

Efeitos das aglomerações urbanas sobre os salários no Brasil (2010-2018): o *Sorting* e o *Match* dos trabalhadores e firmas importam?

Jefferson Douglas da Silva Pereira¹

Keuler Hissa Teixeira²

RESUMO

Este artigo mensura os efeitos das aglomerações urbanas sobre os salários dos trabalhadores formais brasileiros no período de 2010 a 2018, controlando, simultaneamente, os efeitos do *sorting* espacial dos trabalhadores, das firmas e do *match* trabalhador-firma. Este estudo é, até então, o primeiro a fazer uma análise que controla, de maneira simultânea, esses efeitos para o Brasil. Para realizar essa análise, utiliza-se um painel de dados pareado trabalhador-firma, oriundos da RAIS (2010-2018), e estima-se por múltiplos efeitos fixos. Mesmo após as estimativas por múltiplos efeitos fixos, é possível que estimativas estejam enviesadas devido a possível endogeneidade entre a densidade do emprego e os salários. Para lidar com esse problema, estima-se por variáveis instrumentais. Os resultados encontrados no modelo principal deste estudo mostram que existe um efeito líquido das aglomerações sobre os salários de 1,69%, o que é um coeficiente bem abaixo dos encontrados na maioria dos estudos. Além disso, os resultados apontam para importância do *sorting* espacial dos trabalhadores e firmas e para pouca relevância 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.

ABSTRACT

This paper estimates the effects of urban agglomerations on the wages of Brazilian formal workers in the period from 2010 to 2018, simultaneously controlling the effects of the spatial sorting of workers and the firms, and the worker-firm match. This study is, so far, the first to carry out an analysis that simultaneously controls these effects to Brazil. To carry out this analysis, the study uses a panel of paired worker-firm data from RAIS (2010-2018), and it is estimated by multiple fixed effects. Even after estimating by fixed multiple effects, it is possible the existence of endogeneity between the density of employment and wages. To deal with this issue, the model is estimated by Instrumental Variables. The results found in the main model of this paper show that there is a net effect of agglomerations on wages of 1.69%, which is a coefficient well below those found in most studies. Furthermore, the results point to the importance of the spatial sorting of workers and the firms, and to the little relevance of the worker-firm match in determining the wages of the Brazilian local labor market areas.

Keywords: Agglomeration economies. Wages. *Sorting*. Worker-firm match.

JEL Classification: R23; E24; J31; C23; C26

Área 10: Economia Regional e Urbana

¹ Doutorando do Programa de Pós-Graduação em Economia da Universidade Federal de Juiz de Fora (PPGE-UFJF).

² Doutor em Economia. Professor Adjunto do Programa de Mestrado em Economia Aplicada da Universidade Federal de Alagoas (UFAL).

1 INTRODUÇÃO

A relação entre a densidade urbana e a produtividade é entendida como sendo a razão principal para existência das cidades, portanto, compreender essa relação é fundamental nos estudos sobre aglomeração (GLAESER; RESSEGER, 2010). As grandes disparidades regionais (crescimento, salários, renda *per capita*, por exemplo) que são observadas em diferentes territórios ou, muitas vezes, dentro de um mesmo território são a base para os estudos sobre economias de aglomeração.

Essas disparidades podem ser observadas tanto em países desenvolvidos quanto nos países em desenvolvimento. A Ilha de França, região metropolitana de Paris, apesar de ocupar apenas 2,2% da área total da França, possui 18,9% da população do total do país e produz 30% do PIB francês (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 ocupava 0,018% do território brasileiro, sendo responsável por 10,6% do PIB nacional. Esses resultados indicam a importância de um componente espacial na determinação da produtividade local.

A aglomeração 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, 1991). 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).

A partir do trabalho seminal de Mincer (1974), muitos estudos, em especial, na Economia do Trabalho, têm abordado os diferenciais de salariais. Uma regularidade desses estudos é que, mesmo controlando características observáveis e não observáveis dos trabalhadores, diferenças salariais persistem em favor dos grandes centros urbanos. Isto tem motivado diversos estudos, no âmbito da Economia Regional e Urbana, a investigar o papel das aglomerações nas análises dos diferenciais salariais. A inclusão da aglomeração como determinante salarial é o ponto de integração entre a Economia do Trabalho e a Economia Regional e Urbana, que originou uma nova vertente de estudos sobre o mercado de trabalho, conhecida como Economias de Aglomeração.

A classificação atual dos mecanismos por trás das economias de aglomeração (*sharing, matching e learning*) foi formalizada por Duranton e Puga (2004), com base em Marshall (1890). A previsão de que a produtividade aumenta com a escala de atividade em um local torna difícil mensurar o impacto individual de cada um desses mecanismos sobre a produtividade (PUGA, 2010). A fim de contornar essa dificuldade, muitos trabalhos têm buscado quantificar o efeito geral das economias de aglomeração sobre a produtividade por meio de diferentes escopos³.

Utilizando a densidade do emprego como medida de aglomeração, e os salários como forma de mensurar a produtividade⁴, Glaeser e Maré (2001) encontraram um prêmio salarial de 33% a favor das áreas metropolitanas dos EUA quando comparadas às áreas não metropolitanas. A partir desse estudo, muitos outros buscaram investigar os efeitos das aglomerações sobre os salários, porém com foco nos EUA e em países da Europa (WHEATON; LEWIS, 2002; FINGLETON, 2003; COMBES; DURANTON; GOBILLON, 2008; KLAESSON E LARSSON, 2009; GROOT; GROOT; SMIT, 2014).

Recentemente, alguns estudos têm buscado analisar a relação entre aglomerações urbanas e salários para países em desenvolvimento (DURANTON, 2016; CHAUVIN *et al.*, 2017). Chauvin *et al* (2017) chamam atenção para o fato de que o prêmio salarial decorrente das aglomerações em países em desenvolvimento, como é o caso do Brasil, pode ser diferente do encontrado para os países desenvolvidos, tendo em vista os processos diferentes de urbanização desses dois tipos de países. Por esse motivo, não se pode, *à priori*, estender uma evidência documentada frequentemente na literatura sobre ganhos de aglomeração aos países em desenvolvimento, já que esta conclusão é feita a partir de estudos para países desenvolvidos.

³ Rosenthal e Strange (2004) fornece um rico detalhamento sobre os escopos das economias de aglomeração.

⁴ 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).

As pesquisas têm avançado metodologicamente para evitar a superestimação dos efeitos das economias de aglomeração sobre os salários. Os estudos têm buscado controlar o *sorting* dos trabalhadores (GLAESER; MARÉ, 2001; COMBES; DURANTON; GOBILLON, 2008; MATANO; NATICCHIONI, 2016), das firmas (MION; NATICCHIONI, 2009; SILVA, 2017; CAMPOS, 2018; NEVES JÚNIOR, 2018) e do *match* trabalhador-firma (WOODCOCK, 2008; 2015; CAMPOS, 2018; NEVES JÚNIOR, 2018) a fim de obter estimativas mais robustas do efeito das aglomerações sobre os salários.

Ainda que de maneira tímida, a literatura brasileira sobre aglomerações revela que há um prêmio salarial positivo a favor dos grandes centros urbanos (ROCHA; SILVEIRA NETO; GOMES, 2011; SILVA; SANTOS; FREGUGLIA, 2016; BARURUFI; HADDAD; NIJKAMP, 2016). O estudo mais robusto sobre os efeitos das aglomerações na determinação salarial do Brasil foi realizado por Silva (2017). A autora controla em sua análise, de maneira simultânea, os efeitos do *sorting* espacial dos trabalhadores e das firmas sobre os salários. Os resultados mostraram que 4,9% da diferença salarial individual observada são decorrentes dos efeitos puros de aglomeração.

Considerando a escassez de estudos para países em desenvolvimento e, em especial, a pouca literatura sobre o Brasil, este artigo tem por objetivo mensurar os efeitos das aglomerações sobre os salários dos trabalhadores formais brasileiros no período de 2010 a 2018. Este estudo apresenta avanços quanto à abordagem empírica, pois a partir de uma grande base microdados longitudinais controla, simultaneamente, os efeitos do *sorting* espacial dos trabalhadores, das firmas e do *match* trabalhador-firma. Essa abordagem é, até então, inédita para o Brasil. Além disso, destaca-se que a unidade espacial de análise utilizada neste estudo, as áreas de mercado de trabalho local, são construídas a partir das definições de arranjos populacionais e concentrações urbanas, de regiões de influência das cidades e de municípios isolados propostas pelo IBGE.

É importante destacar também que este artigo utiliza uma base de dados para um período até então nunca utilizado na literatura brasileira sobre economias de aglomeração. No período de 2010 a 2018, o Brasil passou por importantes mudanças econômicas. De acordo com o IBGE, nesse período, o PIB anual brasileiro cresceu, em média, 1,67%, tendo experimentado, 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 brasileiro, o que pode afetar as estimativas do efeito das aglomerações sobre os salários dos trabalhadores formais do Brasil.

Além desta introdução, este artigo conta com mais quatro seções. A seção 2 diz respeito aos dados e as variáveis consideradas neste estudo. A seção 3 descreve a metodologia do estudo. A seção 4 apresenta os resultados e as discussões das estimações. A seção 5 apresenta as considerações finais deste artigo. Por fim, são apresentadas as referências usadas.

2 DADOS E VARIÁVEIS

Esta seção apresenta a base de dados e está dividida em três subseções. A subseção 2.1 traz informações sobre a base de dados utilizada neste artigo. A subseção 2.2 apresenta a definição de áreas de mercado de trabalho, unidade espacial de análise deste estudo. A última subseção, a 2.3, faz uma descrição das variáveis consideradas para as análises aqui propostas.

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)⁵. A RAIS⁶ 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, cobrindo aproximadamente 97% do setor formal da economia. A principal vantagem da RAIS é que o seu

⁵ 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.

⁶ A RAIS foi instituída pelo decreto nº 76.900, de 23/12/75.

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.

A partir dos dados da RAIS, foi gerado um painel de dados trabalhador-firma com 31.308.255 observações, relativas a 3.478.695 indivíduos e a 858.953 estabelecimentos, cobrindo o período de 2010 a 2018. Desse painel de dados, extraiu-se uma amostra aleatória de 5% de observações, mantendo a representatividade das variáveis e a estrutura do painel, de modo que a amostra final contém 11.585.637 de observações, correspondentes a 520.455 firmas e 1.287.293 indivíduos e 2.232.883 combinações únicas de indivíduos e firmas.

2.2 Definição das áreas de mercado de trabalho local (*Labour Market Areas - LMA*)

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, doravante LMA. Para definição das LMA utilizadas neste estudo, 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).

Ao todo foram definidas 371 LMA, 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 LMA definidas 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 brasileiro no ano de 2018.

2.3 Descrição das variáveis

Esta subseção apresenta as variáveis utilizadas nesta pesquisa. Elas se referem aos trabalhadores, às firmas e à localização do emprego, isto é, às LMA. A variável dependente utilizada nos modelos estimados neste estudo é o logaritmo natural do salário-hora real do trabalhador em dezembro de cada ano, que foi deflacionado pelo Índice de Preços Amplo ao Consumidor- IPCA⁷, calculado pelo IBGE, tendo 2018 como ano base. O logaritmo natural do salário-hora real é dado da seguinte forma:

$$\ln_{\text{Salário}_{\text{hora-real}}} = \ln \left(\frac{\text{Salário real}_{\text{dezembro}}}{(5 \times \text{Quantidade de horas semanais contratadas})} \right) \quad (1)$$

No nível do trabalhador as variáveis utilizadas são: idade (18-65 anos), idade ao quadrado, experiência (tempo de emprego, em meses, no mesmo vínculo), experiência ao quadrado, gênero, raça (raça/cor), educação e grupo ocupacional⁸.

⁷ O IPCA é calculado para as seguintes regiões metropolitanas: Belém, Fortaleza, Recife, Salvador, Belo Horizonte, Vitória (desde dezembro de 2013), Rio de Janeiro, São Paulo, Curitiba, Porto Alegre e Brasília, e para os seguintes municípios: Goiânia, Campo Grande (desde dezembro de 2013), Rio Branco, São Luís e Aracaju (estes três últimos desde março de 2018).

⁸ 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).

Em relação à firma, as variáveis consideradas neste estudo são duas: setor de atividade econômica da firma⁹ e tamanho da firma¹⁰: Microempresa; Empresa Pequena; Empresa Média e Empresa Grande.

Quanto às LMA as variáveis são: logaritmo natural da densidade do emprego, que é calculada a partir do logaritmo da razão entre o número total de empregos e a área em quilômetros quadrados (Km²) da LMA; logaritmo natural da área, em km², da LMA; tamanho da LMA, que está classificado em: pequeno, médio e grande; macrorregião em que a LMA está localizada: Sudeste, Nordeste, Norte, Sul e Centro-oeste.

3 METODOLOGIA

Esta subseção aborda a metodologia utilizada neste trabalho e está dividida em três subseções. A subseção 3.1 apresenta o modelo teórico em que a análise deste estudo está baseada. A subseção 3.2 traz a especificação econométrica adotada neste estudo. Por fim, a subseção 3.3 apresenta as estratégias de identificação e estimação aqui adotadas.

3.1 Modelo teórico

A fim de capturar os efeitos das aglomerações 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 *sorting* espacial proposto por Combes, Duranton e Gobillon (2008).

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.

O modelo proposto por Combes, Duranton e Gobillon (2008) inova, ao considerar em uma única estrutura de determinação salarial esses três fatores:

$$w_i = A(E_{a(i)}, I_{a(i)}, I_{a(i),k(i)})s_i. \quad (2)$$

Em que w_i é a remuneração do trabalhador i , que é determinada pelas habilidades, s_i , e pela produtividade, A , deste trabalhador, em função das dotações não humanas da região a , $E_{a(i)}$ e das interações locais entre trabalhadores ou entre firmas, $I_{a(i)}, I_{a(i),k(i)}$.

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. Desse modo, os efeitos das aglomerações sobre os salários são capturados de maneira agregada na Equação (2).

⁹ O setor da firma foi definido a partir da Classificação Nacional de Atividades Econômicas 2.0 (CNAE 2.0) do IBGE, os setores são: 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).

¹⁰ O tamanho da firma é determinado a partir da quantidade de trabalhadores ativos que o estabelecimento possui e do seu setor de atuação conforme o estudo Anuário do Trabalho na Micro e Pequena Empresa, organizado pelo Serviço de Apoio às micro e Pequenas Empresas (SEBRAE), e elaborado pelo Departamento Intersindical de Estatísticas e Estudos Socioeconômicos (DIEESE).

O modelo microeconômico desenvolvido por Combes, Duranton e Gobillon (2008) assume 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:

$$\pi_{a,k,t} = p_{a,k,t} y_{a,k,t} - \sum_{i \in (a,k,t)} w_{i,t} l_{i,t} - r_{a,k,t} z_{a,k,t}, \quad (3)$$

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 do modelo é uma função Cobb-Douglas em trabalho efetivo e os outros fatores de produção, como a seguinte:

$$y_{a,k,t} = A_{a,k,t} \left(\sum_{i \in (a,k,t)} s_{i,t} l_{i,t} \right)^b (z_{a,k,t})^{1-b} \quad (4)$$

Essa função (4) possui retornos constantes de escala, $0 < b \leq 1$; $s_{i,t}$ 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 $a(i, t)$, no setor $k(i, t)$ e no ano t , recebe um salário igual ao seu produto marginal:

$$w_{i,t} = b p_{a(i,t),k(i,t),t} A_{a(i,t),k(i,t),t} \left(\frac{z_{a(i,t),k(i,t),t}}{\sum_{i \in (a,k,t)} s_{i,t} l_{i,t}} \right)^{1-b} s_{i,t} \quad (5)$$

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 (5):

$$w_{i,t} = b(1-b)^{\frac{(1-b)}{b}} \left(p_{a(i,t),k(i,t),t} \frac{A_{a(i,t),k(i,t),t}}{(r_{a(i,t),k(i,t),t})^{1-b}} \right)^{\frac{1}{b}} s_{i,t} = B_{a(i,t),k(i,t),t} s_{i,t} \quad (6)$$

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 ($r_{a,k,t}$).

Neste contexto, os autores sugerem que características locais, como a densidade do emprego, podem ter efeitos sobre a produtividade ($B_{a(i,t),k(i,t),t}$), tanto positivos (economias de aglomeração) quanto negativos (congestionamento), não sendo possível identificá-los separadamente. Apesar de não permitir identificar os efeitos de preço e de tecnologia separadamente sobre os salários, o modelo de Combes, Duranton e Gobillon (2008) é uma sofisticada ferramenta para a identificação dos diferenciais salariais por habilidades, especialmente, em uma perspectiva espacial.

3.2 Especificação econométrica

A fim de poder testar empiricamente a Equação (6), 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, $B_{a(i,t),k(i,t),t}$. A especificação econométrica final do primeiro estágio corresponde à demanda inversa por trabalho:

$$\log w_{i,t} = \beta_{a(i,t),t} + \mu_{k(i,t),t} + \tilde{I}_{a(i,t),k(i,t),t} \gamma_k(i,t) + \tilde{X}_{i,t} \varphi + \delta_i + \epsilon_{i,t}, \quad (71)$$

Em que $\beta_{a(i,t),t}$ representa o efeito fixo área-ano (efeito puro de área); $\mu_{k(i,t),t}$ denota o efeito fixo setor-ano (efeito puro do setor); $\tilde{I}_{a(i,t),k(i,t),t}$ é o vetor de variáveis de interações intrasetoriais para cada área-setor-ano; $\tilde{X}_{i,t}$ é um vetor de características do trabalhador variáveis no tempo; δ_i são os efeitos fixos do trabalhador e $\epsilon_{i,t}$ é o termo de erro.

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 intrasetoriais ($\tilde{I}_{a,k,t}\gamma_k$) e do conjunto de dotações e interações entre os setores ($\beta_{a,t}$), sendo considerados os efeitos fixo de indivíduo, setor e área.

A partir da Equação (5), 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 intersetoriais na explicação dos efeitos fixos de área-ano. Assim, a especificação econométrica do segundo estágio é dada da seguinte maneira:

$$\beta_{a,t} = \omega_0 + \theta_t + I_{a,t}\gamma + E_{a,t}\alpha + v_{a,t}, \quad (82)$$

em que θ_t representa os efeitos fixos de tempo; α é um vetor de coeficientes associados com as variáveis das dotações do local, $E_{a,t}$; γ é um vetor de coeficientes ligados com as interações locais intersetoriais, $I_{a,t}$ 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 ω_0 em zero, que é o coeficiente do período, isto é, o ano de referência. A estimação do modelo de Combes, Duranton e Gobillon (2008) é feita em dois estágios a partir das equações (7) e (8).

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, já que esses efeitos são controlados na especificação final.

O modelo desenvolvido pelos autores não controla, contudo, o efeito fixo de firma, apenas o de setor. Essa limitação pode ser superada se ao modelo for adicionado o efeito fixo de firma de modo a controlar o *sorting* espacial das firmas. Essa abordagem é baseada em Abowd, Kramarz e Margolis (1999) que desenvolveram um modelo seminal de decomposição salarial¹¹, a fim de verificar o diferencial setorial de salários na França. O modelo se baseia nos resultados heterogêneos do mercado de trabalho francês percebidos nas diferenças de remunerações e de histórico de emprego que indivíduos notavelmente equivalentes possuem. Segundo os autores, 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.

Mesmo controlando o *sorting* das firmas, a especificação baseada no modelo de Combes, Duranton e Gobillon (2008) pode ser enviesada se não levar em conta as interações simultâneas das características dos trabalhadores e das firmas no espaço. A omissão do *match* trabalhador-firma pode enviesar as estimativas do prêmio salarial (WOODCOCK; 2008; 2015). Com a inclusão desse controle, objetiva-se obter estimativas menos enviesadas possíveis do efeito das aglomerações urbanas sobre os salários. O controle simultâneo do *sorting* espacial dos trabalhadores e das firmas e do *match* trabalhador-firma tem estado presente em poucos estudos, conforme cita Neves Júnior (2018), já que os dados que permitem tal controle, até pouco tempo, não estavam ao alcance dos pesquisadores.

A principal especificação econométrica utilizada neste artigo é, portanto, uma versão expandida do modelo proposto por Combes, Duranton e Gobillon (2008) que inclui além dos efeitos fixos de indivíduos, os efeitos fixos de firma e do *match* trabalhador-firma. Além disso, essa especificação é estimada, diferente do modelo do *sorting* apresentado nesta seção, em um único estágio, como feito em Rosenthal e Strange (2008) e Campos (2018). Sendo assim, a principal especificação econométrica deste artigo é dada da seguinte maneira:

$$\ln(W_{it}) = \beta_1 \ln(DensEmp)_{at} + \beta_2 X_{it} + \beta_3 F_{jt} + \beta_4 LMA_{at} + \beta_5 T_t + \delta_i + \varphi_j + \vartheta_{ij} + \varepsilon_{it} \quad (93)$$

¹¹ Para mais detalhes sobre o modelo ver Abowd, Kramarz e Margolis (1999).

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, idade², experiência, experiência², 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 km², tamanho da LMA e macrorregião) no tempo t ; T_t representam *dummies* de tempo; δ_i representa o efeito fixo de indivíduo; φ_j captura o *sorting* espacial das firmas; ϑ_{ij} captura o efeito do *match* trabalhador-firma e ε_{it} é o erro aleatório.

3.3 Estratégias de identificação e estimação

A literatura tem avançado na identificação dos efeitos das aglomerações sobre os salários. Contudo, como abordado na subseção anterior, a autosseleção (*sorting*) de indivíduos e de firmas para mercados de trabalho mais produtivos pode afetar os salários e enviesar as estimativas, bem como o *match* trabalhador-firma. Outro 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 aglomerações sobre os salários.

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), das firmas (GROOT; GROOT; SMIT, 2014; SILVA, 2017; CAMPOS, 2018; NEVES JÚNIOR, 2018), e do *match* trabalhador-firma (WOODCOCK, 2008; 2015; CAMPOS, 2018; NEVES JÚNIOR, 2018). A inclusão dos efeitos fixos dos indivíduos e das firmas também permite controlar a autosseleção dos indivíduos e firmas para centros urbanos mais densos e mais produtivos.

A identificação dos efeitos fixos dos trabalhadores requer mobilidade destes entre firmas e entre áreas de mercado de trabalho local. Analogamente, a identificação dos efeitos fixos das firmas requer mobilidade destas entre as LMA. Para identificação do efeito fixo do *match* trabalhador-firma, as hipóteses de mobilidade de trabalhadores e de firmas são suficientes conforme Woodcock (2015). A Tabela 1 a seguir mostra a mobilidade de trabalhadores e firmas:

Tabela 1: Mobilidade das firmas e dos trabalhadores entre 2010 e 2018

Mobilidade	Quantidade única	Frequência (%)
Firmas entre LMA	5.999	1,15%
Total de firmas na amostra	520.455	100%
Trabalhadores entre firmas	569.958	44,28%
Trabalhadores entre LMA	142.353	10,67%
Total de trabalhadores na amostra	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 a Tabela 1, 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*.

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). Contudo, esta subseção abordará apenas o método de EF¹², pois considerando os objetivos deste estudo e os resultados do teste de Hausman, ele é o mais adequado¹³.

O estimador de EF é o método mais adequado para especificações que lidam com efeitos fixos não observados, pois, assim como o método EA, assegura a validade da hipótese de exogeneidade estrita entre o termo de erro e as variáveis explicativas condicionadas aos efeitos não observados ($E(\varepsilon_{it}|\mathbf{x}_i, c_i) = 0, t = 1, 2, \dots, T$) e ainda permite que os efeitos não observados constantes no tempo possam estar correlacionados com as variáveis explicativas observadas ($E(c_i|\mathbf{x}_i) = E(c_i) \neq 0, t = 1, 2, \dots, T$), sem que isso comprometa a consistência das estimativas encontradas (WOOLDRIDGE, 2010). Por este motivo, a análise de EF é mais robusta do que a análise por POLS ou EA.

Como já apresentado ao longo deste estudo, a especificação principal do modelo requer a estimação, simultânea, de mais de um efeito fixo. Em decorrência disso, é necessário realizar as estimações a partir de uma metodologia que capture esses múltiplos efeitos fixos.

Para lidar com essa questão, Correia (2014) desenvolveu um algoritmo que implementa a abordagem iterativa proposta por Guimarães e Portugal (2010)¹⁴ de modo a permitir a estimação por múltiplos efeitos fixos. Considerando a forma matricial do modelo principal adotado neste artigo e especificado na Equação (9), a solução de MQO para o modelo é a resolução do sistema de equações normais a seguir:

$$\beta = (X'X)^{-1}X'(W - \delta A - \varphi B - \vartheta C) \quad (104)$$

$$\delta = (A'A)^{-1}A'(W - \beta X - \varphi B - \vartheta C) \quad (115)$$

$$\varphi = (B'B)^{-1}B'(W - \beta X - \delta A - \vartheta C) \quad (126)$$

$$\vartheta = (C'C)^{-1}C'(W - \beta X - \delta A - \varphi B) \quad (137)$$

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 β é um vetor $k \times 1$ de parâmetros a ser estimado para as variáveis explicativas, incluindo a variável densidade do emprego, δ , φ e ϑ 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).

Controlar, simultaneamente, 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, há a possibilidade de causalidade reversa entre a densidade do emprego e os salários. Para Ciccone e Hall (1996), essa causalidade pode existir se escolha do local do emprego por parte do trabalhador não for exógena. Se isso ocorre, 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).

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) que fornece uma solução geral para o problema de uma variável explicativa endógena (WOOLDRIDGE, 2010), como é o caso. 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 tratar a endogeneidade presente nas análises sobre economias de aglomeração e a produtividade. 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) e variáveis geográficas (CICCONE, 2002; ROSENTHAL; STRANGE, 2008; BARUFI, 2015).

¹² Para mais detalhes sobre os métodos POLS e EA ver Wooldridge (2010).

¹³ Para maiores detalhes do teste de Hausman ver Gujarati e Porter (2011).

¹⁴ Para mais detalhes sobre o algoritmo ver Correia (2014).

Este artigo utiliza como instrumento para densidade do emprego, o logaritmo da densidade da população defasada das LMA. Para isto, utiliza-se os dados populacionais e de território, fornecidos pelo IBGE, dos anos 1920, 1940 e 1950. Para que a densidade defasada seja um bom instrumento, ela precisa atender a dois critérios:

$$Cov(Z_{at}, \varepsilon_{at}) = 0 \quad (14)$$

$$Cov(Z_{at}, DensEmp_{at}) \neq 0 \quad (15)$$

A condição (14) exige que a densidade da população defasada não seja correlacionada 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 (15), que pode ser testada, requer que a densidade populacional defasada seja fortemente correlacionada com a densidade do emprego atual.

Uma vez definido o instrumento, 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 e 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:

$$DensEmp_{a,t} = \beta_i \ln DensPop_{a,t} + \phi_i X_{a,t} + \varepsilon_{a,t} \quad (168)$$

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 de todas as variáveis explicativas exógenas do modelo, e $\varepsilon_{a,t}$ 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:

$$\ln(W_{it}) = \beta_1 \widehat{DensEmp}_{a,t} + \tau_i X_{a,t} + v_{it} \quad (17)$$

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

Esta subseção mostrou os principais métodos e as principais estratégias de identificação adotadas para estimação adequada dos efeitos das aglomerações sobre os salários. Em relação às duas principais fontes de viés aqui discutidas, Combes, Duranton e Gobillon (2011) destacam que o *sorting* espacial afeta fortemente as estimativas entre aglomerações e salários, enquanto a endogeneidade afeta os resultados em menor grau.

4 RESULTADOS E DISCUSSÕES

Esta seção apresenta os resultados e está dividida em duas subseções. A subseção 4.1 apresenta as estatísticas descritivas da amostra. A subseção 4.2 apresenta e discute os resultados obtidos das estimações em painel.

4.1 Estatísticas descritivas

As estatísticas descritivas apresentadas nesta subseção são referentes à amostra de dados extraída dos microdados identificados da RAIS (2010-2018), que conta com 11.585.637 de observações de 1.287.293 indivíduos e de 520.455 firmas. O objetivo desta subseção é apresentar como se comportam as informações que aqui estão sendo manipuladas.

A Tabela 2 mostra que média do salário-hora-real dos trabalhadores presentes na amostra é de R\$ 19,054. A média da densidade do emprego nas LMA é de 30,223 hab/km² e o desvio padrão é de 39,174 hab/km². Esse alto desvio pode refletir na composição da mão de obra nas LMA, reforçando ou dispersando as forças aglomerativas, o que pode influenciar na determinação salarial local. A idade média dos trabalhadores da amostra é de 38 anos e a experiência média, isto é, o tempo médio de emprego na mesma firma, é de 83 meses. De resto, a Tabela 2 mostra que a média da área das LMA é de 10.566,1 km².

Tabela 2: Estatísticas descritivas das variáveis contínuas

Variáveis	Média	Desvio-padrão	Min.	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 ²	1566,327	761,429	324	4225
Experiência	83,886	72,731	0,1	597,9
Experiência ²	12326,671	21592,149	0,01	357484,41
<i>Ln</i> (Área da LMA, em km ²)	17,848	2,181	5,545	20,773
Área da LMA (em km ²)	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.

A Tabela 3 mostra a distribuição salarial conforme as variáveis ligadas diretamente ao indivíduo. Quanto ao gênero, a amostra é composta, majoritariamente, por trabalhadores do gênero masculino (aproximadamente 70%), que recebem os maiores salários (R\$ 20,09) quando comparados aos do gênero feminino (R\$ 16,69). Trabalhadores brancos, com média de salário-hora de R\$ 22,36, representam a maioria das observações na amostra, 52,3%. Já, 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.

Tabela 3: Rendimento médio e distribuição dos trabalhadores conforme suas características

Variável	Média	Desvio-padrão	Observações	Frequência (%)
Gênero				
Feminino	16,686	18,692	3.527.982	30,50%
Masculino	20,092	21,627	8.057.655	69,50%
Raça				
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%
Educação				
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%
Ocupação				
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.

A Tabela 3 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 à 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.

Por fim, a Tabela 3 mostra que em relação à ocupação, a maioria dos trabalhadores são da produção de bens e serviços industriais 1 (23,8%), dos serviços e vendedores do comércio em lojas e mercado (21,4%) e dos serviços administrativos (18,2%). 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 (R\$ 41,19). Já os trabalhadores agropecuários, florestais e da pesca recebem os menores salários (R\$ 10,62), seguidos dos trabalhadores dos serviços e vendedores do comércio em lojas e mercado (R\$ 10,96) quando comparados aos trabalhadores das demais ocupações.

Na Tabela 4 apresenta-se uma análise da estrutura salarial com base nas variáveis relacionadas às firmas. A maioria dos trabalhadores está em firmas que atuam no setor da indústria da transformação (32,4%) e no setor do comércio; reparação de veículos automotores e motocicletas (26,4%). Em termos de salário-hora real, 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 melhores remunerados em relação aos outros setores, respectivamente. Enquanto os trabalhadores dos 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.

Tabela 4: Rendimento médio e distribuição dos trabalhadores conforme características das firmas

Variável	Média	Desvio-padrão	Observações	Frequência (%)
Setor				
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 aquic.	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%
Tamanho da firma				
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.

Ainda de acordo com a Tabela 4, os trabalhadores estão empregados, em sua maioria, em firmas de grandes (41,9%), seguidos dos trabalhadores das pequenas firmas (25,1%), das firmas médias (17,9%) e das microempresas (15,1%). Em termos salariais, os trabalhadores das firmas maiores recebem a maior remuneração média (R\$ 22,69), seguidos dos trabalhadores das firmas médias (R\$ 20,81), das firmas pequenas (R\$ 16,21) e das microempresas (R\$ 11,63).

A Tabela 5 faz uma análise da distribuição salarial conforme as variáveis ligadas diretamente às LMA. A maioria dos trabalhadores está empregada em LMA de tamanho grande (69,1%). Os indivíduos das LMA grandes auferem as maiores médias salariais (R\$ 20,53), seguidos dos trabalhadores das médias LMA (R\$ 15,96) e das pequenas LMA (R\$ 14,4).

Tabela 5: Rendimento médio e distribuição dos trabalhadores conforme características das LMA

Variável	Média	Desvio-padrão	Observações	Frequência (%)
Tamanho da LMA				
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%
Região				
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%

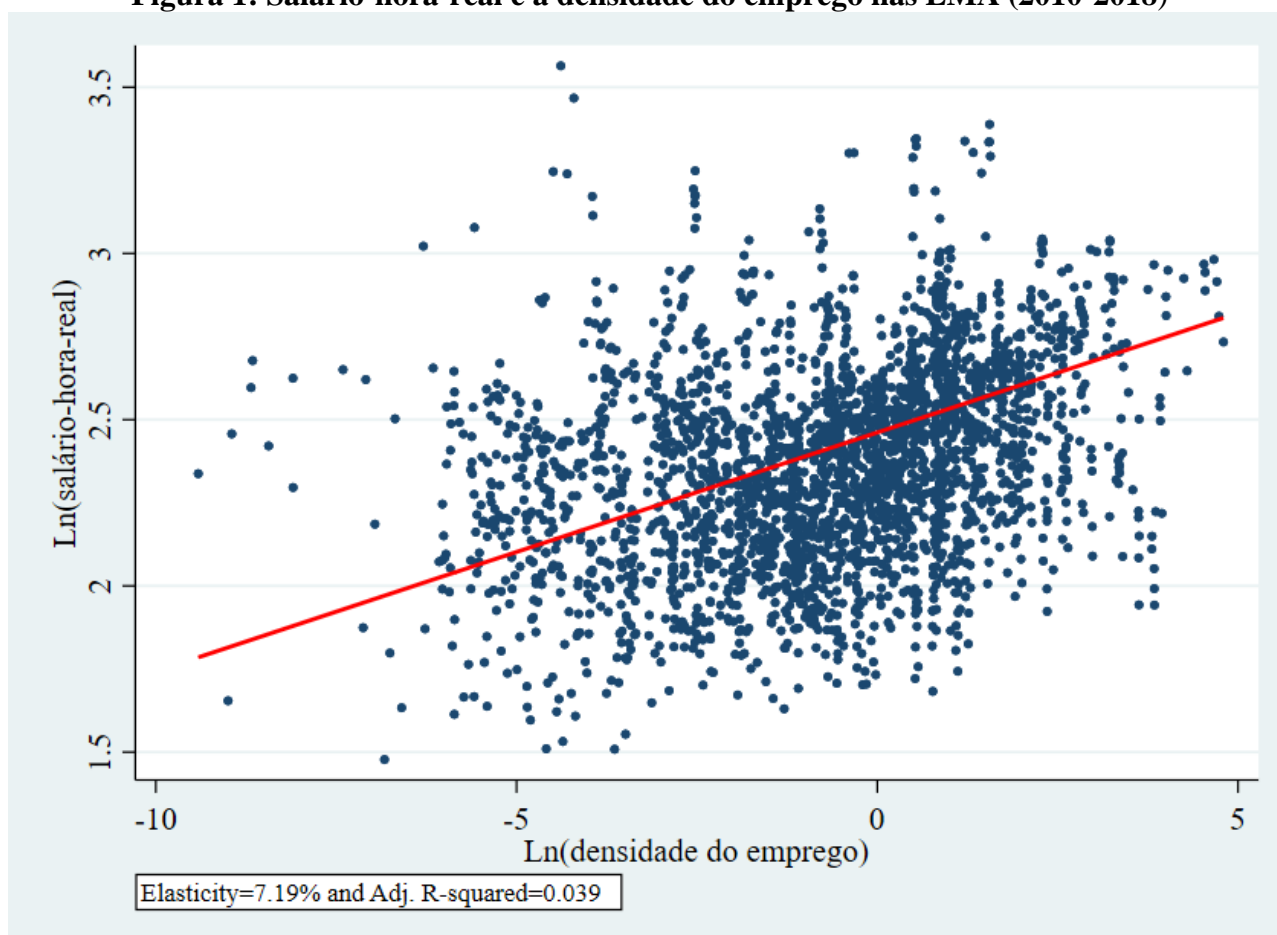
Fonte: Elaborada pelo próprio autor a partir dos microdados identificados da RAIS (2010-2018).

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

Por fim, a Tabela 5 mostra ainda que 61% dos trabalhadores estão empregados em LMA da 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 possuem as LMA com a menor parcela de trabalhadores, 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).

A figura Figura 1 mostra a relação entre a variável de interesse deste estudo, a densidade do emprego, e os salários por meio de um ajuste linear. A partir da figura observa-se que existe uma relação positiva entre as aglomerações e os salários dos trabalhadores formais das LMA. A Figura 1 mostra que a parcela do salário ofertado nas LMA do Brasil explicada pelas aglomerações é de 7,19% no 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%. Esse aumento é superior aos 4,9% encontrados por Combes, Duranton e Gobillon (2008) para França e aos 6,7% encontrados por Silva (2017) para o Brasil.

Figura 1: 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).

4.2 Resultados das estimações em painel

A Tabela 6 apresenta os resultados do modelo POLS (coluna 1), que considera as características observáveis dos trabalhadores, das firmas e das LMA, além de incluir *dummies* de tempo na estimação. O resultado encontrado revela que a variação em 1% nos níveis de densidade do emprego nas LMA resulta em uma variação de 4,07% nos salários locais de maneira diretamente proporcional. Essa variação é próxima dos 4% encontrados por Groot, Groot e Smit (2014) para Holanda, acima dos 3,22% encontrados por Combes, Gobillon e Duranton (2008) para França e dentro do intervalo de 3- 8%, consenso da literatura sobre economias de aglomeração (ROSENTHAL; STRANGE, 2004).

Como já discutido na seção 3, os resultados do modelo POLS apresentam estimativas viesadas da relação entre aglomeração e salário, pois não considera as heterogeneidades não observadas na sua análise. 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 6, ao se controlar apenas pelo efeito fixo dos trabalhadores (coluna 2), 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 3), 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 4), o prêmio salarial associado a uma variação de 1% na densidade do emprego é de 1,91%. A coluna 4 da Tabela 6 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 4. 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.

As estimativas encontradas para o modelo estimado por múltiplos efeitos fixos (coluna 5 da Tabela 6 são certamente mais robustas do que aquelas em que esses efeitos não considerados. 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, o modelo completo é estimado por VI.

Tabela 6: Efeitos das economias de aglomeração sobre os salários individuais nas LMA (2010-2018)

Variáveis	(1) POLS	(2) EF-ID	(3) EF-Firma	(4) EF-ID e Firma	(5) EF-ID, Firma e ID- Firma	(6) VI-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)	0.0169*** (0.0013)
Carac. Trabalhadores	Sim	Sim	Sim	Sim	Sim	Sim
Carac. Firms	Sim	Sim	Sim	Sim	Sim	Sim
Carc. LMA	Sim	Sim	Sim	Sim	Sim	Sim
Ano	Sim	Sim	Sim	Sim	Sim	Sim
Efeito Fixo ID	Não	Sim	Não	Sim	Sim	Sim
Efeito Fixo Firma	Não	Não	Sim	Sim	Sim	Sim
Efeito Fixo ID x Firma	Não	Não	Não	Não	Sim	Sim
F 1° estágio						370985
Prob> F 1° estágio						0
R ²	0.5668	0.9301	0.7826	0.9431	0.9501	0.2259
R ² ajustado	0.567	0.921	0.774	0.933	0.941	0.0761
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	10,933,643
F	268604	27540	3393	58930	22028	58214
Prob>F	0	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 coluna 6 da Tabela 6 apresenta as estimativas do segundo estágio das estimações por VI. O *ln* da densidade populacional defasada das LMA dos anos de 1920, 1940 e 1950 é o instrumento para a densidade do emprego. Como já documentado na literatura discutida neste artigo, 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 o modelo estimado por VI, mostra que o viés de instrumento fraco é pequeno, o que indica que os resultados encontrados são os menos enviesados possíveis.

De acordo com os resultados da principal estimativa deste artigo, disposta na coluna 6 da Tabela 6, quando se considera os efeitos fixos múltiplos (trabalhador, firma, *match* trabalhador-firma) e se trata a endogeneidade, ao mesmo tempo, o prêmio salarial às aglomerações é de 1,69%. Esse prêmio é 11,1% menor quando comparado ao modelo da coluna 5, que não trata a endogeneidade, indicando que ela age superestimando o efeito da densidade do emprego sobre os salários.

Quando se compara o principal resultado deste estudo, constata-se que o prêmio salarial às aglomerações, no valor de 1,69%, é inferior aos encontrados nos trabalhos citados neste artigo e inferior à escala proposta pela literatura (3-8%). Os prêmios salariais de 3,02% encontrado por Combes, Duranton e Gobillon (2008) para França, de 4,8% estimado por Groot, Groot e Smit (2014) para Holanda e de 4,31% encontrado por Chauvin *et al* (2017) para os EUA mostram que os ganhos de aglomeração no Brasil seguem o mesmo sentido dos países desenvolvidos, porém com magnitude, consideravelmente, inferior.

Quando se compara o prêmio salarial brasileiro aos dos países em desenvolvimento, a diferença de magnitude é ainda mais acentuada. O retorno salarial às aglomerações encontrado por

Durantón (2016) para Colômbia é de 5,4%, o encontrado para Índia e China, estimado por Chauvin *et al* (2017), é de 8,28% e de 16,9%, respectivamente.

Em relação estudos feitos para o Brasil, a magnitude do prêmio salarial aqui estimado também é inferior. Chauvin *et al* (2017) encontrou uma elasticidade de 2,6% para o país. Já Silva (2017) encontrou um prêmio salarial de 4,9%, que se assemelha aos resultados encontrados para os países desenvolvidos e para Colômbia.

Como se vê, o prêmio salarial decorrente das aglomerações encontrado neste artigo quando comparado aos encontrados nos estudos para o Brasil é, substancialmente, inferior. Esse resultado pode estar atrelado a cinco fatores principais. O primeiro diz respeito ao período analisado, nenhum dos estudos, para o Brasil, analisou os diferenciais salariais entre 2010 e 2018, período de grandes mudanças econômicas no país. O segundo fator pode estar ligado à qualidade dos microdados fornecidos pela RAIS, é possível que as informações prestadas tenham apresentado uma acurácia maior do que em períodos anteriores.

O terceiro fator pode estar ligado ao fato de este trabalho controlar, simultaneamente, o *sorting* espacial dos trabalhadores e das firmas e o *match* trabalhador-firma, diferentemente, dos demais apresentados neste artigo. É possível que uma especificação mais restrita possa reduzir o efeito líquido das aglomerações sobre os salários, já que parte desse efeito é capturado pelos efeitos fixos. O quarto fato pode estar relacionado ao processo de construção das LMA, que difere dos processos adotados nos estudos para o Brasil aqui apresentados.

Por fim, o último fator e, talvez, um dos mais importantes, diz respeito ao fato dessa estimativa (1,69%) poder estar refletindo os resultados das políticas territoriais implementadas no Brasil durante o período analisado, e que não foram captadas nos demais estudos feitos para o país. É possível que essas políticas não estejam favorecendo a aglomeração, o que pode afetar a produtividade das LMA e, conseqüentemente, o salário dos seus trabalhadores. Esta hipótese, contudo, deve ser melhor investigada, o que é não é objetivo deste estudo.

Ainda sobre a principal estimativa do prêmio salarial encontrada neste artigo (1,69%), ela se aproxima ao prêmio de 1,78% encontrado para as regiões metropolitanas de São Paulo, no estudo feito por Campos (2018). Em sua abordagem, o autor também utilizou uma especificação mais restrita semelhante a feita por este artigo. Guardadas as devidas proporções e tendo em vista que o trabalho de Campos (2018) é feito para regiões metropolitanas de um estado, e o presente artigo considera a totalidade do país, é razoável apontar como uma das possíveis justificativas para resultados tão próximos a adoção de uma especificação mais restrita do que aquelas feitas nos demais estudos apresentados neste trabalho.

Em resumo, 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 por múltiplos efeitos fixos, e para 1,69%, ao se adicionar o tratamento da endogeneidade via VI-MQ2E. Os resultados, de modo geral, mostram uma maior importância do controle do *sorting* espacial dos trabalhadores, principalmente, e das firmas nas estimações dos efeitos das aglomerações sobre os salários em detrimento do *match* trabalhador-firma. Além disso, esses resultados apontam para necessidade de se tratar de maneira adequada a endogeneidade entre a densidade do emprego e os salários para evitar estimativas superestimadas da relação aglomerações e salários.

5 CONSIDERAÇÕES FINAIS

Este artigo mensura em que magnitude as aglomerações explicam os diferenciais salariais existentes no Brasil. Faz isso por meio de uma abordagem, até então, nunca feita nos estudos sobre economias de aglomeração para o país, controlando, simultaneamente, os efeitos do *sorting* espacial dos trabalhadores e das firmas e do *match* trabalhador-firma.

A revisão da literatura apresentada mostra que há uma relação positiva entre aglomerações e salários, contudo a magnitude dessa relação pode ser diferente entre países desenvolvidos e em desenvolvimento, como é o caso do Brasil. É consenso nestes estudos que a autoseleção 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). Neste sentido, este trabalho lida com essas fontes de vies por meio de estimações por múltiplos efeitos fixos.

A literatura também aponta para uma possível endogeneidade entre a variável densidade do emprego e os salários. Para lidar com este problema, este estudo adotou o método de VI via MQ2E. Seguindo a literatura, este estudo utilizou a densidade populacional defasada como instrumento da densidade do emprego atual. 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.

A especificação principal adotada neste artigo foi baseada, 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), e o controle do *match* trabalhador-firma, conforme recomendado por Woodcock (2008; 2015), que destaca o possível vies das estimativas se esse *match* não for considerado nas análises. Portanto, a principal especificação desta pesquisa, que controla, simultaneamente, os múltiplos efeitos fixos, e é estimada por VI-MQ2E, leva em consideração as contribuições desses três estudos.

Os resultados encontrados neste artigo 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 deste estudo, estimado por VI-MQ2E, que foi 1,69%.

O prêmio salarial às aglomerações do modelo de referência (1,69%) é inferior ao intervalo proposto na literatura e ao encontrado em estudos para países desenvolvidos e em desenvolvimento, bem como em estudos feitos para o Brasil. A diferença na magnitude desse prêmio pode ser fruto, entre outras coisas, de uma melhor acurácia dos dados aqui analisados, do período escolhido para análise, do processo de construção das LMA, de uma especificação econométrica mais restritiva e/ou de mudanças nas políticas territoriais brasileiras.

Em resumo, os resultados encontrados neste artigo 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.

Pesquisas subjacentes a este estudo são requeridas. O que pode explicar uma magnitude do prêmio salarial decorrente das aglomerações tão inferior às observadas na literatura como um todo? Diferenças no padrão de aglomeração, melhor qualidade dos dados, políticas territoriais no sentido oposto à aglomeração, podem justificar a baixa magnitude dos efeitos das aglomerações sobre os salários nas LMA do Brasil?

REFERÊNCIAS

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.

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.

- CAMERON, Adrian Colin; TRIVEDI, Pravin K. **Microeconometrics Using Stata**. Stata Press, College Station, 2009. 692 p.
- CAMPOS, Rodger Barros Antunes. **Subcentralidades 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.
- CHAUVIN, Juan Pablo et al. What is different about urbanization in rich and poor countries? Cities in Brazil, China, India and the United States. **Journal of Urban Economics**, [s. l.], v. 98, Urbanization in Developing Countries: Past and Present, p. 17–49, 2017. Disponível em: <https://doi.org/10.1016/j.jue.2016.05.003>.
- CICCONE, Antonio. Agglomeration effects in Europe. **European Economic Review**, [s. l.], v. 46, n. 2, p. 213–227, 2002.
- CICCONE, 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; 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.
- CORREIA, Sergio. **Explanation of the HDFE iteration with 3 FEs**. p.4. 2014. Disponível em: <<http://scorreia.com/research/hdfe.pdf>>. Acesso em: 15 jan. 2021.
- DURANTON, Gilles. Agglomeration Effects in Colombia. **Journal of Regional Science**, [s. l.], v. 56, n. 2, p. 210–238, 2016. Disponível em: <https://doi.org/10.1111/jors.12239>.
- DURANTON, Gilles; PUGA, Diego. Chapter 48 Micro-foundations of urban agglomeration economies. **Handbook of Regional and Urban Economics**. [S. l.]: Elsevier, 2004. v. 4, p. 2063–2117
- FINGLETON, Bernard. 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>.
- FUJITA, Masahisa; THISSE, Jacques-François. **Economics of Agglomeration: Cities, Industrial Location, and Regional Growth**. [S. l.]: Cambridge University Press, 2002.
- GLAESER, Edward L.; MARÉ, David C. Cities and Skills. **Journal of Labor Economics**, [s. l.], v. 19, n. 2, p. 316–342, 2001.
- GLAESER, Edward L.; RESSEGER, Matthew G. The complementarity between cities and skills*. **Journal of Regional Science**, [s. l.], v. 50, n. 1, p. 221–244, 2010. Disponível em: <https://doi.org/10.1111/j.1467-9787.2009.00635.x>.
- 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.
- 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.

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

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 2010-2018: 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 >. 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.

PUGA, Diego. 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, Stuart S.; STRANGE, William 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, 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.

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.