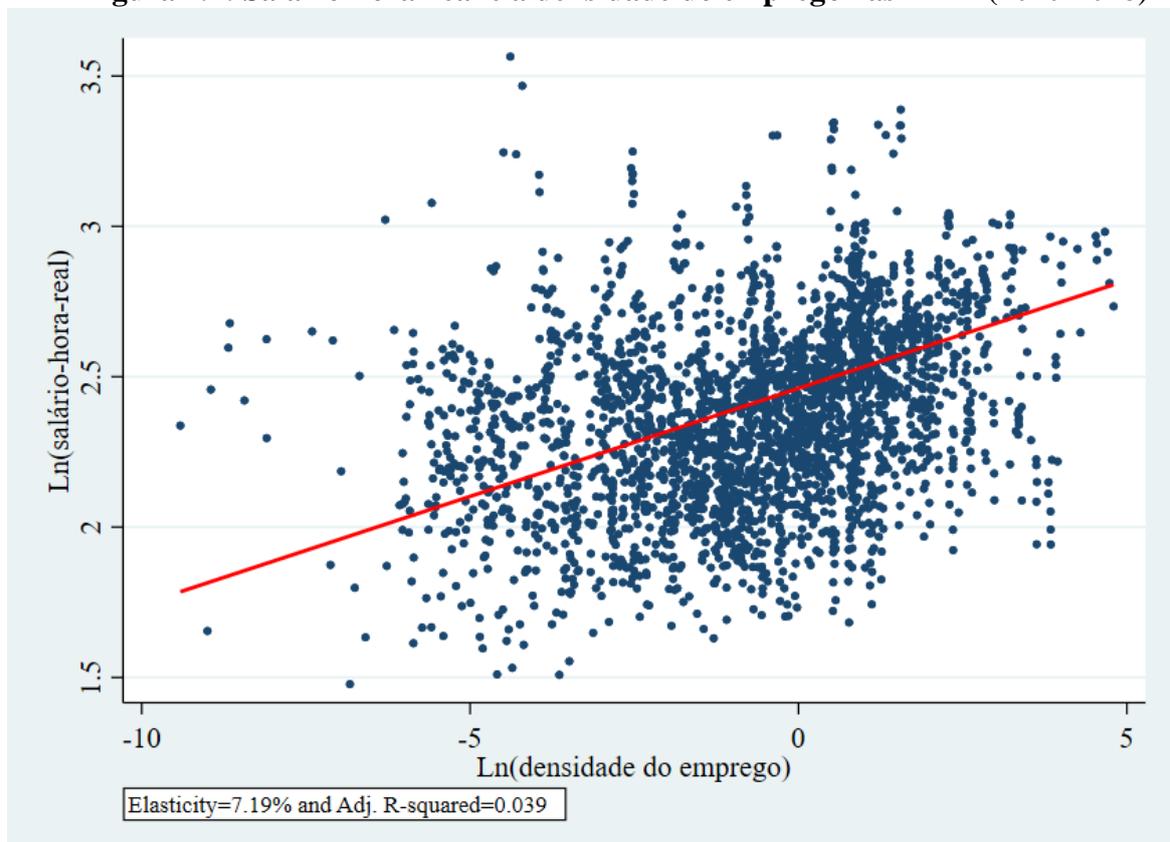
A maior média, R\$ 26,064, é auferida pelos trabalhadores que se moveram entre LMA e firmas ao mesmo tempo. Em seguida, os trabalhadores que se mudaram entre as LMA recebem a segunda maior média do salário-hora-real, no valor R\$ 25,673. Os trabalhadores das firmas que mudaram de área de mercado de trabalho auferem média salarial superior (R\$

21,734) aos trabalhadores que apenas mudaram de firmas (R\$ 20,349). Esses dados mostram que não considerar os efeitos fixos dos trabalhadores e das firmas e suas interação pode enviesar as estimativas sobre os diferenciais salariais.

Por fim, a **Figura 4.1** mostra que os trabalhadores que não realizaram alguma mobilidade no decorrer do tempo analisado foram aqueles que receberam as menores médias do salário-hora-real. Isso reforça ainda mais a importância da inclusão dos múltiplos efeitos fixos nas análises econométricas sobre os diferenciais de salário.

Uma vez mostrado que os múltiplos efeitos fixos podem guardar importante relação com a determinação dos salários, a figura **Figura 4.2**, por meio de um ajuste linear mostra a relação entre a variável de interesse deste estudo, a densidade do emprego, e os salários. A partir da figura observa-se que existe uma relação positiva entre as economias de aglomeração e os salários dos trabalhadores formais das LMA.

**Figura 4.2: Salário-hora-real e a densidade do emprego nas LMA (2010-2018)**



Fonte: Elaborada pelo próprio autor a partir dos microdados identificados da RAIS (2010-2018).

Nota: Salário-hora-real deflacionado pelo IPCA (ano base: 2018).

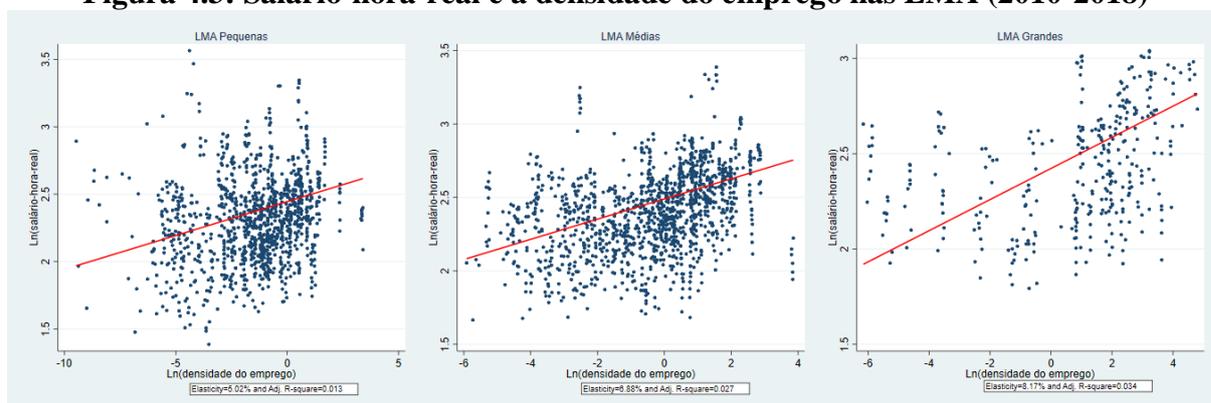
A **Figura 4.2** mostra que a parcela do salário que é explicada pelas economias de aglomeração é de 3,9%, valor próximo aos 4,8% encontrados por Campos (2018) e superior aos 3% encontrados por Silva (2019). A estimativa da elasticidade dos salários em relação às economias de aglomeração, isto é, em relação à densidade do emprego das LMA do Brasil,

foi de 7,19% para o período entre 2010 e 2018. Isto quer dizer que o aumento de 1% na densidade do emprego eleva o salário-hora-real em 7,19%.

A elasticidade disposta na **Figura 4.2** é superior aos 4,9% encontrados por Combes, Duranton e Gobillon (2008) para França e aos 1,51% encontrados por Campos (2018) para as regiões metropolitanas de São Paulo. É também superior aos 5,74% encontrados por Silva (2019) e aos 6,7% encontrados por Silva (2017) para o Brasil. A elasticidade mais elevada deste estudo em relação aos dois últimos pode ser fruto da diferença na construção das áreas de mercado de trabalho local construídas para as análises desta dissertação, esta construção já foi detalhada no capítulo da metodologia.

A **Figura 4.3** mostra que para todos os tamanhos de LMA há uma relação positiva entre os salários e a densidade do emprego. Conforme o tamanho da LMA aumenta, a parcela da variação salarial explicada pela densidade do emprego também se eleva, de modo que para as LMA pequenas o percentual de explicação, medido pelo  $R^2$  ajustado, é de 1,3%, esse percentual é de 2,7% para as LMA médias e de 3,4% para as LMA grandes. O aumento de 1% da densidade do emprego nas pequenas LMA acarreta em uma elevação salarial de 5,02%. Quando se considera o efeito nas LMA médias, o aumento de salário é de 6,88%. Nas LMA grandes, esse aumento salarial é de 8,17%. O que fica evidente é que tanto a parcela do salário que é explicada pela densidade do emprego quanto o aumento salarial decorrente da elevação dessa densidade se elevam conforme aumenta o tamanho da LMA.

**Figura 4.3: Salário-hora-real e a densidade do emprego nas LMA (2010-2018)**



Fonte: Elaborada pelo próprio autor a partir dos microdados identificados da RAIS (2010-2018).

Nota: Salário-hora-real deflacionado pelo IPCA (ano base: 2018).

Os resultados desta seção mostram que a obtenção de estimativas mais robustas para a relação entre os salários e as economias de aglomeração requer o controle dos múltiplos efeitos fixos dos trabalhadores, das firmas e da interação desses efeitos. Contudo, é possível que mesmo após o controle desses múltiplos efeitos fixos, os resultados continuem viesados em decorrência de uma possível endogeneidade entre os salários e a densidade do emprego.

Neste sentido, é preciso tratar essa fonte de viés, o que será feito via estimação por VI-MQ2E. Na próxima seção estão dispostas as estimativas considerando essas duas estratégias de identificação na relação entre as economias de aglomeração e os salários do mercado de trabalho formal brasileiro.

## 4.2 RESULTADOS DAS ESTIMAÇÕES EM PAINEL

Esta seção está dividida em três subseções. A subseção 4.2.1 apresenta os resultados das estimações econométricas das especificações propostas neste trabalho para capturar o efeito das economias de aglomeração sobre os salários individuais nas áreas de mercado de trabalho local (LMA). São apresentados os resultados do modelo POLS, sem o controle dos efeitos fixos, mas incluindo as *dummies* de ano, dos modelos estimados por efeitos fixos dos trabalhadores e das firmas, isoladamente e simultaneamente, dos modelos que consideram os múltiplos efeitos fixos (trabalhador, firma, trabalhador-firma), simultaneamente. O foco da subseção 4.2.1 está em apresentar e discutir as estimativas da variável de interesse deste estudo, a densidade do emprego, contudo, Apêndice D traz as estimativas para todas as variáveis consideradas nas equações. Importante ressaltar que o teste de Hausmann indicou que a estimação por efeitos fixos é a mais adequada para as especificações aqui definidas, a hipótese nula, de que não há diferenças entre EA e EF, foi rejeitada a um nível de significância de 1%.

Na subseção 4.2.2 são apresentadas as estimativas dos diferenciais salariais para os modelos estimados por VI via MQ2E. Como já citado na metodologia, o  $\ln$  da densidade populacional é a variável que instrumentaliza a densidade do emprego. O método de VI-MQ2E busca tratar a possível endogeneidade entre a densidade do emprego e os salários. Além disso, apenas os resultados do segundo estágio da estimação são apresentados. Por fim, a subseção 4.2.3 apresenta como o prêmio salarial oriundo da densidade do emprego se comporta nas LMA de acordo com o seu tamanho e com a sua macrolocalização.

### **4.2.1 Economias de aglomeração e prêmio salarial: resultado da estimação por efeitos fixos**

Como já mencionado, o objetivo desta subseção é apresentar e discutir os resultados das estimações para a variável de interesse desta pesquisa, a densidade do emprego. Porém, de maneira resumida, pode-se dizer que os resultados encontrados para as variáveis observáveis inclusas nas regressões seguem o padrão da literatura. As estimativas via POLS mostram que os indivíduos do gênero masculino recebem maiores salários quando comparados aos do gênero feminino, os trabalhadores com nível Básico 1 de ensino recebem menores salários do

que suas contrapartes, principalmente, em relação aos de nível Superior. Além disso, os indivíduos de raça branca são os que possuem maiores rendimentos quando comparados aos indivíduos das demais raças. Também no modelo POLS, verifica-se que os salários aumentam conforme aumentam a experiência e a idade dos trabalhadores. No que diz respeito às firmas, aquelas de maior porte pagam os maiores salários quando comparadas às Microempresas. Por fim, em relação às LMA, os salários são mais elevados na região Sudeste quando comparados às demais macrorregiões, exceto a região Sul, e aumentam conforme aumenta a área da LMA. Ver **Tabela D.1** do Apêndice D para mais detalhes

A **Tabela 4.7** apresenta os resultados do modelo POLS (coluna I), que considera as características observáveis dos trabalhadores, das firmas e das LMA, além de incluir *dummies* de tempo, mostram que há um prêmio salarial em favor das áreas de trabalho mais densas. De maneira diretamente proporcional, a variação em 1% nos níveis de densidade do emprego, *proxy* para as economias de aglomeração, resulta em uma variação de 4,07% nos salários locais, como pode ser visto na **Tabela 4.7**. Essa variação é próxima dos 4% encontrados por Groot, Groot e Smit (2014) para Holanda, abaixo dos 3,22% encontrados por Combes, Gobillon e Duranton (2008) para França e dentro do intervalo de 3- 8%, consenso da literatura (ROSENTHAL; STRANGE, 2004).

Como já discutido no capítulo 3, os resultados do modelo POLS apresentam estimativas enviesadas da relação entre economias de aglomeração e produtividade, medida pelo salário, já que este não considera as heterogeneidades não observadas. Para lidar com o *sorting* decorrente dessas heterogeneidades, foram estimados modelos que buscam analisar, separadamente e simultaneamente, os efeitos fixos dos trabalhadores, das firmas e do *match* trabalhador-firma.

Como se observa na **Tabela 4.7**, ao se controlar apenas pelo o efeito fixo dos trabalhadores (coluna II), a elasticidade da densidade do emprego se reduz 6,7 vezes em comparação com o modelo POLS, ficando em ficando em 0,61%. Se apenas o efeito fixo de firma é controlado (coluna III), o prêmio salarial associado à densidade do emprego é de 2,16%, superior ao estimado por EF de indivíduos, mas inferior ao coeficiente do modelo POLS. Ao se controlar, simultaneamente, os efeitos fixos dos indivíduos e das firmas (coluna IV), o prêmio salarial associado a uma variação de 1% na densidade do emprego é de 1,91%. Por fim, a coluna IV traz os resultados do modelo em que são considerados, simultaneamente, os efeitos dos trabalhadores, das firmas e do *match* trabalhador-firma. Neste caso, o prêmio salarial de 1,9% apresenta apenas uma mudança marginal em relação ao modelo da coluna IV.

Isto indica, portanto, que não é o efeito do *match* trabalhador-firma a principal fonte de viés nas estimativas da densidade do emprego sobre os salários.

A **Tabela 4.7** mostra ainda que a variação salarial explicada, medida pelo  $R^2$  ajustado, é de 56,7% no POLS; de 92,1% no EF de indivíduos; de 77,4% considerando apenas o EF de firmas; de 93,3% quando se considera, simultaneamente, os efeitos fixos dos trabalhadores e das firmas; e de 94,1% no modelo principal desta dissertação em que se consideram, ao mesmo tempo, os múltiplos efeitos fixos. O poder explicativo do modelo principal (coluna V) se eleva em 2 pontos percentuais quando comparado ao modelo que considera apenas os efeitos fixos dos trabalhadores e em 0,8 pontos percentuais quando comparado ao modelo em que se controla os efeitos fixos dos trabalhadores e das firmas (coluna IV), simultaneamente. Esses resultados apontam para a importância do *sorting* espacial dos trabalhadores e firmas na explicação dos diferenciais salariais locais em detrimento do *match* trabalhador-firma.

**Tabela 4.7: Efeitos das economias de aglomeração sobre os salários individuais nas LMA (2010-2018)**

Variáveis	(I) POLS	(II) EF-ID	(III) EF-Firma	(IV) EF-ID e Firma	(V) EF-ID, Firma e ID-Firma
<i>Ln.</i> Dens. Emprego	0.0407*** (0.0003)	0.0061*** (0.0014)	0.0216*** (0.0084)	0.0191*** (0.0072)	0.0190*** (0.0019)
Carac. Trabalhadores	Sim	Sim	Sim	Sim	Sim
Carac. Firmas	Sim	Sim	Sim	Sim	Sim
Carc. LMA	Sim	Sim	Sim	Sim	Sim
Ano	Sim	Sim	Sim	Sim	Sim
Efeito Fixo ID	Não	Sim	Não	Sim	Sim
Efeito Fixo Firma	Não	Não	Sim	Sim	Sim
Efeito Fixo ID x Firma	Não	Não	Não	Não	Sim
$R^2$	0.5668	0.9301	0.7826	0.9431	0.9501
$R^2$ ajustado	0.567	0.921	0.774	0.933	0.941
$R^2$ <i>Within</i>	-	0.258	0.365	0.223	0.228
Observações	11,585,637	11,585,637	11,516,579	11,516,578	11,160,443
F	268604	27540	3393	58930	22028
Prob>F	0	0	0	0	0

Fonte: Elaboração própria com base nos dados identificados da RAIS 2010-2018.

Nota: Erro-padrão robusto entre parênteses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

A partir dos resultados da **Tabela 4.7**, verifica-se a importância do controle simultâneo do *sorting* espacial dos trabalhadores e firmas e do *match* trabalhador-firma para a explicação dos diferenciais salariais nas LMA. Contudo, o *sorting* espacial dos trabalhadores parece ser o efeito mais importante para a análise dos diferenciais salariais, o que corrobora com as evidências documentadas na literatura (GLASER; MARÉ, 2001; SILVA, 2017; CAMPOS; 2018; NEVES JÚNIOR, 2018).

Em linhas gerais, os resultados da **Tabela 4.7** indicam que, o *match* trabalhador-firma não aparenta ser a principal fonte dos diferenciais salariais nas LMA do Brasil. Na verdade, esses diferenciais parecem estar atrelados, substancialmente, ao *sorting* espacial dos trabalhadores e firmas. Contudo, mesmo após o controle dessas fontes de diferenças de salário, há um prêmio salarial decorrente dos efeitos puros das economias de aglomeração. Esse prêmio, no valor de 1,9%, é inferior aos 2,6% encontrados por Silva (2017) quando a autora considera os efeitos fixos das firmas e dos trabalhadores de maneira simultânea e aos 3,02% encontrados por Combes, Duranton e Gobillon (2008). Esse menor prêmio pode resultar do fato de que uma parcela do efeito de aglomeração é também componente dos efeitos fixos (CAMPOS, 2018). Portanto, ao se controlar por mais efeitos fixos, o efeito puro das economias de aglomeração é reduzido.

As estimativas encontradas para o modelo principal desta dissertação (coluna V da **Tabela 4.7**) são certamente mais robustas do que aquelas em que não se consideram os múltiplos efeitos fixos. Contudo, o efeito causal entre as economias de aglomeração e os salários pode estar comprometido em decorrência da possível endogeneidade entre a densidade do emprego e os salários. Para lidar com isso, a próxima subseção apresenta os resultados do segundo estágio das estimações por Variáveis Instrumentais (VI). O *ln* da densidade populacional defasada das LMA é o instrumento para a densidade do emprego.

#### **4.2.2 Economias de aglomeração e prêmio salarial: resultado da estimação por efeitos fixos e variáveis instrumentais**

Nesta subseção são apresentados os resultados da estimação por Variáveis Instrumentais e Efeitos Fixos. Como já citado na metodologia deste estudo, a estimação por VI é feita via MQ2E. E, assim como na subseção anterior, as estimativas completas de cada variável explicativa estão dispostas na **Tabela D.2** do Apêndice D. O foco aqui será analisar como se comporta a variável de interesse, a densidade do emprego. Estimar por VI é necessário para tratar a possível endogeneidade entre a densidade do emprego e os salários. Para tanto, o instrumento utilizado é o logaritmo da densidade populacional defasada dos anos de 1920, 1940 e 1950.

Para uma variável ser considerada um bom instrumento é preciso que atenda a dois critérios. O primeiro é que esteja correlacionada com a variável endógena. O segundo é que não seja correlacionada com o termo de erro, ou seja, é preciso garantir a exogeneidade do instrumento. Como já documentado na literatura discutida nesta dissertação, a densidade populacional passada não deve impactar os salários contemporâneos, mas deve guardar

relações importantes com a densidade do emprego atual, o que é desejável para que essa variável seja um bom instrumento. Adicionalmente, o teste F do primeiro estágio, para cada um dos modelos, mostra que o viés de instrumento fraco é pequeno, o que indica que os resultados encontrados com a estimação por VI são os menos enviesados possíveis.

De acordo com a **Tabela 4.8**, o prêmio salarial fruto da densidade do emprego apresenta dois sentidos diferentes em relação aos prêmios estimados nos modelos sem tratamento da endogeneidade, dispostos na **Tabela 4.7**. No modelo de EF dos trabalhadores (coluna II) esse prêmio é de 1,68%, no modelo de EF das firmas (coluna III) é de 2,69% e no modelo de EF de trabalhador e firma (coluna IV) é de 3,62%. Esses resultados são, respectivamente, 63,7%, 19,7% e 47,2% maiores do que os encontrados para estas mesmas especificações nos modelos estimados apenas por efeitos fixos. A ampliação dos coeficientes indica que a endogeneidade age subestimando o efeito entre a densidade do emprego e os salários nas LMA, o que corrobora com os resultados encontrados por Campos (2018) e Groot, Groot, Smit (2014), ainda que neste último o aumento tenha sido apenas marginal.

Todavia, quando se considera as especificações dos modelos POLS (coluna I) e do modelo principal desta dissertação (coluna V), a endogeneidade age superestimando os efeitos da densidade do emprego sobre os salários, quando comparados às suas especificações análogas, dispostas na **Tabela 4.7**. O prêmio salarial no modelo POLS por VI é de 1,76%, 56,7% menor do que o modelo sem VI, e no principal modelo desta dissertação, que considera os efeitos fixos múltiplos (trabalhador, firma, *match* trabalhador-firma), esse prêmio é de 1,69%, 11,1% menor do que o modelo que não trata a endogeneidade. Esses resultados corroboram com Combes, Duranton e Gobillon (2008) e Silva (2017).

Os resultados encontrados na **Tabela 4.8** indicam que a endogeneidade, se não tratada, pode enviesar os resultados das estimativas dos efeitos das economias de aglomeração sobre os salários. Os sentidos opostos observados na variação do coeficiente da densidade do emprego nos modelos que incluem os efeitos fixos (**Tabela 4.7**) e nos modelos que tratam a endogeneidade (**Tabela 4.8**) encontram respaldo na literatura. Barurufi, Haddad e Nijkamp (2016) também encontraram variação em sentidos opostos do retorno salarial à densidade do emprego quando se compara o modelo estimado por efeitos fixos e o modelo estimado por VI a depender da especificação adotada. Isto que dizer que a endogeneidade pode agir superestimando (COMBES; DURANTON; GOBILLON, 2008; SILVA, 2017) ou subestimando (CAMPOS, 2018) os efeitos da densidade do emprego sobre os salários.

**Tabela 4.8: Efeitos das economias de aglomeração sobre os salários individuais nas LMA (2010-2018) – VI-MQ2E**

Variáveis	(I) POLS	(II) EF-ID	(III) EF-Firma	(IV) EF-ID e Firma	(V) EF-ID, Firma e ID-Firma
<i>Ln. Dens. Emprego</i>	0.0176*** (0.0005)	0.0168*** (0.0024)	0.0269* (0.0148)	0.0362*** (0.0025)	0.0169*** (0.0013)
Carac. Trabalhadores	Sim	Sim	Sim	Sim	Sim
Carac. Firmas	Sim	Sim	Sim	Sim	Sim
Carc. LMA	Sim	Sim	Sim	Sim	Sim
Ano	Sim	Sim	Sim	Sim	Sim
Efeito Fixo ID	Não	Sim	Não	Sim	Sim
Efeito Fixo Firma	Não	Não	Sim	Sim	Sim
Efeito Fixo ID x Firma	Não	Não	Não	Não	Sim
F 1º estágio	7427	53037	85.30	6033	370985
Prob> F 1º estágio	0	0	0	0	0
R <sup>2</sup>	0.5685	0.2566	0.3654	0.2220	0.2259
R <sup>2</sup> ajustado	0.568	0.257	0.365	0.0994	0.0761
Observações	11,352,518	11,351,500	11,516,579	11,284,391	10,933,643
F	276752	26831	3402	59201	58214
Prob>F	0	0	0	0	0

Fonte: Elaboração própria com base nos dados identificados da RAIS 2010-2018.

Nota: Erro-padrão entre parênteses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

O modelo de referência desta dissertação quando estimado por VI (coluna V da **Tabela 4.8**), apresenta um prêmio salarial à densidade do emprego no valor de 1,69%. Esse prêmio salarial quando comparado aos encontrados por outros estudos, também por meio de estimação por variáveis instrumentais e efeitos fixos, apresenta algumas diferenças em sua magnitude, embora a direção seja a mesma, isto é, um efeito positivo das economias de aglomeração sobre os salários. O prêmio encontrado por Combes, Duranton e Gobillon (2008) é de 3,02%, o prêmio estimado por Groot, Groot e Smit (2014) é de 4,8%, o prêmio encontrado por Silva (2017) é de 4,9% e o encontrado por Silva (2019) é de 2,46%, com um nível de significância de 5%.

A diferença na magnitude do prêmio salarial encontrado neste estudo e nos citados no parágrafo anterior pode decorrer do fato de que na análise aqui proposta são controlados os efeitos fixos dos trabalhadores, das firmas e do *match* trabalhador-firma, simultaneamente, na especificação por VI. Isso parece de fato ser a principal razão para essas diferenças, tendo em vista que o prêmio salarial nas LMA do Brasil de 1,69% decorrente da densidade do emprego é muito próximo ao prêmio de 1,78% encontrado para as regiões metropolitanas de São Paulo no estudo de Campos (2008), em sua especificação principal, que inclui os efeitos fixos aqui considerados, além dos efeitos da própria célula. Isso mostra que a inclusão dos múltiplos efeitos fixos feita nesta dissertação parece adequada para capturar o efeito puro das economias de aglomeração sobre os salários, tendo em vista que estudos que não consideram esses

efeitos encontraram prêmios salariais superiores aos aqui encontrados. Sendo assim, a especificação preferida desta dissertação é a que está disposta na coluna V da **Tabela 4.8**, tendo em vista que ela lida com o viés causado pelo *sorting* espacial de trabalhadores e firmas e também lida com os efeitos fixos do *match* do trabalhador-firma, ao mesmo tempo em que trata a endogeneidade entre a densidade do emprego e os salários.

Os resultados descritos nesta seção mostram a importância do controle do *sorting* espacial dos trabalhadores e firmas nas estimações dos efeitos das economias de aglomeração sobre os salários e a pouca relevância do *match* trabalhador-firma na determinação salarial. Além disso, aponta para necessidade de se tratar de maneira adequada a endogeneidade entre a densidade do emprego e os salários. Com esses procedimentos adotados, a estimativa encontrada para o prêmio salarial decorrente das economias de aglomeração diminui de 4,07% no modelo POLS, sem nenhum controle de efeitos fixos e sem o tratamento da endogeneidade, para 1,9% ao se controlar os efeitos fixos múltiplos e para 1,69% ao se adicionar o tratamento da endogeneidade via VI.

A subseção 4.2.3 apresenta como o prêmio salarial oriundo da densidade do emprego se comporta nas LMA de acordo com o seu tamanho e com a sua macrolocalização. Os resultados da subseção seguinte são dispostos apenas para os modelos principais desta dissertação, isto é, apenas para o modelo da equação 3.19 estimado por múltiplos efeitos fixos e por variável instrumental.

#### **4.2.3 Efeitos das economias de aglomeração sobre os salários individuais: Tamanho e Macrolocalização das LMA importam?**

Considerando que a composição dos mercados de trabalho local pode variar de acordo com o tamanho das LMA e sua macrolocalização, esta subseção apresenta as estimativas do prêmio salarial à densidade do emprego considerando esses aspectos. Assim como nas subseções anteriores, os resultados são dispostos apenas para a variável de interesse, estando os resultados completos dispostos na **Tabela D.3** do Apêndice D. Além disso, as estimações desta subseção foram realizadas considerando apenas a especificação principal deste estudo.

As colunas I, III e V da **Tabela 4.9** apresentam as estimativas do modelo da equação 3.19 considerando os múltiplos efeitos fixos (trabalhador, firma e *match* trabalhador-firma) e as colunas II, IV e VI apresentam as estimativas do modelo preferido desta dissertação, aquele que inclui não só os múltiplos efeitos fixos, mas também que trata a endogeneidade por meio da estimação via VI-MQ2E. Como se vê, para todos os tamanhos os resultados da relação entre a densidade do emprego e os salários são significativos, com exceção das estimativas

das LMA pequenas (coluna II) e grandes (coluna VI) no modelo de referência, estimado via VI-MQ2E.

Quando se controla apenas o *sorting* espacial dos trabalhadores e firmas e do *match* trabalhador-firma, o efeito da densidade do emprego sobre os salários é positivo em 3,34% nas pequenas LMA, em 3,65% nas médias LMA e negativo em 1,1% nas grandes LMA. Esses resultados indicam que há prevalência de um efeito de congestão das aglomerações nas grandes LMA, o que estaria mitigando o ganho salarial fruto das economias de aglomeração.

Ao se tratar a endogeneidade, as estimativas do prêmio salarial são significativas apenas para as médias LMA, como se vê na coluna IV da **Tabela 4.9**. Além disso, a endogeneidade entre a densidade do emprego e os salários nessas LMA parece subestimar, assim como em Campos (2018), os efeitos das economias de aglomeração sobre os salários. O efeito da densidade sobre os salários vai de 3,65%, no modelo apenas com os múltiplos efeitos fixos, para 5,86%, no modelo em que se trata a endogeneidade.

**Tabela 4.9: Efeitos das economias de aglomeração sobre os salários individuais conforme tamanho das LMA (2010-2018)**

Variáveis	(I)	(II)	(III)	(IV)	(V)	(VI)
	Pequena EF-ID, Firma e ID-Firma	Pequena VI e EF-ID, Firma e ID-Firma	Média EF-ID, Firma e ID-Firma	Média VI e EF-ID, Firma e ID-Firma	Grande EF-ID, Firma e ID-Firma	Grande VI e EF-ID, Firma e ID-Firma
Ln. Dens. Emprego	0.0334*** (0.0112)	-0.0113 (0.0411)	0.0365*** (0.0047)	0.0586*** (0.0167)	-0.0109*** (0.0028)	-0.0016 (0.0035)
Carac. Trabalhadores	Sim	Sim	Sim	Sim	Sim	Sim
Carac. Firmas	Sim	Sim	Sim	Sim	Sim	Sim
Carc. LMA	Sim	Sim	Sim	Sim	Sim	Sim
Ano	Sim	Sim	Sim	Sim	Sim	Sim
Efeito Fixo ID	Sim	Sim	Sim	Sim	Sim	Sim
Efeito Firma	Sim	Sim	Sim	Sim	Sim	Sim
Efeito ID x Firma	Sim	Sim	Sim	Sim	Sim	Sim
F 1º estágio		18823		147621		101000
Prob> F 1º estágio		0		0		0
R <sup>2</sup>	0.9337	0.2729	0.9373	0.2448	0.9537	0.2177
R <sup>2</sup> -ajustado	0.922	0.151	0.926	0.108	0.944	0.0609
R <sup>2</sup> within	0.270		0.249		0.218	
Observações	453,294	377,905	3,017,863	2,867,034	7,677,036	7,677,036
F	1300.46	3035	6878.57	17883	15026.97	40447
Prob>F	0	0	0	0	0	0

Fonte: Elaboração própria com base nos dados identificados da RAIS 2010-2018.

Nota: Erro-padrão entre parênteses (colunas II, IV e VI). Erro-padrão robusto entre parênteses (colunas I, III, e V). \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Os resultados do modelo de referência dispostos na **Tabela 4.9** indicam que o tamanho das LMA pode influenciar na maneira como os salários são afetados pelas aglomerações. Nas pequenas e grandes LMA, isto é, nos extremos, o efeito sobre os salários não é significativo quando se estima por VI-MQ2E, indicando que a endogeneidade entre a densidade do emprego e a produtividade nestas LMA pode confundir as estimativas do prêmio salarial se

não for tratada. O diferencial de salários só é impactado significativamente e de maneira positiva pela densidade do emprego nas médias LMA.

De maneira geral, o que a **Tabela 4.9** reporta é que em um país com dimensões continentais, é possível que os diferenciais salariais existentes entre as áreas de mercado de trabalho local não possuam os mesmos determinantes. Como se vê, a densidade é significativa para explicar as diferenças salariais apenas nas médias LMA. Esses resultados são importantes, pois os estudos realizados até então não investigaram como o prêmio salarial às economias de aglomeração se comporta de acordo com o tamanho de cada LMA.

Tendo em vista que a densidade do emprego pode ser ou não significativa para explicar as diferenças salariais nas LMA de tamanhos diferentes, este estudo também busca verificar o seu comportamento entre as regiões brasileiras. Neste sentido, a **Tabela 4.10** apresenta as estimativas do modelo principal, estimando por múltiplos efeitos fixos e também por VI. Os resultados completos destas estimações estão na **Tabela D.4** do Apêndice D.

Como pode ser observado na **Tabela 4.10**, a região Centro-Oeste (colunas IX e X) não apresenta estimativas significativas para densidade do emprego em nenhum dos modelos considerados. As regiões Nordeste e Norte, por sua vez, só apresentam estimativas significativas no modelo estimado por VI, como pode ser visto nas colunas IV e VI. As regiões Sudeste (colunas I e II) e Sul (colunas VII e VIII) apresentam retornos salariais significativos e positivos para as economias de aglomeração em ambos os modelos.

Ao analisar a **Tabela 4.10**, percebe-se que a endogeneidade superestima os efeitos das economias de aglomeração sobre os salários nas regiões Sudeste e Sul. No Sudeste, esse efeito cai de 3,07%, no modelo estimado por múltiplos efeitos fixos, para 1,64%, no modelo estimado por VI. Na região Sul, esse efeito cai 4,82% para 2,24%. As economias de aglomeração possuem efeitos positivos sobre os salários da região Norte na ordem de 3,82%, no modelo de referência desta dissertação (coluna VI). Esse efeito é de 0,7% para região Nordeste, considerando um nível de significância de 10%. Esses resultados indicam que essas regiões podem possuir diferentes características não observadas variantes no tempo, tais como, tecnologia da firma e/ou habilidades do trabalhador (CAMPOS, 2018), que podem explicar as diferenças na magnitude do coeficiente da densidade do emprego nelas observado.

Em resumo, os resultados desta subseção mostram que em um país tão heterogêneo como o Brasil, não se pode esperar que uma variável afete os salários de uma mesma maneira em todas as regiões ou em áreas com tamanhos distintos. Estar ciente disto é importante, sobretudo, no que diz respeito ao planejamento de políticas públicas para o mercado de trabalho brasileiro. No capítulo 5 são apresentadas as considerações finais desta dissertação.

**Tabela 4.10: Efeitos das economias de aglomeração sobre os salários individuais conforme macrolocalização das LMA (2010-2018)**

Variáveis	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)	(IX)	(X)
	Sudeste EF-ID, Firma e ID-Firma	Sudeste VI e EF- ID, Firma e ID- Firma	Nordeste EF-ID, Firma e ID-Firma	Nordeste VI e EF- ID, Firma e ID- Firma	Norte EF-ID, Firma e ID-Firma	Norte VI e EF- ID, Firma e ID- Firma	Sul EF-ID, Firma e ID-Firma	Sul VI e EF- ID, Firma e ID- Firma	Centro- Oeste EF-ID, Firma e ID-Firma	Centro- Oeste VI e EF- ID, Firma e ID- Firma
<i>Ln.</i> Dens. Emprego	0.0307*** (0.0059)	0.0164*** (0.0032)	0.0014 (0.0044)	0.0070* (0.0038)	-0.0301 (0.0205)	0.0382** (0.0165)	0.0482*** (0.0064)	0.0224*** (0.0050)	0.0277 (0.0241)	-0.0091 (0.0073)
Carac. Trabalhadores	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Carac. Firms	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Carac. LMA	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Ano	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Efeito Fixo ID	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Efeito Firma	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Efeito ID x Firma	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Teste F 1º estágio		1.363e+06		2.226e+06		160621		932934		37876
Prob>F 1º estágio		0		0		0		0		0
R <sup>2</sup>	0.943	0.2076	0.931	0.2335	0.921	0.2291	0.926	0.2852	0.914	0.2615
R <sup>2</sup> -ajustado	0.208	0.0531	0.233	0.0851	0.230	0.0834	0.288	0.152	0.262	0.110
Observações	6,792,579	6,773,223	1,426,397	1,423,628	303,666	282,845	2,057,985	1,901,324	575,479	548,626
F	13374	35359	3130	8649	692.9	1684	5860	15228	1479	3838
Prob>F	0	0	0	0	0	0	0	0	0	0

Fonte: Elaboração própria com base nos dados identificados da RAIS 2010-2018.

Nota: Erro-padrão entre parênteses (colunas II, IV, VI, VIII e X). Erro-padrão robusto entre parênteses (colunas I, III, V, VII e IX). \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

## 5 CONSIDERAÇÕES FINAIS

O objetivo desta dissertação foi mensurar os efeitos das economias de aglomeração sobre os salários individuais no mercado de trabalho formal brasileiro no período de 2010 a 2018, considerando os efeitos do *sorting* espacial dos trabalhadores, das firmas e do *match* trabalhador-firma, que foram identificados pelos atributos não observáveis dos trabalhadores, das firmas e do *matching* assortivo entre trabalhadores e firmas nas áreas de mercado de trabalho local. Áreas estas que foram definidas a partir dos estudos sobre arranjos populacionais e concentrações urbanas no Brasil e sobre as regiões de influência das cidades – REGIC- realizados pelo IBGE, considerando as versões mais recentes divulgadas pelo instituto. A construção dessas LMA adota uma metodologia nova, baseada em Neves Júnior (2018), que inclui a ideia de município influenciado na construção das LMA. Foram delimitadas 371 LMA, compostas por 73,8% da população brasileira residente em 2018. O estudo foi conduzido apenas para as LMA aqui definidas.

Os dados principais são oriundos dos microdados identificados da RAIS (2010-2018). Após vários blocos de tratamento, a amostra final desta dissertação conta com 11.585.637 de observações correspondentes a 520.455 firmas e a 1.287.293 indivíduos, que formam 2.232.883 de combinações únicas de trabalhadores e firmas.

A revisão da literatura mostrou que há um efeito positivo das economias de aglomeração sobre os salários. Além disso, é consenso nestes estudos que a autosseleção dos trabalhadores e das firmas para LMA mais produtivas pode enviesar as estimativas da densidade do emprego sobre os salários (GLAESER; MARÉ, 2001; COMBES; DURANTON; GOBILLON, 2008; MION; NATICCHIONI, 2009; GROOT; GROOT; SMIT, 2014; MATANO; NATICCHIONI, 2016; SILVA, 2017; CAMPOS, 2018; NEVES JÚNIOR, 2018; SILVA 2019). Neste sentido, este trabalho lida com essas fontes de viés por meio de estimações por múltiplos efeitos fixos. A estratégia de identificação para controle desses vieses requer mobilidade de trabalhadores entre LMA e firmas, e de firmas entre LMA. Os resultados encontrados mostram que estas condições foram satisfeitas.

Tendo em vista a possível endogeneidade entre a variável densidade do emprego e os salários, esta dissertação adotou o método de Variáveis Instrumentais via MQ2E para tratar esta outra fonte de viés. Seguindo a literatura, este estudo utilizou a densidade populacional defasada como instrumento da densidade do emprego atual (CICCONE; HALL, 1996; BARUFI, 2015; DAUTH *et al*, 2018; PEREIRA, 2018; LEKNES; RATTSSØ; STOKKE, 2020). Embora não se possa testar a hipótese de exogeneidade entre o instrumento e o termo

de erro, as estatísticas encontradas mostraram que a hipótese de instrumento fraco é descartada. Baseado nisto e na ampla literatura disponível, a densidade populacional defasada se mostrou um bom instrumento para a densidade do emprego atual.

As especificações dos modelos foram baseadas, principalmente, no modelo de *sorting* desenvolvido por Combes, Duranton e Gobillon (2008). A esta especificação, foi adicionado o controle do *sorting* espacial das firmas com base no modelo desenvolvido por Abowd, Kramarz e Margolis (1999). Além disso, adicionou-se ao modelo o controle do *match* trabalhador-firma conforme recomendação de Woodcock (2008; 2015), que destaca o possível viés das estimativas se esse *match* não for considerado nas análises. Portanto, a especificação principal do modelo de referência desta pesquisa, que controla, simultaneamente, os múltiplos efeitos fixos, e é estimada por VI-MQ2E, leva em consideração as contribuições desses três estudos.

A variação salarial explicada ( $R^2$  ajustado) foi de 56,7% no modelo POLS. No modelo em que se consideram apenas os efeitos fixos de indivíduo, esse percentual sobe para 92,1%, sendo bem superior ao modelo que apenas leva em conta os efeitos fixos de firma (77,4%). O poder explicativo do modelo principal, em que se controla por múltiplos efeitos fixos (trabalhador, firma, *match* trabalhador-firma), é de 94,1%. Esses resultados mostram a importância do *sorting* espacial dos trabalhadores e firmas e do efeito fixo do *match* trabalhador-firma para a produtividade local e, conseqüentemente, para os salários nas LMA.

Os resultados encontrados nesta dissertação mostram que o efeito bruto das economias de aglomeração sobre os salários dos trabalhadores das LMA brasileiras, no período de 2010 a 2018, é de 4,07%, dentro do intervalo proposto na literatura, que é entre 3% e 8%. Contudo, ao se incluir nas estimações os múltiplos efeitos fixos, esse efeito cai para 1,9%, valor bem próximo ao encontrado no modelo principal desta dissertação, estimado por VI-MQ2E, que foi 1,69%. O prêmio salarial do modelo de referência adotado neste estudo (1,69%) é inferior ao intervalo proposto na literatura e também ao encontrado em Combes, Duranton e Gobillon (2008), 3,02%, em Groot, Groot e Smit (2014), 4,8%, em Silva (2017), 4,9%, e em Silva (2019), 2,46%. Essa diferença na magnitude do prêmio salarial pode decorrer do fato de que a análise aqui proposta, além de tratar a endogeneidade, controla os efeitos fixos dos trabalhadores, das firmas e do *match* trabalhador-firma, simultaneamente. Essa parece ser a principal razão para essas diferenças, tendo em vista que o prêmio salarial decorrente da densidade do emprego encontrado neste estudo é muito próximo ao encontrado por Campos (2018), 1,78%, que utiliza metodologia semelhante à utilizada nesta dissertação.

Quando se analisa o impacto das economias de aglomeração sobre os salários de acordo com os diferentes tamanho das LMA e com as diferentes regiões do Brasil, a estimação do modelo principal mostra que o prêmio salarial em decorrência da densidade do emprego se comporta de maneira diferente, sendo, em alguns casos, não significativo.

Considerando a análise por tamanho da LMA, o efeito da densidade sobre os salários só é significativo apenas nas médias LMA, no valor de 5,86%. Em relação às regiões apenas o coeficiente relativo à região Centro-Oeste não é significativo na especificação principal aqui adotada. O prêmio salarial na região Sudeste decorrente das economias de aglomeração é de 1,64%, na região Nordeste 0,7%, na região Norte 3,82% e na região Sul 2,24%. Esses resultados mostram a densidade do emprego pode afetar de maneira diferente os salários a depender da região e do tamanho da LMA.

De maneira geral, os resultados aqui encontrados mostram que existe um prêmio salarial líquido nas LMA do Brasil atrelado às economias de aglomeração. Mesmo após a estimação mais restritiva, em que se controla por múltiplos efeitos fixos e trata a endogeneidade, o coeficiente da densidade do emprego permaneceu significativo e positivo. Ao se considera diferentes recortes, foi verificado que o prêmio salarial decorrente das economias de aglomeração pode se comportar, substancialmente, de maneira distinta, indicando que componentes não observados variantes no tempo inerentes a cada recorte pode explicar essas diferenças.

A principal contribuição desta dissertação diz respeito ao controle do *sorting* espacial dos trabalhadores e firmas e do *match* trabalhador-firma, de maneira simultânea, nas estimações do prêmio salarial, a partir de uma nova definição de área de mercado de trabalho local, tratando a endogeneidade entre a densidade do emprego e os salários. Adicionalmente, esta dissertação fornece evidências inéditas do impacto das economias de aglomeração na determinação salarial de cada uma das regiões brasileiras e de cada tamanho das áreas de mercado de trabalho local. Essas evidências podem auxiliar de maneira direta o planejamento de políticas públicas voltadas para o mercado de trabalho brasileiro. Elas indicam que, em um país com as dimensões do Brasil e com as disparidades existentes no mercado de trabalho de cada localidade, ações que considerem as especificidades de cada região podem ser mais eficazes do que ações universais.

Em resumo, os resultados encontrados nesta dissertação apontam para importância do *sorting* espacial dos trabalhadores e firmas e para pouca relevância do *match* trabalhador-firma na determinação dos salários das LMA do Brasil. Ademais, observa-se que as economias de aglomeração, mesmo após o controle dos vieses que poderiam prejudicar as

suas estimativas, apresentam um prêmio salarial líquido positivo. No mais, pode-se afirmar que os objetivos traçados nesta dissertação foram plenamente satisfeitos.

Pesquisas subjacentes a este estudo podem ser realizadas de maneira a incluir nas análises o efeito fixo das LMA e suas interações com os demais efeitos fixos incluídos neste estudo, e a partir disso verificar o comportamento da densidade do emprego diante de uma especificação mais restrita. Como se viu, as regiões apresentam diferentes retornos para o impacto da densidade do emprego sobre os salários, portanto, investigar de maneira mais detalhada o que pode estar por trás dessas diferenças pode ser mais uma extensão deste estudo. Por fim, adotar outra variável instrumental, como o *Bartik* setorial, e comparar com os resultados encontrados com o instrumento utilizado nesta pesquisa pode ser útil no entendimento de como a escolha do instrumento, satisfeitas todas as suposições por trás dessa escolha, pode influenciar as estimativas do prêmio salarial.

## REFERÊNCIAS

- ABOWD, John M.; CREECY, Robert H.; KRAMARZ, Francis. **Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data** *Longitudinal Employer-Household Dynamics Technical Papers*: Longitudinal Employer-Household Dynamics Technical Papers. [S. l.]: Center for Economic Studies, U.S. Census Bureau, 2002.
- ABOWD, John M.; KRAMARZ, Francis; MARGOLIS, David N. High Wage Workers and High Wage Firms. *Econometrica*, [s. l.], v. 67, n. 2, p. 251–333, 1999.
- AMARANTE, Patrícia Araújo. **Economias de aglomeração nos municípios brasileiros: uma abordagem utilizando nascimento de firmas, mobilidade dos trabalhadores e equações de salários**. João Pessoa: Tese (Doutorado em Economia) - Centro de Ciências Sociais Aplicadas, Universidade Federal da Paraíba (UFPB). 2017. 125 p.
- ARBACHE, Jorge Saba. Do unions always decrease wage dispersion? The case of Brazilian manufacturing. *Journal of Labor Research*, [s. l.], v. 20, n. 3, p. 425–436, 1999.
- AZZONI, Carlos Roberto; SERVO, Luciana Mendes Santos. Education, cost of living and regional wage inequality in Brazil. *Papers in Regional Science*, v. 81, n. 2, p. 157–175, 1 abr. 2002.
- BARUFI, Ana Maria Bonomi; HADDAD, Eduardo Amaral; NIJKAMP, Peter. Industrial scope of agglomeration economies in Brazil. *The Annals of Regional Science*, v. 56, n. 3, p. 707–755, 1 maio 2016.
- BARUFI, Ana Maria Bonomi. **Agglomeration economies and labour markets in Brazil**. Doctoral Dissertation (Doctorate in Economics) School of Economics, Business and Accounting, University of São Paulo (USP), São Paulo. 2015. 185 p.
- BRASIL. **Decreto nº 9.255 , de 29 de Dezembro de 2017**. Disponível em: [http://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2017/decreto/D9255.htm](http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2017/decreto/D9255.htm). Acesso em: 14 jan. 2021.
- BRASIL. **Decreto-Lei nº 5.452, de 1º de Maio de 1943**. Consolidação das Leis do Trabalho. Brasília, Distrito Federal: Presidência da República. Disponível em: [http://www.planalto.gov.br/ccivil\\_03/decreto-lei/del5452.htm#art58](http://www.planalto.gov.br/ccivil_03/decreto-lei/del5452.htm#art58). Acesso em: 13 jan. 2021.
- CAMERON, Adrian Colin; TRIVEDI, Pravin K. **Microeconometrics Using Stata**. Stata Press, College Station, 2009. 692 p.
- CAMPOS, F. M; SILVEIRA NETO, R. M. A. A importância da dimensão do mercado de trabalho para os diferenciais de participação e salários entre gêneros: Uma análise empírica para os centros urbanos brasileiros. In: ENCONTRO NACIONAL DE ECONOMIA, 37, 2009, Foz do Iguaçu. **Anais...** Foz do Iguaçu: ANPEC, 2009. Disponível em: <http://www.anpec.org.br/encontro2009/inscricao.on/arquivos/000c6f925639a711d35a1187c1b9c5483f3.pdf>. Acesso em: 19, Jun. 2020.
- CAMPOS, Rodger Barros Antunes. **Subcentralidades e e prêmio salarial intra-urbano na região metropolitana de São Paulo**. São Paulo: Tese (Doutorado em Economia) - Faculdade de Economia, Administração e Contabilidade (FEA), Universidade de São Paulo (USP). 2018. 230 p.

CICCONI, Antonio. Agglomeration effects in Europe. **European Economic Review**, [s. l.], v. 46, n. 2, p. 213–227, 2002.

CICCONI, Antonio; HALL, Robert E. Productivity and the Density of Economic Activity. **The American Economic Review**, [s. l.], v. 86, n. 1, p. 54–70, 1996.

COMBES, Pierre-Philippe et al. Sorting and local wage and skill distributions in France. **Regional Science and Urban Economics**, [s. l.], v. 42, n. 6, special issue ‘In honor of Jacques Thisse’, p. 913–930, 2012.

COMBES, Pierre-Philippe; DURANTON, Gilles; GOBILLON, Laurent. Spatial wage disparities: Sorting matters! **Journal of Urban Economics**, [s. l.], v. 63, n. 2, p. 723–742, 2008.

COMBES, Pierre-Philippe; DURANTON, Gilles; GOBILLON, Laurent. The identification of agglomeration economies. **Journal of Economic Geography**, [s. l.], v. 11, n. 2, p. 253–266, 2011.

COMBES, Pierre-Philippe; GOBILLON, Laurent. Chapter 5 - The Empirics of Agglomeration Economies. *In*: DURANTON, Gilles; HENDERSON, J. Vernon; STRANGE, William C. (org.). **Handbook of Regional and Urban Economics**. [S. l.]: Elsevier, 2015. (Handbook of Regional and Urban Economics). v. 5, p. 247–348.

CORREIA, Sergio. **A Feasible Estimator for Linear Models with Multi-Way Fixed Effects**. p. 19. 2016. Disponível em: < <http://scoreia.com/research/hdfe.pdf>>. Acesso em: 15 jan. 2021.

CORREIA, Sergio. **Explanation of the HDFE iteration with 3 FEs**. p.4. 2014. Disponível em: < <http://scoreia.com/research/hdfe.pdf>>. Acesso em: 15 jan. 2021.

DAUTH, Wolfgang; FINDEISEN, Sebastian; MORETTI, Enrico; SUEDEKUM, Jens. Spatial wage disparities – Workers, firms, and assortative matching., p. 29, [s. d.]. .

DURANTON, G.; PUGA, D. Chapter 48 Micro-foundations of urban agglomeration economies. **Handbook of Regional and Urban Economics**. [S. l.]: Elsevier, 2004. v. 4, p. 2063–2117

FINGLETON, B. Increasing returns: evidence from local wage rates in Great Britain. **Oxford Economic Papers**, v. 55, n. 4, p. 716–739, 1 out. 2003. <https://doi.org/10.1093/oeq/55.4.716>.

FREGUGLIA, Ricardo da Silva. **Efeitos da migração sobre os salários no Brasil**. São Paulo: Tese (Doutorado em Economia) - Faculdade de Economia, Administração e Contabilidade (FEA), Universidade de São Paulo (USP). 2007. 126 p.

FUJITA, Masahisa; KRUGMAN, Paul; VENABLES, Anthony James. **The Spatial Economy: Cities, Regions, and International Trade**. Cambridge, Mass.: The MIT Press, 2001.

FUJITA, Masahisa; THISSE, Jacques-François. **Economics of Agglomeration: Cities, Industrial Location, and Regional Growth**. [S. l.]: Cambridge University Press, 2002.

GAURE, Simen. OLS with multiple high dimensional category variables. **Computational Statistics & Data Analysis**, [s. l.], v. 66, p. 8–18, 2013.

GLAESER, Edward L.; GOTTLIEB, Joshua. D. The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States. **Journal of Economic Literature**, v. 47, n. 4, p. 983–1028, dez. 2009.

GLAESER, Edward L.; MARÉ, David C. Cities and Skills. **Journal of Labor Economics**, [s. l.], v. 19, n. 2, p. 316–342, 2001.

GROOT, Stefan P. T.; GROOT, Henri L. F. de; SMIT, Martijn J. Regional Wage Differences in the Netherlands: micro Evidence on Agglomeration Externalities. **Journal of Regional Science**, [s. l.], v. 54, n. 3, p. 503–523, 2014.

GUIMARÃES, Paulo; PORTUGAL, Pedro. A Simple Feasible Procedure to fit Models with High-dimensional Fixed Effects. **The Stata Journal**, [s. l.], v. 10, n. 4, p. 628–649, 2010.

GUJARATI, Damodar. N.; PORTER, Dawn. C. **Econometria básica**. 5. ed. São Paulo: Mcgrawhill Bookman, 2011.

HELSLEY, Robert. William.; STRANGE, W. C. Matching and agglomeration economies in a system of cities. **Regional Science and Urban Economics**, v. 20, n. 2, p. 189–212, 1 set. 1990.

HEUERMAN, D.; HALFDANARSON, B.; SUEDEKUM, J. Human Capital Externalities and the Urban Wage Premium: Two Literatures and their Interrelations: **Urban Studies**, 12 jan. 2010.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA – IBGE. **Arranjos Populacionais e Concentrações Urbanas do Brasil**. 2ª ed. Rio de Janeiro: IBGE, 2016. 160 p. Disponível em: <https://www.ibge.gov.br/geociencias/organizacao-do-territorio/divisao-regional/15782-arranjos-populacionais-e-concentracoes-urbanas-do-brasil.html?=&t=acesso-ao-produto>. Acesso em: 11 jan. 2021.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA – IBGE. **Classificação Nacional de Atividades Econômica - CNAE (versão 2.0)**. Disponível em: <https://concla.ibge.gov.br/documentacao/documentacao-cnae-2-0.html>. Acesso em: 13 jan. 2021.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA – IBGE. Downloads: **Áreas Territoriais**, 2018. Disponível em: <https://www.ibge.gov.br/geociencias/organizacao-do-territorio/estrutura-territorial/15761-areas-dos-municipios.html?=&t=o-que-e>. Acesso em: 10 out. 2020.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA – IBGE. **Estimativas da População**, 2018. Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/populacao/9103-estimativas-de-populacao.html?=&t=o-que-e>. Acesso em: 10 out. 2020.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. **Índice Nacional De Preços Ao Consumidor Amplo - IPCA**. Disponível em:

<https://www.ibge.gov.br/estatisticas/economicas/precos-e-custos/9256-indice-nacional-de-precos-ao-consumidor-amplo.html?=&t=o-que-e>. Acesso em: 14 jan. 2021.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. **Natureza Jurídica**. Disponível em: < <https://concla.ibge.gov.br/estrutura/natjur-estrutura.html>>. Acesso em: 11 jan. 2021.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA – IBGE. **Regiões de Influência das Cidades 2018: Nota Metodológica**. Rio de Janeiro: IBGE, 2020. 55 p. Disponível em: <https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=2101729>. Acesso em: 10 jan. 2021.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA – IBGE. **Regiões de Influência das Cidades - REGIC 2018**. Rio de Janeiro: IBGE, 2020. 196 p. Disponível em: <https://www.ibge.gov.br/geociencias/cartas-e-mapas/redes-geograficas/15798-regioes-de-influencia-das-cidades.html?=&t=downloads>. Acesso em: 11 jan. 2021.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA – IBGE. **Relação Anual de Informações Sociais – RAIS**. Disponível em: <<https://ces.ibge.gov.br/base-dados/metadados/mte/relacao-anual-de-informacoes-sociais-rais.html>>. Acesso em: 09 jan. 2021.

ISARD, Walter. **Location and space economy: a general theory relating to industrial location, market areas, land use, trade, and urban structure**. New York: Vintage, 1969.

JÚNIOR, Edivaldo Constantino das Neves. **Habilidades, mudanças de firma e prêmio salarial urbano**. São Paulo: Tese (Doutorado em Economia) - Faculdade de Economia, Administração e Contabilidade (FEA), Universidade de São Paulo (USP). 2018. 249 p.

KLAESSON, Johan; LARSSON, Hanna. **Wages, Productivity and Industry Composition - agglomeration economies in Swedish regions**. [S. l.]: CESIS, 2009.

KRUGMAN, Paul. Increasing Returns and Economic Geography. **Journal of Political Economy**, v. 99, n. 3, p. 483–499, 1 jun. 1991b.

KRUGMAN, Paul R. **Geography and Trade**. [S. l.]: MIT Press, 1991a.

LEKNES, Stefan; RATTSSØ, Jørn; STOKKE, Hildegunn Ekroll. **Assortative labor matching, city size, and the education level of workers1-33**. [S. l.]: NTNU: Norges teknisk-naturvitenskapelige universitet., 2020. Working paper.

LÖSCH, August. **The Economics of Location**. New Haven: Yale University Press. 1954.

MATANO, Alessia; NATICCHIONI, Paolo. What Drives the Urban Wage Premium? Evidence Along the Wage Distribution. **Journal of Regional Science**, [s. l.], v. 56, n. 2, p. 191–209, 2016.

MINCER, Jacob. **Schooling, Experience, and Earnings**. **Human Behavior & Social Institutions**. No. 2. 1974. Acesso em: 24 jun. 2020.

MINISTÉRIO DA ECONOMIA – ME. **Relação Anual de Informações Sociais Identificada 2010-208: RAIS Identificada**. Brasília, Distrito Federal: ME/CGCIPE-COBTRAB. Mídia Digital. 2020.

MINISTÉRIO DO TRABALHO E EMPREGO – MTE. **Nota Técnica MTE (09/09/2015)**. 2015. Disponível em: <<http://portalfat.mte.gov.br/wp-content/uploads/2016/02/Nota-T%C3%A9cnica-MTE-2014.pdf>>. Acesso em: 08 jan. 2021.

MINISTÉRIO DO TRABALHO E EMPREGO. **Classificação Brasileira de Ocupações**. 3ª ed. Brasília: TEM, SPPE, 2010. 198 p. Disponível em: <[http://portalfat.mte.gov.br/wp-content/uploads/2016/04/CBO2002\\_Liv3.pdf](http://portalfat.mte.gov.br/wp-content/uploads/2016/04/CBO2002_Liv3.pdf)>. Acesso em: 08 jan. 2021.

MION, Giordano; NATICCHIONI, Paolo. The spatial sorting and matching of skills and firms. **Canadian Journal of Economics/Revue canadienne d'économique**, [s. l.], v. 42, n. 1, p. 28–55, 2009.

OVERMAN, H. G.; PUGA, D. Labor Pooling as a Source of Agglomeration: An Empirical Investigation. **NBER Chapters**. [S. l.]: National Bureau of Economic Research, Inc, 2010. p. 133–150.

PEREIRA, Jefferson Douglas da Silva. **Diferenciais salariais no Brasil: até onde a aglomeração explica?**. 2018. 52 f. Trabalho de Conclusão de Curso (Bacharelado em Ciências Econômicas) – Unidade Acadêmica de Serra Talhada, Universidade Federal Rural de Pernambuco, Serra Talhada, 2018.

PUGA, D. The Magnitude and Causes of Agglomeration Economies\*. **Journal of Regional Science**, v. 50, n. 1, p. 203–219, 2010.

ROCHA, R. de M.; NETO, R. da 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. Chapter 49 - Evidence on the Nature and Sources of Agglomeration Economies. In: HENDERSON, J. V.; THISSE, J.-F. (orgs.). **Handbook of Regional and Urban Economics**. Cities and Geography. [S. l.]: Elsevier, 2004. v. 4, p. 2119–2171.

ROSENTHAL, Stuart S.; STRANGE, William C. The attenuation of human capital spillovers. **Journal of Urban Economics**, [s. l.], v. 64, n. 2, p. 373–389, 2008.

SERVIÇO BRASILEIRO DE APOIO ÀS micro E PEQUENAS EMPRESAS - SEBRAE (org). **Anuário do Trabalho nos Pequenos Negócios: 2016**. 9ª ed. São Paulo - SP: DIEESE, 2018. 546 p. Disponível em: <<https://www.sebrae.com.br/Sebrae/Portal%20Sebrae/Anexos/Anu%C3%A1rio%20do%20Trabalho%20nos%20Pequenos%20Neg%C3%B3cios%202016%20VF.pdf>>. Acesso em: 12 jan. 2021.

SILVA, Adelson Santos da. **Economias de aglomeração e matching de ocupação e qualificação no mercado de trabalho brasileiro**. Salvador: Tese (Doutorado em Economia)-Faculdade de Economia, Universidade Federal da Bahia (UFBA). 2019. 147 p.

SILVA, Diana Lúcia Gonzaga da; SANTOS, Gervásio Ferreira dos; FREGUGLIA, Ricardo da Silva. Distribuição espacial dos efeitos de aglomeração sobre os retornos à educação no Brasil entre 1995 e 2008. <http://ppe.ipea.gov.br>, ago. 2016. Disponível em: <http://repositorio.ipea.gov.br/handle/11058/7295>. Acesso em: 24 jun. 2020.

SILVA, Diana Lúcia Gonzaga da. **Economias de aglomeração e heterogeneidade de trabalhador e firma na determinação de salários no Brasil**. São Paulo: Tese (Doutorado em Economia) - Faculdade de Economia, Administração e Contabilidade (FEA), Universidade de São Paulo (USP). 2019. 160 p.

SMITH, Adam. **An Inquiry into the Nature and Causes of the Wealth of Nations**. Methuen, 1950.

THÜNEN, Johan Heinrich Von. **Von Thunen's isolated state**. Oxford, ([1826] 1966).

VAN DER PANNE, Gerben. Agglomeration externalities: Marshall versus Jacobs. **Journal of Evolutionary Economics**, v. 14, n. 5, p. 593–604, 1 dez. 2004.

WEBER, Alfred. [1909]. **Theory of locations of industries**. 2nd ed. Chicago: University of Chicago Press, 1959.

WHEATON, W. C.; LEWIS, M. J. Urban Wages and Labor Market Agglomeration. **Journal of Urban Economics**, v. 51, n. 3, p. 542–562, 1 maio 2002.

WOODCOCK, Simon D. Match effects. **Research in Economics**, v.69, p. 100-121, 2015.

WOODCOCK, Simon D. Wage differentials in the presence of unobserved worker, firm, and match heterogeneity. **Labour Economics**, v. 15, n. 4, p. 771-793, 2008.

WOOLDRIDGE, Jeffrey Marc. **Econometric Analysis of Cross Section and Panel Data**. Cambridge: Massachusetts Institute of Technology - MIT Press, 2º ed, 2010. 1064 p.

## APÊNDICES

### APÊNDICE A – Tratamento Inicial dos microdados da RAIS Identificada (2010-2018)

A **Tabela A.1** mostra a quantidade de observações excluídas a cada etapa do bloco 1 de tratamento dos microdados identificados da RAIS. Na Etapa I, foram excluídos os vínculos inativos em 31 dezembro de cada ano; em seguida excluíram-se os vínculos sem registros ou com inconsistências no PIS, com renda nula ou com inconsistências, com tempo de emprego zero ou faltante e com carga horária semanal contratada inferior a 20 horas (Etapa II). Na etapa III, excluíram-se os trabalhadores da Administração Pública e as entidades públicas e de economia mista; mais adiante, na Etapa IV, excluiu-se os trabalhadores sem vínculo CLT e cujo tipo de admissão é registrado como proveniente de transferências, de contratos temporários ou por tempo determinado, bem como os registrados como aprendiz. A Etapa V excluiu os trabalhadores fora da faixa etária produtiva (18-65 anos). Por fim, na Etapa VI, definiu-se um único vínculo trabalhador-firma que caracteriza a relação de trabalho principal do indivíduo, seguindo os seguintes procedimentos: i) exclusão de trabalhadores com mais de 5 vínculos empregatício; ii) exclusão dos vínculos com divergências de gênero e de idade; iii) manutenção dos vínculos com maior carga horária semanal contratada; iv) permanência dos contratos cuja data de admissão é a mais antiga; e v) manutenção apenas de um único vínculo por trabalhador, por meio de escolha aleatória. Na **Tabela A.1**, observa-se que, após o bloco 1 de tratamento, foram mantidas, em média, 39,5% das observações totais.

**Tabela A.1: Total de vínculos ativos e inativos registrados na RAIS: 2010-2018**

Ano	Total	Etapa 1	Etapa 2	Etapa 3	Etapa 4	Etapa 5	Etapa 6	RAIS Limpa	RAIS (%)
2010	66.747.302	22.678.947	2.274.405	14.761.610	335.254	342.944	147.455	26.206.687	<b>39,26</b>
2011	70.971.125	24.660.494	2.488.338	15.531.874	372.041	384.503	170.879	27.665.869	<b>38,98</b>
2012	73.326.485	25.867.773	2.592.780	15.200.715	395.127	402.141	180.724	28.687.225	<b>39,12</b>
2013	75.400.510	26.452.077	2.568.192	15.917.740	403.338	408.768	181.573	29.468.822	<b>39,08</b>
2014	76.107.279	26.535.769	2.654.833	16.225.171	413.955	397.792	180.765	29.698.994	<b>39,02</b>
2015	73.629.639	24.644.090	2.829.115	16.453.706	387.153	314.835	702.468	28.298.272	<b>38,43</b>
2016	67.144.598	21.084.400	2.638.510	15.709.776	373.378	251.922	125.382	26.961.230	<b>40,15</b>
2017	65.655.882	19.374.292	2.599.928	16.064.366	416.767	230.600	116.556	26.853.373	<b>40,90</b>
2018	66.214.692	19.583.577	2.572.440	16.128.759	470.500	226.250	113.596	27.119.570	<b>40,96</b>
<b>Total</b>	<b>635.197.512</b>	<b>210.881.419</b>	<b>23.218.541</b>	<b>141.993.717</b>	<b>3.567.513</b>	<b>2.959.755</b>	<b>1.919.398</b>	<b>250.960.042</b>	<b>39,51</b>
<b>Média</b>	<b>70.577.501</b>	<b>23.431.269</b>	<b>2.579.838</b>	<b>15.777.080</b>	<b>396.390</b>	<b>328.862</b>	<b>213.266</b>	<b>27.884.449</b>	<b>39,51</b>

Fonte: Elaboração própria a partir dos microdados identificados da RAIS (2010-2018).

## APÊNDICE B – Determinação dos Municípios Influenciados Baseados no REGIC (2018)

A seguir será mostrado um exemplo de como os municípios influenciados são determinados a partir dos seus municípios imediatos. Os municípios imediatos são determinados no estudo REGIC do IBGE (2018) que estabelece a rede de encadeamento que o município compõe a partir dos seus vínculos com outros municípios. O encadeamento é estabelecido por meio do destino principal (vínculo imediato) para busca de bens e serviços e/ou para gestão do território até se chegar em uma metrópole, os elos finais da rede urbana, definidos no estudo REGIC (IBGE, 2018). A cidade de Palestina (AL), por exemplo, tem o seguinte encadeamento até chegar na metrópole à qual se vincula:

Palestina (AL) → Santana do Ipanema (AL) → Arapiraca (AL) → Arranjo Populacional de Maceió/AL → Arranjo Populacional de Recife/PE

Palestina (AL) compõe a região de influência de Santana do Ipanema (AL), que é seu vínculo imediato. Santana do Ipanema (AL), por sua vez, compõe a região de influência de Arapiraca (AL), por ser seu vínculo imediato. Seguindo o encadeamento de vínculos imediatos que iniciou em Palestina/AL, a primeira metrópole que aparece no sequenciamento é o arranjo populacional de Recife/PE, constituindo, assim, o fim da sequência.

Para determinar o município influenciado, isto é, o município que sofre influência de um outro município, este estudo considerou apenas o município vinculado aos municípios isolados, definidos no estudo sobre arranjos populacionais e concentrações urbanas (IBGE, 2016). Com base no sequenciamento descrito neste apêndice e sabendo que Arapiraca (AL) é definido como um município isolado no estudo sobre arranjos populacionais e concentrações urbanas do IBGE (2016), o município de Santana do Ipanema (AL) é definido como um município influenciado, pois seu município imediato (vínculo imediato) é um município isolado. Palestina (AL), neste caso, de acordo com a definição de áreas de mercado de trabalho local proposta neste estudo, não é considerado um município influenciado de Arapiraca (AL), pois seu vínculo imediato é Santana do Ipanema (AL), que não com um município isolado conforme o IBGE (2016). É importante dizer que, cidades cujo vínculo imediato é um arranjo populacional não são consideradas municípios influenciados, assim como os municípios isolados, que não são municípios influenciados sob nenhuma perspectiva. Por fim, caso um município influenciado tenha como vínculo imediato mais de um município isolado, ele será considerado município influenciado do município isolado pertencente ao seu estado e, se no mesmo estado, considerar-se-á o vínculo de menor distância.

## APÊNDICE C: Descrição das Variáveis do Modelo

**Tabela C.1: Índice de correção do IPCA (IBGE) - Ano Base 2018**

<b>Unidade da Federação</b>	<b>2010</b>	<b>2011</b>	<b>2012</b>	<b>2013</b>	<b>2014</b>	<b>2015</b>	<b>2016</b>	<b>2017</b>
RO	1,5571	1,4866	1,3727	1,3032	1,2226	1,1120	1,0415	1,0300
AC	1,5571	1,4866	1,3727	1,3032	1,2226	1,1120	1,0415	1,0300
AM	1,5571	1,4866	1,3727	1,3032	1,2226	1,1120	1,0415	1,0300
RR	1,5571	1,4866	1,3727	1,3032	1,2226	1,1120	1,0415	1,0300
PA	1,5571	1,4866	1,3727	1,3032	1,2226	1,1120	1,0415	1,0300
AP	1,5571	1,4866	1,3727	1,3032	1,2226	1,1120	1,0415	1,0300
TO	1,5571	1,4866	1,3727	1,3032	1,2226	1,1120	1,0415	1,0300
MA	1,6283	1,5296	1,4335	1,3477	1,2709	1,1405	1,0528	1,0291
PI	1,6283	1,5296	1,4335	1,3477	1,2709	1,1405	1,0528	1,0291
CE	1,6283	1,5296	1,4335	1,3477	1,2709	1,1405	1,0528	1,0291
RN	1,6283	1,5296	1,4335	1,3477	1,2709	1,1405	1,0528	1,0291
PB	1,6122	1,5200	1,4235	1,3322	1,2531	1,1377	1,0623	1,0283
PE	1,6122	1,5200	1,4235	1,3322	1,2531	1,1377	1,0623	1,0283
AL	1,6122	1,5200	1,4235	1,3322	1,2531	1,1377	1,0623	1,0283
SE	1,6122	1,5200	1,4235	1,3322	1,2531	1,1377	1,0623	1,0283
BA	1,5649	1,4694	1,3835	1,3174	1,2456	1,1340	1,0626	1,0403
MG	1,5659	1,4664	1,3830	1,3078	1,2356	1,1314	1,0613	1,0401
ES	1,5846	1,4868	1,3852	1,3048	1,2291	1,1230	1,0683	1,0418
RJ	1,6500	1,5482	1,4424	1,3587	1,2627	1,1425	1,0744	1,0429
SP	1,5904	1,4934	1,4262	1,3443	1,2669	1,1403	1,0745	1,0367
PR	1,6050	1,4982	1,4169	1,3409	1,2571	1,1167	1,0691	1,0338
SC	1,6050	1,4982	1,4169	1,3409	1,2571	1,1167	1,0691	1,0338
RS	1,6203	1,5212	1,4409	1,3622	1,2758	1,1472	1,0726	1,0461
MS	1,5686	1,4772	1,4015	1,3271	1,2430	1,1305	1,0513	1,0296
MT	1,5686	1,4772	1,4015	1,3271	1,2430	1,1305	1,0513	1,0296
GO	1,5863	1,4939	1,4173	1,3421	1,2517	1,1267	1,0702	1,0316
DF	1,5736	1,4706	1,3951	1,3163	1,2384	1,1294	1,0693	1,0305
<b>Brasil</b>	<b>1,5905</b>	<b>1,5002</b>	<b>1,4054</b>	<b>1,3271</b>	<b>1,2467</b>	<b>1,1284</b>	<b>1,0572</b>	<b>1,0325</b>

Fonte: Elaboração própria com base no IPCA (IBGE). Índice do ano base: 2018=1.

Tabela C.2: Descrição das Variáveis do Modelo e Fonte de Dados

Variável	Descrição	Tipo de Variável	Nível	Fonte dos Dados
<i>Ln</i> (salário-hora real)	Logaritmo natural do salário-hora real deflacionado pelo IPCA (ano base: 2018)	Contínua	Individual	RAIS/IBGE
Idade	Idade do trabalhador: 18 a 65 anos	Discreta	Individual	RAIS
Idade <sup>2</sup>	Idade do trabalhador (18-65 anos) ao quadrado	Discreta	Individual	RAIS
Experiência	Tempo de emprego, em meses, no mesmo vínculo empregatício	Discreta	Individual	RAIS
Experiência <sup>2</sup>	Tempo de emprego, em meses, no mesmo vínculo empregatício quadrado	Discreta	Individual	RAIS
Gênero	1 para masculino e 0 para feminino	<i>Dummy</i>	Individual	RAIS
Raça	Raça: indígena, branca, preta amarela, parda, não identificada e multirracial Categoria de referência: Branca	<i>Dummy</i>	Individual	RAIS
Educação	Nível de escolaridade: Básico 1 (analfabeto, até o 5º ano incompleto do ensino fundamental e 5º ano completo do ensino fundamental); Básico 2 (do 6º ao 9º ano incompleto do ensino fundamental, ensino fundamental completo e ensino médio incompleto); Médio (ensino médio completo e ensino superior incompleto); Superior (educação superior completa, mestrado completo e doutorado completo) Categoria de referência: Básico 1	<i>Dummy</i>	Individual	RAIS
Ocupação	Grande grupo ocupacional: Grupo 1 (Dirigentes de empresas, gerentes, aqui se excluiu as categorias do setor público); Grupo 2 (Profissionais das ciências e das artes); Grupo 3 (Técnicos de nível médio); Grupo 4 (Trabalhadores de serviços administrativos); Grupo 5 (Trabalhadores dos serviços, vendedores do comércio em lojas e mercado); Grupo 6 (Trabalhadores agropecuários, florestais e da pesca); Grupo 7 (Trabalhadores da produção de bens e serviços industriais 1); Grupo 8 (Trabalhadores da produção de bens e serviços industriais 2); Grupo 9 (Trabalhadores em serviços de reparação e manutenção). Categoria de referência: Trabalhadores da produção de bens e serviços industriais 1	<i>Dummy</i>	Individual	RAIS/MTE (CBO 2002)
Setor da firma	Classe de atividade econômica da firma: Setor 1 (Agricultura, pecuária, produção florestal, pesca e aquicultura); Setor 2 (Indústrias extrativistas); Setor 3 (Indústrias de Transformação); Setor 4 (Eletricidade e Gás); Setor 5 (Água, esgoto, atividades de gestão de resíduos e descontaminação); Setor 6 (Construção); Setor 7 (Comércio; reparação de veículos automotores e motocicletas); Setor 8 (Transporte, armazenagem e correio); Setor 9 (Alojamento e alimentação); Setor 10 (Informação e Comunicação); Setor 11 (Atividades financeiras, de seguros e serviços relacionados); Setor 12 (Atividades imobiliárias); Setor 13 (Atividades profissionais, científicas e técnicas); Setor 14 (Atividades administrativas e serviços complementares). Categoria de referência: Indústria da transformação	<i>Dummy</i>	Firma	RAIS/IBGE (CNAE 2.0)

Variável	Descrição	Tipo de Variável	Nível	Conclusão Fonte dos Dados
Tamanho da firma	Tamanho do estabelecimento com base no número de empregados da firma: Microempresa (se comércio e serviços: até 9 empregados, se indústria e construção: até 19); Empresa Pequena (se comércio e serviços: de 10 a 49 empregados, se indústria e construção: de 20 a 99); Empresa Média (se comércio e serviços: de 50 a 99 empregados, se indústria e construção: de 100 a 499); Empresa Grande (se comércio e serviços: 100 ou mais empregados, se indústria e construção: 500 ou mais). Categoria de referência: Microempresa	<i>Dummy</i>	Firma	RAIS/DIEESE-SEBRAE
<i>Ln</i> (densidade do emprego)	Logaritmo natural da razão entre o número total de empregos e a área em quilômetros quadrados (Km <sup>2</sup> ) da LMA	Contínua	Localização do emprego	RAIS/IBGE
<i>Ln</i> (área da LMA)	Logaritmo natural da área da LMA (em Km <sup>2</sup> )	Contínua	Localização do emprego	RAIS/IBGE
Tamanho da LMA	Tamanho da LMA de acordo com o contingente populacional: Pequeno (população ≤ 100 mil habitantes); Médio (100 mil habitantes < população ≤ 750 mil habitantes); Grande (população > 750 mil habitantes). Categoria de referência: Pequeno	<i>Dummy</i>	Localização do emprego	RAIS/IBGE
Região	Define a macrorregião onde a LMA está localizada: Norte, Nordeste, Sudeste, Sul e Centro – Oeste. Categoria de referência: Sudeste.	<i>Dummy</i>	Localização do emprego	RAIS

Fonte: Elaboração própria.

## APÊNDICE D: Tabelas dos Resultados

**Tabela D.1: Efeitos das economias de aglomeração sobre os salários individuais nas LMA (2010-2018)**

Variáveis	Continua				
	(I) POLS	(II) EF-ID	(III) EF-Firma	(IV) EF-ID e Firma	(V) EF-ID, Firma e ID-Firma
<i>Ln. Dens. Emprego</i>	0.0407*** (0.0003)	0.0061*** (0.0014)	0.0216*** (0.0084)	0.0191*** (0.0072)	0.0190*** (0.0019)
Idade	0.0420*** (0.0001)	0.0830*** (0.0002)	0.0392*** (0.0004)	0.0720 (0.3116)	0.0729*** (0.0002)
Idade <sup>2</sup>	-0.0005*** (0.0000)	-0.0006*** (0.0000)	-0.0004*** (0.0000)	-0.0005*** (0.0000)	-0.0005*** (0.0000)
Experiência	0.0009*** (0.0000)	0.0009*** (0.0000)	0.0017*** (0.0000)	0.0010*** (0.0000)	0.0008*** (0.0000)
Experiência <sup>2</sup>					
Masculino	0.2529*** (0.0003)		0.1960*** (0.0016)		
Indígena	-0.0880*** (0.0058)		-0.0593*** (0.0142)		
Preta	-0.1358*** (0.0008)		-0.1042*** (0.0022)		
Amarela	0.0458*** (0.0025)		0.0298*** (0.0066)		
Parda	-0.1253*** (0.0004)		-0.0745*** (0.0015)		
Não identificada	-0.0189*** (0.0013)		-0.0501*** (0.0061)		
Multirracial	-0.1049*** (0.0004)		-0.0655*** (0.0014)		
Básico 2	0.1140*** (0.0006)	-0.0009 (0.0016)	0.1009*** (0.0021)	0.0020 (0.0022)	-0.0001 (0.0017)
Médio	0.3277*** (0.0006)	0.0031* (0.0017)	0.2117*** (0.0025)	-0.0003 (0.0025)	0.0009 (0.0018)
Superior	0.9176*** (0.0009)	0.0795*** (0.0019)	0.5918*** (0.0044)	0.0556*** (0.0032)	0.0478*** (0.0020)
Dir. de empresas e gerentes	0.6253*** (0.0008)	0.1241*** (0.0013)	0.6703*** (0.0050)	0.1239*** (0.0022)	0.1091*** (0.0013)
Prof. das ciên. e das artes	0.4288*** (0.0009)	0.0570*** (0.0012)	0.3718*** (0.0058)	0.0555*** (0.0021)	0.0511*** (0.0012)
Téc. nível médio	0.2706*** (0.0006)	0.0190*** (0.0010)	0.1998*** (0.0038)	0.0198*** (0.0020)	0.0186*** (0.0010)
Trab. serviços admin.	0.0125*** (0.0005)	-0.0248*** (0.0009)	0.0358*** (0.0037)	-0.0154*** (0.0017)	-0.0140*** (0.0009)
Trab. serv e vend. do comérc. em lojas e mercad.	-0.0577*** (0.0005)	-0.0350*** (0.0010)	-0.0219*** (0.0057)	-0.0253*** (0.0024)	-0.0238*** (0.0011)
Trab. agropec., florest. e da pesca	-0.0853*** (0.0014)	-0.0061** (0.0028)	-0.1170*** (0.0108)	-0.0057 (0.0036)	-0.0090*** (0.0028)
Trab. da prod. de bens e serv. Industriais 2	0.0139*** (0.0007)	0.0192*** (0.0011)	0.0207*** (0.0044)	0.0176*** (0.0019)	0.0141*** (0.0010)
Trab. em serv. de reparação e manutenção	0.2216*** (0.0008)	0.0206*** (0.0014)	0.1928*** (0.0052)	0.0225*** (0.0022)	0.0225*** (0.0014)

Variáveis	Continua				
	(I) POLS	(II) EF-ID	(III) EF-Firma	(IV) EF-ID e Firma	(V) EF-ID, Firma e ID-Firma
Agricultura, pecuária, prod. florestal, pesca e aquic.	-0.0106*** (0.0017)	-0.0161*** (0.0042)	-0.0210* (0.0112)	-0.0117 (0.0105)	-0.0124*** (0.0045)
Eletricidade e gás	0.2232*** (0.0019)	0.1008*** (0.0065)	-0.0128 (0.0129)	-0.0045 (0.0078)	-0.0062 (0.0052)
Água, esgoto, ativ. de gestão de resíd. e descont.	-0.2063*** (0.0017)	-0.0253*** (0.0038)	-0.0046 (0.0136)	0.0090 (0.0071)	0.0084** (0.0041)
Construção	-0.0628*** (0.0008)	-0.0456*** (0.0019)	-0.0036 (0.0074)	0.0091 (0.0060)	0.0078*** (0.0023)
Comércio; Reparação. de veículos auto. e motocic.	-0.2076*** (0.0004)	-0.0658*** (0.0011)	-0.0169*** (0.0036)	-0.0127*** (0.0029)	-0.0154*** (0.0012)
Transporte, armazenagem e correio	-0.1433*** (0.0006)	-0.0680*** (0.0018)	-0.0152** (0.0061)	-0.0075* (0.0041)	-0.0102*** (0.0025)
Alojamento e alimentação	-0.3482*** (0.0008)	-0.0908*** (0.0028)	-0.0127** (0.0063)	-0.0068 (0.0049)	-0.0092*** (0.0034)
Informação e comunicação	-0.0920*** (0.0010)	-0.0634*** (0.0022)	-0.0122 (0.0080)	-0.0070 (0.0078)	-0.0085*** (0.0024)
Atividades financ., de seguros e serv. relacion.	0.0503*** (0.0009)	0.0691*** (0.0031)	-0.0057 (0.0112)	-0.0107* (0.0061)	-0.0125*** (0.0032)
Atividades imobiliárias	-0.1906*** (0.0023)	-0.0698*** (0.0048)	0.0069 (0.0098)	0.0128 (0.0099)	0.0119** (0.0050)
Atividades profissionais, científicas e técnicas	-0.1357*** (0.0010)	-0.0916*** (0.0020)	-0.0142** (0.0070)	-0.0021 (0.0049)	-0.0043* (0.0025)
Atividades administrativas e serv. complementares	-0.3332*** (0.0006)	-0.0893*** (0.0013)	-0.0146** (0.0058)	-0.0047 (0.0040)	-0.0090*** (0.0016)
Pequena	0.1925*** (0.0005)	0.0509*** (0.0005)	0.0203*** (0.0009)	0.0209*** (0.0008)	0.0233*** (0.0005)
Média	0.2976*** (0.0005)	0.0904*** (0.0008)	0.0408*** (0.0017)	0.0474*** (0.0015)	0.0514*** (0.0007)
Grande	0.3543*** (0.0005)	0.1203*** (0.0009)	0.0563*** (0.0026)	0.0700*** (0.0022)	0.0746*** (0.0009)
<i>Ln.</i> área LMA	0.0436*** (0.0002)	0.0061*** (0.0006)	0.0053* (0.0031)	0.0032 (0.0025)	0.0021** (0.0008)
Média	-0.0227*** (0.0008)	0.0162*** (0.0036)	0.0119 (0.0180)	0.0245 (0.0158)	0.0280*** (0.0050)
Grande	-0.0303*** (0.0010)	0.0129*** (0.0043)	0.0030 (0.0214)	0.0131 (0.0190)	0.0187*** (0.0060)
Nordeste	-0.2606*** (0.0005)	-0.0305*** (0.0035)	-0.0835* (0.0480)	-0.0087 (0.0144)	0.0002 (0.0110)
Norte	-0.0798*** (0.0011)	0.0167** (0.0066)	-0.0617 (0.0712)	0.0289 (0.0250)	0.0441*** (0.0120)
Sul	0.0062*** (0.0004)	-0.0271*** (0.0031)	-0.0383 (0.0254)	-0.0124 (0.0158)	-0.0132* (0.0068)
Centro-Oeste	-0.0172*** (0.0008)	0.0220*** (0.0041)	0.1512*** (0.0519)	0.1595*** (0.0446)	0.1640*** (0.0047)
2011.ano	-0.0020*** (0.0006)	0.0214*** (0.0002)	0.0260*** (0.0009)	0.0246 (0.3116)	0.0191*** (0.0002)
2012.ano	0.0450*** (0.0006)	0.0593*** (0.0002)	0.0648*** (0.0012)	0.0618 (0.6233)	0.0525*** (0.0002)

Variáveis	Conclusão				
	(I) POLS	(II) EF-ID	(III) EF-Firma	(IV) EF-ID e Firma	(V) EF-ID, Firma e ID-Firma
2014.ano	0.0896*** (0.0006)	0.0875*** (0.0002)	0.1009*** (0.0024)	0.0945 (1.2465)	0.0787*** (0.0002)
2015.ano	0.0545*** (0.0006)	0.0399*** (0.0002)	0.0604*** (0.0022)	0.0506 (1.5581)	0.0317*** (0.0002)
2016.ano	0.0597*** (0.0006)	0.0310*** (0.0002)	0.0612*** (0.0025)	0.0467 (1.8697)	0.0266*** (0.0002)
2017.ano	0.0640*** (0.0006)	0.0269*** (0.0002)	0.0627*** (0.0026)	0.0471 (2.1813)	0.0253*** (0.0002)
2018.ano	0.0468*** (0.0007)		0.0421*** (0.0027)	0.0242 (2.4930)	
Constante	0.4259*** (0.0029)	0.1914*** (0.0089)	1.1369*** (0.0405)	0.6050 (10.7378)	0.5341*** (0.0127)
Ano	Não	Sim	Sim	Sim	Sim
Efeito Fixo ID	Não	Sim	Não	Sim	Sim
Efeito Fixo Firma	Não	Não	Sim	Sim	Sim
Efeito Fixo ID x Firma	Não	Não	Não	Não	Sim
R <sup>2</sup>	0.5668	0.9301	0.7826	0.9431	0.9501
R <sup>2</sup> ajustado	0.567	0.921	0.774	0.933	0.941
R <sup>2</sup> Within		0.258	0.365	0.223	0.228
Observações	11,585,637	11,585,637	11,516,579	11,516,578	11,160,443
F	268604	27540	3393		22028
Prob>F	0	0	0		0

Fonte: Elaboração própria com base nos dados identificados da RAIS 2010-2018.

Nota: Erro-padrão robusto entre parênteses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Tabela D.2: Efeitos das economias de aglomeração sobre os salários individuais nas LMA (2010-2018) – VI-MQ2E**

Variáveis	(I)	(II)	(III)	(IV)	Continua (V)
	POLS	EF-ID	EF-Firma	EF-ID e Firma	EF-ID, Firma e ID- Firma
<i>Ln. Dens. Emprego</i>	0.0176*** (0.0005)	0.0168*** (0.0024)	0.0269* (0.0148)	0.0362*** (0.0025)	0.0169*** (0.0013)
Idade	0.0420*** (0.0001)	0.0829*** (0.0002)	0.0392*** (0.0004)	0.0749 (0.4668)	0.0725*** (0.0001)
Idade <sup>2</sup>	-0.0005*** (0.0000)	-0.0006*** (0.0000)	-0.0004*** (0.0000)	-0.0005*** (0.0000)	-0.0005*** (0.0000)
Experiência	0.0009*** (0.0000)	0.0009*** (0.0000)	0.0017*** (0.0000)	0.0010*** (0.0000)	0.0008*** (0.0000)
Experiência <sup>2</sup>	0.0000*** (0.0000)		-0.0000*** (0.0000)		-0.0000*** (0.0000)
Masculino	0.2518*** (0.0003)		0.1960*** (0.0016)		0.0000 (0.0000)
Indígena	-0.0917*** (0.0056)		-0.0593*** (0.0142)		0.0000 (0.0000)
Preta	-0.1368*** (0.0009)		-0.1042*** (0.0022)		0.0000 (0.0000)
Amarela	0.0469*** (0.0022)		0.0298*** (0.0066)		0.0000 (0.0000)
Parda	-0.1262*** (0.0004)		-0.0745*** (0.0015)		0.0000 (0.0000)
Não identificada	-0.0175*** (0.0013)		-0.0501*** (0.0061)		0.0000 (0.0000)
Multirracial	-0.1057*** (0.0004)		-0.0655*** (0.0014)		0.0000 (0.0000)
Básico 2	0.1139*** (0.0008)		0.1009*** (0.0021)	0.0026** (0.0011)	0.0007 (0.0013)
Médio	0.3287*** (0.0007)	0.0032* (0.0017)	0.2117*** (0.0025)	-0.0000 (0.0012)	0.0013 (0.0013)
Superior	0.9216*** (0.0009)	0.0796*** (0.0019)	0.5918*** (0.0044)	0.0558*** (0.0013)	0.0480*** (0.0014)
Dir. de empresas e gerentes	0.6271*** (0.0007)	0.1237*** (0.0013)	0.6703*** (0.0050)	0.1235*** (0.0007)	0.1085*** (0.0007)
Prof. das ciên. e das artes	0.4289*** (0.0008)	0.0568*** (0.0012)	0.3718*** (0.0058)	0.0552*** (0.0007)	0.0507*** (0.0007)
Téc. nível médio	0.2712*** (0.0006)	0.0189*** (0.0010)	0.1998*** (0.0038)	0.0197*** (0.0006)	0.0185*** (0.0006)
Trab. serviços admin.	0.0123*** (0.0005)	-0.0249*** (0.0009)	0.0358*** (0.0037)	-0.0155*** (0.0005)	-0.0140*** (0.0006)
Trab. serv e vend. do comérc. em lojas e mercad.	-0.0591*** (0.0005)	-0.0355*** (0.0010)	-0.0219*** (0.0057)	-0.0256*** (0.0006)	-0.0241*** (0.0006)
Trab. agropec., florest. e da pesca	-0.0868*** (0.0017)	-0.0086*** (0.0029)	-0.1170*** (0.0108)	-0.0065*** (0.0019)	-0.0099*** (0.0019)
Trab.da prod. de bens e serv. Industriais 2	0.0151*** (0.0007)	0.0191*** (0.0011)	0.0207*** (0.0044)	0.0176*** (0.0007)	0.0140*** (0.0007)
Trab. em serv. de reparação e manutenção	0.2210*** (0.0008)	0.0202*** (0.0014)	0.1928*** (0.0052)	0.0222*** (0.0008)	0.0223*** (0.0009)

Variáveis	(I)	(II)	(III)	(IV)	Continua (V)
	POLS	EF-ID	EF-Firma	EF-ID e Firma	EF-ID, Firma e ID- Firma
	(0.0017)	(0.0049)	(0.0150)	(0.0040)	(0.0039)
Agricultura, pecuária, prod. florestal, pesca e aquic.	-0.0263***	-0.0169***	-0.0209*	-0.0125***	-0.0133***
	(0.0019)	(0.0043)	(0.0112)	(0.0037)	(0.0035)
Eletricidade e gás	0.2148***	0.0936***	-0.0129	-0.0049	-0.0039
	(0.0017)	(0.0066)	(0.0129)	(0.0052)	(0.0050)
Água, esgoto, ativ. de gestão de resíd. e descont.	-0.2085***	-0.0267***	-0.0047	0.0089**	0.0086**
	(0.0019)	(0.0038)	(0.0136)	(0.0037)	(0.0035)
Construção	-0.0656***	-0.0464***	-0.0036	0.0091***	0.0081***
	(0.0008)	(0.0019)	(0.0075)	(0.0017)	(0.0016)
Comércio; Reparação. de veículos auto. e motocic.	-0.2125***	-0.0664***	-0.0170***	-0.0138***	-0.0160***
	(0.0005)	(0.0011)	(0.0035)	(0.0009)	(0.0009)
Transporte, armazenagem e correio	-0.1456***	-0.0686***	-0.0153**	-0.0073***	-0.0097***
	(0.0006)	(0.0018)	(0.0061)	(0.0020)	(0.0020)
Alojamento e alimentação	-0.3511***	-0.0914***	-0.0127**	-0.0070**	-0.0097***
	(0.0009)	(0.0028)	(0.0063)	(0.0030)	(0.0029)
Informação e comunicação	-0.0949***	-0.0644***	-0.0122	-0.0079***	-0.0096***
	(0.0009)	(0.0022)	(0.0080)	(0.0017)	(0.0017)
Atividades financ., de seguros e serv. relacion.	0.0459***	0.0679***	-0.0057	-0.0117***	-0.0123***
	(0.0008)	(0.0031)	(0.0112)	(0.0027)	(0.0027)
Atividades imobiliárias	-0.1914***	-0.0702***	0.0068	0.0109***	0.0102***
	(0.0022)	(0.0049)	(0.0098)	(0.0039)	(0.0038)
Atividades profissionais, científicas e técnicas	-0.1366***	-0.0925***	-0.0142**	-0.0027	-0.0047***
	(0.0009)	(0.0020)	(0.0070)	(0.0018)	(0.0018)
Atividades administrativas e serv. complementares	-0.3352***	-0.0903***	-0.0147**	-0.0051***	-0.0094***
	(0.0006)	(0.0014)	(0.0058)	(0.0012)	(0.0012)
Pequena	0.1932***	0.0506***	0.0203***	0.0205***	0.0229***
	(0.0005)	(0.0006)	(0.0009)	(0.0004)	(0.0004)
Média	0.2984***	0.0899***	0.0408***	0.0468***	0.0508***
	(0.0005)	(0.0008)	(0.0017)	(0.0006)	(0.0006)
Grande	0.3568***	0.1197***	0.0563***	0.0696***	0.0743***
	(0.0005)	(0.0009)	(0.0026)	(0.0007)	(0.0007)
Ln. área LMA	0.0354***	0.0096***	0.0068	0.0083***	-0.0099***
	(0.0002)	(0.0009)	(0.0048)	(0.0009)	(0.0007)
Média	-0.0168***	0.0137***	0.0101	0.0221***	0.0416***
	(0.0009)	(0.0039)	(0.0182)	(0.0042)	(0.0040)
Grande	0.0014	-0.0007	-0.0029	-0.0036	0.0279***
	(0.0012)	(0.0050)	(0.0254)	(0.0053)	(0.0046)
Nordeste	-0.2676***	-0.0279***	-0.0829*	-0.0051	0.0247***
	(0.0006)	(0.0036)	(0.0481)	(0.0085)	(0.0089)
Norte	-0.1338***	0.0329***	-0.0518	0.0526***	0.0979***
	(0.0014)	(0.0078)	(0.0762)	(0.0132)	(0.0145)
Sul	0.0103***	-0.0234***	-0.0393	-0.0140**	0.0034
	(0.0004)	(0.0031)	(0.0246)	(0.0057)	(0.0061)

Variáveis	Conclusão				
	(I) POLS	(II) EF-ID	(III) EF-Firma	(IV) EF-ID e Firma	(V) EF-ID, Firma e ID- Firma
	(0.0010)	(0.0049)	(0.0527)	(0.0045)	(0.0055)
2011.ano	-0.0023***	0.0216***	0.0260***	0.0218	0.0178***
	(0.0006)	(0.0002)	(0.0009)	(0.4668)	(0.0003)
2012.ano	0.0443***	0.0594***	0.0648***	0.0558	0.0511***
	(0.0006)	(0.0002)	(0.0012)	(0.9337)	(0.0003)
2013.ano	0.0772***	0.0857***	0.0943***	0.0812	0.0774***
	(0.0006)	(0.0002)	(0.0018)	(1.4005)	(0.0002)
2014.ano	0.0882***	0.0873***	0.1009***	0.0822	0.0794***
	(0.0006)	(0.0002)	(0.0024)	(1.8673)	(0.0002)
2015.ano	0.0531***	0.0399***	0.0604***	0.0356	0.0313***
	(0.0006)	(0.0002)	(0.0022)	(2.3341)	(0.0002)
2016.ano	0.0579***	0.0309***	0.0612***	0.0285	0.0265***
	(0.0006)	(0.0002)	(0.0025)	(2.8010)	(0.0002)
2017.ano	0.0621***	0.0268***	0.0627***	0.0259	0.0254***
	(0.0007)	(0.0002)	(0.0026)	(3.2678)	(0.0002)
2018.ano	0.0447***	0.0000	0.0421***	0.0000	0.0000
	(0.0007)	(0.0000)	(0.0027)	(3.7346)	(0.0000)
Constante	0.3957***				
	(0.0031)				
Ano	Sim	Sim	Sim	Sim	Sim
Efeito Fixo ID	Não	Sim	Não	Sim	Sim
Efeito Fixo Firma	Não	Não	Sim	Sim	Sim
Efeito Fixo ID x Firma	Não	Não	Não	Não	Sim
R <sup>2</sup>	0.5685	0.2566	0.3654	0.2220	0.2259
R <sup>2</sup> ajustado	0.568	0.257	0.365	0.0994	0.0761
F 1º estágio	7427	53037	85.30	6033	370985
Prob> F 1º estágio	0	0	0	0	0
Observações	11,352,518	11,351,500	11,516,579	11,284,391	10,933,643
F	276752	26831	3402	59201	58214
Prob>F	0	0	0	0	0

Fonte: Elaboração própria com base nos dados identificados da RAIS 2010-2018.

Nota: Erro-padrão entre parênteses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Tabela D.3: Efeitos das economias de aglomeração sobre os salários individuais conforme tamanho das LMA (2010-2018)**

Variáveis	Continua					
	(I) Pequena EF-ID, Firma e ID-Firma	(II) Pequena VI e EF- ID, Firma e ID-Firma	(III) Média EF-ID, Firma e ID-Firma	(IV) Média VI e EF-ID, Firma e ID- Firma	(V) Grande EF-ID, Firma e ID- Firma	(VI) Grande VI e EF-ID, Firma e ID- Firma
<i>Ln. Dens. Emprego</i>	0.0334*** (0.0112)	-0.0113 (0.0411)	0.0365*** (0.0047)	0.0586*** (0.0167)	-0.0109*** (0.0028)	-0.0016 (0.0035)
Idade	0.0687*** (0.0009)	0.0680*** (0.0006)	0.0687*** (0.0004)	0.0679*** (0.0002)	0.0750*** (0.0003)	0.0750*** (0.0002)
Idade <sup>2</sup>	-0.0004*** (0.0000)	-0.0004*** (0.0000)	-0.0004*** (0.0000)	-0.0004*** (0.0000)	-0.0005*** (0.0000)	-0.0005*** (0.0000)
Experiência	0.0008*** (0.0000)	0.0008*** (0.0000)	0.0008*** (0.0000)	0.0008*** (0.0000)	0.0008*** (0.0000)	0.0008*** (0.0000)
Experiência <sup>2</sup>						
Masculino						
Indígena						
Preta						
Amarela						
Parda						
Não identificada						
Multirracial						
Básico 2	-0.0044 (0.0061)	-0.0013 (0.0050)	-0.0009 (0.0032)	0.0012 (0.0023)	0.0005 (0.0022)	0.0005 (0.0016)
Médio	-0.0010 (0.0063)	0.0004 (0.0051)	0.0031 (0.0033)	0.0049** (0.0023)	-0.0001 (0.0022)	-0.0001 (0.0016)
Superior	0.0434*** (0.0078)	0.0418*** (0.0059)	0.0571*** (0.0038)	0.0584*** (0.0025)	0.0442*** (0.0025)	0.0442*** (0.0018)
Dir. de empresas e gerentes	0.1453*** (0.0072)	0.1468*** (0.0041)	0.1309*** (0.0026)	0.1304*** (0.0014)	0.0990*** (0.0015)	0.0990*** (0.0009)
Prof. das ciên. e das artes	0.0817*** (0.0076)	0.0833*** (0.0045)	0.0665*** (0.0025)	0.0664*** (0.0015)	0.0434*** (0.0014)	0.0434*** (0.0009)
Téc. nível médio	0.0421*** (0.0048)	0.0438*** (0.0030)	0.0272*** (0.0017)	0.0278*** (0.0010)	0.0129*** (0.0012)	0.0129*** (0.0007)
Trab. serviços admin.	-0.0041 (0.0044)	-0.0049* (0.0029)	-0.0104*** (0.0017)	-0.0095*** (0.0010)	-0.0172*** (0.0012)	-0.0172*** (0.0007)
Trab. serv e vend. do comérc. em lojas e mercad.	-0.0172*** (0.0051)	-0.0193*** (0.0032)	-0.0112*** (0.0021)	-0.0115*** (0.0012)	-0.0293*** (0.0014)	-0.0293*** (0.0008)
Trab. agropec., florest. e da pesca	-0.0096 (0.0065)	-0.0105** (0.0046)	-0.0038 (0.0042)	-0.0059** (0.0027)	-0.0072 (0.0045)	-0.0071** (0.0033)
Trab. da prod. de bens e serv. Industriais 2	0.0058 (0.0039)	0.0034 (0.0027)	0.0180*** (0.0017)	0.0181*** (0.0011)	0.0135*** (0.0014)	0.0135*** (0.0009)
Trab. em serv. de reparação e manutenção	0.0370*** (0.0063)	0.0391*** (0.0041)	0.0323*** (0.0025)	0.0321*** (0.0015)	0.0165*** (0.0017)	0.0165*** (0.0011)

Variáveis	(I)	(II)	(III)	(IV)	(V)	Continua (VI)
	Pequena EF-ID, Firma e ID-Firma	Pequena VI e EF- ID, Firma e ID-Firma	Média EF-ID, Firma e ID-Firma	Média VI e EF-ID, Firma e ID- Firma	Grande EF-ID, Firma e ID- Firma	Grande VI e EF-ID, Firma e ID- Firma
	(0.0267)	(0.0181)	(0.0084)	(0.0059)	(0.0071)	(0.0054)
Agricultura, pecuária, prod. florestal, pesca e aquic.	0.0067	0.0076	0.0015	0.0010	-0.0338***	-0.0338***
	(0.0107)	(0.0086)	(0.0078)	(0.0055)	(0.0066)	(0.0052)
Eletricidade e gás	0.0019	0.0044	0.0235	0.0233	-0.0124**	-0.0125**
	(0.0273)	(0.0403)	(0.0341)	(0.0237)	(0.0053)	(0.0053)
Água, esgoto, ativ. de gestão de resíd. e descont.	-0.0211	-0.0643**	0.0246**	0.0288***	0.0040	0.0040
	(0.0229)	(0.0277)	(0.0113)	(0.0087)	(0.0045)	(0.0040)
Construção	-0.0117	-0.0330**	-0.0012	0.0009	0.0103***	0.0103***
	(0.0142)	(0.0129)	(0.0050)	(0.0034)	(0.0026)	(0.0019)
Comércio; Reparação. de veículos auto. e motocic.	-0.0261***	-0.0278***	-0.0204***	-0.0221***	-0.0183***	-0.0183***
	(0.0069)	(0.0054)	(0.0024)	(0.0018)	(0.0015)	(0.0011)
Transporte, armazenagem e correio	0.0064	0.0161	-0.0140***	-0.0109***	-0.0123***	-0.0123***
	(0.0123)	(0.0104)	(0.0046)	(0.0036)	(0.0032)	(0.0025)
Alojamento e alimentação	-0.0244	-0.0276*	-0.0080	-0.0062	-0.0121***	-0.0121***
	(0.0222)	(0.0160)	(0.0078)	(0.0062)	(0.0038)	(0.0033)
Informação e comunicação	-0.0200	-0.0268**	0.0065	0.0066*	-0.0171***	-0.0171***
	(0.0165)	(0.0124)	(0.0053)	(0.0039)	(0.0028)	(0.0019)
Atividades financ., de seguros e serv. relacion.	0.0478	0.0597	-0.0496***	-0.0525***	-0.0109***	-0.0109***
	(0.0457)	(0.0363)	(0.0097)	(0.0070)	(0.0034)	(0.0029)
Atividades imobiliárias	-0.0206	-0.0204	0.0287**	0.0195**	0.0070	0.0070
	(0.0265)	(0.0285)	(0.0132)	(0.0091)	(0.0054)	(0.0043)
Atividades profissionais, científicas e técnicas	0.0060	-0.0112	-0.0012	0.0004	-0.0078***	-0.0078***
	(0.0196)	(0.0154)	(0.0065)	(0.0044)	(0.0028)	(0.0020)
Atividades administrativas e serv. complementares	-0.0032	-0.0083	-0.0013	-0.0012	-0.0126***	-0.0126***
	(0.0105)	(0.0093)	(0.0037)	(0.0029)	(0.0018)	(0.0014)
Pequena	0.0206***	0.0194***	0.0279***	0.0272***	0.0213***	0.0213***
	(0.0021)	(0.0018)	(0.0008)	(0.0007)	(0.0006)	(0.0005)
Média	0.0494***	0.0480***	0.0593***	0.0584***	0.0478***	0.0478***
	(0.0032)	(0.0028)	(0.0013)	(0.0010)	(0.0009)	(0.0007)
Grande	0.0677***	0.0695***	0.0838***	0.0831***	0.0712***	0.0712***
	(0.0041)	(0.0035)	(0.0016)	(0.0013)	(0.0011)	(0.0008)
Ln. área LMA	0.0161	-0.0336	0.0185***	0.0343***	-0.0016	0.0005
	(0.0163)	(0.0491)	(0.0046)	(0.0094)	(0.0010)	(0.0010)
Nordeste	-0.1163***	0.0000	-0.0583	-0.0039	-0.0014	-0.0028
	(0.0040)	(0.0000)	(0.0492)	(0.0456)	(0.0118)	(0.0097)
Norte		0.0000	0.0203	-0.1810	-0.0265**	-0.0089
		(0.0000)	(0.0390)	(0.1538)	(0.0130)	(0.0149)
Sul		0.0000	0.0004	-0.0737**	-0.0178**	-0.0202***
		(0.0000)	(0.0199)	(0.0294)	(0.0086)	(0.0074)
Centro-Oeste		0.0000	0.0855***	0.1307***	-0.0333***	-0.0203*
		(0.0000)	(0.0154)	(0.0373)	(0.0121)	(0.0119)
2011.ano	0.0225***	0.0234***	0.0165***	0.0168***	0.0198***	0.0198***
	(0.0010)	(0.0012)	(0.0004)	(0.0005)	(0.0003)	(0.0003)

Variáveis	Conclusão					
	(I) Pequena EF-ID, Firma e ID-Firma	(II) Pequena VI e EF- ID, Firma e ID-Firma	(III) Média EF-ID, Firma e ID-Firma	(IV) Média VI e EF-ID, Firma e ID- Firma	(V) Grande EF-ID, Firma e ID- Firma	(VI) Grande VI e EF-ID, Firma e ID- Firma
2013.ano	(0.0011) 0.0815***	(0.0012) 0.0808***	(0.0004) 0.0783***	(0.0004) 0.0776***	(0.0003) 0.0771***	(0.0003) 0.0771***
2014.ano	(0.0011) 0.0829***	(0.0012) 0.0809***	(0.0004) 0.0822***	(0.0004) 0.0817***	(0.0003) 0.0770***	(0.0003) 0.0770***
2015.ano	(0.0011) 0.0317***	(0.0012) 0.0321***	(0.0004) 0.0329***	(0.0004) 0.0328***	(0.0003) 0.0312***	(0.0003) 0.0312***
2016.ano	(0.0010) 0.0359***	(0.0011) 0.0348***	(0.0004) 0.0262***	(0.0004) 0.0259***	(0.0003) 0.0262***	(0.0003) 0.0262***
2017.ano	(0.0009) 0.0321***	(0.0011) 0.0319***	(0.0004) 0.0262***	(0.0004) 0.0261***	(0.0002) 0.0245***	(0.0003) 0.0245***
2018.ano	(0.0008)	(0.0011)	(0.0003)	(0.0004)	(0.0002)	(0.0003)
Constante	0.3574** (0.1661)	0.0000 (0.0000)	0.4170*** (0.0635)	0.0000 (0.0000)	0.4372*** (0.0196)	0.0000 (0.0000)
Ano	Sim	Sim	Sim	Sim	Sim	Sim
Efeito Fixo ID	Sim	Sim	Sim	Sim	Sim	Sim
Efeito Fixo Firma	Sim	Sim	Sim	Sim	Sim	Sim
Efeito Fixo ID x Firma	Sim	Sim	Sim	Sim	Sim	Sim
R <sup>2</sup>	0.9337	0.2729	0.9373	0.2448	0.9537	0.2177
R <sup>2</sup> ajustado	0.922	0.151	0.926	0.108	0.944	0.0609
R <sup>2</sup> <i>within</i>	0.270		0.249		0.218	
F 1º estágio		18823		147621		101000
Prob> F 1º estágio		0		0		0
Observações	453,294	377,905	3,017,863	2,867,034	7,677,036	7,677,036
F	1300.46	3035	6878.57	17883	15026.97	40447
Prob>F	0	0	0	0	0	0

Fonte: Elaboração própria com base nos dados identificados da RAIS 2010-2018.

Nota: Erro-padrão entre parênteses (colunas II, IV e VI). Erro-padrão robusto entre parênteses (colunas I, III, e V). \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Tabela D.4: Efeitos das economias de aglomeração sobre os salários individuais conforme macrolocalização das LMA (2010-2018)

Continua

Variáveis	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)	(IX)	(X)
	Sudeste EF-ID, Firma e ID- Firma	Sudeste VI e EF-ID, Firma e ID- Firma	Norte EF-ID, Firma e ID- Firma	Norte VI e EF-ID, Firma e ID- Firma	Nordeste EF-ID, Firma e ID- Firma	Nordeste VI e EF-ID, Firma e ID- Firma	Sul EF-ID, Firma e ID- Firma	Sul VI e EF-ID, Firma e ID- Firma	Centro-Oeste EF-ID, Firma e ID- Firma	Centro-Oeste VI e EF-ID, Firma e ID- Firma
<i>Ln. Dens.</i>	0.0164***	0.0307***	-0.0070*	0.0014	0.0382**	-0.0301	0.0224***	0.0482***	-0.0091	0.0277
Emprego	(0.0032)	(0.0059)	(0.0038)	(0.0044)	(0.0165)	(0.0205)	(0.0050)	(0.0064)	(0.0073)	(0.0241)
Idade	0.0722***	0.0722***	0.0662***	0.0663***	0.0789***	0.0786***	0.0753***	0.0748***	0.0806***	0.0801***
	(0.0003)	(0.0002)	(0.0006)	(0.0004)	(0.0014)	(0.0009)	(0.0005)	(0.0003)	(0.0010)	(0.0006)
Idade <sup>2</sup>	-0.0005***	-0.0005***	-0.0004***	-0.0004***	-0.0005***	-0.0005***	-0.0005***	-0.0005***	-0.0005***	-0.0005***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Experiência	0.0008***	0.0008***	0.0006***	0.0006***	0.0005***	0.0005***	0.0010***	0.0010***	0.0009***	0.0009***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Experiência <sup>2</sup>		-0.0000***		-0.0000***		-0.0000***		-0.0000***		-0.0000***
		(0.0000)		(0.0000)		(0.0000)		(0.0000)		(0.0000)
Masculino		0.0000		0.0000		0.0000		0.0000		0.0000
		(0.0000)		(0.0000)		(0.0000)		(0.0000)		(0.0000)
Indígena		0.0000		0.0000		0.0000		0.0000		0.0000
		(0.0000)		(0.0000)		(0.0000)		(0.0000)		(0.0000)
Preta		0.0000		0.0000		0.0000		0.0000		0.0000
		(0.0000)		(0.0000)		(0.0000)		(0.0000)		(0.0000)
Amarela		0.0000		0.0000		0.0000		0.0000		0.0000
		(0.0000)		(0.0000)		(0.0000)		(0.0000)		(0.0000)
Parda		0.0000		0.0000		0.0000		0.0000		0.0000
		(0.0000)		(0.0000)		(0.0000)		(0.0000)		(0.0000)
Não identificada		0.0000		0.0000		0.0000		0.0000		0.0000
		(0.0000)		(0.0000)		(0.0000)		(0.0000)		(0.0000)
Multirracial		0.0000		0.0000		0.0000		0.0000		0.0000
		(0.0000)		(0.0000)		(0.0000)		(0.0000)		(0.0000)
Básico 2	0.0042*	0.0042***	-0.0024	-0.0024	0.0077	0.0080	-0.0024	0.0010	-0.0294***	-0.0263***
	(0.0022)	(0.0016)	(0.0046)	(0.0033)	(0.0123)	(0.0095)	(0.0042)	(0.0032)	(0.0072)	(0.0059)
Médio	0.0052**	0.0053***	-0.0021	-0.0021	0.0048	0.0023	-0.0059	-0.0043	-0.0150**	-0.0121**
	(0.0023)	(0.0016)	(0.0046)	(0.0033)	(0.0124)	(0.0096)	(0.0043)	(0.0033)	(0.0073)	(0.0060)

Variáveis	<b>Continua</b>									
	(I) Sudeste EF-ID, Firma e ID- Firma	(II) Sudeste VI e EF-ID, Firma e ID- Firma	(III) Norte EF-ID, Firma e ID- Firma	(IV) Norte VI e EF-ID, Firma e ID- Firma	(V) Nordeste EF-ID, Firma e ID- Firma	(VI) Nordeste VI e EF-ID, Firma e ID- Firma	(VII) Sul EF-ID, Firma e ID- Firma	(VIII) Sul VI e EF-ID, Firma e ID- Firma	(IX) Centro-Oeste EF-ID, Firma e ID- Firma	(X) Centro-Oeste VI e EF-ID, Firma e ID- Firma
Dir. de empresas e gerentes	0.1012*** (0.0016)	0.1011*** (0.0009)	0.1249*** (0.0041)	0.1246*** (0.0021)	0.1365*** (0.0086)	0.1347*** (0.0048)	0.1196*** (0.0030)	0.1188*** (0.0017)	0.1310*** (0.0064)	0.1315*** (0.0036)
Prof. das ciên. e das artes	0.0464*** (0.0015)	0.0462*** (0.0009)	0.0509*** (0.0041)	0.0507*** (0.0023)	0.0736*** (0.0086)	0.0781*** (0.0051)	0.0570*** (0.0027)	0.0558*** (0.0017)	0.0752*** (0.0064)	0.0759*** (0.0038)
Téc. nível médio	0.0168*** (0.0013)	0.0166*** (0.0008)	0.0178*** (0.0030)	0.0177*** (0.0017)	0.0386*** (0.0058)	0.0406*** (0.0034)	0.0189*** (0.0021)	0.0187*** (0.0013)	0.0238*** (0.0052)	0.0258*** (0.0031)
Trab. serviços admin.	-0.0134*** (0.0012)	-0.0136*** (0.0007)	-0.0202*** (0.0026)	-0.0205*** (0.0015)	-0.0019 (0.0057)	0.0011 (0.0034)	-0.0132*** (0.0020)	-0.0130*** (0.0013)	-0.0192*** (0.0047)	-0.0184*** (0.0029)
Trab. serv e vend. do comérc. em lojas e mercad.	-0.0304*** (0.0014)	-0.0306*** (0.0008)	-0.0249*** (0.0028)	-0.0252*** (0.0016)	0.0161** (0.0063)	0.0175*** (0.0036)	-0.0138*** (0.0026)	-0.0133*** (0.0015)	-0.0059 (0.0052)	-0.0065** (0.0031)
Trab. agropec., florest. e da pesca	-0.0153*** (0.0038)	-0.0152*** (0.0025)	-0.0125* (0.0073)	-0.0123*** (0.0048)	-0.0110 (0.0148)	-0.0116 (0.0112)	0.0058 (0.0062)	-0.0029 (0.0053)	0.0041 (0.0085)	0.0145** (0.0064)
Trab.da prod. de bens e serv. Industriais 2	0.0143*** (0.0013)	0.0143*** (0.0009)	0.0094*** (0.0030)	0.0088*** (0.0019)	0.0394*** (0.0075)	0.0391*** (0.0048)	0.0122*** (0.0022)	0.0117*** (0.0015)	0.0158*** (0.0051)	0.0177*** (0.0035)

Variáveis	<b>Continua</b>									
	(I) Sudeste EF-ID, Firma e ID- Firma	(II) Sudeste VI e EF-ID, Firma e ID- Firma	(III) Norte EF-ID, Firma e ID- Firma	(IV) Norte VI e EF-ID, Firma e ID- Firma	(V) Nordeste EF-ID, Firma e ID- Firma	(VI) Nordeste VI e EF-ID, Firma e ID- Firma	(VII) Sul EF-ID, Firma e ID- Firma	(VIII) Sul VI e EF-ID, Firma e ID- Firma	(IX) Centro-Oeste EF-ID, Firma e ID- Firma	(X) Centro-Oeste VI e EF-ID, Firma e ID- Firma
Indústria extrativista	-0.0023 (0.0067)	-0.0027 (0.0049)	-0.0433** (0.0183)	-0.0434*** (0.0120)	-0.0028 (0.0188)	-0.0018 (0.0159)	-0.0196 (0.0157)	-0.0214* (0.0117)	0.0906*** (0.0163)	0.0732*** (0.0151)
Agricultura, pecuária, prod. florestal, pesca e aquic.	-0.0042 (0.0066)	-0.0048 (0.0051)	-0.0636*** (0.0085)	-0.0638*** (0.0069)	0.0094 (0.0309)	0.0060 (0.0218)	-0.0078 (0.0113)	-0.0051 (0.0090)	0.0285** (0.0145)	0.0283** (0.0111)
Eletricidade e gás	-0.0042 (0.0055)	-0.0043 (0.0056)	-0.1208*** (0.0237)	-0.1279*** (0.0237)	0.0736 (0.0469)	0.0745** (0.0312)	0.0718*** (0.0186)	0.0689*** (0.0156)	0.0095 (0.0523)	0.0069 (0.0552)
Água, esgoto, ativ. de gestão de resíd. e descont.	-0.0006 (0.0046)	-0.0006 (0.0042)	0.0048 (0.0132)	0.0046 (0.0101)	-0.0038 (0.0342)	0.0123 (0.0311)	0.0557*** (0.0130)	0.0609*** (0.0107)	0.0143 (0.0273)	0.0137 (0.0212)
Construção	0.0028 (0.0027)	0.0027 (0.0020)	0.0048 (0.0059)	0.0047 (0.0043)	-0.0316** (0.0155)	-0.0058 (0.0129)	0.0234*** (0.0069)	0.0237*** (0.0043)	0.0376*** (0.0122)	0.0352*** (0.0074)
Comércio; Reparação de veículos auto. e motocic.	-0.0178*** (0.0016)	-0.0179*** (0.0012)	-0.0253*** (0.0038)	-0.0255*** (0.0026)	-0.0111 (0.0089)	-0.0145* (0.0076)	-0.0158*** (0.0030)	-0.0159*** (0.0021)	-0.0079 (0.0066)	-0.0097** (0.0048)
Transporte, armazenagem e correio	-0.0100*** (0.0033)	-0.0102*** (0.0027)	-0.0147** (0.0073)	-0.0148*** (0.0054)	-0.0366** (0.0172)	-0.0307** (0.0142)	-0.0075 (0.0057)	-0.0044 (0.0042)	-0.0115 (0.0113)	-0.0135 (0.0090)

Variáveis	Conclusão									
	(I) Sudeste EF-ID, Firma e ID- Firma	(II) Sudeste VI e EF-ID, Firma e ID- Firma	(III) Norte EF-ID, Firma e ID- Firma	(IV) Norte VI e EF-ID, Firma e ID- Firma	(V) Nordeste EF-ID, Firma e ID- Firma	(VI) Nordeste VI e EF-ID, Firma e ID- Firma	(VII) Sul EF-ID, Firma e ID- Firma	(VIII) Sul VI e EF-ID, Firma e ID- Firma	(IX) Centro-Oeste EF-ID, Firma e ID- Firma	(X) Centro-Oeste VI e EF-ID, Firma e ID- Firma
Informação e comunicação	-0.0081*** (0.0030)	-0.0083*** (0.0021)	-0.0129 (0.0079)	-0.0130** (0.0058)	-0.0052 (0.0244)	-0.0010 (0.0176)	-0.0162*** (0.0054)	-0.0183*** (0.0038)	-0.0137 (0.0123)	-0.0176** (0.0087)
Atividades financ., de seguros e serv. relacion.	-0.0158*** (0.0034)	-0.0160*** (0.0030)	-0.0280** (0.0109)	-0.0282*** (0.0094)	0.0002 (0.0264)	0.0030 (0.0218)	-0.0103 (0.0146)	-0.0141 (0.0088)	0.0184 (0.0282)	0.0157 (0.0155)
Atividades imobiliárias	0.0157** (0.0070)	0.0155*** (0.0051)	-0.0141 (0.0103)	-0.0143 (0.0089)	-0.0509* (0.0307)	-0.0477 (0.0299)	0.0280** (0.0129)	0.0195** (0.0093)	0.0174 (0.0158)	0.0122 (0.0142)
Atividades profissionais, científicas e técnicas	-0.0023 (0.0032)	-0.0025 (0.0023)	-0.0070 (0.0064)	-0.0073 (0.0045)	-0.0018 (0.0192)	-0.0029 (0.0161)	-0.0240*** (0.0068)	-0.0258*** (0.0048)	0.0067 (0.0129)	0.0032 (0.0086)
Atividades administrativ as e serv. complementa res	-0.0088*** (0.0019)	-0.0089*** (0.0014)	-0.0195*** (0.0053)	-0.0197*** (0.0038)	0.0027 (0.0136)	0.0096 (0.0113)	-0.0076* (0.0040)	-0.0089*** (0.0030)	-0.0034 (0.0091)	-0.0052 (0.0067)
Pequena	0.0225*** (0.0006)	0.0225*** (0.0005)	0.0134*** (0.0013)	0.0134*** (0.0010)	0.0300*** (0.0031)	0.0289*** (0.0026)	0.0283*** (0.0010)	0.0281*** (0.0008)	0.0242*** (0.0022)	0.0224*** (0.0018)
Média	0.0501*** (0.0009)	0.0501*** (0.0007)	0.0432*** (0.0019)	0.0431*** (0.0015)	0.0570*** (0.0045)	0.0538*** (0.0036)	0.0551*** (0.0015)	0.0545*** (0.0012)	0.0533*** (0.0034)	0.0512*** (0.0026)
Grande	0.0726*** (0.0011)	0.0727*** (0.0009)	0.0715*** (0.0024)	0.0714*** (0.0018)	0.0949*** (0.0053)	0.0902*** (0.0042)	0.0744*** (0.0020)	0.0744*** (0.0016)	0.0701*** (0.0042)	0.0702*** (0.0032)
Ln. área LMA	0.0048*** (0.0011)	0.0086*** (0.0016)	-0.0180*** (0.0031)	-0.0196*** (0.0028)	0.0239 (0.0147)	-0.0057 (0.0136)	-0.0062 (0.0044)	0.0091* (0.0048)	-0.0068 (0.0099)	0.0297 (0.0201)

Variáveis	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)	(IX)	(X)
	Sudeste EF-ID, Firma e ID- Firma	Sudeste VI e EF-ID, Firma e ID- Firma	Norte EF-ID, Firma e ID- Firma	Norte VI e EF-ID, Firma e ID- Firma	Nordeste EF-ID, Firma e ID- Firma	Nordeste VI e EF-ID, Firma e ID- Firma	Sul EF-ID, Firma e ID- Firma	Sul VI e EF-ID, Firma e ID- Firma	Centro-Oeste EF-ID, Firma e ID- Firma	Centro-Oeste VI e EF-ID, Firma e ID- Firma
Grande	(0.0058) -0.0200***	(0.0052) -0.0305***	(0.0248) 0.0761***	(0.0154) 0.0802***	(0.0663) -0.2020*	(0.0869) 0.0511	(0.0120) 0.1165***	(0.0108) 0.0593***	(0.0419) -0.0016	(0.0551) -0.1046
2011.ano	(0.0074) 0.0203***	(0.0076) 0.0203***	(0.0258) 0.0134***	(0.0166) 0.0133***	(0.1073) 0.0260***	(0.1213) 0.0284***	(0.0170) 0.0192***	(0.0182) 0.0197***	(0.0481) 0.0158***	(0.0766) 0.0157***
2012.ano	(0.0003) 0.0546***	(0.0003) 0.0547***	(0.0006) 0.0440***	(0.0006) 0.0439***	(0.0014) 0.0372***	(0.0016) 0.0378***	(0.0005) 0.0526***	(0.0005) 0.0525***	(0.0011) 0.0579***	(0.0011) 0.0563***
2013.ano	(0.0003) 0.0779***	(0.0003) 0.0779***	(0.0006) 0.0639***	(0.0006) 0.0638***	(0.0015) 0.0677***	(0.0015) 0.0683***	(0.0005) 0.0854***	(0.0005) 0.0846***	(0.0011) 0.0902***	(0.0012) 0.0903***
2014.ano	(0.0003) 0.0773***	(0.0003) 0.0772***	(0.0006) 0.0753***	(0.0006) 0.0753***	(0.0015) 0.0639***	(0.0014) 0.0647***	(0.0005) 0.0860***	(0.0005) 0.0850***	(0.0011) 0.0879***	(0.0010) 0.0874***
2015.ano	(0.0003) 0.0312***	(0.0003) 0.0312***	(0.0006) 0.0318***	(0.0006) 0.0317***	(0.0014) 0.0166***	(0.0014) 0.0177***	(0.0005) 0.0310***	(0.0005) 0.0307***	(0.0010) 0.0495***	(0.0010) 0.0495***
2016.ano	(0.0003) 0.0249***	(0.0003) 0.0249***	(0.0005) 0.0248***	(0.0006) 0.0249***	(0.0012) 0.0081***	(0.0014) 0.0088***	(0.0004) 0.0322***	(0.0005) 0.0312***	(0.0009) 0.0414***	(0.0010) 0.0418***
2017.ano	(0.0002) 0.0236***	(0.0003) 0.0236***	(0.0005) 0.0261***	(0.0006) 0.0261***	(0.0011) 0.0380***	(0.0015) 0.0391***	(0.0004) 0.0270***	(0.0005) 0.0266***	(0.0008) 0.0312***	(0.0011) 0.0312***
2018.ano		0.0000		0.0000		0.0000		0.0000		0.0000
Constante	0.6523*** (0.0156)		0.4003*** (0.0552)		0.1686 (0.1593)		0.5234*** (0.0521)		0.1516 (0.1287)	
Ano	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Efeito Fixo ID	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Efeito Firma	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Efeito ID x Firma	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Teste F 1º estágio		1.363e+06		2.226e+06		160621		932934		37876
Prob>F 1º estágio		0		0		0		0		0
R <sup>2</sup>	0.943	0.2076	0.931	0.2335	0.921	0.2291	0.926	0.2852	0.914	0.2615
R <sup>2</sup> - ajustado	0.208	0.0531	0.233	0.0851	0.230	0.0834	0.288	0.152	0.262	0.110

Continua

Observações	6,792,579	6,773,223	1,426,397	1,423,628	303,666	282,845	2,057,985	1,901,324	575,479	548,626
F	13374	35359	3130	8649	692.9	1684	5860	15228	1479	3838
Prob>F	0	0	0	0	0	0	0	0	0	0

Fonte: Elaboração própria com base nos dados identificados da RAIS 2010-2018.

Nota: Erro-padrão entre parênteses (colunas II, IV, VI, VIII e X). Erro-padrão robusto entre parênteses (colunas I, III, V, VII e IX). \*\*\* p<0.01, \*\* p<0.05, \* p<0.1