

AGLOMERAÇÃO E DIFERENCIAIS SALARIAIS NO BRASIL

AGGLOMERATION AND WAGE DIFFERENTIALS IN BRAZIL

Jefferson Douglas da Silva Pereira¹

Adelson Santos da Silva²

Roberta de Moraes Rocha³

RESUMO

O objetivo principal deste estudo é identificar a contribuição das economias de aglomeração, mensurada pela densidade do emprego, sobre o diferencial salarial nas regiões brasileiras no período de 2010 a 2014. As estimações foram feitas a partir das equações salariais mincerianas, com uso dos dados provenientes da Relação Anual de Informações Sociais do Ministério do Trabalho e Previdência Social (RAIS/MTPS). Os resultados encontrados em todos os modelos indicaram que trabalhadores com elevada escolaridade, empregados de setores mais industrializados e de maiores firmas, assim como residentes das regiões Centro-Oeste e Sudeste, recebem os maiores salários quando comparados às suas contrapartes. Por fim, para a variável de interesse – densidade do emprego – os resultados mostraram que as economias de aglomeração influenciam positivamente os diferenciais salariais dos trabalhadores nas regiões brasileiras.

Palavras-chave: diferencial salarial, economias de aglomeração, *sorting* de habilidades.

Classificação JEL: R23; E24; J31; C23; C26.

ABSTRACT

The main objective of this study is to identify the contribution of agglomeration economies, captured by the employment density, on the wage differential in the Brazilian regions from 2010 to 2014. The regressions are based on the model proposed by Combes, Duranton, and Gobillon (2008), from the mincerian wage equations by using data from the Ministry of Labor (RAIS-MTE). The results found in all the regression models indicate that the most qualified workers, the employees in more industrialized sectors, and in the largest firms, as well as the residents of the Midwest and Southeast regions, they receive the highest salaries when compared to their counterparts. Finally, for the variable of interest – employment density – the results show that agglomeration economies positively influence the wage differentials of workers in Brazilian regions.

Keywords: wage differentials, agglomeration economies, sorting skills.

JEL Code: R23; E24; J31; C23; C26.

¹ Doutorando do Programa de Pós-Graduação em Economia da Universidade Federal de Juiz de Fora (PPGE/UFJF), Juiz de Fora, MG, Brasil. E-mail: jeffersondoglas08@hotmail.com

² Doutor em Economia pela Universidade Federal da Bahia (UFBA), Salvador, BA, Brasil. Professor do Departamento de Economia da Unidade Acadêmica de Serra Talhada da Universidade Federal Rural de Pernambuco (UAST/UFRPE), Serra Talhada, PE, Brasil. E-mail: adelsonsantos@gmail.com

³ Doutora em Economia pela Universidade Federal de Pernambuco (UFPE), Recife, PE, Brasil. Professora Associada da UFPE – Campus do Agreste (UFPE/CAA), Caruaru, PE, Brasil. E-mail: roberta_rocha_pe@yahoo.com.br

1. INTRODUÇÃO

A questão espacial tem sido foco da Economia Regional e Urbana, que tem direcionado esforços para compreender a relação entre a concentração espacial das atividades econômicas e os ganhos de produtividade por ela gerados. O aumento de produtividade oriundo da disposição espacial pode refletir em ganhos de rendimentos, já que 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). Nessa perspectiva, a investigação dos diferenciais salariais é cada vez mais recorrente em estudos de economia urbana, principalmente considerando o papel da aglomeração do ambiente urbano na distribuição dos rendimentos.

Marshall (1890) propõe três mecanismos para explicar o surgimento das aglomerações: transbordamento de conhecimentos (*knowledge spillovers*); agrupamentos de mão-de-obra (*labor pooling*); e encadeamento intersetorial (*intermediate input linkages*) (COMBES; GOBILLON, 2015). Com base nisso, Puga (2010) realça que um mercado maior permite um *sharing* (compartilhamento) mais eficiente da infraestrutura e instalações locais, uma variedade de fornecedores intermediários de insumos e/ou um grupo de trabalhadores com habilidades semelhantes. Além disso, estimula um melhor *matching* (combinação) entre empregadores e funcionários, compradores e fornecedores ou parceiros de negócios. Por fim, pode facilitar o *learning* (aprendizado), por exemplo, promovendo o desenvolvimento e a ampla adoção de novas tecnologias e práticas de negócios.

Os estudos empíricos sobre economias de aglomeração urbana ganham destaque na literatura econômica a partir do modelo centro-periferia desenvolvido por Krugman (1991), que buscou microfundamentar os mecanismos de aglomeração. Contudo, como destaca Puga (2010), a análise empírica desses mecanismos não é simples, dado que todas as fontes de aglomeração compartilham a previsão de que a produtividade aumenta com a escala da atividade local. Isso torna difícil mensurar o impacto individual de cada mecanismo sobre a produtividade. Com o intuito de superar essa limitação, Glaeser e Maré (1994; 2001) usaram a densidade do emprego para capturar o efeito geral das economias de aglomeração sobre a determinação dos salários. Os autores identificaram a existência de um prêmio salarial a favor das maiores áreas metropolitanas dos EUA, na ordem de 32% (1994) e de 33% (2001). Dentre as principais críticas, paira a de que Glaeser e Maré não controlaram a provável endogeneidade resultante da simultaneidade entre salários e densidade urbana.

Nos últimos anos, entretanto, as pesquisas têm avançado metodologicamente para evitar a superestimação dos efeitos das economias de aglomeração sobre os salários (COMBES; DURANTON; GOBILLON, 2008; ROSENTHAL; STRANGE, 2008), o que pode ocorrer caso trabalhadores e firmas mais produtivos se aglomerem em mercados de trabalhos mais densos e profícuos. Nesse sentido, é importante isolar os efeitos da concentração de trabalhadores mais habilidosos (*sorting*) na estimação dos efeitos das economias de aglomeração sobre os salários (GLAESER; MARÉ, 2001; COMBES; DURANTON; GOBILLON, 2008; MATANO; NATICCHIONI, 2016).

Para o Brasil, as evidências sugerem que a aglomeração desempenha um papel fundamental na determinação dos salários. Rocha, Silveira Neto e Gomes (2011) constataram um diferencial positivo de 9,4% para os trabalhadores das regiões metropolitanas brasileiras, no período entre 2000 e 2008. Já Silva, Santos e Freguglia (2016) encontraram um prêmio salarial urbano de 3% para essa mesma região, no período de 1995 a 2008. Por outro lado, apesar da persistência de ganhos salariais em favor das maiores cidades, uma gama de trabalhos

aponta para uma considerável redução na magnitude do prêmio salarial urbano quando se controla a heterogeneidade não-observada (FREGUGLIA, 2007, ROCHA; SILVEIRA NETO; GOMES, 2011, BARUFI; HADDAD; NIJKAMP, 2016).

Em estudo mais recente, Silva (2017) encontrou um efeito positivo da densidade do emprego (utilizada como medida de aglomeração) sobre os salários na ordem de 4,9% nos arranjos populacionais brasileiros. Em sua análise, a autora leva em conta o *sorting* espacial tanto dos trabalhadores quanto das firmas na determinação salarial e na identificação dos efeitos das economias de aglomeração. No referido estudo, fica evidenciado que o efeito trabalhador é mais relevante do que o efeito firma na variação individual dos salários e ressalta-se a importância desses controles para a mensuração adequada dos efeitos de aglomeração sobre os salários.

De todo modo, há indícios de ganhos de aglomeração no Brasil, que, em parte, é causa das disparidades regionais. Dados da Relação Anual de Informações Sociais (RAIS), do Ministério do Trabalho e Previdência Social (MTPS) indicam que a média salarial da Região Nordeste foi 77,7% da média salarial real⁴ do Brasil, a menor entre todas as regiões. Por sua vez, a Região Sudeste apresentou uma média salarial 20,9% maior que a nacional. Comparando as duas regiões, os ganhos salariais médios do Nordeste correspondem a 64,3% da média do Sudeste. Paralelamente, de acordo com o Instituto Brasileiro de Geografia e Estatística (IBGE), a despeito de uma menor extensão territorial, a região Sudeste apresenta uma densidade populacional aproximadamente 2,5 vezes maior do que a observada no Nordeste. A combinação de uma maior densidade de emprego e maiores retornos salariais, nesse contexto, sugere a presença de ganhos de aglomeração atuantes na determinação dos salários nas regiões brasileiras.

Nesses termos, este trabalho objetiva analisar a influência das economias de aglomeração sobre os diferenciais salariais no Brasil, no período de 2010 a 2014. O estudo contribui com a literatura ao implementar um amplo conjunto de variáveis para controlar simultaneamente os efeitos de indivíduos, firmas e regiões nas equações de salários. A densidade do emprego é utilizada como medida de aglomeração e construída a nível municipal, e, levando em consideração que não houve mudanças significativas na distribuição de municípios no Brasil no intervalo 2010-2014, o período escolhido é homogêneo em termos de desempenho da atividade econômica, o que reverbera em um mercado de trabalho mais estável.

Para tanto, este trabalho faz uso de equações mincerianas de salário a partir de um modelo com dados em painel, cuja principal vantagem é a possibilidade de controle das características não-observadas de trabalhadores. Os dados são oriundos da RAIS que, por seu formato longitudinal, permite o acompanhamento do trabalhador ao longo do tempo, possibilitando estudos de mobilidade geográfica, setorial e ocupacional de indivíduos. Nesse contexto, a Relação provê um monitoramento mais robusto da evolução das características do trabalhador, sendo, portanto, uma ferramenta útil em estudos sobre o comportamento do mercado de trabalho.

A relevância deste estudo se dá pelo fato de que, a despeito dos esforços da literatura nacional em descortinar a influência das cidades na determinação salarial, poucas pesquisas têm se debruçado na análise dos efeitos da combinação das características observáveis e não observáveis de indivíduos, atributos regionais e firmas sobre a produtividade dos trabalhadores. Ademais, em um país de dimensões continentais como o Brasil, heterogeneidades de funcionários e firmas podem ter efeitos distintos sobre a produtividade dos trabalhadores em escopo regional, na medida em que interagem com variáveis de aglomeração. Estudos brasileiros como os de Freguglia (2007) e Rocha, Silveira Neto e Gomes (2011), assim como este estudo,

⁴ O ano base para cálculo do salário real foi 2014, considerando o Índice Nacional de Preços ao Consumidor Amplo (IPCA), calculado pelo IBGE. Disponível em: <https://sidra.ibge.gov.br/tabela/6691>. Acesso em: 15 ago. 2018.

controlam a heterogeneidade não observada dos trabalhadores, contudo, não incluem em suas análises o componente espacial como um determinante salarial, o que é feito neste artigo.

Nestes termos, este trabalho avança na medida em que busca identificar a influência das economias de aglomeração sobre a produtividade nos municípios brasileiros, considerando fundamentalmente os efeitos da concentração local de habilidades. Ao concentrar a análise em nível municipal, é permitida uma comparação mais desagregada de como esses efeitos de produtividade se distribuem. Silva (2017) faz análise similar a este estudo, contudo, a unidade espacial considerada em seu trabalho são os arranjos populacionais, e não os municípios. Outra contribuição deste trabalho recai sobre o uso da população defasada como instrumento para a densidade atual, partindo do princípio de que cidades com salários locais mais elevados devem atrair trabalhadores mais habilidosos, aumentando a força de trabalho local e, portanto, influenciando a densidade do emprego (COMBES; DURANTON; GOBILLON, 2008, ABEL; DEITZ, 2012, BERLINGERI, 2018).

Além desta introdução, este trabalho conta com mais quatro seções. A seção 2 discorre e apresenta o modelo teórico sobre o *sorting* de habilidades como fonte de aglomeração produtiva, fundamento central deste estudo. A seção 3 apresenta a metodologia, os dados e as variáveis utilizadas para estimação do modelo econométrico proposto. Na seção 4 são descritos os principais resultados obtidos. Por fim, a seção 5 traz as considerações finais deste trabalho. Na sequência estão dispostas as referências utilizadas.

2. ***SORTING* DE HABILIDADES COMO FONTE DE AGLOMERAÇÃO PRODUTIVA: MODELO TEÓRICO**

Na economia urbana, os efeitos originários da associação entre maior produtividade nos mercados de trabalho mais densos, trabalhadores mais habilidosos e firmas mais produtivas são conhecidos como *sorting* (COMBES; DURANTON; GOBILLON, 2008; SILVA, 2017). A ocorrência desse fenômeno é frequentemente associada às diversas configurações espaciais na distribuição de habilidades por regiões e, nesse sentido, constitui elemento fundamental para a compreensão das disparidades regionais de produtividade, principalmente considerando que divergências salariais resultam das diferenças regionais na composição de atributos dos trabalhadores, nas dotações não humanas e nas interações produtivas locais.

Nesse sentido, esse efeito representa a associação entre produtividade, localização, trabalhadores e firmas. A questão é que a distribuição de habilidades dos trabalhadores depende da configuração setorial do mercado de trabalho, que influencia a demanda por qualificações. A distribuição de firmas, por sua vez, guarda relação com variáveis como proximidade dos mercados de insumos e produtos, tamanho do mercado de trabalho e disposição de mão-de-obra qualificada. À medida que interações entre essas variáveis são promovidas como fonte de ganhos de aglomeração, também os de produtividade podem ser gerados da mesma forma (MARSHALL, 1890; GLAESER; MARÉ, 2001; COMBES; DURANTON; GOBILLON, 2008; ROCHA; SILVEIRA NETO; GOMES, 2011; SILVA, 2017).

O modelo desenvolvido por Combes, Duranton e Gobillon (2008) sistematiza a ideia de *sorting* preconizada por Marshall (1890). Nele, os salários locais são determinados pela disposição local de habilidades. Considerando que o salário de um trabalhador i independe da localização, a remuneração é determinada pelo produto das habilidades individuais (S_i) e da produtividade do trabalho (A):

$$(1) \quad W_i = AS_i$$

No modelo proposto pelos autores, a produtividade pode ser influenciada por dotações não-humanas: os trabalhadores em algumas áreas podem ter um produto marginal mais alto do que em outras devido a características geográficas, como localização favorável, clima mais adequado para a atividade econômica ou presença de recursos naturais. As dotações locais, porém, não podem ser restritas a esses recursos e devem abranger fatores de produção, como capital público ou privado, instituições locais e tecnologia (COMBES; DURANTON; GOBILLON, 2008). Dessa forma, o salário de uma dada região a com dotações E_a é dado por:

$$(2) \quad W_a = A(E_a)$$

Além disso, interações locais entre trabalhadores habilitados e firmas produtivas em mercados de trabalho mais densos podem influenciar os salários (SILVA, 2017). Na verdade, os benefícios gerados por tais interações refletem em economias de urbanização, efeitos de localização e, ainda, na concentração setorial de firmas, os quais respaldam importantes fontes de economias de aglomeração, como o *matching* de trabalhadores ou o *sharing* regional. Assim, Combes, Duranton e Gobillon (2008) definem o salário médio na região a e setor k como $W_{ak} = A(I_a, I_{ak})$, em que I_a, I_{ak} representam as interações entre as economias de urbanização e localização. Desse modo, o salário de equilíbrio de um trabalhador i na área $a(i)$ e setor $k(i)$ será dado por:

$$(3) \quad W_i = A(E_{a(i)}, I_{a(i)}, I_{a(i)}, K_{(i)}) S_i$$

Conforme Combes, Duranton e Gobillon (2008), a estrutura da equação (3) – que engloba habilidades, dotações e interações – deve fornecer um senso de magnitude sobre a importância desses três tipos de explicações na determinação salarial entre as áreas. Ainda, o modelo microeconômico desenvolvido pelos autores 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:

$$(4) \quad \pi_{i,a,k,t} = p_{a,k,t} Y_{a,k,t} - \sum_{i \in (a,k,t)} W_{i,t} L_{i,t} - r_{a,k,t} Z_{a,k,t}$$

Na função, $p_{a,k,t}$ é o preço do produto $Y_{a,k,t}$; $W_{i,t}$ é o salário por dia; $L_{i,t}$ é o número de dias trabalhado; $Z_{a,k,t}$ representa outro fator de produção; e $r_{a,k,t}$ é o preço desse fator.

O modelo assume uma função de produção Cobb-Douglas em trabalho efetivo, em que $0 < b \leq 1$, $S_{i,t}$ são as habilidades do trabalhador i no período t ; $A_{a,k,t}$ é a produtividade total dos fatores em (a,k,t) ; e $S_{i,t}$ representa as habilidades do trabalhador i no período t :

$$(5) \quad Y_{a,k,t} = A_{a,k,t} \left(\sum_{i \in (a,k,t)} S_{i,t} L_{i,t} \right)^b (Z_{a,k,t})^{1-b}$$

No equilíbrio competitivo desse modelo, Combes, Duranton e Gobillon (2008) destacam que o trabalhador recebe salário igual a sua produtividade marginal:

$$(6) \quad W_{i,t} = b p_{a(i),k(i),t} A_{a(i),k(i),t} X \left(\frac{Z_{a(i),k(i),t}}{\sum_{i \in (a,k,t)} S_{i,t} L_{i,t}} \right)^{1-b} S_{i,t}$$

Maximizando o lucro em relação a $Z_{a,k,t}$ e o substituindo em $W_{i,t}$, tem-se:

$$(7) \quad W_{i,t} = b(1-b)^{\frac{(1-b)}{b}} X \left(p_{a(i,t),k(i,t),t} X \frac{A_{a(i,t),k(i,t),t}}{(r_{a(i,t),k(i,t),t})^{1-b}} \right)^{\frac{1}{b}} S_{i,t}$$

e

$$(8) \quad W_{i,t} = B_{a(i,t),k(i,t),t} S_{i,t}$$

Nesse caso, $B_{a(i,t),k(i,t),t} = b(1-b)^{\frac{(1-b)}{b}} (p_{a(i,t),k(i,t),t} A_{a(i,t),k(i,t),t} r_{a(i,t),k(i,t),t}^{1-b})^{\frac{1}{b}}$.

Pela expressão para $W_{i,t}$, fica evidenciado que diferenças regionais tanto de habilidade quanto de produtividade são fatores potencialmente explicativos para os diferenciais de salário entre áreas. Por outro lado, o diferencial de produtividade entre regiões está associado à produtividade dos fatores (A_{akt}) e à relação de preços e custos dos insumos ($p_{akt}; r_{akt}$).

Nesse contexto, tanto as forças aglomerativas quanto as características locais podem afetar a produtividade. Assim, apesar de não conseguir identificar como os efeitos tecnológicos e os preços afetam os salários isoladamente, o modelo proposto por Combes, Duranton e Gobillon (2008) é uma sofisticada ferramenta para a compreensão dos diferenciais salariais, especialmente numa perspectiva espacial.

3. METODOLOGIA: DADOS, VARIÁVEIS E MODELO ECONOMÉTRICO

Esta seção apresenta os dados, as variáveis e o modelo econométrico adotados nesta pesquisa. Ela está dividida em duas subseções. A primeira apresenta a base de dados e as variáveis e a segunda descreve o modelo econométrico.

3.1 Dados e variáveis

Os dados utilizados neste estudo são provenientes da RAIS, base instituída em 1975, que recolhe informações, de forma compulsória e anual, de todos os estabelecimentos do mercado de trabalho formal brasileiro. A principal vantagem da RAIS é o seu formato longitudinal, que permite construir dados em painel combinados para trabalhadores e firmas. Como forma de utilizar uma amostra mais homogênea, de modo a obter comparações mais pertinentes, foram considerados apenas indivíduos do sexo masculino que estavam presentes na base em todos os anos e com vínculo ativo em 31 de dezembro do respectivo ano. Foram excluídos trabalhadores sem identificador, sem registro de remuneração e com jornada de trabalho inferior a 20 horas. Além disso, os trabalhadores da administração pública foram excluídos da análise, tendo em vista que a dinâmica de contratação desse setor é diferente daquela inerente ao âmbito privado. Foram mantidos apenas trabalhadores em idade produtiva (18 a 65 anos) e sem inconsistências nas variáveis educação e sexo. Ainda, excluíram-se os perfis de trabalhadores sem informação de município. Após a limpeza, o painel constou 236.357.641 observações. Porém, dada a limitação computacional, foi gerada uma amostra aleatória estratificada por município de 1,67%, com 3.946.910 observações, equivalente a 789.382 anuais. É preciso frisar que buscou-se atender ao critério de suficiente mobilidade entre municípios, de modo a assegurar a relevância da amostra gerada e garantir a estrutura do painel.

A variável dependente – logaritmo natural do salário-hora-real do trabalhador – foi construída tomando por base o salário nominal dos trabalhadores recebido no mês de dezembro

de cada ano deflacionado pelo Índice Nacional de Preços ao Consumidor Amplo (IPCA), tendo dezembro de 2014 como o período base⁵. A variável de interesse neste trabalho é a densidade do emprego, *proxy* para aglomeração, e foi construída a partir da razão população empregada no local pela área em quilômetros do município⁶. As demais variáveis explicativas são oriundas da RAIS e estão expostas no Quadro 1. É preciso mencionar que, neste estudo, as categorias de ensino fundamental, médio e superior da variável escolaridade englobam tanto os indivíduos com a etapa completa quanto incompleta.

QUADRO 1 – DESCRIÇÃO DAS VARIÁVEIS EXPLICATIVAS

Variável	Descrição
Escolaridade	Representa o nível de escolaridade do trabalhador. Categoria base: trabalhador analfabeto
Sexo	<i>Dummy</i> : 1 para homem 0 e para mulher
Cor ou raça	Representa a raça ou a cor do indivíduo. Categoria base: trabalhador branco
Idade	A idade do indivíduo
Idade ²	Quadrado da idade do indivíduo
Tempo no emprego	<i>Proxy</i> para experiência no emprego
Tempo no emprego ²	Quadrado da experiência no emprego
Grupo ocupacional	Representa a ocupação desempenhada pelo trabalhador. Categoria base: trabalhadores agropecuários, florestais, da caça e da pesca.
Setor da atividade econômica	Representa o setor no qual a firma está inserida. Categoria base: agricultura, pecuária e produção florestal.
Tamanho da firma	Representa o tamanho do estabelecimento. Categoria base: microfimas
Região	Macrorregião em que o emprego está localizado. Categoria base: Norte

Fonte: Elaboração própria.

Por fim, cabe mencionar que este trabalho utilizou a população dos anos de 1910, 1920, 1930 e 1940⁷ como instrumento para densidade do emprego⁸.

3.2 Modelo econométrico

O modelo econométrico adotado para as estimações é baseado naquele proposto por Combes, Duranton e Gobillon (2008), a partir das equações salariais mincerianas (MINCER, 1974), considerando um modelo com dados em painel, no período de 2010 a 2014. A especificação segue a forma funcional

$$(9) \quad \ln w_{ijt} = \beta Emp_{jt} + \delta X_{ijt} + \varphi F_{kt} + \gamma Z_j + \varepsilon_{it},$$

⁵ Disponível em: <https://sidra.ibge.gov.br/tabela/6691>. Acesso em: 15 ago.2018.

⁶ Disponível em: <https://www.ibge.gov.br/geociencias-novoportal/organizacao-do-territorio/estrutura-territorial/15761-areas-dos-municipios.html?=&t=downloads>. Acesso em: 15 jun. 2020.

⁷ Disponível em: <http://cnae.ibge.gov.br/busca-online-cnae.html?view=estrutura>. Acesso em: 15 jun. 2020.

⁸ A seção 3.2 discute o porquê de se utilizar a população defasada dos anos 1910, 1920, 1930 e 1940 como instrumento.

onde $\ln w_{ijt}$ é o logaritmo do salário-hora real do indivíduo i , no município j e no tempo t ; X é o vetor de características individuais do trabalhador, que incorpora as variáveis idade, idade ao quadrado, educação, raça, tempo no emprego e tempo no emprego ao quadrado. O vetor F_{kt} incorpora as características da firma k no tempo t , tais quais tamanho e setor de atividade; Z é um conjunto de cinco *dummies* representando as cinco macrorregiões brasileiras, com o Norte sendo a categoria de referência. A variável de interesse nesse trabalho é Emp_{jt} , o logaritmo natural da densidade do emprego na região j no tempo t , que capta diferenças de aglomeração entre os municípios e tem como objetivo identificar a influência do tamanho do mercado de trabalho sobre os diferenciais salariais. Os coeficientes β , δ , φ e γ representam os parâmetros a serem estimados, e ε_{it} é o erro estocástico.

Para evitar problemas de correlação entre ε_i e as variáveis explicativas do modelo, é necessário que as hipóteses abaixo sejam mantidas:

$$(10) \quad E(\varepsilon_i|X_i)=0; E(\varepsilon_i|F_k)=0; E(\varepsilon_i|Z_j)=0; e E(\varepsilon_i|Emp_j)=0;$$

Sem essas hipóteses, não há como manter a relação de causalidade, o que pode ocasionar problemas de endogeneidade no modelo, gerando estimativas dos parâmetros viesados e inconsistentes. Em termos práticos, isso implica na correlação entre alguma variável explicativa e o erro estocástico, trazendo estimativas viesadas e inconsistentes (WOOLDRIDGE, 2011).

Neste contexto, Freguglia (2007) argumenta que a existência de heterogeneidades não observadas de indivíduos pode acarretar endogeneidade, tornando os estimadores inconsistentes. A hipótese é que o diferencial salarial está relacionado a um conjunto de características não observáveis dos trabalhadores – tais como força de vontade, criatividade, motivação – e, nesses termos, não considerá-las implica em viés de variável omitida. Ademais, se trabalhadores mais habilidosos forem, de fato, atraídos para mercados de trabalho mais densos, pode haver viés de seleção, implicando em estimativas viesadas.

A fim de controlar esse problema, procedeu-se à estimação do modelo anterior com efeito fixo. Após essa inclusão, os atributos não observáveis individuais fixos no tempo são captados separadamente, evitando a correlação entre termo do erro e as variáveis explicativas. Dessa forma, a endogeneidade atribuída às características invariantes no tempo é controlada, permitindo estimação mais verossímil do efeito da aglomeração sobre os salários individuais (SILVA, 2013). Assim, a especificação do modelo segue a forma funcional adiante:

$$(11) \quad \ln w_{ijt} = \beta Emp_{jt} + \delta X_{ijt} + \varphi F_{kt} + \gamma Z_j + \varnothing_i + \varepsilon_{it}$$

Com \varnothing_i sendo o efeito fixo de indivíduo, a hipótese de identificação deste modelo exige que $E(\varepsilon_i | \varnothing_i)=0$. Essa expressão denota que a correlação entre X_i e ε_i é captada por variáveis explicativas fixas (\varnothing_i), ou seja, que não variam no decorrer do tempo. A inclusão desse efeito torna, portanto, o modelo mais consistente e não viesado.

A inclusão dos efeitos fixos na estimação não garante a total correção da endogeneidade, principalmente se há evidências de simultaneidade no modelo. Segundo Wooldridge (2011), a simultaneidade ocorre quando há determinação mútua entre uma variável explicativa e a dependente do modelo. É consenso na literatura da economia urbana que a relação entre densidade do emprego e produtividade é naturalmente endógena, já que choques contemporâneos em uma região podem ampliar os salários locais, gerando causalidade reversa (CICCONE; HALL, 1996, COMBES; DURANTON; GOBILLON, 2008, SILVA, 2017). Logo, a densidade

do emprego e os salários locais podem se influenciar reciprocamente e, se isso ocorre, a hipótese de exogeneidade ($Cov(EMP_{jt}, \varepsilon_{it})=0$) não se sustenta, o que torna o estimador ineficiente.

Nessas condições, Wooldridge (2011) propõe o uso de variável instrumental (VI). Instrumentos históricos têm sido amplamente defendidos nos estudos de economia urbana (CICCONE; HALL 1996; COMBES, DURANTON; GOBILLON, 2008; MELO; GRAHAM, 2009). Particularmente, a população defasada se mostra instrumento válido no controle da simultaneidade existente entre produtividade e densidade do emprego, uma vez que a densidade atual deve estar correlacionada com o padrão de concentração populacional do passado, mas os níveis de produtividade atuais não guardam relações com a distribuição passada da população.

Dessa forma, seguindo o consenso da literatura, este trabalho faz uso do logaritmo natural da população a nível municipal como instrumento para densidade urbana. Especificamente, faz-se uso do logaritmo da população dos anos de 1910, 1920, 1930 e 1940, como instrumentos para a densidade do emprego. Como pode ser visto, os instrumentos são relevantes devido à inércia da população local, no entanto, dadas as transformações que afetaram o mercado de trabalho brasileiro nas últimas décadas, espera-se que sejam exógenos aos salários.

Uma vez definida a variável instrumental, procedeu-se a estimação pelo método de mínimos quadrados em dois estágios (MQ2E). Combes, Duranton e Gobillon (2008) apresentam duas razões principais para o uso da estimação em dois estágios. A primeira diz respeito ao fato de que em estágio único não é possível calcular a variação dos choques locais, já que o modelo é projetado na dimensão *within* da equação salarial⁹ e os trabalhadores podem se mover entre áreas. Caso não seja estimado em dois estágios, o modelo terá vieses nos erros-padrão para os coeficientes estimados de variáveis explicativas agregadas. A segunda leva em consideração que a estimação em estágio único não permite a distinção entre os choques locais e aqueles puramente idiossincráticos no nível do trabalhador, o que é importante, tendo em vista a ausência das variáveis de dotação local.

4. RESULTADOS

Esta seção se organiza da seguinte maneira: primeiro, é feito um breve perfil do trabalhador brasileiro no período de análise e, posteriormente, são estimados modelos em dados de painel, com vistas a identificar os efeitos da aglomeração, bem como dos vetores de características de firmas e trabalhadores sobre os diferenciais de salários. Por fim, é estimado um modelo com variável instrumental com o objetivo de tratar a endogeneidade da densidade do emprego.

4.1 Perfil do trabalhador brasileiro no período de 2010 a 2014: estatísticas básicas

A amostra selecionada possui 3.946.910 observações divididas igualmente para cada um dos anos analisados, que vão de 2010 a 2014. A Tabela 1 e a Tabela 2 trazem as principais características relacionadas ao trabalhador. O salário real médio no território brasileiro (W_real), a preços de 2014, foi de R\$ 2.496,72, com variações de R\$ 153,00 até R\$ 108.468,80. A variável de salário, em logaritmo natural ($lnw-hora-real$), apresentou média de 7,53, com uma dispersão salarial de 0,78, indicando razoável distorção de salários, apesar de serem considerados apenas indivíduos do sexo masculino. Em parte, esse dado pode refletir a existência de segmentação no mercado de trabalho, de forma que características distintas como raça podem estar por trás

⁹ Na dimensão *within*, todas as variáveis são centralizadas em relação a sua média para cada indivíduo, o que resulta nos coeficientes de todas essas variáveis, exceto o efeito fixo do trabalhador. Em estágio único, não seria possível capturar esse efeito fixo.

desse resultado. Ademais, o estoque formal de capital humano (educação e experiência) e os atributos inerentes à ocupação e à região podem ser importantes elementos explicativos da trajetória dos salários no Brasil.

TABELA 1 – CARACTERÍSTICAS DOS TRABALHADORES COM BASE NA RAIS (2010 A 2014)

Variável	Observações	Média	Desvio	Min.	Max.
<i>W_real</i>	3.946.910	2.496,72	3.188,84	153,00	108.468,80
<i>Lnw-hora-real</i>	3.946.910	7,53	0,78	5,03	11,59
Tempo no emprego	3.946.910	70,19	75,11	0	597,90
Idade	3.946.910	37,87	10,30	18,00	65,00

Fonte: Elaboração própria a partir de dados da amostra da RAIS (2010 a 2014).

Quanto ao tempo de emprego no mesmo vínculo, *proxy* para experiência neste estudo, a média observada é 70,19 meses no período, com desvio de 75,11 meses. Isso sinaliza para trabalhadores com distintos níveis de experiência no mercado de trabalho, o que pode corroborar a ideia de que tempo maior no emprego induz a ganho de habilidades e, portanto, leva a salários mais elevados. No que tange à idade dos indivíduos observados, a média é de 37,87 anos, com desvio de 10,30 anos. Esse resultado aponta para duas importantes situações: o trabalhador médio brasileiro não é tão jovem, o que incide sobre a produtividade individual e, por conseguinte, nos salários; as distorções salariais refletem parcialmente a diversidade na composição etária da força de trabalho no Brasil.

Na Tabela 2, verifica-se que a grande maioria dos trabalhadores formais brasileiros tem ensino médio (50,8%) e fundamental (32,10%). A parcela desses trabalhadores que possui ensino superior é de 16,60%. A variável educação é particularmente relevante nos estudos de economia urbana e do trabalho; autores como Mincer (1974) argumentam que ela é a principal aproximação para o capital humano formal dos indivíduos e deve ter importante efeito sobre os salários. Se a força de trabalho mais qualificada está heterogeneamente distribuída no Brasil, isto pode sinalizar desequilíbrios regionais na distribuição de oportunidades (ineficiência de *matching*). Desse modo, atributos regionais poderiam atuar como mecanismo de seleção de indivíduos mais habilidosos (*sorting*) e, assim, diferenças regionais de salários refletiriam parcialmente na contribuição de características da própria região.

Já no que tange à distribuição por raça/cor, nota-se que o número de trabalhadores brancos representa 60% da amostra, seguida por pardos (28,7%) e pretos (5,5%), em porcentagem bem menor. Isso sugere que há uma predominância de homens brancos no mercado de trabalho formal brasileiro, o que pode acarretar diferenças de salário. Quanto à ocupação¹⁰, os grupos ocupacionais 5 (trabalhadores dos serviços, vendedores do comércio em lojas e mercado) e 8 (trabalhadores da produção de bens e serviços industriais [1]) representam 29,29% e 20,62% da amostra total de trabalhadores, respectivamente. De modo geral, os trabalhadores estão alocados, em sua maioria, em ocupações com baixa demanda de habilidades, o que pode afetar a própria configuração dos salários nos municípios brasileiros.

¹⁰ A ocupação foi definida de acordo com a Classificação Brasileira de Ocupações 2002 (CBO 2002), que define dez grandes grupos ocupacionais. Disponível em: <http://www.mtecbo.gov.br/cbosite/pages/downloads.jsf> Acesso em: 15 ago. 2018.

TABELA 2 – CARACTERÍSTICAS DOS TRABALHADORES COM BASE NA RAIS (2010 A 2014)

Variável	Frequência absoluta	Frequência relativa
Educação		
Analfabeto	19.769	0,50%
Ensino Fundamental	1.265.995	32,10%
Ensino Médio	2.006.316	50,80%
Ensino Superior	654.830	16,60%
Raça/cor		
Indígena	8.505	0,22%
Branco	2.368.140	60%
Preto	217.270	5,50%
Amarelo	30.282	0,77%
Pardo	1.132.209	28,70%
Não identificado	190.504	4,83%
Grupo de ocupação		
(1) Forças armadas, polícia [...]	161	0,00%
(2) Membros do poder público [...]	187.645	4,76%
(3) Profissionais das ciências e das artes	227.081	5,75%
(4) Técnicos de nível médio	383.864	9,73%
(5) Trabalhadores do serviço administrativo	539.681	13,68%
(6) Trabalhadores dos serviços [...]	813.894	20,62%
(7) Trabalhadores agropecuários [...]	191.153	4,84%
(8) Trabalhadores da produção [...] (1)	1.155.829	29,29%
(9) Trabalhadores da produção [...] (2)	250.408	6,35%
(10) Trabalhadores de manutenção [...]	196.735	4,99%
Observações totais	3.946.910	-

Fonte: Elaboração própria a partir de dados da amostra da RAIS (2010 a 2014).

A Tabela 3 mostra a média salarial de acordo com as regiões do Brasil. Os dados mostram que a maior média de salários real, no período de análise, foi na região Sudeste (R\$ 3.017,67), seguida pelas regiões Centro-Oeste, com R\$ 2.606,85, e Sul, com R\$ 2.506,55. Os indivíduos da região Nordeste apresentaram a menor média salarial real, R\$ 1.939,981, seguidos dos trabalhadores do Norte, cuja média salarial foi de R\$ 2.412,575. Os desvios-padrão dos salários em todas as regiões foram maiores que a média, e isso revela que existem grandes diferenças de ganhos não só entre os indivíduos de diferentes regiões brasileiras, mas também dentro das próprias regiões. Esse comportamento da média salarial só reforça mais ainda a necessidade de compreender esses diferenciais.

TABELA 3 – RENDIMENTO MÉDIO POR REGIÃO NO PERÍODO DE ANÁLISE

Região	Observações	Média (R\$)	Des. Pad.(R\$)	Min. (R\$)	Max. (R\$)
Nordeste	592.859	1.939,98	2.855,33	153,00	104.881,60
Norte	145.034	2.412,57	3.149,09	156,00	10.069,90
Sudeste	2.052.645	3.017,67	3.757,75	154,00	108.468,80
Sul	717.553	2.506,55	2.815,41	154,58	106.103,00
Centro- Oeste	264.102	2.606,85	3.366,63	156,30	83.977,22
Total	3.772.193	-	-	-	-

Fonte: Elaboração própria a partir de dados da amostra da RAIS (2010 a 2014).

4.2 Resultados das estimações de dados em painel

Regularidades empíricas apontam para a existência de um prêmio salarial em favor dos grandes centros urbanos (GLAESER; MARÉ, 2001; FREGUGLIA, 2007; ROCHA; SILVEIRA NETO; GOMES, 2011). Ao mesmo tempo, a maioria desses estudos destaca a relevante contribuição das características observáveis e não observáveis dos trabalhadores na explicação dos diferenciais salariais. Nesse sentido, assumir que variáveis de aglomeração constituem causa única para os diferenciais de salário é controverso. Ainda assim, a influência de variáveis regionais nas diferenças geográficas de produtividade é um aspecto relevante e, portanto, deve ser estudado, sobretudo em um país heterogêneo como o Brasil.

Tendo esses aspectos como base, a Tabela 4 apresenta os resultados das estimações dos modelos com dados em painel: *pooled data* (POLS); efeitos aleatórios (EA); efeitos fixos (EF); variáveis instrumentais por efeitos aleatórios; e variáveis instrumentais por efeitos fixos¹¹. Antes de tudo, é importante destacar que a escolha dos modelos mais consistentes se deu considerando os testes de Breusch-Pagan e Hausman, pelos quais ficou evidenciado que o modelo de efeito fixo é o mais adequado, uma vez que há indícios de que os componentes individuais não observados são invariantes no tempo.

Ao final da Tabela 4 estão dispostas as estatísticas do primeiro estágio dos modelos estimados por variáveis instrumentais e o coeficiente estimado no primeiro estágio para o instrumento do logaritmo da densidade do emprego (logaritmo da população defasada), que se mostrou elevado e significativo a 1%. A estatística F do primeiro estágio também é elevada e mostra que a hipótese de instrumento fraco é descartada. Embora não seja possível testar a hipótese de exogeneidade entre o instrumento e o termo de erro, os resultados do primeiro estágio e a ampla literatura disponível apontam para a adequabilidade do logaritmo da população defasada como instrumento para a densidade do emprego atual¹².

As colunas (1), (2) e (4) mostram os modelos POLS, EA e variável instrumental por EA, respectivamente, sem o controle pela heterogeneidade não-observada. Os resultados mostram que apenas ao dar o tratamento de painel há considerável relevância no efeito da densidade do emprego sobre os salários. A estimativa do efeito de aglomeração caiu 33% no modelo com efeitos aleatórios e, quando estimado por variável instrumental, o coeficiente saltou para 1,41%. A leve oscilação sinaliza que a endogeneidade afeta a identificação do modelo e que devem ser consideradas suas outras fontes potenciais.

¹¹ Os coeficientes para variável raça/cor não estão dispostos na Tabela 4, pois, com exceção do POLS e EA, todos os outros modelos são estimados controlando os efeitos fixos de trabalhador e, uma vez que raça/cor é variável fixa, nesses modelos sua média é zero.

¹² A correlação entre o instrumento e o logaritmo da densidade do emprego foi de 0,7672, reforçando, portanto, a adequabilidade do logaritmo da população defasada como instrumento para variável endôgena.

Já na comparação entre os modelos *pooled* e com efeito fixo (3), houve queda de 2,01% para 0,224%, com total aproximado de 88%. A validade dos testes de Hausman e Breusch-Pagan sugerem que as principais características não-observadas são invariantes no tempo. Ao mesmo tempo, a expressiva queda no coeficiente estimado para o efeito da densidade do emprego sobre os salários atesta que a principal fonte de endogeneidade é a concentração de indivíduos habilitados nos maiores centros urbanos. Nesse contexto, os salários mais elevados a favor das grandes cidades são consequência da maior disposição de habilidades, que se reflete em produtividade mais expressiva.

A Tabela 4 mostra que, quando os efeitos fixos e a causalidade reversa (5) são tratados simultaneamente, há uma oscilação positiva do coeficiente estimado em relação ao modelo que incorpora apenas os efeitos fixos (3), saltando de 0,00201 para 0,00404. Apesar disso, é notória a mudança em comparação ao modelo *pooled*, e isso indica a importância do controle pela heterogeneidade não-observada. Ademais, fica evidente que a causalidade reversa, ainda que em menor proporção em relação à disposição das características individuais, é uma fonte de endogeneidade; não tratá-la, portanto, pode levar a problemas de viés de estimação, como a subestimação dos efeitos da densidade do emprego sobre os salários, conforme observado no modelo estimado neste artigo.

TABELA 4 – RESULTADO DAS ESTIMAÇÕES ECONÔMICAS: EFEITO DA AGLOMERAÇÃO, COM CONTROLES POR CARACTERÍSTICAS INDIVIDUAIS, REGIÃO E FIRMA, SOBRE OS DIFERENCIAIS SALARIAIS DOS TRABALHADORES BRASILEIROS ENTRE 2010 E 2014 (VARIÁVEL DEPENDENTE: LN DA RENDA MÉDIA REAL)

	(1) POLS	(2) EA	(3) EF	(4) VI e EA	(5) VI e EF
Ln densidade emprego	0,0201*** (177,25)	0,0133*** (61,69)	0,00224*** (7,68)	0,0141*** (45,42)	0,0040*** (8,92)
Característica do trabalhador					
Idade	0,0517*** (275,81)	0,0649*** (227,95)	0,0595*** (107,48)	0,0631*** (223,88)	0,0599*** (133,35)
Idade ²	-0,0005*** (-229,05)	-0,0007*** (-192,91)	-0,0008*** (-160,82)	-0,0007*** (-196,84)	-0,0008*** (180,30)
Tempo no emprego	0,0132*** (128,14)	0,0008*** (68,06)	0,0008*** (-57,78)	0,0009*** (85,36)	0,0008*** (76,64)
Tempo no emprego ²	0,0000*** (35,98)	0,0000*** (2,69)	-0,0000*** (-6,76)	0,0000*** (10,29)	0,0000*** (-72,75)
Controles para e educação					
Fundamental	0,177*** (47,32)	0,0567*** (13,67)	-0,00672 (-1,50)	0,2563 (54,7)	-0,0189 (-3,76)
Médio	0,366*** (97,59)	0,148*** (35,16)	-0,00688 (-1,52)	0,32*** (68,84)	-0,0176* (-3,46)
Superior	0,912*** (237,12)	0,414*** (92,10)	0,0384*** (7,96)	2,414*** (357,18)	0,2001 (0,41)

	(1) POLS	(2) EA	(3) EF	(4) VI e EA	(5) VI e EF
Controles para ocupações					
Ciências/Artes	0,647*** (299,39)	0,375*** (115,00)	0,0995*** (28,19)	0,108*** (3,76)	0,0442 (1,58)
Téc. nível médio	0,412*** (216,77)	0,232*** (83,16)	0,0488*** (15,56)	0,0034 (0,12)	-0,0027 (-0,10)
Serviço admin.	0,158*** (85,16)	0,133*** (49,74)	0,000557 (0,18)	-0,0937 (-3,28)	-0,0154 (-0,55)
Trab. dos serviços	0,0464*** (25,69)	0,0418*** (15,90)	-0,0133*** (-4,44)	-0,162 (-5,63)	-0,0072 (-0,25)
Trab. produção (1)	0,149*** (85,92)	0,0920*** (36,46)	0,0219*** (7,58)	-0,0444 (-1,55)	0,0190 (0,68)
Trab. produção (2)	0,157*** (80,00)	0,123*** (43,86)	0,0452*** (14,31)	-0,0130 (-0,45)	0,0426 (1,52)
Trab. manutenção	0,165*** (163,46)	0,0861*** (58,03)	0,0230*** (13,63)	0,0426 (1,49)	0,0406 (1,45)
Controle para tamanho da firma					
Pequenas empresas	0,157*** (186,46)	0,0755*** (97,86)	0,0438*** (54,46)	0,0775*** (103,84)	0,0394*** (52,19)
Médias empresas	0,283*** (326,28)	0,154*** (151,55)	0,0919*** (82,59)	0,157*** (176,75)	0,0842*** (90,34)
Grandes empresas	0,382*** (448,30)	0,217*** (181,86)	0,128*** (94,39)	0,220*** (222,17)	0,116*** (108,41)
Controles por setor de atividade da firma					
Pesca e aquicultura	-0,0558*** (-4,52)	-0,0181 (-0,91)	-0,0300 (-1,25)	0,0070 (0,40)	-0,0062 (-0,30)
Indústria extrativa	0,506*** (184,96)	0,311*** (59,74)	0,0994*** (16,38)	0,357*** (90,02)	0,104*** (22,24)
Ind. de transform.	0,0640*** (37,23)	0,109*** (39,92)	0,0436*** (12,66)	0,141*** (52,99)	0,0567*** (17,55)
Elétrica, gás e água	0,159*** (65,90)	0,185*** (41,53)	0,0452*** (8,25)	0,209*** (56,79)	0,0531*** (11,83)
Construção civil	0,0450*** (22,50)	0,0628*** (19,83)	0,0181*** (4,70)	0,0864*** (29,37)	0,0264*** (7,60)
Comércio	-0,103*** (-58,25)	0,000472 (0,17)	-0,0134*** (-3,80)	0,0137 (5,00)	-0,0052 (-1,59)

	(1) POLS	(2) EA	(3) EF	(4) VI e EA	(5) VI e EF
Transpor. e Correios	0,000113 (0,06)	0,0357*** (11,79)	-0,00877* (-2,30)	0,0550 (19,10)	-0,0005 (-0,13)
Aloj. e alimentação	-0,265*** (-109,91)	-0,119*** (-29,66)	-0,0557*** (-10,71)	-0,111*** (-31,12)	-0,0486*** (-11,36)
Inf. e comunicação	-0,0953*** (-40,15)	0,0572*** (14,19)	-0,0284*** (-6,05)	0,0829*** (25,22)	-0,0173*** (-4,54)
Ativ. financeiras	0,305*** (133,11)	0,373*** (75,49)	0,102*** (15,50)	0,400*** (112,12)	0,107*** (23,74)
Ativ. imobiliárias	-0,142*** (-25,61)	-0,00572 (-0,77)	-0,0185* (-2,33)	0,00913 (1,56)	-0,0118 (-1,88)
Técnico	-0,00637** (-2,64)	0,0626*** (16,24)	-0,0152*** (-3,46)	0,0905*** (28,09)	-0,0048 (-1,28)
Ativ. administrativas	-0,163*** (-85,00)	-0,0446*** (-14,76)	-0,0368*** (-9,86)	-0,0332*** (-11,54)	-0,0292*** (-8,56)
Educação	-0,330*** (-119,59)	-0,0613*** (-11,17)	-0,0541*** (-7,55)	-0,0549*** (-13,51)	-0,0395*** (-8,04)
Saúde e ser. sociais	-0,306*** (-113,23)	-0,0464*** (-9,79)	-0,0181** (-2,94)	-0,0401*** (-10,24)	-0,0087 (-1,86)
Artes, cult. e esportes	-0,346*** (-83,31)	-0,110*** (-14,10)	-0,0324** (-3,22)	-0,0973*** (-16,41)	-0,0140* (-1,96)
Outros ser. coletivos	-0,206*** (-80,57)	-0,0356*** (-8,88)	-0,0296*** (-6,14)	-0,0257*** (-7,35)	-0,0219*** (-5,41)
Serviços domésticos	-0,261*** (-8,45)	-0,0881** (-2,93)	-0,0392 (-1,32)	-0,0739* (-2,52)	-0,0363 (-1,22)
Org. internacionais	0,425*** (22,27)	0,0905*** (5,51)	-0,0591*** (-4,07)	-0,0212* (-0,93)	-0,0595* (-2,56)
Controle de Macrorregião					
Nordeste	-0,210*** (-245,58)	-0,192*** (-87,12)	-0,0495*** (-10,85)	-0,204*** (-112,35)	-0,0359*** (-11,84)
Sudeste	0,146*** (204,11)	0,126*** (74,43)	0,0191*** (5,76)	0,118*** (94,77)	0,0208*** (9,58)
Centro oeste	0,0919*** (80,38)	0,0675*** (25,27)	0,0338*** (7,24)	0,0512*** (21,30)	0,0354*** (8,35)
Sul	-0,0904*** (-114,22)	-0,0843*** (-46,63)	-0,0399*** (-8,77)	0,0040 (1,27)	0,0846*** (14,80)
Controles de raça	Sim				

	(1) POLLS	(2) EA	(3) EF	(4) VI e EA	(5) VI e EF
<i>Dummies</i> de ano			Sim		
Número de Observações			3.946.910		
R2	0,5640				
<i>R(overall)</i>		0,5233	0,2235	0,1973	0,5426
Estatísticas do 1º estágio					
<i>Ln</i> (população defasada)				0,8167*** (787,85)	0,9466*** (1.920,86)
F (49,3.560.430)				181.042,38	125.910,06
Prob. > F				0,0000	0,0000
R2 ajustado					0,6341

Fonte: Elaboração própria a partir de dados da amostra da RAIS (2010 a 2014).

Nota: Estatística *t* entre parênteses.

**p* < 0,05

***p* < 0,01

****p* < 0,001

Em relação às características dos trabalhadores e seus impactos sobre os salários, idade e tempo no emprego, apresentaram coeficientes positivos e significativos, seguindo o padrão da literatura. Os coeficientes negativos para a forma quadrática dessas variáveis, no entanto, sugerem que os ganhos de produtividade delas derivados são cada vez menores à medida que o indivíduo fica mais velho e permanece por mais tempo no vínculo. Indivíduos com nível superior têm ganhos quase três vezes maiores em comparação aos trabalhadores de nível médio, enquanto esses últimos ganham aproximadamente duas vezes mais do que aqueles com nível fundamental, apenas.

No que tange à ocupação, verifica-se na Tabela 4 que os profissionais das ciências e das artes (9,95%), os técnicos de nível médio (4,88%) e os trabalhadores da produção de bens e serviços industriais (2) (4,52%) apresentaram maior diferencial salarial em comparação aos trabalhadores agropecuários, florestais e da pesca, principalmente no modelo com efeitos fixos. Os resultados são válidos, sobretudo se for considerado que as ocupações desses grupos exigem um nível educacional mais elevado, bem como requerem maior dotação de habilidades cognitivas (COELHO; CORSEUIL, 2002; MACIENTE, 2012).

Em todos os modelos há diferenças salariais positivas a favor dos trabalhadores de maiores estabelecimentos, mesmo nos modelos com efeito fixo, nos quais fica evidenciado um ganho de 12,8% em comparação às suas contrapartes nas microfirms. Provavelmente as maiores firmas são aquelas de elevada complexidade no processo produtivo e, conseqüentemente, demanda por trabalhadores mais habilitados. Estes resultados são consistentes com os achados de Topel e Ward (1992). Em consonância, os setores de atividades financeiras (10,7%), da indústria extrativa (10,4%), da indústria de transformação (5,67%), e de eletricidade, gás e água (5,31%) são os que apresentam os maiores salários, respectivamente. Isso revela que os setores mais complexos da economia guardam relação fundamental com a natureza dos ganhos. Na prática, setores mais dinâmicos requerem ocupações de maior complexidade, o que respalda a demanda por trabalhadores mais habilitados.

Nesse sentido, as evidências aqui apresentadas encontram suporte na literatura e apontam para a importância das características dos trabalhadores e das firmas sobre a produtividade local no mercado de trabalho, mas, por outro lado, distorções produtivas inter-regionais podem refletir debilidades na matriz produtiva local. Em outros termos, algumas regiões concentram as firmas mais produtivas, enquanto outras agrupam as atividades mais rudimentares. As disparidades setoriais, então, podem atuar como mecanismo indutor de desigualdades regionais, na medida em que geram diferentes oportunidades para trabalhadores com as mesmas habilidades. Ademais, se as firmas mais complexas estão localizadas nas regiões mais populosas, trabalhadores mais habilidosos devem estar em mercados de trabalhos também mais densos, nos quais possam se beneficiar de melhores interações com níveis distintos de capital humano ou com melhor dotação tecnológica.

Por fim, os controles por macrolocalização revelam ganhos salariais positivos a favor dos trabalhadores do Sudeste e Centro-Oeste em relação aos trabalhadores da Região Norte em todos os modelos estimados. Nos modelos EF, VI e EF, os ganhos do Sudeste são de 1,91% e de 2,08%, respectivamente, e os ganhos do Centro-Oeste são de 3,38% e de 3,54%, quando comparados à região Norte. Já o Nordeste apresentou diferenciais negativos em todas as análises do trabalho em relação ao Norte. No modelo EF, VI e EF esses diferenciais foram negativos em 4,95% e 3,59%, respectivamente, resultados que indicam a existência de diferenças estruturais produtivas entre as regiões do Brasil, que estão por trás da geração de oportunidades produtivas.

De modo geral, os modelos estimados neste trabalho revelam que as economias de aglomeração, apesar de influenciarem a distribuição dos salários no Brasil, têm papel modesto na composição dos diferenciais salariais. O efeito da densidade do emprego se reduz, consideravelmente, conforme se controlam as fontes de endogeneidade da relação com os salários. Isso indica, portanto, que as características dos trabalhadores e firmas, aliadas às particularidades e configurações produtivas da região, se mostram mais relevantes que a densidade do emprego na explicação dos diferenciais salariais, refletindo, aparentemente, melhores interações entre habilidades individuais e o desempenho da atividade produtiva local.

5. CONCLUSÃO

Este trabalho teve por objetivo principal identificar a contribuição dos efeitos de trabalhadores, firmas e regiões – com foco na densidade do emprego como medida de aglomeração – sobre o diferencial salarial nas regiões brasileiras no período de 2010 a 2014. A base de dados extraída da RAIS correspondeu a uma amostra com 3.946.910 observações, equivalendo a 789.382 observações anuais, para trabalhadores do sexo masculino do setor formal brasileiro. A partir desses dados, e baseando-se no modelo proposto por Combes, Duranton e Gobillon (2008), foram estimados modelos econométricos a partir das equações mincerianas dos salários. A principal estratégia de estimação adotada foi a regressão por meio de dados em painel, por permitir o acompanhamento do indivíduo no decorrer do tempo e o controle das heterogeneidades não observadas.

O método de estimação mais adequado foi aquele em que se controlou tanto as heterogeneidades não observadas quanto a endogeneidade (VI e EF). Os resultados para esse modelo mostraram que a contribuição marginal das economias de aglomeração para a determinação salarial foi de 0,404%, resultado bem inferior ao encontrado no modelo endógeno sem controle das heterogeneidades não observadas (POLS), que foi de 2,01%. Portanto, a endogeneidade e o *sorting* espacial das heterogeneidades não observadas podem superestimar as estimativas da densidade do emprego sobre os salários.

Em linhas gerais, conclui-se que o trabalho atingiu os objetivos almejados ao mensurar como as economias de aglomeração influenciam na existência dos diferenciais de salários, controlando o *sorting* espacial das heterogeneidades não observadas e tratando o problema da causalidade reversa entre a densidade do emprego e os salários, por meio do uso do *log* defasado da população como instrumento. Esta pesquisa, ao adotar uma metodologia robusta, respaldada pela literatura internacional e nacional, vem para acrescentar às evidências empíricas, ainda tão escassas sobre o tema, sobretudo na literatura nacional. Este estudo mostra que é preciso tratar o problema da endogeneidade e de controle do *sorting* para que as estimativas sejam consistentes, eficientes e não-viesadas. Por fim, a partir de todos os modelos estimados, é possível considerar que as economias de aglomeração influenciam positivamente os diferenciais de rendimentos entre os trabalhadores das regiões brasileiras.

REFERÊNCIAS

- ABEL, Jaison; DEITZ, Richard. Agglomeration and job matching among college graduates. **Federal Reserve Bank of New York Staff Reports**, New York, n. 587, 2012.
- BARUFI, Ana Maria Bonomi; HADDAD, Eduardo Amaral; NIJKAMP, Peter. Industrial scope of agglomeration economies in Brazil. **The Annals of Regional Science**, New York, v. 56, n. 3, p. 707-755, 2016.
- BERLINGERI, Matheus Mascioli. **Competências socioemocionais e mercado de trabalho: um estudo para o caso brasileiro**. 2018. Dissertação (Mestrado em Economia) – Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto, Universidade de São Paulo, São Paulo, 2018.
- CICCONI, Antonio; HALL, Robert Ernest. Productivity and the Density of Economic Activity. **The American Economic Review**, Saratoga Springs, v. 86, n. 1, p. 54-70, 1996.
- COELHO, Allexandro Mori; CORSEUIL, Carlos Henrique. **Diferenciais salariais no Brasil: um breve panorama**. Rio de Janeiro: Ipea, 2002.
- COMBES, Pierre-Philippe; DURANTON, Gilles; GOBILLON, Laurent. Spatial wage disparities: Sorting matters! **Journal of Urban Economics**, Amsterdam, v. 63, n. 2, p. 723-742, 2008.
- _____; GOBILLON, Laurent. Chapter 5 – The Empirics of Agglomeration Economies. In: DURANTON, Gilles; HENDERSON, Vernon; STRANGE, William (eds.). **Handbook of Regional and Urban Economics**. Amsterdam: Elsevier, 2015. Volume 5. p. 247-348.
- FREGUGLIA, Ricardo da Silva. **Efeitos da migração sobre os salários no Brasil**. 2007. Tese (Doutorado em Economia) – Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, São Paulo, 2007.
- GLAESER, Edward; MARÉ, David. Cities and Skills. **Journal of Labor Economics**, Chicago, v. 19, n. 2, p. 316-342, 2001.
- _____; _____. **Cities and Skills**. Working Paper n. 4728. Cambridge: NBER, 1994.
- KRUGMAN, Paul. Increasing Returns and Economic Geography. **Journal of Political Economy**, Chicago, v. 99, n. 3, p. 483-499, 1991.
- MARSHALL, Alfred. **Principles of Economics**. London: Macmillan, 1890.
- MATANO, Alessia; NATICCHIONI, Paolo. What Drives the Urban Wage Premium? Evidence Along the Wage Distribution. **Journal of Regional Science**, Hoboken, v. 56, n. 2, p. 191-209, 2016.

MELO, Patricia; GRAHAM, Daniel; NOLAND, Robert. A meta-analysis of estimates of urban agglomeration economies. **Regional Science and Urban Economics**, Amsterdam, v. 39, n. 3, p. 332-342, 2009.

MINCER, Jacob. **Schooling, Experience, and Earnings**. Human Behavior and Social Institutions No. 2. Cambridge: NBER, 1974.

PUGA, Diego. The Magnitude and Causes of Agglomeration Economies. **Journal of Regional Science**, Hoboken, v. 50, n. 1, p. 203-219, 2010.

ROCHA, Roberta de Moraes; SILVEIRA NETO, Raul da Mota; GOMES, Sônia Maria Fonseca Pereira Oliveira. Maiores Cidades, Maiores Habilidades Produtivas: Ganhos de Aglomeração ou Atração de Habilidade? Uma Análise para as Cidades Brasileiras. **Revista Econômica do Nordeste**, Fortaleza, v. 42, n. 4, p. 675-696, 2011.

ROSENTHAL, Stuart; STRANGE, William. Chapter 49 – Evidence on the Nature and Sources of Agglomeration Economies. In: HENDERSON, Vernon; THISSE, Jacques-François (eds.). **Handbook of Regional and Urban Economics**. Cities and Geography. Amsterdam: Elsevier, 2004. Volume 4. p. 2119-2171.

_____; _____. The attenuation of human capital spillovers. **Journal of Urban Economics**, Amsterdam, v. 64, n. 2, p. 373-389, 2008.

SILVA, Adelson Santos da. **Migração e Prêmio Salarial**: evidências para os trabalhadores do agreste pernambucano. 2013. Dissertação (Mestrado em Economia) – Universidade Federal de Pernambuco, Centro Acadêmico do Agreste, Caruaru, 2013.

SILVA, Diana Lúcia Gonzaga da. **Economias de aglomeração e heterogeneidade de trabalhador e firma na determinação de salários no Brasil**. 2017. Tese (Doutorado em Ciências) – Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, São Paulo, 2017.

_____; 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. **Pesquisa e Planejamento Econômico**, Rio de Janeiro, v. 46, n. 2, p. 7-46, 2016.

TOPEL, Robert H.; WARD, Michael P. Job Mobility and the Careers of Young Men. **The Quarterly Journal of Economics**, Ann Arbor, v. 107, n. 2, p. 439-479, 1992.

WOOLDRIDGE, Jeffrey Marc. **Econometric analysis of cross section and panel data**. 2. ed. Cambridge: MIT Press, 2011.

Recebido em: 31/08/2020

Aceito para publicação em: 08/09/2021