UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE
DEPARTAMENTO DE ECONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

**ECONOMIAS DE AGLOMERAÇÃO E HETEROGENEIDADE DE TRABALHADOR
E FIRMA NA DETERMINAÇÃO DE SALÁRIOS NO BRASIL**

Diana Lúcia Gonzaga da Silva

Orientador: Prof. Dr. Carlos Roberto Azzoni

SÃO PAULO

2017

Prof. Dr. Marco Antonio Zago
Reitor da Universidade de São Paulo

Prof. Dr. Adalberto Américo Fischmann
Diretor da Faculdade de Economia, Administração e Contabilidade

Prof. Dr. Hélio Nogueira da Cruz
Chefe do Departamento de Economia

Prof. Dr. Márcio Issao Nakane
Coordenador do Programa de Pós-Graduação em Economia

DIANA LÚCIA GONZAGA DA SILVA

**ECONOMIAS DE AGLOMERAÇÃO E HETEROGENEIDADE DE TRABALHADOR
E FIRMA NA DETERMINAÇÃO DE SALÁRIOS NO BRASIL**

Tese apresentada ao Programa de Pós-Graduação em Economia do Departamento de Economia da Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo, como requisito parcial para a obtenção do título de Doutor em Ciências.

Orientador: Prof. Dr. Carlos Roberto Azzoni

Versão Corrigida

(versão original disponível na Biblioteca da Faculdade de Economia, Administração e Contabilidade)

SÃO PAULO

2017

FICHA CATALOGRÁFICA

Elaborada pela Seção de Processamento Técnico do SBD/FEA/USP

Silva, Diana Lúcia Gonzaga da
Economias de aglomeração e heterogeneidade de trabalhador e
firma na determinação de salários no Brasil / Diana Lúcia Gonzaga
da Silva. – São Paulo, 2017.
160 p.

Tese(Doutorado) – Universidade de São Paulo, 2017.
Orientador: Carlos Roberto Azzoni.

1. Salários 2. Economias de aglomeração 3. Sorting 4. Heterogenei-
dade de firma 5. Heterogeneidade individual I. Universidade de São
Paulo. Faculdade de Economia, Administração e Contabilidade II. Título.

CDD – 331.2

AGRADECIMENTOS

Agradeço à minha família por toda a formação e incentivo, em especial aos meus pais, irmãs e sobrinha. Agradeço aos amigos, colegas, funcionários e professores da FEA-USP. Agradeço à FIPE pela parceria com os dados da RAIS e por tornar possível a realização deste trabalho com a disponibilização do suporte computacional necessário para lidar com os micro dados da RAIS. Agradeço ao Ministério do Trabalho e Emprego por autorizar o uso dos micro dados da RAIS neste trabalho. Agradeço à Universidade Federal da Bahia por toda a minha formação desde a Graduação até o Mestrado, bem como pelos amigos e parceiros que construí ao longo da trajetória nessa instituição, em especial ao meu orientador de Mestrado, Gervásio Santos.

Agradeço a todo o grupo do NEREUS, professores, alunos e secretária Adriana, em especial ao meu orientador Carlos Azzoni pela excelente orientação e por todo o apoio e incentivo no curso e nas questões pessoais e profissionais. Agradeço aos professores da FEA-USP, André Chagas, Naercio Menezes Filho, Eduardo Haddad, Paula Pereda e Renata Narita pelas valiosas contribuições para este trabalho de tese. Agradeço às queridas amigas Marta Rossi, Solange Ledi, Jaqueline Visentin e Juliana Scriptore, pelas conversas, contribuições, incentivos e momentos de descontração. Deixo um agradecimento especial à amiga Ana Barufi por todas as contribuições fundamentais para essa tese e pela disponibilidade de sempre e à Keyi Ussami por todas as orientações e suporte no ArcGis. Agradeço a Gleide Seixas por todo o incentivo, apoio e carinho. Agradeço ao querido companheiro de todos os momentos, e ao longo do Doutorado, Geidson Santana, por todos os incentivos e apoio, pelos conselhos e contribuições e pelo carinho de sempre.

Por fim, agradeço ao suporte financeiro da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), através da concessão de bolsa de Doutorado, sem a qual não seria possível prosseguir no curso. Esta tese foi desenvolvida com o suporte financeiro de bolsa da FAPESP, no âmbito do convênio FAPESP/CAPES, vinculada ao processo nº 2014/03077-7¹.

¹ O projeto aprovado apresenta o seguinte título: Economias de aglomeração e *sorting* na determinação espacial de salários dos trabalhadores do Brasil. As opiniões, hipóteses e conclusões ou recomendações expressas neste material são de responsabilidade dos autores e não necessariamente refletem a visão da FAPESP.

RESUMO

O objetivo desta tese é identificar a contribuição dos efeitos de aglomeração e do *sorting* espacial, associado às heterogeneidades não observadas dos trabalhadores e firmas, para a determinação dos salários individuais e dos salários locais nos arranjos populacionais do Brasil. Os dados do mercado de trabalho (RAIS-MTE) mostram que existe um diferencial espacial de salários, o qual pode ser explicado pelas distintas composições produtivas e de trabalhadores entre os locais e pelos diferenciais de custo de vida. A disponibilidade crescente de micro dados longitudinais permitiu a inclusão das habilidades não observadas individuais na equação de salários. Os estudos da Economia do Trabalho mostram que as habilidades são responsáveis por uma grande parcela dos diferenciais de salários. No entanto, os estudos nacionais ainda encontram um diferencial significativo, mesmo após o controle dos componentes individuais e do custo de vida, sugerindo a existência de efeitos específicos associados à localização das firmas e dos trabalhadores. A Economia Urbana considera as economias de aglomeração como um determinante salarial relevante nos mercados de trabalhos densos, particularmente a partir dos trabalhos de Glaeser e Maré (1994; 2001). Por sua vez, a maior produtividade das áreas densas pode ser atribuída à concentração de trabalhadores e firmas mais produtivos, o que ficou conhecido nessa literatura como *sorting*. Os estudos da Economia Urbana controlam somente o *sorting* dos atributos individuais não observados. Este trabalho contribui com a literatura ao considerar o *sorting* espacial dos atributos não observados das firmas e dos trabalhadores na determinação dos salários e dos efeitos de aglomeração. O estudo utiliza um modelo de decomposição salarial para lidar com múltiplos efeitos fixos no painel pareado de trabalhadores e firmas. Os efeitos puros da aglomeração (densidade) sobre os salários locais serão estimados em um modelo de dois estágios. O primeiro estágio estima uma equação salarial incluindo as características observadas dos trabalhadores e do emprego e os efeitos de localização, com um painel de micro dados da RAIS (2002-2014). O segundo estágio realiza a decomposição dos efeitos de localização em componentes associados às características locais dos arranjos e aos atributos não observados das firmas e dos trabalhadores. A estratégia de identificação propõe o controle dos efeitos fixos dos trabalhadores e firmas e o uso de variável instrumental para identificar os efeitos da aglomeração. Ademais, os dados de satélite sobre a luminosidade noturna são usados para estimar a proporção da área total dos arranjos habitada, a qual é utilizada para calcular a densidade. Os resultados mostraram que os efeitos do trabalhador foram mais relevantes do que os efeitos da firma para explicar a variação dos salários individuais e locais. O modelo principal, que utiliza o instrumento *Bartik* e a área iluminada, encontrou um efeito da densidade sobre os salários locais de 4,9%, o qual é superior ao *lower bound* da literatura prévia (3%). Os resultados sugerem que ignorar as limitações indicadas neste estudo pode levar a uma subestimação nas estimativas dos efeitos da densidade.

Palavras-chave: Salários. Economias de aglomeração. *Sorting*. Heterogeneidade de firma. Heterogeneidade individual.

ABSTRACT

The goal of this study is to identify the contribution of agglomeration effects and spatial sorting for the determination of individual and local wages in Brazilian urban agglomerations. Administrative records from the Ministry of Labor (RAIS-MTE) show a spatial differential in wages, which can be explained by the different productive structures and compositions of workers across cities, and by differentials in cost-of-living. The longitudinal microdata allowed the inclusion of unobserved individual skills in the wage equation. Studies in Labor Economics show that skills are responsible for a large portion of the wage differential. However, the available studies on Brazil still find a significant differential, even after controlling for the individual components and the cost of living. This suggests the existence of specific effects associated with the location of firms and workers. The literature on Urban Economics considers the economies of agglomeration as a relevant wage determinant in dense labor markets. The higher productivity of dense areas can be attributed to the concentration of more productive workers and firms more productive, which became known in this literature as sorting. Studies in Urban Economics only control the sorting of unobserved individual attributes. This dissertation contributes to the literature by considering the spatial sorting of unobserved attributes of firms and of workers in the determination of wages and of the effects of agglomeration. The study uses a wage decomposition model to deal with multiple fixed effects in a matched panel of workers and firms. The pure effects of agglomeration (density) on local wages are estimated in a two-stage model. The first stage estimates a wage equation including the observed characteristics of workers and firms and the effects of location, with a microdata panel of RAIS (2002-2014). The second stage decomposes the location effects into components associated to local characteristics and to unobserved attributes of firms and workers. The identification strategy involves controlling for fixed effects of workers and firms, and using an instrumental variable to identifying the effects of agglomeration. Satellite data on illumination are used to estimate the proportion of the overall area occupied with population and firms in each local labour markets. The results indicate that the worker effects are more relevant to explain wage variation than the firm's effects. The model of preference indicates a density effect on wages of 4.9%, much higher than the literature lower bound (3%). This suggests that ignoring the variables included in this study can lead to an underestimation of the effects of agglomeration.

Keywords: Wages. Economies of agglomeration. *Sorting*. Firm heterogeneity. Worker heterogeneity.

LISTA DE TABELAS

Tabela 3.1 – Arranjos Populacionais no Brasil.....	53
Tabela 3.2 – Decomposição da variação total na mobilidade de trabalhador e firma.....	58
Tabela 3.3 – Distribuição da área total e iluminada nos grandes arranjos: 2007 e 2013....	68
Tabela 4.1 – Estatísticas descritivas das variáveis do primeiro estágio (2002-2014).....	77
Tabela 4.2 – Distribuição dos trabalhadores por tamanho do arranjo populacional e setor e mobilidade dos trabalhadores e firmas.....	79
Tabela 4.3 – Estatísticas descritivas das variáveis do segundo estágio (2002–2014).....	80
Tabela 4.4 – Efeitos de localização nos grandes arranjos sobre o salário-hora real: primeiro estágio.....	84
Tabela 4.5 – Decomposição dos efeitos de localização dos arranjos: segundo estágio.....	93
Tabela 4.6 – Decomposição dos efeitos de localização e dos efeitos médios da firma e do trabalhador nos arranjos: segundo estágio.....	97
Tabela 4.7 – Decomposição dos salários reais médios dos arranjos: segundo estágio.....	98
Tabela 4.8 – Distribuição da área oficial e da área iluminada no Brasil: 2007 e 2013.....	99
Tabela 4.9 – Decomposição dos efeitos de localização dos arranjos: segundo estágio com a área iluminada (2013).....	101
Tabela 4.10 – Estatísticas descritivas da variável endógena e do instrumento <i>Bartik</i>	103
Tabela 4.11 – Decomposição dos efeitos de localização dos arranjos (segundo estágio): <i>Bartik</i> IV para a densidade do emprego.....	105
Tabela 4.12 – Decomposição dos efeitos de localização dos arranjos (segundo estágio): <i>Bartik</i> IV e densidade do emprego da área iluminada (2013).....	108
Tabela 4.13 – Estimativas dos efeitos de aglomeração (densidade do emprego).....	110

LISTA DE FIGURAS

Figura 4.1 – Salário real e densidade do emprego nos arranjos (2002–2014).....	81
Figura 4.2 – Distribuição espacial do salário-hora real (log), prêmio salarial urbano e efeitos fixos do trabalhador e da firma nos arranjos populacionais.....	87
Figura 4.3 – Efeito da localização sobre o salário-hora real dos indivíduos nos grandes arranjos populacionais.....	88
Figura 4.4 – Correlação entre os efeitos de localização dos grandes arranjos e os efeitos da firma e do trabalhador.....	94
Figura 4.5 – Emprego médio observado e estimado (<i>Bartik</i> IV) dos arranjos: 2002-2014.....	104

LISTA DE SIGLAS

AMC	Áreas Mínimas Comparáveis
ANP	Agência Nacional do Petróleo, Gás Natural e Biocombustíveis
BS	<i>Bootstrap</i>
CEI	Cadastro Específico do Instituto Nacional do Seguro Social (INSS)
CNAE	Classificação Nacional de Atividades Econômicas
CNPJ	Cadastro Nacional de Pessoas Jurídicas
COND	<i>Conditional Method</i>
CPS	<i>Current Population Survey</i>
DADS	<i>Déclarations Annuelles des Données Sociales</i>
DMSP-OLS	<i>Defense Meteorological Satellite Program - Operational Linescan System</i>
EUA	Estados Unidos da América
FE	<i>Fixed Effects</i>
FGLS	<i>Feasible Generalized Least Squares</i>
FGTS	Fundo de Garantia do Tempo de Serviço
FWL	Frisch-Waugh-Lovell
IBGE	Instituto Brasileiro de Geografia e Estatística
IPCA	Índice Nacional de Preços ao Consumidor Amplo
IPEA	Instituto de Pesquisa Econômica Aplicada
IV	<i>Instrumental Variable</i>
LMA	<i>Labor Market Area</i>
MQO	Mínimos Quadrados Ordinários
MTE	Ministério do Trabalho e Emprego
NCEI	<i>National Centers for Environmental Information</i>
NEG	<i>New Economic Geography</i>
NLSY	<i>National Longitudinal Survey of Youth</i>
OLS	<i>Ordinary Least Squares</i>
PIA	Pesquisa Industrial Anual
PIB	Produto Interno Bruto
PIS	Programa de Integração Social
PNAD	Pesquisa Nacional por Amostras de Domicílios
POLS	<i>Pooled Ordinary Least Squares</i>

PSID	<i>Panel Study of Income Dynamics</i>
RAIS	Relação Anual de Informações Sociais
RAIS-Migra	Relação Anual de Informações Sociais e Migração
RE	<i>Random Effects</i>
REGIC	Regiões de Influência das Cidades
RM	Região Metropolitana
USAF	<i>United States Air Force</i>
2SLS	<i>Two-Stage Least Squares</i>

SUMÁRIO

1	INTRODUÇÃO.....	11
2	ECONOMIAS DE AGLOMERAÇÃO URBANA, <i>SORTING</i> ESPACIAL DE TRABALHADOR E FIRMA E DETERMINAÇÃO SALARIAL.....	17
2.1	ECONOMIAS DE AGLOMERAÇÃO URBANA E MERCADO DE TRABALHO.....	17
2.2	<i>SORTING</i> ESPACIAL DE TRABALHADOR E FIRMA E DETERMINAÇÃO DE SALÁRIOS.....	24
2.2.1	Modelo de <i>Sorting</i> Espacial por Habilidades e Determinação Salarial.....	26
2.2.2	Evidências Empíricas de <i>Sorting</i> na Determinação de Salários com Heterogeneidade de Trabalhador e Firma.....	31
3	METODOLOGIA.....	39
3.1	MODELO DE DECOMPOSIÇÃO SALARIAL.....	39
3.2	BASE DE DADOS.....	47
3.3	MODELO ECONOMETRICO.....	54
3.4	MÉTODOS DE ESTIMAÇÃO E ESTRATÉGIA DE IDENTIFICAÇÃO.....	57
3.4.1	Estratégia de Identificação.....	57
3.4.1.1	Endogeneidade e Variável Instrumental.....	60
3.4.1.2	Densidade do Emprego e Área Iluminada.....	65
3.4.2	Métodos de Estimação Econométrica.....	71
4	RESULTADOS.....	77
4.1	CARACTERÍSTICAS DA AMOSTRA.....	77
4.2	PRÊMIO SALARIAL URBANO DOS ARRANJOS POPULACIONAIS.....	81
4.3	DECOMPOSIÇÃO DOS EFEITOS DE LOCALIZAÇÃO.....	92
4.3.1	Densidade do Emprego e Luminosidade Noturna.....	99
4.3.2	Densidade do Emprego e Variável Instrumental.....	102
4.4	CONSIDERAÇÕES FINAIS.....	109
5	CONCLUSÃO.....	111
	REFERÊNCIAS.....	117
	APÊNDICE.....	125

1 INTRODUÇÃO

A desigualdade salarial é um tema recorrente nos estudos em Economia e pode ser analisada a partir de diversas esferas, tais como a individual e a espacial. Na literatura internacional, o trabalho pioneiro de Mincer (1974) apresentou uma equação de determinação de salários, a partir dos componentes de escolaridade e experiência. No Brasil, diversos trabalhos buscaram compreender os determinantes e as disparidades salariais. A partir da equação minceriana, as pesquisas em Economia do Trabalho avançaram no controle de características observáveis e não observáveis dos trabalhadores. Entretanto, permanecia um diferencial salarial favorável aos grandes centros urbanos. Isso motivou o surgimento de estudos que integram Economia Urbana e Economia do Trabalho para explicar o diferencial salarial espacial, através das economias de aglomeração urbana. Nessa abordagem, a localização é incluída como um determinante salarial.

A distribuição espacial de salários no Brasil apresenta um comportamento considerado historicamente desigual². Os dados da Relação Anual de Informações Sociais (RAIS), do Ministério do Trabalho e Emprego (MTE), mostram que o estado da Paraíba apresentou o menor salário médio do Brasil em 2014, o qual correspondia a 65% do salário médio de São Paulo e 41% do salário de Brasília. Entre as regiões metropolitanas (RMs), o salário médio de João Pessoa correspondeu a 65% do salário médio da RM de São Paulo e 49% do salário da RM de Brasília. Por outro lado, os salários nas regiões metropolitanas são maiores do que as médias dos seus estados. As RMs de Florianópolis e Belo Horizonte apresentaram salários médios 38% e 29% maiores do que seus respectivos estados em 2014. Isso pode sugerir a existência de ganhos de aglomeração nos mercados de trabalho mais densos, se o diferencial de custo de vida não for contabilizado.

No entanto, o custo de vida mais elevado pode ser o resultado da maior concentração populacional e das atividades produtivas, que torna a disponibilidade de terra e de outros fatores mais escassa. Nesse sentido, Menezes e Azzoni (2006) encontraram que, embora relevante, o diferencial de custo de vida não foi individualmente capaz de explicar o diferencial de salários entre as regiões metropolitanas do Brasil, usando dados da Pesquisa Nacional por Amostras de Domicílios (PNAD) do Instituto Brasileiro de Geografia e

² Ver Langoni (2005).

Estatística (IBGE), para o período de 1981 a 2003. As características regionais e do capital humano foram mais relevantes para explicar o diferencial salarial entre as regiões metropolitanas.

As diferenças observadas de salários podem ocorrer por diversos fatores, dentre eles os diferenciais de custo de vida e de mobilidade, os diferenciais de produtividade, as dotações de amenidades locais, as segmentações regionais nos mercados de trabalho, tais como, formal, informal, setorial e por tamanho de empresas, dentre outros (SAVEDOFF, 1990; SERVO; AZZONI, 2002; SILVEIRA NETO; AZZONI, 2004; ROCHA; SILVEIRA NETO; GOMES, 2011). Azzoni (1999) mostrou que 40% do diferencial de renda pessoal e familiar, entre os estados do Brasil, era explicado pelo componente regional, usando dados dos Censos Demográficos de 1960-1991. A composição desigual de trabalhadores é um fator recorrente na explicação dos diferenciais salariais no Brasil. Silveira Neto e Azzoni (2004) indicaram que as características pessoais, ocupacionais e os custos de vida eram mais importantes do que as amenidades locais para o diferencial de renda entre áreas urbanas das regiões Nordeste e Sudeste em 2002. A educação apresentou a maior contribuição para as desigualdades salariais entre grandes áreas metropolitanas do Brasil, 17,6% e 31%, entre 1992 e 1997, em Servo e Azzoni (2002).

De acordo com Barros, Franco e Mendonça (2007), a desigualdade no rendimento do trabalho tem declinado no Brasil desde 1995, apresentando queda mais intensa entre 2001 e 2005. Menezes-Filho e Komatsu (2015) apontam que o crescimento real do salário mínimo, mais intenso nos anos 2000, contribuiu para a redução da dispersão salarial entre 2004 e 2014. Diversos estudos mostraram uma redução na desigualdade de renda do país³ a partir dos anos 2000, apontando a educação como a principal determinante. Barros, Henriques e Mendonça (2000) mostraram que a educação é responsável por 40% dos diferenciais salariais no Brasil. O impacto de um ano adicional de educação sobre os salários chega a 16%, de acordo com o estudo.

A redução na desigualdade educacional dos trabalhadores brasileiros contribuiu em 17% para a redução da desigualdade salarial, entre 2001 e 2005. No mesmo período, houve uma intensificação na queda dos diferenciais de remuneração por nível educacional, que vinha

³ Ver Barros, Henriques e Mendonça (2000) e Menezes-Filho, Fernandes e Picchetti (2007).

ocorrendo desde 1995 (BARROS; FRANCO; MENDONÇA, 2007). Menezes-Filho, Fernandes e Picchetti (2007) mostraram que a expansão do ensino médio, associada à queda nos retornos à educação, reduziu a desigualdade salarial. Apesar das evidências da influência da educação (ou habilidade observada), o controle das características individuais observadas não elimina as desigualdades salariais.

Estudos mais recentes em Economia do Trabalho consideram a presença de habilidades não observadas dos trabalhadores na diferenciação salarial. Freguglia (2007; 2009) destacou que, embora a literatura *mainstream* aponte para o diferencial salarial espacial como o principal estímulo à migração, existe uma auto seleção entre os trabalhadores migrantes. O controle das heterogeneidades individuais não observadas na equação salarial pode reduzir ou até eliminar grande parte da vantagem salarial aparente associada às áreas urbanas densas. Considerar as habilidades não observadas possibilita corrigir o viés de seleção associado à atração de trabalhadores mais habilidosos para os grandes centros urbanos.

Combes, Duranton e Gobillon (2008) mostraram que as diferenças na composição de habilidades foram responsáveis por 40 a 50% das disparidades espaciais de salários de trabalhadores da França, sugerindo evidências de *sorting* por habilidades. Combes et al. (2012b) mostraram uma seleção negativa por habilidades na migração de trabalhadores, para áreas menos densas, e positiva, para as mais densas da França. No Brasil, Freguglia e Menezes-Filho (2012) mostraram que, entre 1995 e 2002, 63% do diferencial salarial dos estados era explicado pelas características não observadas dos trabalhadores. Contudo, persistiu um diferencial salarial, mesmo após considerar o custo de vida, as habilidades observadas e não observadas dos trabalhadores e suas características ocupacionais. A persistência do diferencial salarial pode estar associada aos efeitos específicos de localização sobre os salários dos trabalhadores, o que sugere a necessidade de compreender os diferenciais salariais a partir das aglomerações urbanas.

Em um estudo recente, o IBGE (2015) mostrou que, aproximadamente 56% da população do Brasil residia em arranjos populacionais, definidos a partir da intensidade dos fluxos para trabalho e estudo, com base nos dados do Censo Populacional de 2010. Por sua vez, os arranjos concentravam em torno de 82% do emprego formal em 2010, sendo 57% somente nos grandes arranjos (população acima de 750 mil). De acordo com os dados da RAIS, os

salários médios nos arranjos populacionais de dezembro de 2014, eram aproximadamente 20% maiores do que nos municípios que não pertenciam a algum arranjo e 40% maiores nos grandes arranjos. Esses dados conjuntamente reforçam os indícios da existência de fatores específicos locais em áreas urbanas densas, que dinamizam os mercados de trabalho locais e elevam a produtividade nessas áreas.

A literatura sobre economias de aglomeração prevê maiores salários e produtividade em áreas urbanas densas, decorrentes das externalidades positivas dessas áreas. Glaeser e Maré (1994; 2001) identificaram um prêmio salarial urbano de 32 a 33% para os trabalhadores de áreas metropolitanas dos Estados Unidos (EUA), entre 1968 e 1993. A partir desses estudos, surgiu uma abordagem na Economia Urbana que incorpora os efeitos de aglomeração na determinação salarial. Os efeitos da densidade urbana sobre a produtividade local variam entre 4% e 8% na literatura internacional. No Brasil, já existem evidências de ganhos no mercado de trabalho associados à aglomeração urbana. Rocha, Silveira Neto e Gomes (2011) mostraram que os trabalhadores de regiões metropolitanas apresentaram um ganho salarial de 9,4%, após controle por características observáveis e não observáveis, entre 2000 e 2008.

Por outro lado, os ganhos associados às características de trabalhadores e firmas que escolhem se localizar nas áreas densas podem ser confundidos com ganhos de aglomeração. Se a heterogeneidade das firmas e trabalhadores dos grandes centros urbanos não é considerada, deve existir um viés nas estimativas dos efeitos de localização, associado ao *sorting* espacial na determinação de salários. O *sorting* por habilidades surge se trabalhadores com melhores atributos tendem a se aglomerar em mercados de trabalho locais densos e qualificados. A seleção de firma pode surgir se apenas as mais produtivas sobrevivem nos grandes centros urbanos. No entanto, até o momento não foi encontrado algum estudo no Brasil que tenha considerado os efeitos de firma na determinação dos efeitos de localização sobre os salários.

Os estudos no Brasil controlam somente uma parte do *sorting* espacial salarial, associado às heterogeneidades não observadas dos trabalhadores. Entretanto, o controle da seleção de trabalhadores mais produtivos para as firmas que pagam maiores salários requer a disponibilidade de dados longitudinais pareados (*matched employer-employee data*) de trabalhador e firma. A limitação empírica para estimar modelos com heterogeneidade não observada de trabalhador e firma, usando o método de Mínimos Quadrados Ordinários (MQO), surge da complexidade computacional para incluir uma grande dimensão de

identificadores nas estimações econométricas com dados longitudinais. Abowd, Kramarz e Margolis (1999) desenvolveram abordagens alternativas para estimar os parâmetros de equações salariais, dada a limitação computacional da solução de MQO. Desde o estudo de Abowd, Kramarz e Margolis (1999), a literatura internacional em Economia do Trabalho tem buscado soluções estruturais e estatísticas factíveis computacionalmente para estimar os parâmetros de interesse, em modelos que lidam com múltiplos efeitos fixos (ABOWD; KRAMARZ; MARGOLIS, 1999; ABOWD; CREECY; KRAMARZ, 2002; ANDREWS; SCHANK; UPWARD, 2006; GUIMARÃES; PORTUGAL, 2010; McCAFFREY, 2012; MITTAG, 2012; GAURE, 2013).

Este trabalho tem como justificativa contribuir para a literatura internacional e nacional sobre os determinantes salariais espaciais. Em relação à literatura internacional, contribui com a inclusão de um grupo mais amplo de controles, a exemplo dos efeitos de firma, e com a identificação de evidências para um país em desenvolvimento, que possui expressiva heterogeneidade entre os mercados de trabalho locais. Além disso, o estudo sugere uma medida mais adequada para calcular a área e a respectiva densidade do emprego. Nesse sentido, a área com luminosidade noturna será utilizada para mensurar a área com efetiva ocupação humana e corrigir a subestimação dos efeitos de aglomeração. Por outro lado, o estudo também contribui com o uso do instrumento *Bartik*, tradicionalmente utilizado nos estudos empíricos da Economia do Trabalho, para lidar com a endogeneidade da densidade e identificar o efeito exógeno da aglomeração. Para a literatura nacional, o avanço corresponde ao controle simultâneo do *sorting* espacial da firma e do trabalhador na determinação de salários dos indivíduos e na identificação dos efeitos de aglomeração. Ademais, o trabalho utilizará a definição recente do IBGE (2015) de arranjos populacionais, a qual se mostra mais adequada para a identificação dos mercados de trabalho locais aglomerados, tendo em vista a intensidade dos fluxos para trabalho e estudo e as características de urbanização adotadas na definição dessas categorias.

O objetivo da tese é identificar a contribuição dos efeitos de aglomeração e do *sorting* espacial de trabalhador e firma para a determinação de salários individuais e para o diferencial salarial entre os arranjos populacionais do Brasil. O estudo propõe identificar a contribuição do *sorting* espacial a partir do controle dos efeitos de firma e trabalhador, recorrendo aos avanços metodológicos da literatura internacional. Essa pesquisa propõe o uso de um painel

de micro dados combinados de trabalhador e firma, extraídos da RAIS (MTE), para estimar as equações de salário com as características observadas dos trabalhadores e do emprego, a heterogeneidade não observada dos trabalhadores e firmas e os efeitos de localização em arranjos populacionais, para o período entre 2002 e 2014. A estratégia de identificação proposta nesta tese propõe lidar com o *sorting* das heterogeneidades não observadas que resulta da atração e concentração dos trabalhadores e firmas mais produtivos nos mercados de trabalho locais densos. Ao mesmo tempo, o estudo utiliza uma variável instrumental para lidar com a endogeneidade da densidade na determinação dos salários locais.

Além desta introdução, este trabalho de tese é composto por mais quatro capítulos. O segundo capítulo apresenta uma revisão da literatura de economias de aglomeração urbana, *sorting* espacial e determinação de salários. O terceiro capítulo apresenta a metodologia da pesquisa, incluindo o modelo de decomposição estatística de salários, o banco de dados, a estratégia de identificação e os métodos de estimação econométrica utilizados. O capítulo quatro analisa as estatísticas descritivas da amostra utilizada e os resultados encontrados neste estudo. Por fim, o capítulo cinco apresenta as conclusões do trabalho de tese.

2 ECONOMIAS DE AGLOMERAÇÃO URBANA, *SORTING* ESPACIAL DE TRABALHADOR E FIRMA E DETERMINAÇÃO SALARIAL

Este capítulo apresenta uma revisão da literatura sobre as economias de aglomeração urbana e o *sorting* espacial no mercado de trabalho. Além disso, o capítulo propõe apresentar as principais evidências encontradas na literatura empírica nacional e internacional.

2.1 ECONOMIAS DE AGLOMERAÇÃO URBANA E MERCADO DE TRABALHO

Com o surgimento da Nova Geografia Econômica (*New Economic Geography*-NEG), a partir do modelo Centro-Periferia de Krugman (1991), o espaço passou a ser incorporado formalmente na teoria econômica, embora já houvesse extensa literatura na área da Ciência Regional sobre o tema. Os teóricos tradicionais da Economia Urbana, como Von Thünen, já haviam introduzido os elementos iniciais para discutir a importância da localização na economia (geografia econômica). No entanto, dadas as limitações de modelagem para a introdução das dimensões espaciais, muitos destes estudos simplesmente assumiam a existência das cidades como dada. A NEG busca explicar os mecanismos econômicos que geram os padrões de concentração e dispersão espacial. Os principais elementos explicativos dessa desigualdade são os custos de transporte, os retornos crescentes e a concorrência imperfeita. As economias de aglomeração urbana emergem nesse contexto como fatores de reforço da concentração espacial (FUJITA; KRUGMAN; VENABLES, 2002).

As pesquisas sobre economias de aglomeração iniciaram a partir dos estudos sobre a existência das cidades e a formação de aglomerações urbanas. A aglomeração é um agrupamento de atividades econômicas e de pessoas, gerado e sustentando por uma lógica circular (FUJITA; KRUGMAN; VENABLES, 2002). As bases da Economia Espacial e das economias de aglomeração urbana são encontradas em trabalhos seminais, como: Dixit e Stiglitz (1977), Krugman (1979), Krugman (1991), Krugman (1998), Fujita, Krugman e Venables (1999) e Fujita e Thisse (2002).

A literatura sobre as economias de aglomeração prevê maiores salários e produtividade em áreas urbanas densas. Os ganhos de aglomeração no mercado de trabalho podem emergir do aumento da produtividade dos trabalhadores na presença de externalidades positivas. São

exemplos destas externalidades o acesso a maior diversidade de ocupações, o acúmulo de capital humano e *spillovers* de conhecimento, as interações, melhor tecnologia das firmas, trabalhos mais produtivos e outros. As economias de aglomeração ou economias externas de escala são as forças que levam à concentração das indústrias em *clusters* e das atividades produtivas nas cidades. A abordagem marshalliana clássica sugere três principais fontes das economias de aglomeração. A primeira é o *sharing* (compartilhamento) de insumos cuja produção envolve retornos internos crescentes a escala. A segunda é o *pooling* (agrupamento) no mercado de trabalho, onde a aglomeração permite um melhor *matching* (associação) entre a habilidade do trabalhador e a demanda da firma. A última fonte é o *spillover* de conhecimentos que ocorre quando uma indústria é localizada, permitindo a troca de aprendizado entre os trabalhadores (ROSENTHAL; STRANGE, 2004; COMBES; DURANTON; GOBILLON, 2011).

Os trabalhos mais recentes sugerem outras fontes para a existência das economias de aglomeração. O efeito do mercado doméstico surge da concentração da demanda, que estimula a aglomeração, enquanto o efeito das economias de consumo emerge das preferências pela cidade (ROSENTHAL; STRANGE, 2004). De modo similar à abordagem clássica, a literatura recente da Economia Urbana identifica as causas micro-fundamentadas das economias de aglomeração. A primeira causa considera que um mercado mais amplo permite o compartilhamento (*sharing*) de infraestrutura, de fornecedores e de um *pool* de trabalhadores com habilidades similares. A segunda assegura que um grande mercado permite melhor *matching* entre trabalhadores e firmas ou entre compradores e fornecedores. Por último, a terceira indica que um mercado maior pode facilitar a aprendizagem (*learning*) e o desenvolvimento e disseminação de novas tecnologias e práticas empresariais (PUGA, 2010).

Os estudos de Glaeser e Maré (1994; 2001) identificaram uma relação positiva entre os salários e o tamanho das cidades, denominada de prêmio salarial urbano. Os trabalhos encontraram evidências de salários 32 a 33% maiores para trabalhadores de áreas metropolitanas dos EUA, entre 1968 e 1993. A partir dos estudos de Glaeser (1994; 2001), emergiu um ramo de pesquisa que integra a Economia do Trabalho e a Economia Urbana, ao incorporar as economias de aglomeração urbana como fatores relevantes para a determinação salarial. Os maiores salários recebidos pelos trabalhadores nas grandes cidades representam um prêmio urbano. Na literatura internacional, esse prêmio gera um aumento na magnitude de 5 a 10% nos salários reais, a partir da estimação de equações mincerianas (HEUERMANN;

HALFDANARSON; SUEDEKUM, 2010). O benefício da proximidade é um suposto subjacente ao prêmio salarial urbano. Quando possível, é descontado do prêmio o viés de seleção associado à heterogeneidade individual ou espacial não observada.

Alguns estudos da Economia Urbana buscam possíveis interpretações para o prêmio salarial urbano. A primeira interpretação sugere que o prêmio representa um viés de habilidades omitidas, já que os trabalhadores mais habilidosos apresentam maior produtividade e estão concentrados nas grandes cidades. A segunda sugere que o prêmio pode ser interpretado como um efeito em nível, já que as cidades geram um aumento na produtividade do trabalhador, que eleva os salários imediatamente através das externalidades urbanas. Por último, o prêmio pode representar um efeito de crescimento salarial acelerado nas cidades, ao longo do tempo, em decorrência do maior aprendizado e aquisição de habilidades ou do melhor *matching* nos empregos das cidades (GLAESER; MARÉ, 1994; GLAESER; MARÉ, 2001; YANKOW, 2006; HEUERMAN; HALFDANARSON; SUEDEKUM, 2010).

As pesquisas recentes sobre aglomeração e mercado de trabalho consideram a importância das habilidades ou heterogeneidades individuais não observadas. Embora algumas abordagens definam a habilidade pelo nível de educação do trabalhador, existem aspectos da habilidade que podem ser desconsiderados ao assumir essa igualdade. A habilidade pode ser definida a partir das manifestações comportamentais dos traços individuais e das inteligências. Os traços são as características pessoais estáveis, tais como o temperamento e a personalidade, que são determinados pelas dotações genéticas e pelo ambiente a qual o indivíduo foi exposto na fase inicial de desenvolvimento. As inteligências são definidas pela capacidade de processar conteúdos do mundo, incluindo o processamento de palavras e relações lógicas (inteligência verbal *vs.* proposicional) e a capacidade de organizar e mover objetos no espaço (inteligência espacial). As inteligências também são geradas por dotação genética ou pelo ambiente, e podem ser influenciadas pelos traços (BACOLOD; BLUM; STRANGE, 2010).

Os traços individuais e a inteligência interagem com o ambiente para gerar habilidades. A educação e a aglomeração urbana contribuem para o desenvolvimento e manifestação das habilidades. A aglomeração impacta no desenvolvimento das habilidades e na sua utilização na produção. O pensamento criativo e o planejamento flexível são exemplos de habilidades produtivas. Portanto, a aglomeração impacta a determinação de salários indiretamente através

do desenvolvimento de habilidades e diretamente por influenciar a aplicação das habilidades no mercado de trabalho (BACOLOD; BLUM; STRANGE, 2010).

Os estudos empíricos buscam estimar as economias de aglomeração com regressões que captam os efeitos de medidas de concentração setorial ou urbana sobre as medidas de produtividade (GRAHAM et al., 2010). Os possíveis escopos ou dimensões de identificação das economias de aglomeração são: industrial; temporal; geográfico; e organizacional ou de competição. A estratégia de estimação direta do efeito da aglomeração sobre a produtividade utiliza uma função de produção. As estratégias indiretas estimam a influência da aglomeração sobre o crescimento do emprego, o nascimento de novas firmas, os salários e os preços dos imóveis. A discussão existente nos estudos empíricos que tratam do escopo industrial é se as economias de aglomeração são relacionadas à concentração de uma indústria ou ao tamanho de uma cidade, independente da competição. A primeira explicação é denominada de economias de localização e a segunda é conhecida como economias de urbanização. Os estudos de Marshall (1920) tratavam das vantagens localizadas da concentração da indústria e especialização urbana, enquanto Jacobs (1969) sugeria a importância da diversidade urbana para a geração e difusão de ideias. Portanto, o debate do escopo industrial da aglomeração ficou caracterizado como Marshall *versus* Jacobs (ROSENTHAL; STRANGE, 2004).

Os estudos empíricos encontram ganhos de produtividade oriundos da aglomeração urbana, mas há muita variação na magnitude das estimativas encontradas. As evidências dos estudos empíricos sobre o impacto das economias de urbanização mostram que a produtividade aumenta entre 3% e 8% ao dobrar o tamanho da cidade (ROSENTHAL; STRANGE, 2004; MELO; GRAHAM; NOLAND, 2009). No entanto, Graham et al. (2010) destacaram a dificuldade de identificação dos efeitos de aglomeração, particularmente devido à potencial causalidade reversa dos efeitos de aglomeração sobre a produtividade. Nesse sentido, muitos estudos nessa literatura utilizam variáveis instrumentais para identificar os efeitos de aglomeração.

A abordagem mais comum é utilizar longas defasagens temporais da população ou densidade como instrumento para os níveis atuais de aglomeração. A explicação é que a densidade observada atualmente é determinada pelos padrões anteriores de concentração da população e essa, por sua vez, não apresenta correlação com a produtividade contemporânea. Por outro lado, características geográficas e geológicas contemporâneas também já foram utilizadas

como instrumentos em alguns estudos. No entanto, o efeito da densidade após a correção da endogeneidade não é claro. Alguns estudos encontram um efeito menor ou maior do que o de MQO, enquanto outros estudos mostram que a endogeneidade não gera um viés substancial nas estimativas do efeito de aglomeração (MELO; GRAHAM; NOLAND, 2009; COMBES; DURANTON; GOBILLON, 2011). De acordo com Combes, Duranton, Gobillon (2011), houve um significativo progresso na estimação das economias de aglomeração nos últimos anos, de modo que as estimativas mais atuais da elasticidade dos salários em relação à densidade variam entre 0,02 e 0,05.

A literatura sobre as economias de aglomeração urbana tem apresentado substancial progresso nos anos recentes, com a crescente disponibilidade de dados, cujo nível de desagregação permite investigar questões que não poderiam ser respondidas anteriormente (ROSENTHAL; STRANGE, 2004). A literatura internacional em Economia Urbana que trata da existência das economias de aglomeração e seus efeitos no mercado de trabalho pode ser encontrada, dentre outros trabalhos, em: Rauch (1993), Glaeser e Maré (1994 e 2001), Glaeser e Resseger (2010), Rosenthal e Strange (2004; 2008), Yankow (2006) e Andersson, Burgess e Lane (2007). A identificação das economias de aglomeração no mercado de trabalho do Brasil pode ser encontrada nos estudos de Campos e Silveira Neto (2009), Falcão e Silveira Neto (2007), Rocha, Silveira Neto e Gomes (2011), Cruz e Naticchioni (2012), Barufi (2015), Barufi, Haddad e Nijkamp (2016) e Silva, Santos e Freguglia (2016).

Melo, Graham e Noland (2009) realizaram uma meta-análise das estimativas das economias de aglomeração urbana encontradas em diferentes estudos empíricos da literatura. A revisão foi baseada em estudos que estimaram a elasticidade das economias de aglomeração usando uma função de produção ou um modelo salarial, cobrindo o período entre 1965 e 2002. A análise mostrou que, embora as evidências dos ganhos de produtividade da aglomeração urbana sejam geralmente positivas, as características de cada estudo importam para o resultado encontrado. Nesse sentido, os efeitos específicos dos países, o escopo setorial, a especificação das economias de aglomeração e a presença de controles para a heterogeneidade não observada e para as diferenças na qualidade do trabalhador podem dar origem a grandes diferenças nos resultados. Por sua vez, a correção da causalidade reversa da aglomeração não parece gerar mudanças significantes nas estimativas.

Diversos estudos na literatura internacional encontraram evidências de prêmio salarial urbano. Glaeser e Maré (1994; 2001) encontraram um prêmio de 32% a 33% para os trabalhadores de áreas metropolitanas dos EUA, entre 1968-1993 (PSID, CPS, NLSY)⁴, sem controlar as habilidades individuais não observadas e de 3% a 4% quando controlou essas heterogeneidades. Yankow (2006) identificou um prêmio salarial urbano de 19% para os trabalhadores de grandes áreas urbanas dos EUA, entre 1979-1994 (NLSY), e de 6% após o controle dos efeitos fixos individuais.

Embora essa literatura seja relativamente recente, já existem evidências empíricas da existência de ganhos de aglomeração urbana no mercado de trabalho do Brasil. Rocha, Silveira Neto e Gomes (2011) encontraram evidências de que os trabalhadores de regiões metropolitanas possuíam um diferencial de 0,12 salário mínimo ou 9,4%, após controle por características observáveis e não observáveis, com a RAIS-Migra no período de 2000 a 2008. Os trabalhadores das áreas densas apresentaram um diferencial de 0,14 salário mínimo. Campos e Silveira Neto (2009) encontraram um ganho salarial de até 16% para os trabalhadores de regiões metropolitanas, em 2000 (Censo-IBGE), em relação aos residentes em aglomerados urbanos não metropolitanos. Estes ganhos foram reduzidos a 9,5% com a inclusão do tamanho da população.

Falcão e Silveira Neto (2007) encontraram uma externalidade positiva da aglomeração, associada à concentração de trabalhadores mais qualificados nos municípios com população acima de 100 mil habitantes, variando entre 0,9% e 12%, com micro dados dos Censos de 1991 e 2000. Cruz e Naticchioni (2012) identificaram um prêmio salarial urbano de 22,8% para os trabalhadores de áreas metropolitanas, em 2002 (PNAD), o qual declinou para 17,4% em 2009. Silva, Santos e Freguglia (2016) encontraram um prêmio salarial urbano de 3,4% nas regiões metropolitanas, a partir da estimação de uma equação de salários com as características observadas e não observadas dos trabalhadores, utilizando os micro dados da RAIS, para o período de 1995 a 2008. O prêmio urbano foi de 22,4% sem o controle dos efeitos individuais e variou de 16,5% (RM de Macapá) para 1,9% (RM da Baixa Santista), com o controle desses efeitos.

⁴ *Panel Study of Income Dynamics-PSID; Current Population Survey-CPS; National Longitudinal Survey of Youth-NLSY.*

Barufi (2015) e Barufi, Haddad e Nijkamp (2016) estimaram as economias de urbanização entre os setores produtivos, com micro dados da RAIS entre os anos de 2004, 2008 e 2012, e encontraram efeitos da densidade do emprego sobre os salários locais variando entre 0,05 e 0,09. Os efeitos de aglomeração foram mais intensos nas indústrias de alta e baixa tecnologia, seguidos pelos serviços de alta intensidade de conhecimento. O estudo incluiu defasagens temporais da população, a distância da costa litorânea e indicadores da produção de cana de açúcar, café e extração de ouro, no período colonial, como variáveis instrumentais para a densidade do emprego nos mercados de trabalho locais. A modelo de variável instrumental encontrou efeitos da aglomeração variando entre 0,03 e 0,09. As evidências empíricas reforçam o papel dos atributos locais na determinação de salários das áreas urbanas do Brasil.

Por sua vez, existem estudos aplicados ao Brasil que utilizam modelos da NEG para captar possíveis efeitos de aglomeração. Fally, Paillacar e Terra (2010) estimaram um modelo da NEG para encontrar o efeito do acesso ao mercado sobre os salários, cujo mercado potencial foi mensurado a partir de um modelo gravitacional para o comércio nacional e internacional. O estudo encontrou um efeito de 0,14 do acesso ao mercado sobre os salários entre estados e setores, em 1999. Do mesmo modo, Amaral et al. (2010) utilizou um modelo da NEG para estimar o efeito do mercado potencial, mensurado pela renda de todas as fontes, sobre os salários municipais, com os Censos de 1980, 1991 e 2000. O estudo utilizou modelagem de dados em painel com erros espacialmente correlacionados e variável instrumental para o mercado potencial baseada na latitude dos municípios. Os resultados mostraram que quanto maior o mercado potencial maior é o salário do município. O aumento previsto no salário foi de 0,35% para cada 1% de aumento no mercado potencial.

Portanto, as evidências na literatura de Economia Espacial (Regional e Urbana) mostram que existe uma relação positiva entre densidade urbana e salários. Isso reforça a busca da explicação dos diferenciais salariais persistentes no Brasil, a partir das economias de aglomeração urbana. No entanto, a presença de externalidades específicas locais pode garantir retornos salariais diferenciados espacialmente para habilidades similares. Ao mesmo tempo, se a distribuição de trabalhadores e firmas mais produtivos estiver concentrada em grandes centros urbanos, a vantagem salarial associada à localização pode estar sobrestimada. Neste sentido, surgem novas abordagens na literatura internacional de aglomeração urbana e mercado de trabalho buscando separar os efeitos de aglomeração daqueles decorrentes dos

atributos de trabalhadores e firmas, para a correta identificação das externalidades salariais positivas das aglomerações urbanas.

2.2 *SORTING* ESPACIAL DE TRABALHADOR E FIRMA E DETERMINAÇÃO DE SALÁRIOS

Os retornos salariais associados às características de trabalhadores e firmas que escolhem se localizar nas áreas urbanas densas podem ser confundidos com os ganhos de aglomeração. Se as escolhas de localização não são exógenas, os atributos de firmas e trabalhadores localizados em áreas densas podem enviesar as estimativas do prêmio salarial urbano, se não forem considerados. A maior produtividade em áreas urbanas densas pode ser atribuída à concentração de trabalhadores mais habilidosos e firmas mais produtivas, um fenômeno conhecido na literatura de Economia Urbana como *sorting* espacial nos mercados de trabalho locais. A seleção espacial por características observadas e não observadas de trabalhadores e firmas pode ser atribuída à existência de um *sorting* entre localizações.

A maior produtividade das cidades pode ser explicada, em parte, pelo mecanismo de auto seleção gerado pelo ambiente urbano. A auto seleção melhora a qualidade dos *matches* nessas áreas. O mecanismo pode ser reforçado pelo desenvolvimento de redes informacionais nas cidades que apresentam uma grande parcela de trabalhadores de alta habilidade. Esse processo pode gerar um aumento na produtividade das cidades para os trabalhadores de todos os tipos de habilidades. Por sua vez, os trabalhadores de baixa habilidade podem permanecer afastados dos *matches*, particularmente se as cidades fortalecem as parcerias entre os trabalhadores de habilidades similares. Além disso, as diferenças de custo de vida entre as cidades pode induzir à auto seleção entre diferentes tipos de habilidades, de modo que os trabalhadores de alta habilidade escolhem viver nas cidades mais caras, enquanto os de baixa habilidade permanecem nas cidades de menor custo (VENABLES, 2011).

O *sorting* por habilidades pode surgir se os trabalhadores com os melhores atributos produtivos tendem a permanecer espacialmente aglomerados nos mercados de trabalho mais densos e qualificados. Bacolod, Blum e Strange (2010) encontraram uma relação positiva da educação e da urbanização com as habilidades pessoais e cognitivas e uma relação negativa da urbanização com as habilidades motoras, usando dados dos EUA para o período de 1982 a 1998. Os autores sugerem que as grandes cidades permitem desenvolver melhores habilidades

cognitivas e pessoais, enquanto as pequenas reforçam as habilidades motoras. O estudo também encontrou fortes evidências de *sorting* das habilidades pessoais e cognitivas, onde os trabalhadores mais habilidosos escolhiam desproporcionalmente as grandes cidades.

Combes et al. (2012b) sugeriram a existência de quatro fatores explicando as diferenças de salários entre áreas: as diferenças nas distribuições iniciais de habilidades dos trabalhadores; o *sorting* por habilidades dos trabalhadores; os benefícios da aglomeração, que aumentam os salários dos trabalhadores em qualquer nível de habilidade; e a atuação diferenciada destes benefícios de aglomeração sobre os distintos níveis de habilidades. Se os trabalhadores mais habilidosos se beneficiam mais da aglomeração, o efeito fixo de trabalhador pode superestimar as habilidades não observadas deste trabalhador. A seleção resulta das escolhas de localização dos trabalhadores, a depender de suas habilidades, enquanto o *sorting* representa as diferentes distribuições espaciais de habilidades, ou seja, o *sorting* envolve a atração de trabalhadores habilidosos (seleção) e a manutenção dos mais hábeis locais. A seleção é uma condição suficiente para o *sorting*, já que o mesmo também ocorre através de outros canais, tais como as habilidades inatas ou adquiridas localmente. Nesse sentido, as diferenças de habilidade podem ser explicadas pelas diferenças a priori na formação de habilidades, antes da entrada no mercado de trabalho, tais como a melhor educação familiar e formal nas áreas densas, ou pela intensidade do *learning* nas cidades.

De acordo com Combes et al. (2012a), as firmas são, em média, mais produtivas em grandes cidades. A seleção de firmas e as economias de aglomeração foram apontadas como duas explicações possíveis para essa evidência. A seleção de firmas pode surgir se os grandes centros urbanos fortalecem a competição fazendo com que apenas as mais produtivas sobrevivam. Por sua vez, a abordagem das economias de aglomeração mostra que as grandes cidades promovem interações que aumentam a produtividade, além das vantagens naturais localizadas. Por outro lado, Combes, Duranton e Gobillon (2008) destacaram que a identificação de estimativas não enviesadas na equação salarial requer a inclusão dos efeitos fixos de firmas, desde que estes podem estar correlacionados com os efeitos de setores. No entanto, o estudo não realizou este controle.

Combes, Duranton e Gobillon (2008) e Combes et al. (2012b) desenvolveram uma abordagem unificada dos determinantes salariais, incluindo os efeitos de localização, de setores e as

características observadas e não observadas dos trabalhadores. No entanto, por limitações de dados, não adotaram controles suficientes para as características observadas dos trabalhadores, tais como a educação. Usando os micro dados da França, Combes, Duranton e Gobillon (2008) mostraram que as diferenças na composição de habilidades dos trabalhadores foram responsáveis por 40 a 50% das disparidades espaciais de salários, sugerindo evidências de *sorting* espacial por habilidades. O estudo de Combes et al. (2012b) mostrou uma seleção negativa por habilidades na migração de trabalhadores, para áreas menos densas, e positiva, para as mais densas da França. No entanto, a migração explicou somente uma pequena parcela das diferenças de habilidade entre as áreas. A seção a seguir apresenta o modelo de determinação salarial e *sorting* espacial individual que será tomado como referência principal neste trabalho de tese.

2.2.1 Modelo de *Sorting* Espacial por Habilidades e Determinação Salarial

Na literatura de Economia Urbana aplicada aos mercados de trabalhos locais, o modelo teórico desenvolvido no estudo de Combes, Duranton e Gobillon (2008) tornou-se uma referência para analisar o *sorting* espacial salarial. No modelo, as disparidades salariais resultam das diferenças espaciais na composição de habilidades dos trabalhadores, nas dotações não humanas locais e nas interações produtivas locais.

A ocorrência de *sorting* entre localizações pode explicar porque as habilidades produtivas dos trabalhadores variam entre áreas distintas. A composição setorial diferenciada espacialmente requer distintas composições de trabalhadores, de modo que podem ser esperados salários mais altos nas áreas especializadas em setores intensivos em habilidades. No modelo de Combes, Duranton e Gobillon (2008), a hipótese de salários locais determinados pelas habilidades assume que o salário de um trabalhador i é independente da localização e será dado por:

$$w_i = As_i \quad (2.1)$$

onde s_i denota as habilidades individuais e A a produtividade do trabalho. A produtividade do trabalho de um local pode superar a de outro em decorrência de características geográficas, tais como localização favorável em relação a portos e rios, clima mais propício à atividade econômica ou recursos naturais, e dotações locais não naturais, tais como capital público e

privado, instituições e tecnologia. Neste caso, o modelo assume que as dotações locais E_a afetam a produtividade do trabalho de uma área a , tal que o salário nesta área será dado por $w_a = A(E_a)$.

Interações locais entre firmas ou trabalhadores também podem levar a ganhos de produtividade. As interações correspondem às ligações de insumo-produto entre ofertantes e compradores, ao melhor *matching* entre trabalhadores com habilidades específicas e firmas em mercados de trabalho densos e às externalidades tecnológicas. Tais benefícios podem ser originados pelo tamanho do mercado, as economias de urbanização, ou pela concentração geográfica setorial, as economias de localização. Neste caso, o salário médio na área a e setor k será dado por $w_a = A(I_a, I_{a,k})$, onde I_a e $I_{a,k}$ representam as interações capturando economias de urbanização e localização, abordadas em estudos de economia regional e urbana como economias de aglomeração. Portanto, em equilíbrio, o modelo encontra o salário de um trabalhador i , na área $a(i)$ e setor $k(i)$ dado por:

$$w_i = A(E_{a(i)}, I_{a(i)}, I_{a(i),k(i)})S_i. \quad (2.2)$$

Embora o modelo microeconômico apresentado assuma uma firma competitiva, com função de produção Cobb-Douglas, os autores avaliam que as abordagens de concorrência imperfeita levam a resultados similares. O modelo microeconômico assume uma firma representativa competitiva operando em uma área a , no setor k e no período t , com função de lucro dada por

$$\pi_{a,k,t} = p_{a,k,t}y_{a,k,t} - \sum_{i \in (a,k,t)} w_{i,t}l_{i,t} - r_{a,k,t}z_{a,k,t} \quad (2.3)$$

onde, $p_{a,k,t}$ 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 trabalhados, $z_{a,k,t}$ representa outro fator de produção e $r_{a,k,t}$ é o preço desse fator. Assumindo uma função de produção Cobb-Douglas trabalho efetiva:

$$y_{a,k,t} = A_{a,k,t} \left(\sum_{i \in (a,k,t)} S_{i,t} l_{i,t} \right)^b (z_{a,k,t})^{1-b} \quad (2.4)$$

onde $0 < b \leq 1$, $s_{i,t}$ são as habilidades do trabalhador i no período t e $A_{a,k,t}$ é a produtividade total dos fatores em (a, k, t) . É possível, contudo, reformular o modelo para uma firma com função de produção genérica, em uma área a , no setor k e no período t , dada por:

$$y_{a,k,t} = A_{a,t} f\left(\sum_{i \in (a,k,t)} s_{i,t} l_{i,t}, z_{a,k,t}\right) \quad (2.5)$$

onde $s_{i,t}$ e $l_{i,t}$ são as habilidades e horas trabalhadas do trabalhador i no período t , respectivamente, $A_{a,t}$ é um índice de produtividade da área a no período t , o qual pode depender de fatores locacionais, e $z_{a,k,t}$ representa outros fatores de produção. No equilíbrio competitivo, o trabalhador recebe salário igual a sua produtividade marginal:

$$w_{i,t} = b p_{a(i,t),k(i,t),t} A_{a(i,t),k(i,t),t} \times \left(\frac{z_{a(i,t),k(i,t),t}}{\sum_{i \in (a,k,t)} s_{i,t} l_{i,t}} \right)^{1-b} s_{i,t}. \quad (2.6)$$

Usando a condição de primeira ordem da maximização de lucro em relação a $z_{a,k,t}$ e substituindo em $w_{i,t}$, encontra-se:

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

$$w_{i,t} = B_{a(i,t),k(i,t),t} s_{i,t} \quad (2.7)$$

onde $B_{a(i,t),k(i,t),t} \equiv b(1-b)^{\frac{(1-b)}{b}} (p_{a(i,t),k(i,t),t} A_{a(i,t),k(i,t),t} r_{a(i,t),k(i,t),t}^{1-b})^{\frac{1}{b}}$. A expressão para $w_{i,t}$ mostra que as diferenças salariais entre áreas decorrem das diferenças de habilidades ($s_{i,t}$) ou de produtividade da área por setor ($B_{a(i,t),k(i,t),t}$). Por sua vez, o diferencial de produtividade entre áreas pode ser causado pela produtividade dos fatores ($A_{a,k,t}$) ou pelas diferenças em preços e custos de insumos ($p_{a,k,t}$; $r_{a,k,t}$).

Diferenças em preços podem ocorrer quando os produtores das áreas com melhor acesso ao mercado podem cobrar preços relativos maiores do que aqueles nas áreas com pior acesso. Diferenças em custos de fatores podem refletir demandas geradas por motivações não relacionadas à produção. Nesse aspecto, as áreas com melhores amenidades de consumo podem enfrentar uma maior demanda por terra. Por último, as diferenças em tecnologia

podem ser causadas por dotações geográficas locais (vantagens naturais), dotações produtivas não geográficas (instituições e capital) ou por interações locais entre trabalhadores, firmas ou ambos, essas últimas geralmente denominadas de economias de aglomeração (COMBES; DURANTON; GOBILLON, 2008; COMBES et al., 2012).

Características locais, a exemplo da densidade de emprego, podem afetar positivamente ou negativamente a produtividade ($B_{a,k,t}$), atuando como economias de aglomeração ou elemento de congestão. Nesse sentido, uma limitação do modelo é que, embora seja possível identificar os determinantes dos diferenciais espaciais de salários (habilidades, dotações e interações), não é possível identificar separadamente os efeitos tecnológicos e de preço. No entanto, esse é um problema clássico da literatura de aglomeração urbana, em que não é possível identificar os canais exatos pelos quais as economias de aglomeração operam (COMBES; DURANTON; GOBILLON, 2008).

A vantagem deste modelo é que pode ser interpretado com abordagens da NEG e de concorrências imperfeita, com a microfundamentação das economias de aglomeração, em que a aglomeração pode decorrer do tamanho do mercado (ganhos de escala), ou de outros fatores. Nesse sentido, Combes, Duranton e Gobillon (2008) apontaram que qualquer abordagem de competição imperfeita onde o salário é um *mark-up* sobre a produtividade marginal deve levar a resultados similares, uma vez que na especificação logarítmica este termo deve ser acrescido à constante ou aos efeitos fixos setoriais se ele variar entre setores mas não entre áreas.

O estudo de Combes, Duranton e Gobillon (2008) adotou uma abordagem em dois estágios para estimar o modelo apresentado. A especificação micro-econométrica do modelo assume que a produtividade por área, setor e ano pode ser expressa como:

$$\log B_{a,k,t} = \beta_{a,t} + \mu_{k,t} + I_{a,k,t} \gamma_k \quad (2.8)$$

onde $\beta_{a,t}$ é um efeito fixo de área no tempo, $\mu_{k,t}$ é o efeito fixo de setor no tempo, $I_{a,k,t}$ é um vetor de variáveis de interação intra-setor (a,k,t) e γ_k é um vetor de parâmetros. A habilidade para o trabalhador i é dada por:

$$\log s_{it} = X_{it} \varphi + \delta_i + \epsilon_{it} \quad (2.9)$$

onde X_{it} são as características dos trabalhadores variantes no tempo e δ_i é um efeito fixo de trabalhador. A combinação das duas expressões com a equação de salário do modelo resulta na seguinte demanda inversa por trabalho, que foi estimada no primeiro estágio:

$$\log w_{i,t} = \beta_{a(i,t),t} + \mu_{k(i,t),t} + I_{a(i,t),k(i,t),t} \gamma_{k(i,t)} + X_{i,t} \varphi + \delta_i + \epsilon_{i,t} \quad (2.10)$$

onde $X_{i,t}$ são as características dos trabalhadores variantes no tempo, δ_i é um efeito fixo de trabalhador, $\beta_{a(i,t),t}$ é um efeito fixo de área, $\mu_{k(i,t),t}$ é um efeito fixo de setor, $I_{a(i,t),k(i,t),t}$ são as interações intra-setoriais e $\epsilon_{i,t}$ é um termo de erro. Os efeitos fixos de área-ano estimados no primeiro estágio são utilizados como variável dependente no segundo estágio, para explicar a variação de salários entre áreas. A especificação econométrica do segundo estágio será dada por:

$$\beta_{a,t} = \omega_0 + \theta_t + I_{a,t} \gamma + E_{a,t} \alpha + v_{a,t} \quad (2.11)$$

onde θ_t são os efeitos de tempo, $E_{a,t}$ são variáveis de dotação (percentual de municípios na área que possui costa litorânea, montanhas, lagos e patrimônio cultural e arquitetônico) e $I_{a,t}$ são interações locais entre setores (logaritmo da densidade do emprego, da área, da diversidade setorial e do mercado potencial). O termo de erro reflete choques tecnológicos locais e são assumidos independentes e identicamente distribuídos entre áreas e períodos.

A primeira razão para a estimação do modelo em dois estágios é que a estimação do modelo em um único estágio não permite calcular a variância dos choques locais ($v_{a,t}$), dada a dimensão *within* da equação salarial, o que gera viés nas estimativas de erros-padrão das variáveis agregadas. A segunda razão é que não é possível distinguir os choques locais dos choques idiossincráticos no nível do trabalhador. Por sua vez, a estimação do modelo em dois estágios requer a correção, no segundo estágio, da possível heterocedasticidade oriunda da utilização do efeito de área-tempo estimado no primeiro estágio. Isso pode ocorrer se os efeitos das áreas com poucos trabalhadores forem menos precisamente estimados do que os das áreas com muitos trabalhadores na amostra. Utilizando o método *Feasible Generalized Least Squares* (FGLS) para realizar a correção, os resultados foram muito próximos aos encontrados sem a correção, sugerindo erros de amostragem negligenciáveis.

A identificação dos efeitos fixos de trabalhador, de área e de setor requer suficiente observação de trabalhadores e mobilidade entre setor e área. Combes et al. (2012b) assumem alguns pressupostos para identificar as habilidades individuais e os efeitos de área, buscando minimizar as limitações empíricas existentes. Nesse aspecto, a identificação do modelo assume que a produtividade da área afeta os trabalhadores igualmente (efeito monótono da aglomeração sobre as habilidades) e não altera as habilidades individuais (fixas). Como os efeitos de aglomeração podem ser mais intensos para os mais qualificados e os trabalhadores podem aprender mais em áreas mais produtivas, ampliando as habilidades, o estudo sugere ser mais apropriado comparar os trabalhadores dentro das áreas do que entre as áreas.

Por outro lado, os autores assinalam que o pressuposto de migração exógena de trabalhadores entre áreas pode não ser restritivo, já que o modelo controla os efeitos de área e de tempo, amenizando o possível viés decorrente da atração de trabalhadores para locais de elevados salários esperados. De acordo com os autores, o viés na equação salarial existe se a decisão de localização é motivada pelo exato valor do salário que pode ser obtido no ano, mas deixa de existir se a decisão é motivada pelo salário médio de outros trabalhadores (salário esperado) e pelo seu próprio efeito fixo. Por sua vez, os autores acreditam que o viés migratório em resposta às ofertas de salário específicas pode ser atenuado em função de preferências idiossincráticas e de fatores não relacionados ao salário, que influenciam as decisões locais, além das barreiras à mobilidade existentes.

O modelo de Combes, Duranton e Gobillon (2008) será adotado nesse trabalho como o principal modelo de referência para a identificação dos efeitos de aglomeração sobre os salários reais e controle do respetivo *sorting* espacial de trabalhador e firma nos mercados de trabalho locais do Brasil. Para a proposta de identificação sugerida nesse trabalho, o modelo de referência será adaptado com a inclusão de variáveis de controle adicionais, disponíveis nos dados utilizados.

2.2.2 Evidências Empíricas de *Sorting* na Determinação de Salários com Heterogeneidade de Trabalhador e Firma

Combes, Duranton e Gobillon (2008) utilizaram um painel quadrienal, de 1976 a 1996, com micro dados dos trabalhadores da França extraídos das *Déclarations Annuelles des Données*

Sociales (DADS), para estimar o modelo apresentado na seção anterior. A estimação salarial incluiu a idade e o efeito fixo do trabalhador, os efeitos fixos temporais das áreas de emprego, os efeitos fixos de setor e as características do setor na área (logaritmo da parcela de empregos e do número de estabelecimentos e a parcela de trabalhadores em ocupações qualificadas). Os efeitos não observados dos trabalhadores apresentaram o maior poder explicativo na regressão salarial. Os efeitos de área apresentaram grande poder explicativo, superior aos efeitos fixos setoriais e às características intra-setoriais, embora menor do que os efeitos de trabalhador. Os resultados evidenciam a importância do *sorting* individual e da localização para a determinação salarial.

As interações locais entre setores apresentaram forte contribuição para os efeitos de áreas. As economias de urbanização ou economias de aglomeração urbana, mensuradas pela densidade do emprego local, apresentaram a maior contribuição para os efeitos de áreas (3,7%), enquanto as dotações locais apresentaram pequena contribuição. Por sua vez, choques locais contemporâneos podem ampliar a migração e os salários gerando uma possível endogeneidade e causalidade reversa da densidade no segundo estágio. Utilizando *lags* temporais longos da população como instrumento, a estimativa da densidade caiu para 3%, valor abaixo daqueles encontrados na literatura (4% a 8%).

O estudo mais recente de Combes et al. (2012b), para o período entre 1993 e 2007, mostrou que os salários das áreas densas da França eram 13,1% maiores do que nas áreas menos densas. A elasticidade dos salários em relação à densidade do emprego foi estimada em 0,044, enquanto a elasticidade dos efeitos fixos de trabalhadores em relação à mesma densidade foi de 0,018. A diferença entre as elasticidades (0,026) interpretada como uma elasticidade da densidade líquida dos efeitos individuais indicou que, pouco menos da metade do diferencial de salários, atribuído à densidade entre áreas, seria explicado pelos efeitos do trabalhador. As evidências sugerem que as habilidades não observadas são em média mais altas nas áreas mais densas. Por sua vez, foram encontradas evidências de maior dispersão salarial em áreas de emprego maiores. Essa dispersão foi justificada pela maior sobre-representação de trabalhadores com salários altos e, em menor proporção, de trabalhadores com salários baixos.

Por outro lado, Combes et al. (2012a) desenvolveram um modelo para avaliar em que medida as diferenças de produtividade entre áreas de emprego pode ser explicadas pela seleção das firmas ou pelas economias de aglomeração. O estudo utilizou dados de estabelecimentos para

o período de 1994 a 2002 e uma abordagem de regressão quantílica para a distribuição de produtividade das firmas. Os resultados mostraram que a seleção das firmas não explicou a diferença de produtividade espacial, mesmo quando a estimação era realizada por setores, tamanho das cidades ou com diferentes definições de áreas. Esse resultado aponta para a importância das economias de aglomeração como fonte das diferenças de produtividade entre grandes cidades. O estudo encontrou um ganho médio de produtividade de 9,7% atribuído à densidade urbana. Os ganhos foram mais intensos para os estabelecimentos mais produtivos (14,4%) em relação aos de menor produtividade (4,8%).

No Brasil, poucos estudos estimaram o modelo de Combes, Duranton e Gobillon (2008) ou utilizaram alguma abordagem similar (FALLY; PAILLACAR; TERRA, 2010; BARUFI, 2015; BARUFI; HADDAD; NIJKAMP, 2016). Barufi, Haddad e Nijkamp (2016) estimaram o modelo citado com os micro dados da RAIS (2004, 2008; 2012) e as áreas REGIC-Regiões de Influência das Cidades (IBGE, 2008). O primeiro estágio da equação salarial gerou a estimativa do efeito fixo combinado de setor (indústria e serviço), área e ano, que foi utilizada como variável dependente no segundo estágio. O efeito da densidade do emprego, no segundo estágio (*OLS*)⁵, variou entre 0,05 e 0,09 para todos os setores e áreas, sendo mais intenso nas indústrias de alta e baixa tecnologia e nos serviços intensivos em conhecimento.

Fally, Paillacar e Terra (2010) estimaram o impacto do acesso ao mercado (MA) e aos fornecedores (SA) sobre o diferencial de salários entre os estados e os setores do Brasil, em 1999, usando um modelo estrutural da NEG e incluindo os controles individuais. A estratégia empírica do estudo utilizou uma abordagem de três estágios, com dados da RAIS, da PIA⁶ (IBGE) e de outras fontes. O primeiro estágio fez a regressão dos salários sobre características individuais e efeito de setor–estado. O segundo estimou um modelo gravitacional setorial para calcular o acesso ao mercado e aos fornecedores em cada estado. Por fim, as estimativas de acesso ao mercado e fornecedores foram incluídas como variáveis explicativas para o diferencial salarial medido pelos efeitos de setor–estado do primeiro estágio.

Os resultados de Fally, Paillacar e Terra (2010) mostraram um efeito positivo e significativo do acesso ao mercado (0,14) sobre o prêmio salarial do setor–estado. Ao estimar o efeito de MA sobre os salários médios do setor–estado, sem estimar o primeiro estágio que controlava

⁵ *Ordinary Least Squares*.

⁶ Pesquisa Industrial Anual do IBGE.

as características individuais, a estimativa foi ampliada para 0,17, sugerindo que o *sorting* espacial do capital humano explicaria uma parcela do efeito. O efeito de MA (0,10) foi superior ao de SA (0,06). O estudo utilizou um indicador de mercado potencial baseado no Produto Interno Bruto (PIB), a população em 1940 e a data média de registro do município como instrumentos para MA. Os resultados foram robustos à estimação com os instrumentos (0,14), bem como à estimação do prêmio no nível municipal. No entanto, o estudo apontou a necessidade de levar em conta fatores como o *match* das heterogeneidades de trabalhador e firma, para entender os fenômenos mais complexos que têm emergido nessa área.

Por outro lado, utilizando um modelo de *sorting* por habilidade individual, Ehrl e Monasterio (2016) buscaram encontrar evidências para o Brasil de que a concentração espacial das habilidades interpessoais e analíticas em áreas densas geram externalidades positivas sobre os salários dos trabalhadores, a partir do Censo Demográfico de 2010. Para lidar com a endogeneidade do *sorting* de trabalhadores nas áreas de altos salários, os autores utilizaram a distribuição espacial das ocupações industriais e liberais como variável instrumental para as habilidades, com base nos Censos de 1872 e 1920. Os resultados encontraram efeitos positivos das habilidades sociais e analíticas sobre os salários individuais ou locais (áreas mínimas comparáveis). O aumento de um desvio padrão na concentração das habilidades interpessoais gerou um aumento de 10% nos salários médios. As evidências mostraram que os trabalhadores graduados e aqueles sem instrução formal foram os grupos que mais se beneficiaram das economias de aglomeração geradas pelo *sorting* espacial de habilidades.

Os estudos aplicados da Economia Urbana para o Brasil controlam somente uma parte do *sorting*, associada às heterogeneidades observadas e não observadas dos indivíduos. Outra abordagem da literatura empírica nacional, baseada na Economia do Trabalho, busca controlar o *sorting* associado às firmas realizando o controle da mobilidade individual entre firmas (FREGUGLIA; MENEZES-FILHO; SOUZA, 2007; FREGUGLIA; PROCÓPIO, 2013). Entretanto, o controle da mudança de firmas não é suficiente para capturar toda a magnitude dos efeitos específicos de firmas. Essa *proxy* do efeito de firma seria interpretada como um diferencial salarial dos trabalhadores que se movem entre firmas, enquanto os efeitos específicos da firma seriam capturados pela inclusão direta dos seus efeitos fixos.

O estudo de Freguglia e Procópio (2013) para os trabalhadores formais mostrou que a mudança de emprego e a migração intermunicipal, entre 1995 e 2006, eram fatores relevantes

para a correta identificação dos diferenciais salariais interestaduais, mesmo após o controle dos efeitos fixos individuais. O estudo encontrou uma maior intensidade do efeito de mudança de firma em relação ao efeito das migrações. Por sua vez, Freguglia, Menezes-Filho e Souza (2007) analisaram o diferencial salarial de trabalhadores de Belo Horizonte que migraram entre setores e regiões entre 1999 e 2001, e encontraram que embora tais efeitos de migração setorial e espacial fossem significantes, grande parte do diferencial salarial dos trabalhadores seria explicado por seus atributos individuais não observados.

Freguglia e Menezes-Filho (2012) mostraram que 63% do diferencial de salários entre estados, entre 1995 e 2002 (RAIS-Migra), é explicado por características não observadas dos trabalhadores, tais como o talento e a motivação. A educação e as heterogeneidades não observadas explicam grande parte do diferencial salarial. Contudo, mesmo após o controle das heterogeneidades individuais, permaneceu um diferencial salarial em menor magnitude, o que pode sugerir a existência de um efeito de localização específico sobre os salários. A investigação dos efeitos de localização requer o controle simultâneo dos efeitos de firma e trabalhador nas equações mincerianas, além dos controles usuais da literatura de Economia Urbana. Existe uma abordagem relativamente recente na literatura de Economia do Trabalho que busca identificar os parâmetros da equação salarial a partir do controle simultâneo dos efeitos de trabalhador e firma. O modelo de decomposição estatística dos salários desenvolvido por Abowd, Kramarz e Margolis (1999) tornou-se uma referência nessa abordagem.

Abowd, Kramarz e Margolis (1999) propuseram um modelo estatístico para decompor o rendimento anual do trabalhador e explicar o diferencial salarial setorial a partir de componentes associados às características observadas dos trabalhadores e à heterogeneidade não observada de indivíduos e firmas. O estudo desenvolveu abordagens alternativas de estimação para recuperar os parâmetros da equação salarial, dada a limitação computacional da solução de mínimos quadrados. Para a estimação dos resultados, foi adotada a base longitudinal de micro dados da França (DADS, 1976-1987). Os resultados mostraram que a heterogeneidade individual não observada era uma fonte importante da variação salarial, mais relevante do que os efeitos de firma. As especificações que controlavam os efeitos de trabalhador explicaram entre 77% e 83% da variação salarial, enquanto a especificação que controlou somente os efeitos de firma explicou 55%. Por sua vez, os efeitos de trabalhador

explicaram cerca de 90% do diferencial salarial setorial, enquanto os efeitos de firma explicaram entre 7% e 25%. Os resultados sugeriram uma maior contribuição dos efeitos de trabalhador para o diferencial salarial setorial. No entanto, ambos responderam por uma parcela substancial (85-96%) da variação salarial setorial da França.

Somente um trabalho empírico no Brasil realizou o controle simultâneo dos efeitos de trabalhador e firma, embora tenha assumido um efeito combinado. A partir do modelo de decomposição salarial de Abowd et al. (2001), Menezes-Filho, Muendler e Ramey (2008) estimaram uma equação salarial para São Paulo, principal centro industrial do país, utilizando os micro dados da RAIS (1990; 1997). O estudo controlou os efeitos de estabelecimento assumindo que estes eram uma combinação dos efeitos puros de estabelecimento com os efeitos médios de trabalhador no estabelecimento. Os resultados mostraram que as características observadas dos trabalhadores (experiência, gênero e ocupação), condicionais ao local de trabalho, eram mais relevantes para a variação de salários (57-62%) do que o efeito combinado de estabelecimento (26-34%). Nesse sentido, os fatores observados e não observados associados ao trabalhador seriam mais importantes do que os fatores da firma para explicar a alta desigualdade salarial do Brasil em relação aos EUA e à França.

Na literatura internacional, o estudo recente de Dauth et al. (2016) propôs uma abordagem em dois estágios similar à deste trabalho, onde a dimensão local é adicionada ao modelo de Abowd, Kramarz e Margolis (1999) de modo a integrar as literaturas de aglomeração e Economia do Trabalho. O estudo buscou decompor a variação espacial de salários nas áreas de emprego da Alemanha em efeitos de trabalhador, efeitos de firma e efeitos combinados de trabalhador e firma (*worker-firm assortative matching*), no período de 1990 a 2010. Os resultados mostraram que o melhor *matching* de trabalhadores e firmas nos mercados de trabalho densos foi o principal mecanismo explicando o prêmio salarial urbano, sendo tão importante quanto as fontes de explicação baseadas no trabalhador.

A abordagem proposta neste trabalho se diferencia da literatura internacional e nacional na medida em que utiliza uma nova estratégia de identificação para a estimação dos efeitos de aglomeração. O trabalho contribui para a literatura com a utilização de uma medida de densidade mais adequada para as áreas urbanas e com a adoção de uma variável instrumental tradicionalmente usada na literatura de Economia do Trabalho para lidar com a endogeneidade da densidade e identificar os efeitos de aglomeração. Além dessas

contribuições, até onde se tem conhecimento, este estudo é o primeiro no Brasil a lidar simultaneamente com os efeitos não observados de trabalhador e firma na identificação dos efeitos de aglomeração urbana.

A literatura de Economia Espacial mostrou que o espaço é relevante para a determinação e diferenciação salarial. Diversas evidências empíricas sobre a relação entre densidade urbana, aglomeração e salários já foram encontradas. A literatura recente tem buscado avançar na identificação dos efeitos de aglomeração, considerando os diversos fatores correlacionados com esses efeitos. Portanto, com base nos avanços dessa literatura e da Economia do Trabalho esta pesquisa propõe contribuir com os estudos de aglomeração urbana, *sorting* e determinação salarial, recorrendo a uma abordagem que integra os modelos de *sorting* espacial da Economia Urbana com os modelos de controle simultâneo dos efeitos de trabalhador e firma na equação de salários.

3 METODOLOGIA

Este capítulo apresenta a metodologia adotada neste trabalho. Inicialmente será apresentado o modelo de decomposição salarial proposto por Abowd, Kramarz e Margolis (1999) e utilizado neste estudo. As seções seguintes apresentam o banco de dados, o modelo econométrico, os métodos e a estratégia de identificação proposta neste estudo.

3.1 MODELO DE DECOMPOSIÇÃO SALARIAL

Os estudos empíricos em Economia do Trabalho encontram resultados individuais bastante heterogêneos nos mercados de trabalho, o que tem motivado uma extensa literatura para decompor as fontes dessas heterogeneidades. Indivíduos observacionalmente similares podem ganhar salários diferenciados e possuir trajetórias distintas de emprego. A ausência de micro dados relacionando as características de trabalhadores e firmas é um fator limitador nas pesquisas que buscam separar os efeitos decorrentes de decisões de firmas daqueles decorrentes das decisões dos trabalhadores. Uma abordagem dessa literatura, focada na explicação dos diferenciais salariais, enfatiza a heterogeneidade individual não observada (efeito de indivíduo), enquanto outra abordagem avalia a heterogeneidade não observada de firmas (efeito de firma) como um fator explicativo relevante (ABOWD; KRAMARZ; MARGOLIS, 1999).

Para lidar com as diferentes dimensões da heterogeneidade salarial, este trabalho utiliza as estratégias de estimação desenvolvidas a partir do modelo de decomposição estatística de salários de Abowd, Kramarz e Margolis – AKM (1999). O modelo AKM é considerado a primeira análise estatística da heterogeneidade simultânea de trabalhador e firma na determinação de salários. Considere a seguinte equação de salários:

$$y_{it} = \mu_y + (x_{it} - \mu_x)\beta + \theta_i + \psi_{J(i,t)} + \varepsilon_{it}, \quad (3.1)$$

onde y_{it} é o logaritmo do salário do indivíduo $i = 1, \dots, N$, no período $t = 1, \dots, T$, x_{it} é um vetor de P características exógenas do indivíduo variando no tempo, θ_i é o efeito puro de indivíduo, $\psi_{J(i,t)}$ é o efeito puro de firma, para a firma em que o trabalhador i está empregado no tempo t (indicada por $J(i,t)$), μ_y e μ_x são as médias gerais de y_{it} e x_{it} , e ε_{it} é o resíduo estatístico.

Assumindo uma amostra aleatória com N indivíduos observados em T anos, então ε_{it} apresentará as seguintes propriedades:

$$E[\varepsilon_{it}|i, t, J(i, t), x_{it}] = 0 \quad (3.1.1)$$

$$\text{cov}[\varepsilon_{it}, \varepsilon_{ns}|i, t, n, s, J(i, t), J(n, s), x_{it}, x_{ns}] = \begin{cases} \sigma_\varepsilon^2 & \text{para } i = n \text{ e } t = s \\ 0 & \text{caso contrário.} \end{cases} \quad (3.1.2)$$

Em notação matricial, a equação (3.1) torna-se:

$$y = X\beta + D\theta + F\psi + \varepsilon, \quad (3.2)$$

onde, X é uma matriz ($N^* \times P$) de características observáveis (em desvios da média) variantes no tempo, D é uma matriz ($N^* \times N$) de indicadores para os indivíduos, F é uma matriz ($N^* \times J$) de indicadores para os efeitos da firma em que i trabalha em t (J é o número total de firmas), y é o vetor ($N^* \times 1$) de salários (em desvios da média), ε é o vetor de resíduos e $N^* = NT$. Os parâmetros são os vetores β ($P \times 1$), θ ($N \times 1$), ψ ($J \times 1$) e a variância do erro σ_ε^2 .

As equações (3.1) e (3.2) são interpretadas como a esperança condicional dos salários, dadas as características observadas e os indicadores de indivíduos e firmas. Embora alguns estudos já tenham avançado na solução de Mínimos Quadrados (*Full Least Squares*) para a equação (3.2), com amostras pequenas (LEONARD et al., 1996; ENTORF; GOLLAC; KRAMARZ, 1999; GOUX; MAURIN, 1999), o esforço da literatura recente tem sido encontrar soluções factíveis para grandes bancos de dados pareados de firma e trabalhador. Muitos estudos estimam versões incompletas da equação, o que gera interpretações ambíguas sobre os seus parâmetros. A estimação dessa equação sem algum dos seus componentes, ou com combinações lineares de efeitos (viés de agregação), pode modificar o significado dos efeitos subjacentes. Nesse contexto, Abowd, Kramarz e Margolis (1999) descreveram as variações existentes na parametrização da equação de salários na literatura de Economia do Trabalho.

A versão da equação (3.2) sem o efeito puro de firma (ψ) implica que o efeito de indivíduo será a soma do efeito puro de indivíduo (θ) com um termo de viés de variável omitida:

$$\theta^* = \theta + (D'M_X D)^{-1} D'M_X F \psi, \quad (3.3)$$

onde $M_A \equiv I - A(A'A)^{-1}A'$, para uma matriz A arbitrária. O parâmetro associado às características individuais variantes no tempo (β^*) é representado pela soma de β do modelo completo com um termo de viés de variável omitida:

$$\beta^* = \beta + (X'M_D X)^{-1} X'M_D F \psi. \quad (3.4)$$

A versão da equação (3.2) sem o efeito puro de indivíduo (θ) implica em ψ^{**} e β^{**} . Similarmente, o efeito de firma é dado pela soma do puro efeito de firma (ψ) com um termo de viés de variável omitida:

$$\psi^{**} = \psi + (F'M_X F)^{-1} F'M_X D \theta. \quad (3.5)$$

O vetor β^{**} é definido pela soma de β do modelo completo com um termo de viés de variável omitida:

$$\beta^{**} = \beta + (X'M_F X)^{-1} X'M_F D \theta. \quad (3.6)$$

Portanto, estimações de versões reduzidas do modelo completo (3.2) resultam em estimativas e interpretações enviesadas dos efeitos de firma e indivíduo. A literatura sobre os diferenciais salariais entre setores tem apontado que estes diferenciais não podem ser explicados somente por características observadas de trabalhadores e firmas. No modelo AKM, o efeito puro de setor é definido como a agregação dos efeitos puros de firmas dentro do respectivo setor, correspondendo à inclusão de variáveis indicadoras de setor na equação (3.2). O efeito residual de firma é definido como um desvio do efeito de setor. Portanto, a equação (3.2) torna-se:

$$y = X\beta + D\theta + FA\kappa + (F\psi - FA\kappa) + \varepsilon, \quad \kappa \equiv (A'F'FA)^{-1}A'F'F\psi \quad (3.7)$$

onde a matriz A ($J \times K$) classifica cada uma das J firmas em um dos K setores, tal que, $a_{jk} = 1$, se e somente se, $K(j)=k$, onde $K(j)$ denota a classificação setorial da firma j . O vetor de parâmetros κ ($K \times 1$) é interpretado como uma média ponderada do efeito puro de firma. O efeito $(F\psi - FA\kappa)$ pode ser representado como $M_{FA}F\psi$. Os termos $FA\kappa$ e $M_{FA}F\psi$, de (3.7), resultam da decomposição de $F\psi$ em dois componentes ortogonais. Nesse contexto, duas

versões incompletas de (3.7) podem gerar inconsistentes estimativas. A versão da equação (3.7) sem o efeito firma residual ($M_{FA}F\psi$) implicará em κ^* . O efeito de setor será a soma do efeito puro de setor (κ) com um termo de viés de variável omitida:

$$\kappa^* = \kappa + (A'F'M_{[D\ X]}FA)^{-1}A'F'M_{[D\ X]}M_{FA}F\psi, \quad (3.8)$$

onde $M_{[D\ X]}$ é a matriz M_Z com $Z \equiv [D\ X]$. O vetor de observáveis β^* será a soma de β do modelo completo com um viés de variável omitida:

$$\beta^* = \beta + (X'M_{[D\ FA]}X)^{-1}X'M_{[D\ FA]}M_{FA}F\psi. \quad (3.9)$$

A versão da equação (3.7) sem o remanescente efeito de firma ($M_{FA}F\psi$) e o efeito fixo de indivíduo (θ) implicará em κ^{**} .

$$\begin{aligned} \kappa^{**} &= \kappa + (A'F'M_XFA)^{-1}A'F'M_X(M_{FA}F\psi + D\theta) \\ &\equiv (A'F'M_XFA)^{-1}A'F'M_XF\psi + (A'F'M_XFA)^{-1}A'F'M_XD\theta. \end{aligned} \quad (3.10)$$

Portanto, as estimativas de versões incompletas da equação (3.7) também devem resultar em efeitos setoriais inconsistentes. O modelo AKM adotou a equação (3.10) para decompor os efeitos setoriais e verificar a proporção dos diferenciais salariais setoriais que era explicada pelos efeitos de trabalhador, de firma e de ambos.

Com base no modelo geral apresentado na equação de salários (3.2), Abowd, Kramarz e Margolis (1999) desenvolveram abordagens estatísticas alternativas à de mínimos quadrados para estimar o modelo completo, incluindo efeitos fixos de firmas e indivíduos. A matriz (*cross-product*) contendo as submatrizes de variáveis do modelo geral (3.2), que será pré-multiplicada pelo vetor de parâmetros para a sua identificação é dada por:

$$\begin{bmatrix} X'X & X'D & X'F \\ D'X & D'D & D'F \\ F'X & F'D & F'F \end{bmatrix} \quad (3.11)$$

cujas dimensões dependem do tamanho de N (indivíduos) e J (firmas). Métodos computacionais usuais para a estimação dos parâmetros $[\beta' \theta' \psi']$ por mínimos quadrados não

são, em geral, factíveis. Como a estimação do modelo completo e irrestrito seria limitada, Abowd, Kramarz e Margolis (1999) propuseram métodos alternativos para tentar preservar, em alguma medida, a estrutura geral do modelo completo. A mobilidade dos indivíduos entre firmas é uma condição necessária para a identificação estatística do modelo, ou seja, para encontrar os efeitos fixos separadamente, independente da abordagem computacional adotada.

A primeira abordagem, denominada de método consistente, recorre à primeira diferença nos dados intra firma-indivíduo, mantendo os supostos assumidos em (3.1) e usando a definição:

$$\psi_j = \phi_j + \gamma_j s_{it} \quad (3.12)$$

$$F\psi = F_0\phi + F_1\gamma \quad (3.13)$$

onde s_{it} denota a permanência (*seniority*) do indivíduo i na firma $j = J(i, t)$ no ano t , ϕ_j denota o intercepto específico da firma, F_0 e F_1 são matrizes $N \times J$ e γ é $J \times 1$. Desse modo, a primeira diferença, aplicada a todas as informações nas quais $J(i, n_{it}) = J(i, n_{it-1})$, será dada por:

$$\begin{aligned} y_{in_{it}} - y_{in_{it-1}} &= (x_{in_{it}} - x_{in_{it-1}})\beta + \gamma_{J(i, n_{it})}(s_{in_{it}} - s_{in_{it-1}}) + \varepsilon_{in_{it}} - \varepsilon_{in_{it-1}} \\ \Delta y &= \Delta X\beta + \tilde{F}\gamma + \Delta\varepsilon, \end{aligned} \quad (3.14)$$

onde n_{it} corresponde a um indicador da primeira observação até a última observação e ano em que o indivíduo i aparece, Δy é $\tilde{N}^* \times 1$, ΔX é $\tilde{N}^* \times P$, $\tilde{F}$ é $\tilde{N}^* \times J$, $\Delta\varepsilon$ é $\tilde{N}^* \times 1$ e $\tilde{N}^*$ é o número de combinações (i, t) na amostra que satisfazem à condição $J(i, n_{it}) = J(i, n_{it-1})$. Então, as estimativas deste método serão:

$$\tilde{\beta} = (\Delta X' M_{\tilde{F}} \Delta X)^{-1} \Delta X' M_{\tilde{F}} \Delta y \quad (3.15)$$

$$\tilde{\gamma} = (\tilde{F}' \tilde{F})^{-1} \tilde{F}' (\Delta y - \Delta X \tilde{\beta}). \quad (3.16)$$

No entanto, o método consistente é ineficiente para estimar o modelo completo (3.2), pois a primeira diferença vai eliminar os trabalhadores cuja firma em t difere da firma em $t-1$. Além disso, dada a restrição de $J(i, n_{it}) = J(i, n_{it-1})$, o método consistente não pode ser usado para identificar os efeitos fixos de firma e trabalhador separadamente.

A classe de estimadores definidos para o modelo AKM como métodos condicionais não restringe a amostra na estimação, sendo capaz de identificar os efeitos fixos separadamente. A denominação de condicional resulta da sua relação com as técnicas computacionais padrões para modelos lineares e por sua origem na literatura de dados em painel, associada aos modelos de efeitos fixos individuais. Hipóteses adicionais de ortogonalidade são necessárias: as interações entre X , D e F serão usadas como *proxy* para as correlações entre elas. A estimação assumirá ortogonalidade condicional, dadas as interações.

Neste contexto, define-se uma matriz Z ($N \times Q$) a partir de Q funções de informações em X , D e F . Para a construção de Z , Abowd, Kramarz e Margolis (1999) incluíram o tamanho da firma e seu quadrado, o setor, a experiência do trabalhador e sua idade ao final do período de formação escolar. A solução de mínimos quadrados para (3.2) pode ser encontrada assumindo alguns pressupostos de ortogonalidade condicional a Z . Sob a hipótese de que X e D são ortogonais à projeção de F sobre o espaço nulo de Z , a equação (3.2) torna-se:

$$y = X\beta + D\theta + Z\gamma + M_Z F\psi + \varepsilon, \quad (3.17)$$

com $\gamma \equiv (Z'Z)^{-1}Z'F\psi$. O suposto de ortogonalidade condicional entre X e F e entre D e F , dado Z , implica em:

$$X'M_Z F = 0, \quad (3.18)$$

$$D'M_Z F = 0. \quad (3.19)$$

Portanto, dados os supostos, a solução de mínimos quadrados para (3.17) será:

$$\begin{bmatrix} \hat{\beta} \\ \hat{\theta} \\ \hat{\lambda} \end{bmatrix} = \begin{bmatrix} X'X & X'D & X'Z \\ D'X & D'D & D'Z \\ Z'X & Z'D & Z'Z \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ D'y \\ Z'y \end{bmatrix}, \quad (3.20)$$

$$\hat{\psi} = (F'M_Z F)^- F'M_Z y, \quad (3.21)$$

onde $(F'M_Z F)^-$ denota uma inversa generalizada, requerida devido ao $\text{rank}(F'M_Z F) = J - 1 - Q$. A estimação independente da ordem (*order-independent estimation*) é um método condicional para computar as soluções (3.20) e (3.21), realizado em dois estágios

independentes. O primeiro estágio segue uma abordagem de estimação longitudinal intra-indivíduo (*within-D*), em que X e Z são projetados sobre D , obtendo:

$$\begin{bmatrix} \hat{\beta} \\ \hat{\lambda} \end{bmatrix} = \begin{bmatrix} X'M_D X & X'M_D Z \\ Z'M_D X & Z'M_D Z \end{bmatrix}^{-1} \begin{bmatrix} X'M_D y \\ Z'M_D y \end{bmatrix} \quad (3.22)$$

$$\hat{\theta} = (D'D)^{-1} D'(y - X\hat{\beta} - Z\hat{\lambda}) \quad (3.23)$$

O segundo estágio, intra-firma (*within-F*), computa a solução de mínimos quadrados para os parâmetros de F e Z , a partir de:

$$y = F\psi + Z\pi + v, \quad (3.24)$$

onde π ($Q \times 1$) é um vetor de parâmetros auxiliares e $v \sim N(0, \sigma_v^2 I)$, dados as hipóteses de ortogonalidade condicional. A solução no segundo estágio será dada por:

$$\hat{\pi} = (Z'M_F Z)^{-1} Z'M_F y, \quad (3.25)$$

$$\hat{\psi} = (F'F)^{-1} F'(y - Z\hat{\pi}). \quad (3.26)$$

No método condicional dependente da ordem (*order-dependent estimation*), os estágios são dependentes, de modo que as estimativas dos parâmetros podem ser diferentes, a depender de quais efeitos são estimados primeiro. Na abordagem em que os efeitos de indivíduos são estimados primeiro, o primeiro estágio recupera os parâmetros β , θ e λ como no método de independência da ordem, de acordo com (3.22) e (3.23). No segundo estágio, os efeitos de firma são estimados utilizando as equações (3.12) e (3.13). Para tanto, todas as observações sobre os indivíduos empregados na mesma firma são agrupadas no conjunto $\{j\} \equiv \{(i, t) \mid J(i, t) = j\}$, de N_j elementos, tal que:

$$\hat{y}_{\{j\}} \equiv y_{\{j\}} - x_{\{j\}}\hat{\beta} - \hat{\theta}_{\{j\}} \quad (3.27)$$

$$y_{\{j\}} \equiv \begin{bmatrix} \dots \\ y_{ns} \\ \dots \end{bmatrix}, \forall (n, s) \in \{j\}, \quad (3.28)$$

com $x_{\{j\}}$ e $\theta_{\{j\}}$ similares a $y_{\{j\}}$, e utilizando as estimativas de primeiro estágio para $x\hat{\beta}$ e $\hat{\theta}$. A equação no nível de firma será, então, dada por:

$$\begin{aligned}\hat{y}_{\{j\}} &= F_{\{j\}} \begin{bmatrix} \phi_j \\ \gamma_j \end{bmatrix} + \zeta_{\{j\}}, \\ F_{\{j\}} &\equiv \begin{bmatrix} 1 & \cdots \\ & s_{ns} \\ & \cdots \end{bmatrix}, \forall (n, s) \in \{j\}, \\ \zeta_{\{j\}} &\equiv \varepsilon_{\{j\}} + x_{\{j\}}(\beta - \hat{\beta}) + (\theta_{\{j\}} - \hat{\theta}_{\{j\}}).\end{aligned}\tag{3.29}$$

Portanto, o estimador de mínimos quadrados de (3.29) será:

$$\begin{bmatrix} \hat{\phi}_j \\ \hat{\gamma}_j \end{bmatrix} = (F'_{\{j\}} F_{\{j\}})^{-1} F'_{\{j\}} \hat{y}_{\{j\}}, \text{ para } j = 1, \dots, J.\tag{3.30}$$

Por sua vez, na abordagem em que os efeitos de firma são estimados primeiro, o primeiro estágio utiliza o estimador de $\hat{\psi}$ dado por (3.21) ou (3.26), enquanto o segundo estágio encontra β e θ a partir de:

$$\begin{aligned}y - F\hat{\psi} &= X\beta + D\theta + \xi \\ \xi &= \varepsilon + F(\psi - \hat{\psi}).\end{aligned}\tag{3.31}$$

Portanto, os estimadores de β e θ serão:

$$\hat{\beta} = (X' M_D X)^{-1} X' M_D (y - F\hat{\psi})\tag{3.32}$$

$$\hat{\theta} = (D' D)^{-1} D' (y - X\hat{\beta} - F\hat{\psi}).\tag{3.33}$$

Os estimadores de β e θ obtidos pelo método independente da ordem são idênticos aos obtidos pelo método condicional dependente da ordem, em que os efeitos de indivíduos são estimados primeiro. O estimador de ψ independente da ordem é idêntico ao seu estimador condicional dependente da ordem, com o efeito de firma estimado primeiro. Abowd, Kramarz e Margolis (1999) apontaram que a desvantagem do método independente da ordem é não impor a ortogonalidade da solução *Full Least Squares* entre o resíduo estimado e os efeitos fixos do modelo. O método dependente da ordem impõe a ortogonalidade, mas gera diferentes resultados a depender de qual efeito é estimado primeiro. No entanto, o estudo encontrou um melhor desempenho dos modelos em que os efeitos de indivíduo são estimados primeiro. Nesse sentido, este trabalho utiliza o modelo dependente da ordem com os efeitos fixos

individuais sendo estimados primeiro. Este trabalho adotará as estratégias de estimação desenvolvidas no modelo AKM de decomposição salarial, apresentado nessa seção.

3.2 BASE DE DADOS

Este estudo utiliza os micro dados da Relação Anual de Informações Sociais (RAIS) do Ministério do Trabalho e Emprego (MTE)⁷ como principal banco de dados para a estimação dos resultados de interesse. A RAIS é um banco de dados combinado de trabalhador e estabelecimento (*matched employer-employee data*), composto pelos registros administrativos declarados anualmente, de forma compulsória, pelas empresas do mercado de trabalho formal do Brasil. O objetivo da RAIS é fornecer informações sobre o mercado de trabalho, subsidiar o controle da arrecadação de benefícios trabalhistas e viabilizar a concessão de benefícios sociais e trabalhistas. Por sua natureza compulsória, a RAIS funciona como um censo do mercado de trabalho formal do Brasil, apresentando cobertura de mais de 90% deste mercado a partir dos anos 1990. Os registros permitem acompanhar a trajetória geográfica, ocupacional e setorial do trabalhador formal ao longo do tempo e características selecionadas do estabelecimento empregador.

A vantagem da RAIS é que o seu formato longitudinal permite a construção de dados em painel para os trabalhadores e o uso de modelos e métodos específicos para dados combinados de trabalhador e firma. Nesse sentido, o banco de dados permite lidar com os respectivos efeitos não observados de trabalhadores e firmas. Uma desvantagem dessa base é que o seu uso requer um suporte computacional avançado para armazenamento e processamento dos micro dados, particularmente quando o estudo faz uso de dados em painel⁸. Quanto maior o número de anos incluído no painel, maior será a necessidade de geração de amostras para ser possível trabalhar com os micro dados da RAIS.

O Brasil ampliou substancialmente o número de empregos formais recentemente, de cerca de 29 milhões para cerca de 49 milhões no período de 2002 a 2014, de acordo com a RAIS. O número de estabelecimentos também foi ampliado de 2,4 milhões para cerca de 4 milhões no mesmo período. Portanto, a expansão do mercado de trabalho formal não decorreu apenas do crescimento do número de vagas, mas também do número de firmas. Uma característica

⁷ A RAIS foi instituída em 1975, pelo Decreto nº 76.900/75 (MTE, 2010).

⁸ No período de 2002 a 2014, os micro dados da RAIS possuem aproximadamente 160 *Gigabytes*.

comum do mercado de trabalho de países em desenvolvimento, particularmente na América Latina, é a grande parcela de empregos informais. O mercado de trabalho do Brasil apresenta um grande número de trabalhadores informais com desigual distribuição entre as macrorregiões do país (Apêndice A). A região Nordeste apresentou 34% dos ocupados no segmento formal enquanto o Sudeste apresentou 57%, em 2014⁹. Neste sentido, as evidências e conclusões sugeridas a partir dos resultados deste estudo têm seu escopo limitado ao segmento formal de cada localidade e não podem ser extrapoladas para todo o mercado de trabalho.

Os bancos de microdados identificados da RAIS são disponibilizados para cada ano. Para a construção do painel, inicialmente, procedeu-se à organização das bases anuais excluindo os registros sem o identificador do trabalhador, sem remuneração e com carga horária semanal contratada inferior a 20 horas. Os trabalhadores da Administração Pública foram excluídos¹⁰, tendo em vista que a lógica de funcionamento e de determinação salarial desse setor é diferente da lógica de mercado e os salários não devem variar entre localizações. Em 2014, a Administração Pública representava 19% dos vínculos registrados na RAIS e 18% desses empregos estavam no estado de São Paulo, de acordo com a Classificação Nacional de Atividades Econômicas (CNAE 1.0) do IBGE. A RAIS permite que cada trabalhador possa ser registrado em mais de um vínculo empregatício. Nesse sentido, para construir o painel, mantendo apenas um vínculo por trabalhador, foi necessário escolher o emprego principal. Para essa finalidade, foram adotados os critérios de organização da RAIS-Migra (MTE), a qual era extraída da RAIS anual e disponibilizada em formato longitudinal¹¹.

Inicialmente foram excluídos os trabalhadores que apresentavam mais de cinco vínculos no ano, pois é muito provável que esses casos representem erro de registro. O emprego principal foi definido pela data de admissão mais antiga, de acordo com os critérios adotados na RAIS-Migra. Além disso, foram mantidos apenas os vínculos com maior carga horária semanal contratada. Os vínculos cujos registros apresentavam informações divergentes dos demais vínculos foram excluídos, a exemplo do gênero e da data de nascimento. Por fim, após

⁹ Se o trabalhador conta-própria for incluído na categoria informal, a região Nordeste apresenta cerca de 50% dos ocupados nesse segmento, enquanto a região Sudeste apresenta apenas 36%, em 2014.

¹⁰ Foram excluídos os trabalhadores com vínculos Estatutário, Estatutário regido pelo Regime Geral de Previdência Social, Estatutários Não Efetivos e temporários contratados por Legislação Estadual ou Municipal. Além desses, foram excluídos os vínculos registrados no setor da Administração Pública, Defesa e Segurança Social na Classificação Nacional de Atividades Econômicas (CNAE 1.0) do IBGE.

¹¹ Desde 2012, o Ministério do Trabalho e Emprego interrompeu a disponibilização da RAIS-Migra.

proceder às etapas descritas acima, gerou-se um número aleatório para os trabalhadores que ainda apresentavam mais de um vínculo, mantendo aleatoriamente apenas um dos vínculos. O Apêndice B mostra as observações que foram excluídas em cada uma das etapas iniciais de limpeza da base.

Após a organização inicial do banco de dados, procedeu-se à etapa de agregação das bases anuais. Nessa fase, tendo em vista que cada base anual possuía entre 20 e 37 milhões de registros, entre 2002 e 2014, adotou-se a estratégia de agregação inicial em duas bases: 2002-2007 e 2008-2014. Em seguida, foram mantidos apenas os trabalhadores que estavam presentes em todos os anos em cada base. Finalmente, as duas bases foram agregadas e a base completa apresentou 42.140.815 de registros. Como a informação da localização do emprego é fundamental para os objetivos desta pesquisa, foram excluídos os vínculos sem o registro do município, restando apenas 3.440.663 de registros. Em seguida, como a idade do trabalhador é calculada a partir da sua data de nascimento, foram excluídos os indivíduos com datas de nascimento divergentes ao longo dos anos (130.152). A pesquisa propõe avaliar apenas os indivíduos na faixa-etária produtiva, entre 18 e 65 anos, seguindo o padrão utilizado na literatura¹². Essa etapa resultou na exclusão de 74.176 registros. Dessa forma, foi construído o painel inicial necessário para a geração da amostra.

O painel inicial foi composto por 42.072.355 observações, correspondentes a 3.236.335 trabalhadores e 1.036.188 estabelecimentos. A partir do painel inicial, procedeu-se à limpeza das inconsistências do identificador do estabelecimento. Essa etapa é necessária, tendo em vista que este trabalho faz uso dos identificadores de trabalhador e firma. A RAIS define os estabelecimentos empregadores a partir das unidades de cada empresa, separadas espacialmente, que possuem Cadastro Nacional de Pessoas Jurídicas - CNPJ ou Cadastro Específico do Instituto Nacional do Seguro Social (INSS) - CEI. As empresas que declaram a RAIS de forma centralizada devem fornecer as informações separadas para cada estabelecimento. No entanto, o estudo identificou dois problemas nos registros do CNPJ. O primeiro problema é que alguns estabelecimentos apresentavam o mesmo CNPJ em localizações distintas no mesmo ano, inclusive em estados distintos. O segundo problema decorre dos estabelecimentos registrados em setores de atividades distintos no mesmo ano, de acordo com a CNAE 1.0 (IBGE).

¹² Ver Glaeser e Maré (2001), Silveira-Neto e Azzoni (2004) e Freguglia e Menezes-Filho (2012).

O primeiro problema no registro do CNPJ pode ocorrer se os estabelecimentos das filiais são registrados com o mesmo CNPJ da matriz, assim como as empreitadas dos estabelecimentos da Construção Civil. Por sua vez, o segundo problema pode resultar de erros na declaração. Para corrigir os problemas de registro, optou-se por usar o CEI em substituição ao CNPJ desses estabelecimentos (854 no primeiro caso e 453 no segundo) e excluir aqueles que permaneceram com o erro após a correção (11 no primeiro caso e 26 no segundo). Por fim, foram excluídos os estabelecimentos que não apresentavam registro no CNPJ ou CEI (33.697). Após as correções, o painel apresentou 42.038.353 observações, sendo 1.043.006 estabelecimentos e 3.236.231 trabalhadores. O número de estabelecimentos foi ampliado com a correção dos erros, já que havia estabelecimentos distintos sob o mesmo CNPJ.

Existem alguns erros de registros na RAIS, a exemplo da redução da escolaridade, mudança de gênero e código do setor inexistente. A próxima etapa de organização do painel excluiu os registros com inconsistências nas variáveis de educação (9.121.026), gênero (915.597) e setor (1.872) e aqueles cuja remuneração em número de salários mínimos¹³ era menor do que um salário mínimo vigente (916.890), já que esse é o piso salarial legalmente estabelecido para o setor formal. Por fim, o balanceamento do painel excluiu 90.851 observações dos trabalhadores que não apareceram em todos os anos (2002-2014), resultando em um painel com 30.992.117 observações correspondentes a 2.384.009 trabalhadores e 750.414 estabelecimentos¹⁴. A geração da amostra adotou dois critérios necessários para a estratégia de identificação do estudo, a qual será descrita nas próximas seções. O primeiro critério requer uma amostra com suficiente mobilidade de estabelecimentos entre os arranjos populacionais, os quais representam as áreas de emprego aglomeradas adotadas neste estudo. A mobilidade permite identificar os efeitos fixos de firma e de arranjos separadamente. O segundo critério requer uma amostra aleatória representativa para os arranjos populacionais.

O painel completo apresentou 16.704 (2,2%) estabelecimentos com alguma mobilidade entre os arranjos, correspondendo a 2.562.599 observações de 197.123 trabalhadores. Como o percentual de firmas móveis entre os arranjos no mercado de trabalho formal é baixo, foi necessário manter todas as firmas móveis na amostra. Para garantir a aleatoriedade da amostra e a representatividade dos arranjos populacionais, foi gerada uma segunda amostra a partir

¹³ Remuneração referente ao mês de dezembro de cada ano.

¹⁴ No período de 2002 a 2014 o atrito amostral nos empregos privados da RAIS foi em média 25%, variando entre 23,6% e 26,2%.

dos estabelecimentos que não apresentaram mobilidade entre arranjos. Portanto, foi extraída uma amostra de 1% aleatorizada e estratificada entre os arranjos populacionais (3.485.794 observações, de 268.138 trabalhadores), a partir dos 2.186.886 trabalhadores das firmas sem mobilidade. O Apêndice C apresenta a representatividade amostral dos arranjos populacionais em relação aos empregos da RAIS para o mesmo período (2002-2014).

O banco de micro dados final deste estudo é composto por 6.048.393 observações, que correspondem a um painel anual balanceado com 465.261 trabalhadores, 271.951 estabelecimentos e 1.081.990 combinações de trabalhador e estabelecimento, para o período entre 2002 e 2014. Para testar a robustez dos resultados sem a seleção de uma parte da amostra em firmas móveis, extraiu-se uma amostra aleatória de 2%, entre os arranjos populacionais, gerando um painel balanceado que contém 7.158.099 observações de 550.623 trabalhadores, 322.601 estabelecimentos (com 7.213 ou 2,2% móveis) e 1.077.432 combinações de trabalhador e estabelecimento.

O PIS¹⁵ foi utilizado como variável identificadora do trabalhador¹⁶. Os registros dos empregos na RAIS são realizados pelo estabelecimento empregador. Para a construção dos salários foi utilizado o valor da remuneração¹⁷ nominal do trabalhador recebido no mês de dezembro de cada ano. O Índice Nacional de Preços ao Consumidor Amplo – IPCA (IBGE) das regiões metropolitanas foi utilizado para deflacionar os salários nominais, assumindo dezembro de 2014 como o período base. O IPCA é calculado apenas para onze regiões metropolitanas do Brasil: Belém; Fortaleza; Recife; Salvador; São Paulo; Rio de Janeiro; Goiânia; Distrito Federal; Belo Horizonte; Curitiba e Porto Alegre.

Desse modo, os índices das regiões metropolitanas foram extrapolados para os estados correspondentes. O critério de maior aproximação foi utilizado para os estados que não possuíam alguma das regiões metropolitanas contempladas no cálculo do índice¹⁸. Assim, o índice de preços de Belém foi extrapolado para os estados da região Norte, o de Fortaleza para

¹⁵ O Programa de Integração Social-PIS é um registro do Governo Federal para cada trabalhador com o objetivo de permitir o acesso aos benefícios sociais garantidos por lei no Brasil, a exemplo do seguro-desemprego e do Fundo de Garantia do Tempo de Serviço-FGTS (CAIXA, 2012).

¹⁶ O estudo optou por utilizar o PIS e não o Cadastro de Pessoa Física (CPF), pois o PIS apresentou menor número de informações ausentes relativamente ao CPF.

¹⁷ Na RAIS, a remuneração corresponde aos salários, ordenados, vencimentos, honorários, vantagens, adicionais e gratificações, excluindo o 13º salário (MTE, 2010).

¹⁸ Este procedimento foi adotado com base em Freguglia (2007).

os estados do Maranhão, Piauí e Rio Grande do Norte, o de Recife para a Paraíba, Alagoas e Sergipe, o do Rio de Janeiro para o Espírito Santo, o de Curitiba para Santa Catarina e o de Goiânia para os estados do Mato Grosso e Mato Grosso do Sul. O salário mensal real foi transformado em salário real hora, a partir da variável de horas semanais contratadas, disponível na RAIS¹⁹.

As características observadas dos trabalhadores extraídas da RAIS foram: a escolaridade em ciclos de estudo; a idade (18-65 anos); a experiência, em meses no mesmo vínculo; e o gênero. Na RAIS, a variável educação é definida por grau de instrução: (1) analfabetos; (2) até o 5º ano do ensino fundamental incompleto; (3) 5º ano completo do ensino fundamental; (4) do 6º ao 9º ano do ensino fundamental incompleto; (5) ensino fundamental completo; (6) ensino médio incompleto; (7) ensino médio completo; (8) ensino superior incompleto; e (9) ensino superior completo. Para a inclusão da educação no painel foram construídas variáveis *dummies* para os ciclos escolares, descritas na próxima seção. As características observadas extraídas do estabelecimento foram: tamanho, em número de empregos; município; e setor de atividade da CNAE 1.0 do IBGE. Além disso, o CNPJ foi utilizado como variável identificadora da firma.

Para identificar os possíveis efeitos de aglomeração sobre os salários reais, este estudo assumirá como aglomerados urbanos os arranjos populacionais definidos no estudo recente do IBGE (2015), sobre as concentrações urbanas do Brasil. O arranjo populacional é um agrupamento de municípios definido a partir de forte integração populacional, mensurada por movimentos pendulares para trabalho e estudo ou pela contiguidade das principais manchas urbanas²⁰. No estudo, o termo integração sintetiza os diversos contextos a partir dos quais o fluxo de pessoas entre municípios pode ser originado, a exemplo do crescimento de uma cidade e da localização de indústrias. Os fluxos de deslocamentos foram obtidos a partir de amostra do Censo Demográfico (IBGE, 2010). O IBGE identificou 294 arranjos populacionais, os quais eram compostos por 938 municípios e 55,9% (106.782.428) da população (CENSO, 2010). Cerca de 90% dos arranjos eram compostos por até 5 municípios, podendo sugerir a intensificação da formação de aglomerações em centros urbanos menores

¹⁹ O salário real do mês de dezembro foi dividido pelo número de horas mensal contratado, a partir da variável semanal disponível.

²⁰ As manchas urbanizadas foram definidas a partir da presença de edificações típicas de espaços urbanos e padrões de demarcação de ruas que facilitem as relações de vizinhança. Foram utilizadas imagens de satélite do *Google Earth* e *Google Maps* para a medição de distância entre as manchas urbanas (IBGE, 2015).

(IBGE, 2015). Os critérios específicos utilizados pelo IBGE para a definição dos arranjos encontram-se no Apêndice D.

Os 294 arranjos populacionais foram classificados, a partir de faixas populacionais, em 189 pequenos (população ≤ 100 mil habitantes), 81 médios ($100 \text{ mil} < \text{pop.} \leq 750 \text{ mil hab.}$) e 24 grandes ($\text{pop.} > 750 \text{ mil hab.}$). Para a definição das concentrações urbanas, os municípios isolados que não formam arranjos, mas pertencem ao alto escalão de urbanização brasileira, foram agrupados em grandes e médios arranjos, a partir das suas respectivas faixas populacionais. Portanto, foram definidas 26 grandes concentrações urbanas correspondendo a 2 municípios isolados e 24 arranjos. As médias concentrações representam 158 centros intermediários de articulação urbana, correspondendo a 81 arranjos e 77 municípios isolados. Neste sentido, foram definidas 184 médias e grandes concentrações urbanas (IBGE, 2015). Como o interesse deste trabalho é analisar as áreas com intensa dinâmica urbana no mercado de trabalho é suficiente assumir os municípios isolados como arranjos populacionais, mantendo uma única categoria de análise. Portanto, os arranjos e municípios isolados serão adotados como as áreas de emprego de interesse e os demais municípios que não pertencem aos arranjos serão as categorias de referência na estimação do modelo.

O estudo do IBGE somente identifica os arranjos populacionais, mas não propõe indicar as funções assumidas por cada centro urbano, o que já foi realizado no estudo das Regiões de Influência das Cidades – REGIC (IBGE, 2008). No entanto, na perspectiva espacial, os arranjos seguem o padrão da urbanização do Brasil e sua distribuição acompanha as metrópoles e capitais regionais identificadas no estudo sobre as REGIC. O Apêndice E mostra a distribuição e a densidade populacional (2010) dos arranjos no Brasil. A Tabela 3.1 apresenta as características dos arranjos populacionais.

Tabela 3.1 – Arranjos Populacionais no Brasil

Arranjos	Municípios	Arranjo/Brasil (%)	População	Arranjo/Brasil (%)
Pequeno	387	6,9	8.130.580	4,3
Médio	297	5,3	23.431.663	12,3
Grande	254	4,6	75.220.185	39,3
Total dos Arranjos	938	16,8	106.782.428	55,9
Municípios Isolados	79	1,4	15.750.765	8,3
Outros Municípios	4.548	81,8	68.222.606	35,8
Total Brasil	5.565	100	190.755.799	100

Fonte: Elaboração própria a partir do IBGE (2015), com base no Censo do IBGE (2010).

Tendo em vista as mudanças ocorridas nos limites administrativos e geográficos no Brasil, em função da criação e extinção de municípios ao longo do período analisado, o estudo fará uso de Áreas Mínimas Comparáveis (AMC). As AMCs são agregações de municípios que sofreram desmembramentos e alterações ou foram criados. Esse ajuste é realizado para que seja possível fazer comparações intertemporais espacialmente consistentes, considerando as novas áreas agregadas constantes no tempo (REIS et al., 2011). O número de municípios do Brasil evoluiu de 4.491 para 5.507 entre 1991 e 2000, alcançando 5.570 municípios em 2013.

A partir das 4.267 AMCs construídas pelo Instituto de Pesquisa Econômica Aplicada - IPEA (REIS et al., 2011), para o período de 1991 a 2000, foram criadas novas AMCs ao considerar as alterações municipais no período de 2001 a 2013. Nesse sentido, foram definidas 4.256 AMCs, para o período de 2002 a 2014. As AMCs são as unidades de referência para a criação das variáveis municipais deste estudo. Por fim, as AMCs foram compatibilizadas entre os arranjos populacionais para manter a consistência temporal das unidades espaciais analisadas. Portanto, foram definidos 365 arranjos populacionais (1.392 municípios), a partir dos 373 originais (incluindo os municípios isolados aqui assumidos como arranjos), dos quais 177 são de tamanho pequeno (495 municípios), 161 são médios (596 municípios) e 27 são grandes (301 municípios).

3.3 MODELO ECONOMETRICO

Este estudo utiliza como referência o modelo teórico da Economia Urbana desenvolvido por Combes, Duranton e Gobillon (2008) para identificar os efeitos de aglomeração sobre os salários individuais. O modelo AKM de decomposição estatística dos salários será utilizado para identificar a contribuição do *sorting* das características não observadas dos trabalhadores e firmas, para os salários locais e para os efeitos de aglomeração nos arranjos populacionais.

A equação salarial do modelo econométrico completo, que será estimado no primeiro estágio com base em diversas especificações alternativas, é dada por:

$$\begin{aligned} \ln(w_{it}) = & \beta_0 + \sum_e \beta_1^e educ_{it}^e + \beta_2 exp_{it} + \beta_3 exp_{it}^2 + \beta_4 age_{it} + \beta_5 age_{it}^2 + \\ & \beta_6 gender + \sum_s \beta_7^s sector_{it}^s + \beta_8 size_firmsect_{j(i,t)} + \sum_t \beta_9^t T_t + \\ & \sum_a \kappa_a lma_{it}^a + \theta_i + \psi_{j(i,t)} + \varphi_{i,j,t} + \varepsilon_{it} \end{aligned} \quad (3.34)$$

onde, $\ln(w_{it})$ é o logaritmo natural do salário-hora real do trabalhador i no tempo $t = \{2002, \dots, 2014\}$, $educ_{it}^e$ são *dummies* de escolaridade para quatro ciclos escolares $e = \{1, \dots, 4\}$, exp_{it} é a experiência do trabalhador, age_{it} é a idade, $gender$ é uma *dummy* de gênero (=1 se mulher), $sector_{it}^s$ são *dummies* para os setores de atividade $s = \{1, \dots, 16\}$, $size_firmsect_{j(i,t)}$ são *dummies* para o tamanho relativo da empresa j no setor de atividade correspondente e T_t são de *dummies* de tempo. A idade e experiência são incluídas em nível e na forma quadrática, conforme especificações da literatura. Para captar os efeitos de localização sobre os salários dos trabalhadores serão incluídos os efeitos fixos de cada arranjo populacional $a = \{1, \dots, 365\}$. Se o efeito fixo do arranjo populacional (*Labor Market Area-LMA*) for positivo e significativo, o vetor κ_a será a magnitude do prêmio salarial urbano. As heterogeneidades não observadas do trabalhador (i) e da firma (j) são representadas pelos efeitos fixos θ_i e $\psi_{j(i,t)}$, respectivamente. Por sua vez, o efeito fixo combinado de trabalhador e firma $\varphi_{i,j,t}$ representa o *matching* no mercado de trabalho local e ε_{it} é o termo de erro do modelo.

A descrição detalhada das variáveis utilizadas se encontra no Apêndice F. Na estimação do modelo, as categorias de referência para as variáveis indicadoras serão: (i) Educação: até ensino médio incompleto (1); (ii) Gênero: homem; (iii) Setor: indústria de transformação (4); (iv) Tempo: ano de 2002; e (v) Arranjos: municípios que não pertencem a algum arranjo. A equação de salários será estimada no primeiro estágio a partir de especificações incompletas do modelo geral (3.34). Cada especificação difere em relação ao componente não observado incluído, correspondente aos efeitos fixos do trabalhador, da firma ou do *matching*, além das variáveis que são fixas ou apresentam variação constante no tempo (gênero; idade; setor).

Abowd, Kramarz e Margolis (1999) realizaram a decomposição dos efeitos setoriais da França, estimados na equação salarial do primeiro estágio, a partir da equação (3.10) do seu modelo. Tomando o modelo AKM como referência, o segundo estágio efetuará uma decomposição dos efeitos de localização dos arranjos populacionais, em componentes associados às heterogeneidades de trabalhador e firma. Por sua vez, a especificação incluirá os controles típicos da Economia Urbana para identificar os efeitos de aglomeração sobre os salários locais, com base no modelo de Combes, Duranton e Gobillon (2008). Para isso, os efeitos de localização serão estimados na equação salarial do primeiro estágio sem os efeitos fixos de indivíduo e firma. Dessa forma, serão recuperados os efeitos de localização “brutos”. A especificação econométrica do segundo estágio é dada por:

$$\hat{\kappa}_a = \gamma_0 + \ln(\text{employment density}_{a,t}) \gamma_1 + \ln(\text{land area}_a) \gamma_2 + \text{touristic}_a \gamma_3 + \text{frontier}_a \gamma_4 + \text{region}_{r(a)} \gamma_5 + \bar{\theta}_{a,t} \gamma_6 + \bar{\psi}_{a,t} \gamma_7 + \rho_t + v_{a,t}. \quad (3.35)$$

A variável de interesse na estimação do segundo estágio do modelo é a densidade do emprego no arranjo ($\text{employment density}_{a,t}$), a qual capta os possíveis efeitos das economias de aglomeração sobre os salários locais (produtividade). Como sugerido por Combes, Duranton e Gobillon (2008), a área total do arranjo (quilômetros quadrados) é incluída para distinguir os efeitos de densidade dos efeitos de escala. O efeito de escala pode ser interpretado, possivelmente, como um limite para a expansão dos empregos. Por sua vez, esse limite pode ser associado aos custos da terra nos mercados locais.

A especificação incluiu como variável de dotação ou atributo local uma *dummy* indicando se o arranjo litorâneo possui uma função turística²¹ e uma *dummy* indicando se o arranjo localizado na fronteira do país possui integração com algum município do país vizinho²². Combes, Duranton e Gobillon (2008) sugerem que os efeitos das dotações produtivas podem ocorrer por três canais: menor custo de exportação; menor custo de insumos; ou maior produtividade local. A proximidade ao litoral pode afetar o salário local como uma amenidade (dotação não produtiva) ou como uma dotação produtiva se indicar a proximidade a portos de escoamento. Por sua vez, a proximidade à fronteira pode ser uma dotação produtiva, se a distância aos potenciais mercados consumidores é reduzida. As macrorregiões foram incluídas para considerar as significantes heterogeneidades existentes entre as mesmas, a exemplo da estrutura produtiva. A região Sudeste foi assumida como a categoria de referência.

Para captar a contribuição do *sorting* dos atributos não observados sobre os salários locais, a especificação incluiu as médias locais dos efeitos fixos de trabalhador ($\bar{\theta}_{a,t}$) e firma ($\bar{\psi}_{a,t}$) estimados no primeiro estágio. Os efeitos fixos de trabalhador e firma foram estimados no primeiro estágio com os mesmos controles utilizados na estimação dos efeitos fixos das áreas (educação, experiência e seu quadrado, tamanho da firma e efeitos de tempo), seguindo o

²¹ Os arranjos turísticos e de veraneio no litoral foram definidos pelo IBGE (2015) como arranjos de localização litorânea, voltados para as atividades turísticas e para atender à demanda de veraneio de metrópoles próximas. Existem municípios no litoral brasileiro que possuem características turísticas, mas não formam arranjos.

²² Os arranjos fronteiriços definidos pelo IBGE (2015) são aqueles situados na fronteira internacional do Brasil. Esses arranjos foram identificados a partir da contiguidade da mancha urbana entre os respectivos municípios, uma vez que o Censo (2010) não fornece informações sobre a localidade do país vizinho cujos movimentos pendulares para trabalho e/ou estudo são direcionados.

modelo AKM. Essa estratégia é importante para que seja possível realizar a decomposição dos efeitos de áreas em efeitos de trabalhador e firma. Por sua vez, algumas abordagens teóricas da Economia Urbana defendem que o *sorting* corresponde a uma parcela importante do denominado efeito de aglomeração.

Portanto, para encontrar o efeito bruto da aglomeração, medido pela densidade do emprego nos arranjos, o estudo estimará a contribuição da densidade do emprego para os efeitos fixos de trabalhador e firma. O efeito bruto da aglomeração será, então, definido pela soma do efeito de aglomeração residual, estimado na equação (3.35), com a contribuição da densidade para os efeitos de trabalhador e firma. Nesse sentido, a contribuição da densidade para os efeitos de trabalhador e firma será estimada pelas seguintes especificações:

$$\bar{\theta}_{a,t} = \alpha_0 + \ln(\text{employment density}_{a,t}) \alpha_1 + \ln(\text{land area}_a) \alpha_2 + \text{touristic}_a \alpha_3 + \text{frontier}_a \alpha_4 + \text{region}_{r(a)} \alpha_5 + \rho_t + \epsilon_{a,t} \quad (3.36)$$

$$\bar{\psi}_{a,t} = \omega_0 + \ln(\text{employment density}_{a,t}) \omega_1 + \ln(\text{land area}_a) \omega_2 + \text{touristic}_a \omega_3 + \text{frontier}_a \omega_4 + \text{region}_{r(a)} \omega_5 + \rho_t + e_{a,t}. \quad (3.37)$$

Os parâmetros de interesse nas equações (3.36) e (3.37) serão α_1 e ω_1 , os quais serão somados à estimativa γ_1 da equação (3.35) e conjuntamente interpretados como uma estimativa bruta dos efeitos de aglomeração urbana. A próxima seção apresenta mais detalhes sobre a estimação do modelo e a estratégia de identificação do trabalho.

3.4 MÉTODOS DE ESTIMAÇÃO E ESTRATÉGIA DE IDENTIFICAÇÃO

Esta seção tem como objetivo apresentar a estratégia de identificação e os métodos econométricos utilizados para estimar os resultados de interesse deste trabalho de tese.

3.4.1 Estratégia de Identificação

A estratégia de identificação deste trabalho de tese consiste na adoção de modelo longitudinal com dados pareados para controlar a heterogeneidade não observada de firma e trabalhador e o respectivo *sorting* sobre os salários e efeitos de aglomeração. A inclusão do efeito fixo do

indivíduo permite controlar o viés da habilidade não observada existente na estimação da equação salarial pelo método de MQO. Essa estratégia permite, ainda, a correção da auto seleção, associada à atração dos trabalhadores mais habilitados para os grandes centros urbanos, que pode ser confundida com os ganhos puros de aglomeração. Por sua vez, a inclusão dos efeitos fixos de localização e de firma permite controlar o *sorting* ou auto seleção decorrente da atração e manutenção das firmas mais produtivas nas áreas mais desenvolvidas. O controle do efeito de firma permite, ainda, lidar com a seleção dos trabalhadores mais qualificados para as firmas mais produtivas e corrigir o viés de agregação que resulta da inclusão somente dos efeitos setoriais.

As estratégias buscam a identificação robusta de possíveis efeitos de aglomeração sobre os salários locais, desde que os erros não sejam espacialmente correlacionados. A identificação dos efeitos fixos de trabalhador, firma e localização requer mobilidade dos trabalhadores entre firmas e arranjos populacionais. Por sua vez, para identificar os efeitos de firma e arranjos separadamente, é necessário ter alguma mobilidade de firmas entre os arranjos. Do mesmo modo, é necessário ter variabilidade entre os ciclos de educação analisados. A Tabela 3.2 apresenta a variação para as características indicadas das firmas e trabalhadores.

Tabela 3.2 – Decomposição da variação total na mobilidade de trabalhador e firma²³

Categoria	Variável	Variação	Desvio	Observações	Variação (%)
Trabalhador	Firma	<i>Overall</i>	2,63E+13	N = 6.048.393	
		<i>Between</i>	2,34E+13	n = 465.261	79,16
		<i>Within</i>	1,20E+13	T = 13	20,82
	LMA	<i>Overall</i>	125,20	N = 6.048.393	
		<i>Between</i>	116,10	n = 465.261	85,99
		<i>Within</i>	46,86	T = 13	14,01
	Educação	<i>Overall</i>	1,147	N = 6.048.393	
		<i>Between</i>	1,090	n = 465.261	90,4
		<i>Within</i>	0,356	T = 13	9,6
Firma	LMA	<i>Overall</i>	137,40	N = 1.842.890	
		<i>Between</i>	131,88	n = 271.951	92,12
		<i>Within</i>	25,75	T-bar = 6,78	3,51

Fonte: Elaboração própria, a partir da amostra da RAIS (MTE).

Nota: A variação no nível das firmas não soma 100%, pois as firmas não estão presentes em todos os anos do painel. O painel é balanceado para os indivíduos.

A variação total de cada regressor pode ser decomposta na variação intra-indivíduo (*within*) ao longo do tempo e na variação entre indivíduos (*between*). As variáveis que são fixas no

²³ O percentual de variação é construído com o valor da variância *within* (*between*), obtida a partir do desvio-padrão, a qual foi dividida pela variância total (*overall*). A mobilidade entre arranjos inclui, também, os migrantes que estão entrando e saindo de algum arranjo, em algum ano do painel.

tempo devem possuir variação *within* igual a zero (CAMERON; TRIVEDI, 2009). Como se pode observar na tabela, a condição de mobilidade é atendida para a identificação dos efeitos fixos, uma vez que a variação (%) *within* é diferente de zero para todas as variáveis consideradas.

Por outro lado, é possível que algumas características locais, a exemplo da densidade do emprego, sejam endógenas em relação aos salários locais. A endogeneidade na densidade do emprego pode surgir se existem choques locais contemporâneos (tecnológicos ou de outros tipos) que podem ampliar os salários e atrair migrantes, gerando uma causalidade reversa da densidade do emprego com os salários locais. Combes, Duranton e Gobillon (2008) apontaram que a possível endogeneidade no segundo estágio pode ser um problema potencialmente sério, particularmente se as estimativas de interesse são enviesadas e a direção do viés não for clara.

Se o choque amplia os salários locais, a correlação entre o resíduo e a densidade do emprego será positiva. Por sua vez, se existem amenidades de consumo²⁴ correlacionadas com a densidade e essas amenidades não são controladas, é possível encontrar uma correlação negativa com os resíduos do modelo. A causalidade reversa do primeiro caso pode gerar um viés positivo sobre os efeitos da densidade (sobre-estimado), enquanto o viés negativo do segundo caso gera estimativas subestimadas dos efeitos da densidade do emprego. Para lidar com a endogeneidade na densidade do emprego, este trabalho propõe o uso de variável instrumental, a qual será apresentada na próxima seção.

Além disso, a estratégia de identificação assumirá uma nova medida para a área utilizada na densidade do emprego dos arranjos. Portanto, o estudo propõe estimar duas versões adicionais da especificação do segundo estágio para encontrar evidências robustas dos efeitos de interesse. A primeira utilizará uma variável instrumental para a densidade do emprego, enquanto a última utilizará uma nova medida para a densidade do emprego local.

²⁴ As amenidades de consumo são definidas pelas dotações urbanas que não são diretamente relacionadas à produção, a exemplo do patrimônio cultural local, mas afetam a determinação de salários locais. Tais amenidades aumentam a disposição do consumidor a pagar por terra e reduzem a sua disponibilidade para as firmas, o que eleva as rendas da terra. Uma vez que trabalho e terra não são substitutos perfeitos, as amenidades de consumo tendem a reduzir o produto marginal do trabalho (COMBES; DURANTON; GOBILLON, 2008).

3.4.1.1 Endogeneidade e Variável Instrumental

Os efeitos das economias de aglomeração sobre a produtividade são frequentemente tratados como endógenos na literatura empírica da Economia Urbana, devido à causalidade reversa potencial. Para lidar com a possível endogeneidade da densidade do emprego e do respectivo efeito das economias de aglomeração, a literatura propõe o uso de variável instrumental. Se a endogeneidade é causada por choques locais contemporâneos, como pode ser o caso neste estudo, a distribuição populacional no passado não deve ser correlacionada com os níveis de produtividade atuais. Por sua vez, a densidade contemporânea é determinada pelos padrões anteriores de concentração populacional e da densidade. Nesse sentido, variáveis históricas, a exemplo das defasagens temporais longas da população, são comumente usadas nessa literatura, como instrumentos para a endogeneidade da densidade (GRAHAM et al., 2010).

O estudo de Combes, Duranton e Gobillon (2008) assumiu que a aglomeração populacional no passado não se relacionava com as diferenças contemporâneas na produtividade e utilizou a densidade populacional de 1831 a 1921 como instrumento para a densidade do emprego entre 1976 e 1998. Adicionalmente, outros instrumentos temporais foram incluídos no estudo, tais como as características demográficas médias e a parcela de trabalhadores no setor público. Os resultados mostraram que a densidade era somente marginalmente endógena, uma vez que não houve alterações significativas nas estimativas.

Por sua vez, uma variável instrumental tradicionalmente adotada na literatura de Economia do Trabalho para isolar os choques de demanda local por trabalho dos choques de oferta de trabalho ou de terra é conhecida como *Bartik Instrumental Variable (Bartik IV)* ou *Shift-Share Instrument*. O instrumento foi originalmente desenvolvido por Bartik (1991) para estimar o efeito do crescimento do emprego em áreas metropolitanas dos Estados Unidos sobre os resultados dos mercados de trabalho, como o salário real individual e a participação da força de trabalho entre 1980 e 1987. A proposta era tratar a endogeneidade existente na relação entre o crescimento do emprego e os resultados do mercado de trabalho local.

Para garantir a exogeneidade do instrumento, o estudo propôs isolar os componentes locais e nacionais dos choques que afetam o crescimento do emprego local. O principal canal através do qual o instrumento explica a variação na demanda local por trabalho é a mudança na demanda nacional pelos bens produzidos. A variável instrumental propõe estimar o

crescimento no emprego local induzido pelos choques na demanda nacional, tendo em vista que esses choques podem ser tomados como exógenos aos choques locais. Nesse sentido, o crescimento do emprego induzido pelos choques de demanda seria usado como um indicador do tamanho do choque nacional sobre a economia local.

Uma abordagem de *shift-share* propõe decompor o crescimento local em três componentes: crescimento nacional, *share* e *shift*. O instrumento *Bartik* para a demanda local por trabalho foi construído com o componente *share*, o qual é definido pela estrutura setorial do emprego local, em um período base, ponderada pela variação no emprego setorial nacional. O componente *share* calcula o crescimento adicional que teria ocorrido se cada setor local tivesse crescido na média nacional do respectivo setor. Portanto, o instrumento originalmente usado em Bartik (1991) foi definido pela seguinte expressão:

$$G_{mt} = \sum_j R_{jmb} \left[\frac{(R_{jnt} - R_{jnt-1})}{R_{jnb}} \right] \quad (3.38)$$

onde G_{mt} representa o crescimento do emprego na área metropolitana m no período t , R_{jmb} é o nível de emprego no setor j da respectiva área no período base b e $(R_{jnt} - R_{jnt-1})/R_{jnb}$ é o crescimento nacional no setor j como um percentual do emprego nesse setor no ano base. O crescimento nacional de um setor é adotado como uma *proxy* para a mudança na demanda nacional por seu produto, a qual teria efeitos sobre o emprego nos mercados locais. Alguns desses efeitos ocorrem com alguma defasagem temporal. Por sua vez, não é possível garantir que o componente *share* não apresente correlação com as mudanças na oferta de trabalho e de terra. No entanto, Bartik (1991) considera que essa correlação seja improvável, uma vez que as mudanças na demanda nacional dificilmente teriam uma correlação forte com as mudanças exógenas em amenidades locais, que afetam a oferta de trabalho, ou com as mudanças em sistemas de transporte e regras de zoneamento locais, que afetam a oferta de terra.

Após o estudo de Bartik (1991), muitos trabalhos aplicados passaram a adotar o instrumento original ou alguma versão modificada, para estimar o efeito dos choques de demanda por trabalho em diferentes abordagens da teoria econômica: elasticidade da oferta de trabalho local; migração; salários e outros resultados dos mercados de trabalho locais; preço dos imóveis; distribuição de renda, dentre outros. Blanchard e Katz (1992) avaliaram o efeito de choques adversos no emprego sobre a evolução de indicadores do mercado de trabalho

(emprego, desemprego e salário) entre os estados dos EUA no pós-guerra utilizando o instrumento *Bartik share* setorial para o emprego local. Bartik (1994) avaliou os efeitos do crescimento do emprego nas áreas metropolitanas dos EUA sobre a distribuição de renda, entre 1979 e 1988, utilizando um instrumento *Bartik share* setorial.

Card (2001) avaliou os efeitos da entrada de imigrantes entre 1980 e 1990 sobre o salário e o emprego em ocupações específicas nas áreas metropolitanas e grandes cidades dos EUA. O estudo utilizou um *Bartik share* dos imigrantes de cada origem em diferentes ocupações e cidades como instrumento para o crescimento da oferta de trabalhadores na ocupação. Autor e Duggan (2003) buscaram explicar como as mudanças na oferta dos benefícios de invalidez e na demanda por trabalhador de baixa qualificação afetaram a participação da força de trabalho de baixa qualificação nos EUA entre 1979 e 1998. Os autores utilizaram um *Bartik share* setorial, associado às mudanças na demanda por trabalho entre estados para identificar uma variação exógena na demanda pelos benefícios. Por sua vez, o crescimento do emprego setorial nacional inclui o emprego local e, portanto, pode ser movido pelos choques locais. Autor e Duggan (2003) foram pioneiros em solucionar esse problema, ao excluir o emprego do próprio estado no cálculo do crescimento do *share* do emprego setorial nacional para a construção do seu instrumento.

Moretti (2010) estimou o efeito multiplicador local de um aumento exógeno do emprego no setor *tradable* de uma cidade dos EUA sobre o emprego nos setores *tradable* e *nontradable*, entre 1980, 1990 e 2000. Esse estudo utilizou um *Bartik share* setorial para estimar a variação na demanda por trabalho. Diamond (2016) avaliou a divergência na escolha de localização entre cidades dos EUA, pelos indivíduos de alta e baixa qualificação, nos anos de 1980 a 2000 usando um *Bartik IV* para estimar a variação da produtividade local, a demanda por habitação e a preferência por salário relativamente ao preço da habitação e de outros bens. O estudo usou o *Bartik share* setorial ponderado pelo crescimento salarial nacional no setor e no grupo específico de qualificação, excluindo a cidade de referência e o seu entorno. Edlund, Machado e Sviatschi (2016) usaram o *Bartik share* setorial para estimar a demanda por trabalhadores qualificados (nível superior e *full time*) nas 27 maiores cidades dos EUA e seu impacto sobre o preço dos imóveis no período de 1980-2010. O estudo também excluiu a parcela do emprego da cidade em questão no cálculo do instrumento.

Faggio e Overman (2014) avaliaram o impacto do crescimento do emprego público sobre indicadores dos mercados de trabalho locais (emprego privado, desemprego, força de trabalho e população em idade ativa) utilizando um instrumento *Bartik* para esse setor. O estudo usou o *Bartik share* do setor público para instrumentalizar o crescimento do emprego público local nas unidades definidas como *Local Authorities* da Inglaterra. A criação de um emprego em um setor pode gerar novos empregos em outros (efeito multiplicador), ou ser compensada por realocações e efeito *crowding out*. O estudo não encontrou efeitos significantes sobre o crescimento do emprego privado na Inglaterra entre 2003-2007. No entanto, o emprego no setor público alterou a composição setorial do emprego no setor privado: efeito multiplicador de 0,5 empregos nos setores de construção e serviços (*non-tradable*) para cada emprego criado no setor público; efeito *crowding out* de 0,4 empregos na indústria (*tradable*). Entre 1999-2007, somente o efeito *crowding out* sobre o emprego privado foi encontrado.

Para a estimação do segundo estágio do modelo desta tese serão adotadas duas versões do instrumento *Bartik* para a densidade do emprego total nos arranjos: o instrumento *Bartik* do crescimento setorial do emprego e o *Bartik* do crescimento do emprego no setor público, utilizado em Faggio e Overman (2014). Para o cálculo do crescimento do emprego, o ano de 1991 será assumido como referência. O ano base escolhido busca garantir uma maior exogeneidade do instrumento em relação aos choques contemporâneos nos mercados locais. Como o período base corresponde a no mínimo uma década anterior ao primeiro ano de análise (2002-2014), os possíveis choques contemporâneos, agindo sobre o nível de salários dos arranjos (setor privado), não devem afetar a composição prévia do emprego local.

Assim como o original, o instrumento *Bartik* utilizado neste estudo deve prever o crescimento do emprego local a partir do crescimento nacional de cada respectivo setor, com base em um período inicial. Por outro lado, o canal através do qual o crescimento do emprego público, a partir do ano base, pode explicar a densidade do emprego nos arranjos é a tendência de longo prazo à localização ou concentração das firmas nos centros urbanos que já possuem uma estrutura consolidada de oferta de bens e serviços públicos. Como a lógica de expansão do emprego público no Brasil foi orientada pelas necessidades econômicas, sociais e institucionais geradas pelo desenvolvimento industrial e urbano (MATTOS, 2011), a estrutura prévia do emprego público pode representar uma demanda originada no processo de expansão urbana local. Esse desenvolvimento urbano prévio constitui as bases para a formação e

expansão dos principais centros urbanos atuais do país. Esse canal pode ser reforçado se o crescimento do setor público demandar bens e serviços produzidos pelo setor privado.

Por sua vez, o uso do instrumento do setor público como uma medida de dinâmica econômica local para prever o crescimento do emprego total, não seria adequado para os municípios pequenos com baixa dinâmica produtiva, uma vez que nesses casos a dinâmica do setor público não está associada à dinâmica econômica. Isso pode ocorrer, por exemplo, nos municípios pequenos com grande parcela dos empregos no setor público. No entanto, no estudo presente o instrumento do setor público será usado para medir a densidade do emprego nos principais mercados de trabalho do país, cuja dinâmica econômica não se restringe a esse setor.

Uma importante limitação do instrumento *Bartik*, disseminada nos estudos empíricos da literatura, é a possibilidade de que os choques na demanda nacional sejam movidos pelos choques locais das áreas que possuem grande parcela da produção total de um setor. Isso pode ocorrer se o emprego em um setor for muito concentrado em apenas um arranjo populacional do país. Como a estratégia de identificação é baseada na exogeneidade do crescimento do emprego nacional, a identificação não será válida se o instrumento for movido pela variável endógena. A partir do trabalho de Autor e Duggan (2003), esse problema tem sido solucionado com a exclusão de cada localidade em questão no cálculo do crescimento setorial nacional. Nesse sentido, a variação no emprego setorial nacional deve variar entre as localidades. Portanto, o instrumento *Bartik* utilizado neste estudo para a densidade do emprego total local é definido pela seguinte expressão:

$$Bartik_{lma,t} = \sum_s E_{lma,s,t_0} \left(\frac{E_{(n-lma),s,t} - E_{(n-lma),s,t_0}}{E_{(n-lma),s,t_0}} \right) \quad (3.39)$$

onde E_{lma,s,t_0} é o nível de emprego no setor s ²⁵, no ano base t_0 (1991), do arranjo populacional (LMA), e o termo $(E_{(n-lma),s,t} - E_{(n-lma),s,t_0})/E_{(n-lma),s,t_0}$ representa o crescimento do emprego nacional (n) do setor s em cada ano t (2002-2014), a partir do ano de referência t_0 , excluindo o emprego do local de referência ($n-lma$). O termo de crescimento do emprego será específico a cada arranjo, uma vez que será descontado o seu nível de emprego.

²⁵ Os setores foram definidos a partir da classificação de subsetores do IBGE, disponível desde o início da RAIS em 1985. Utilizou-se a divisão dos setores em 2 dígitos, correspondendo a 25 subsetores de atividade econômica.

Portanto, o instrumento *Bartik* utilizado neste estudo representa uma estimativa do nível de emprego contemporâneo em cada mercado de trabalho local, a partir do seu nível inicial. Nesse sentido, a estimativa do emprego local em cada ano requer o somatório do crescimento previsto para o emprego em todos os setores (s). No caso do *Bartik* do emprego público, o somatório não será necessário, uma vez que será utilizado apenas o crescimento previsto nesse setor.

A exclusão do emprego local no cálculo do seu instrumento é necessária para garantir a exogeneidade do crescimento previsto do emprego nacional. No entanto, se existe forte dependência produtiva entre localidades vizinhas mesmo excluindo o emprego local é possível que algum efeito residual do choque local ainda persista sobre o cálculo do crescimento nacional. Em geral, pode ser esperado que esse efeito seja maior quanto menor for o nível espacial de análise, a exemplo de cidades. Por sua vez, a estratégia de identificação deste estudo incide sobre a análise de unidades espaciais mais agregadas (LMA-arranjos), cuja agregação foi definida a partir da intensidade da dinâmica do emprego entre os municípios que as integram.

Nesse sentido, as unidades de análise neste estudo foram definidas conceitualmente a partir da dependência dinâmica entre os municípios membros com base nos fluxos de emprego e educação e na contiguidade da mancha urbana, que caracterizaria a relação de proximidade produtiva. Ademais, o estudo do IBGE (2015) adotou uma estratégia iterativa para a construção dos arranjos, prevendo uma nova definição do arranjo se fosse caracterizada uma integração entre dois ou mais arranjos. Nesse sentido, espera-se que as relações entre os municípios analisados ocorram internamente a cada arranjo e, portanto, a exclusão do arranjo (LMA) seria suficiente para a exogeneidade do seu instrumento. A seção de resultados apresentará as estatísticas descritivas da variável instrumental e o teste para a sua validade.

3.4.1.2 Densidade do Emprego e Área Iluminada

O efeito das economias de aglomeração urbana sobre os salários locais será mensurado neste estudo a partir da densidade do emprego em cada arranjo populacional. A densidade é definida pela razão entre o total de empregos no arranjo e sua área em quilômetros quadrados (km^2), extraída a partir da extensão territorial oficial dos municípios disponível no IBGE

(2013). No entanto, o uso da extensão territorial oficial como uma medida para a área efetivamente utilizada na habitação, produção e respectiva localização do emprego pode levar a uma subestimação dos efeitos de aglomeração. A subestimação dos efeitos da densidade pode ocorrer se a localização da produção e habitação ocorre em uma parcela relativamente restrita da área total do arranjo. Nesse sentido, tomando como exemplo um arranjo localizado na região Norte do Brasil, a qual possui municípios com grandes extensões territoriais, é possível encontrar estimativas do efeito de densidade abaixo do que seria previsto se sua área efetivamente ocupada tivesse sido adotada. A área efetivamente ocupada é definida pela parcela da extensão territorial total que concentra os empregos e as habitações.

Portanto, as estimativas dos estudos empíricos da literatura nacional e internacional podem subestimar os efeitos da densidade e das respectivas economias de aglomeração, se a diferença entre a área efetivamente ocupada e a extensão territorial oficial for muito alta. Nesse sentido, este trabalho propõe adotar uma nova medida para definir a área territorial efetivamente ocupada, a qual será utilizada para mensurar a densidade do emprego. A área ocupada será calculada a partir da área iluminada de cada arranjo populacional, obtida a partir da luminosidade noturna registrada nas imagens de satélites da Força Aérea dos EUA (*United States Air Force - USAF*), extraídas do *Defense Meteorological Satellite Program - Operational Linescan System* (DMSP-OLS). As imagens são processadas e distribuídas pelo *National Centers for Environmental Information - NCEI*, em uma série anual para o período de 1992 a 2013 (NCEI, 2013).

O estudo utilizará as imagens do arquivo *raster* denominado “*Average Visible, Stable Lights & Cloud Free Coverages*” oriundo do DMSP-OLS, o qual contém a luminosidade persistente das cidades ou de outros locais, após a limpeza de nuvens e incêndios, mas inclui a luminosidade da queima de gás. O último relatório anual da Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP, 2016) mostrou que apenas 12,6% da queima e perda de gás natural (milhões de m³) no Brasil em 2014 foi localizada em terra e todo o restante foi realizado no mar, considerando todas as unidades da federação produtoras²⁶. A média anual de queima e perda de gás natural em terra no período de 2006 a 2015 foi de 15,1%, sugerindo que o Brasil tem realizado a maior parte de sua queima no mar. Portanto, esse fator não deve

²⁶ Estados produtores de gás natural no Brasil: Amazonas, Maranhão, Ceará, Rio Grande do Norte, Alagoas, Sergipe, Bahia, Espírito Santo, Rio de Janeiro, São Paulo e Paraná (ANP, 2016).

limitar a medida de luminosidade obtida para os arranjos populacionais²⁷. Em alguns anos da série existe mais de um satélite capturando as imagens de luminosidade noturna. Nestes casos, é sugerido utilizar uma média das informações anuais (LOWE, 2014).

Os dados de um arquivo *raster* são compostos por *pixels*, cujo tamanho corresponde à resolução da imagem do *raster*. Nas imagens do satélite DMSP-OLS, cada *pixel* possui 30x30 *arcsec* (arco–segundos), o que equivale a aproximadamente 0,86 quilômetros quadrados (km²) na linha do Equador. A cada *pixel* é associado um valor de luminosidade variando de 0 a 63 com a intensidade da luz, onde o valor 255 representa os *pixels* sem luminosidade (LOWE, 2014). Os satélites circulam a Terra 14 vezes ao dia, entre 20:30 e 22:00 do horário local, registrando a intensidade da luz em cada localidade. Os registros ocorrem desde a década de 1970, mas os arquivos digitais tornaram-se disponíveis somente a partir de 1992 (HENDERSON; STOREYGARD; WEIL, 2012). Este estudo utilizou o *software* ArcGIS 10 para obter a área iluminada dos arranjos populacionais do Brasil, a partir dos dados contidos no arquivo *raster* do DMSP-OLS *Nighttime Lights*.

Como já foi observado anteriormente, este trabalho assume áreas mínimas comparáveis para os municípios que sofreram alterações em seus limites, com o objetivo de manter constante a área dos arranjos ao longo do tempo. Nesse sentido a área iluminada do ano de 2013 foi adotada como referência na definição da densidade do emprego em cada arranjo populacional ao longo do período de 2002 a 2014. O Apêndice G apresenta a imagem do satélite DMSP-OLS para a luminosidade noturna do Brasil no ano de 2013. Por sua vez, como teste de robustez, uma versão do modelo será estimada considerando a área iluminada do ano de 2007, para verificar se existem mudanças significativas nos resultados, já que a luminosidade não é constante no tempo.

A área iluminada de cada arranjo populacional será calculada a partir do número de *pixels* iluminados sobre o total de *pixels* que compõem o arranjo. A proporção de iluminação dos *pixels* que compõem o arranjo será, então, multiplicada pela área territorial oficial (km²), de modo que a área iluminada em cada arranjo será definida pela seguinte expressão:

²⁷ No entanto, se um país apresentar grande parcela da sua queima de gás localizada em terra é possível realizar a limpeza das imagens de satélite, excluindo os *pixels* aonde ocorrem as queimas. Nesses casos, o *National Centers for Environmental Information* – NCEI disponibiliza *shapefiles* (*shp*) com o georreferenciamento da queima de gás em diversos países (LOWE, 2014).

$$lighted_area_{a,t} = \delta_{a,2013} \times area_{a,t}, \quad (3.40)$$

$$\delta_{a,2013} = \frac{\#pixels\ iluminados_{a,2013}}{\#pixels_{a,2013}}$$

onde $\delta_{a,2013}$ é a razão entre o número de *pixels* iluminados e o número total de *pixels* do arranjo *a* no ano de 2013, obtidos a partir das imagens de satélite. Portanto, se todos os *pixels* apresentam luminosidade noturna em um arranjo ($\delta_{a,2013} = 1$), a sua área iluminada será equivalente à área territorial oficial. Em geral, é mais provável que a área iluminada seja equivalente à área total nos arranjos mais dinâmicos e desenvolvidos do país. A Tabela 3.3 mostra a proporção de iluminação na área total de cada um dos 27 grandes arranjos populacionais.

Tabela 3.3 – Distribuição da área total e iluminada nos grandes arranjos: 2007 e 2013

Categoria	Área (km²)	Sharelight 2007	Sharelight 2013	Light 2007	Light 2013
Sorocaba/SP	1169,93	0,994	0,988	1163,11	1156,29
Ribeirão Preto/SP	3373,58	0,614	0,734	2070,66	2477,40
Campo Grande	8092,95	0,166	0,164	1343,49	1324,11
Cuiabá/MT	4543,63	0,342	0,347	1555,74	1575,36
Florianópolis/SC	3200,75	0,710	0,700	2272,22	2242,08
Aracaju/SE	1629,95	0,973	0,995	1585,76	1621,11
São José dos Campos/SP	2722,06	0,782	0,799	2127,58	2173,97
João Pessoa/PB	1453,75	1,000	0,968	1453,75	1406,73
Teresina/PI	11183,4	0,190	0,215	2127,68	2404,06
Maceió/AL	1590,76	0,948	0,991	1508,03	1575,72
Natal/RN	1190,20	0,999	0,976	1189,24	1161,40
São Luís/MA	1412,27	0,917	0,924	1295,18	1305,09
Vitória/ES	1741,28	0,954	0,934	1661,37	1626,84
Baixada Santista/SP	2095,96	0,847	0,878	1775,14	1840,85
Manaus	11401,09	0,166	0,143	1896,08	1627,32
Campinas/SP	2423,69	1,000	1,000	2423,69	2423,69
Belém/PA	1819,23	0,922	0,828	1677,19	1506,95
Goiânia/GO	5774,65	0,545	0,577	3144,69	3332,13
Curitiba/PR	8541,30	0,550	0,560	4695,09	4779,78
Porto Alegre/RS	12097,42	0,349	0,377	4222,02	4563,80
Fortaleza/CE	3085,97	0,846	0,900	2612,17	2776,22
Brasília/DF	21440,84	0,324	0,322	6947,63	6904,02
Salvador/BA	3107,57	0,937	0,904	2910,32	2809,80
Recife/PE	3047,97	1,000	1,000	3047,97	3047,97
Belo Horizonte/MG	5593,56	0,864	0,894	4831,29	4998,85
Rio de Janeiro/RJ	6034,03	0,981	0,976	5919,23	5890,04
São Paulo/SP	6729,00	0,989	0,992	6652,34	6675,63

Fonte: Elaboração própria a partir das imagens de satélite do *Defense Meteorological Satellite Program-DMSP, Operational Linescan System-OLS* (NCEI, 2013) e IBGE (2013).

Notas: *Sharelight* é a parcela de *pixels* luminosos em cada área e *light* é a área iluminada obtida a partir do *share*.

A distribuição da proporção da área iluminada confirma a grande variação existente em relação às áreas efetivamente habitadas. A proporção iluminada da área total do arranjo de Manaus foi de apenas 14,3%, enquanto os arranjos de Campinas e Recife apresentaram uma proporção de iluminação de 100%, no ano de 2013. Entre os grandes arranjos, aqueles com maior extensão territorial apresentaram uma menor proporção de iluminação, a exemplo de Manaus (14,3%), Brasília (32,2%) e Porto Alegre (37,7%). Esses dados confirmam a hipótese de que a área efetivamente ocupada pode ser muito inferior à área total dos arranjos com grande extensão. Portanto, o uso da área iluminada deve contribuir para uma identificação mais adequada da magnitude dos efeitos da densidade.

De acordo com Lowe (2014) apenas recentemente os economistas começaram a fazer uso dos dados de densidade da luminosidade noturna extraídos de satélites, cujo uso tem demonstrado a sua significativa adequação como *proxy* para a atividade econômica local. Donaldson e Storeygard (2016) indicaram três principais vantagens nos dados de sensoriamento remoto para os economistas, a exemplo da luz noturna, precipitação, cobertura florestal e topografia. A primeira vantagem é o acesso a informações difíceis de serem obtidas por outros meios, com possibilidade de coleta de dados em painel a um custo marginal baixo. As demais vantagens estão associadas à alta resolução espacial e cobertura geográfica das informações. Os autores destacaram que as imagens de satélite em alta resolução têm se tornado uma poderosa ferramenta em Economia.

Entre os estudos recentes da literatura econômica que utilizaram dados de densidade da luminosidade noturna estão Henderson, Storeygard e Weil (2012), Bleakley e Lin (2012) e Michalopoulos e Papaioannou (2013). Henderson, Storeygard e Weil (2012) utilizaram as imagens de satélite da densidade da iluminação noturna como medidas do crescimento econômico associado ao PIB dos países. Os autores destacaram a qualidade dessa medida e sua utilidade particularmente nos países que não possuem registros da renda nacional ou cujos registros são inconsistentes. O estudo mostrou que nos EUA e no Canadá cerca de 69% e 94% dos *pixels*, respectivamente, não possuíam iluminação noturna, considerando a média do período de 1992 a 2008. No Brasil, 94% dos *pixels* não apresentaram luminosidade, de acordo com o estudo.

Bleakley e Lin (2012) utilizaram os dados de intensidade da iluminação noturna obtidos por satélite no ano de 2003 para medir a persistência na distribuição da atividade econômica entre antigas áreas de comércio dos EUA. De modo similar, Michalopoulos e Papaioannou (2013) utilizaram a densidade da luminosidade noturna dos anos de 2007 e 2008 como *proxy* para a variação regional da atividade econômica e do respectivo desenvolvimento entre países africanos. Alguns estudos pioneiros, não diretamente relacionados à literatura da Economia, já destacavam o potencial de uso das imagens de satélite para analisar padrões de densidade populacional, área urbana construída e consumo urbano de energia (WELCH, 1980; SUTTON et al., 1997). O estudo de Small et al. (2011) utilizou a intensidade da luz noturna de 2008, extraída de imagens de satélite, como um indicador urbano para os diversos graus de desenvolvimento antropogênico, uso humano da terra e habitação entre os continentes.

De acordo com Small, Pozzi e Elvidge (2005), as imagens de satélite da luminosidade antropogênica fornecem um meio para mapear a distribuição global das áreas urbanas. Além disso, as imagens oferecem uma métrica para análises comparativas de extensões urbanas. No entanto, o estudo destacou que existem inconsistências na relação entre as áreas iluminadas e os limites administrativos das áreas urbanas consideradas. A maior extensão espacial da área iluminada em relação à extensão geográfica da respectiva área foi denominada em alguns estudos como “*blooming*”. A inconsistência poderia resultar de três fatores: (i) limitações na resolução espacial do sensor e na captação da luminosidade; (ii) sobreposição na cobertura dos *pixels* adjacentes; e (iii) erros de georreferenciamento no processo de composição. A literatura prévia buscou reduzir esse problema mantendo apenas os *pixels* com frequência de detecção acima de um limite definido. No entanto, a solução também atenuava um grande número de valores menores de luminosidade e restringia a detecção de pequenas áreas iluminadas, bem como o conteúdo das informações de luminosidade. Por sua vez, o trabalho destacou que novas versões das imagens de satélite estavam sendo produzidas naquele período, as quais incluiriam melhorias e novas informações.

O presente estudo utilizará a versão mais atual dos dados de luminosidade noturna extraídos das imagens de satélite do DMSP-OLS (NCEI, 2013), de modo a considerar os ajustes realizados em relação às primeiras versões disponíveis. Por sua vez, a área iluminada dos arranjos será calculada a partir da proporção dos *pixels* iluminados, ao invés da agregação da área física correspondente aos *pixels* luminosos em cada arranjo. Nesse sentido, a área será definida pela proporção de *pixels* iluminados em cada unidade geográfica multiplicada pela

área oficial (km²). Essa forma de obter a área iluminada permite atenuar o problema de sobreposição de *pixels* luminosos entre duas áreas distintas.

Por sua vez, é importante ressaltar que a área iluminada será usada neste estudo para atenuar a possível subestimação da densidade do emprego ao utilizar a área oficial, particularmente nos arranjos cuja ocupação humana é concentrada em uma parcela relativamente pequena da sua área total. Nesse sentido, o estudo não propõe usar a área iluminada como uma medida de área de emprego, mas sim como uma medida de área efetivamente ocupada com habitações e/ou empresas. No entanto, é esperado que uma parcela da iluminação seja associada à localização das empresas e não apenas das residências. A seção de resultados apresentará as estatísticas descritivas da luminosidade noturna nos arranjos populacionais do Brasil e da medida de área obtida a partir dessas informações.

3.4.2 Métodos de Estimação Econométrica

Para encontrar os efeitos de localização dos arranjos populacionais, a equação salarial do primeiro estágio será estimada com o método de Mínimos Quadrados Ordinários Agrupado (*Pooled Ordinary Least Squares*-POLS), sem os efeitos de firma e trabalhador. O método POLS garante a estimação consistente dos parâmetros de interesse a partir da sua hipótese de exogeneidade do erro e da condição de posto pleno (WOOLDRIDGE, 2002):

$$E(\mathbf{x}'_{it}\varepsilon_{it}^{POLS}) = 0 \quad (3.41)$$

$$\text{rank}[E(\sum_{t=1}^T \mathbf{x}'_{it}\mathbf{x}_{it})] = K \quad (3.42)$$

onde ε_{it}^{POLS} é o termo de erro do modelo POLS e $\mathbf{x}$ é um vetor das variáveis explicativas, incluindo o efeito fixo de localização. No entanto, as estimativas do método devem ser enviesadas, uma vez que o modelo não reconhece a existência das heterogeneidades não observados da firma e do trabalhador. Nesse sentido, o suposto de exogeneidade não será válido, já que os fatores não observados são correlacionados com as covariadas do modelo, de modo que as hipóteses do método não serão satisfeitas:

$$E(\varepsilon_{it}^{POLS}|\mathbf{x}_{it}, FE_i, FE_j) \neq 0 \quad (3.43)$$

$$E(\mathbf{x}'_{it}FE_{i,j}) \neq 0 \quad (3.44)$$

onde FE_i é o efeito fixo individual, FE_j é o efeito fixo da firma e $\mathbf{FE}_{i,j}$ é um vetor composto por ambos os efeitos fixos. Além disso, a estimação do modelo POLS requer a adoção da matriz de variância robusta, uma vez que os erros serão serialmente correlacionados devido à presença dos fatores não observados em cada período (WOOLDRIDGE, 2002). Embora gere estimativas inconsistentes, o método será necessário para a estimação dos efeitos brutos de localização, os quais devem estar correlacionados com os efeitos não observados da firma e do trabalhador. Nesse sentido, o POLS será utilizado apenas para gerar as estimativas brutas dos efeitos de localização, necessárias para a decomposição do prêmio urbano em efeitos de firma e de trabalhador, no segundo estágio do modelo.

Os efeitos fixos da firma e do trabalhador, utilizados como variáveis explicativas na decomposição do segundo estágio, serão estimados a partir do modelo completo (3.34), excluindo apenas as características individuais cuja variação temporal é constante (gênero e idade) e o setor de atividade. A estimação do modelo econométrico completo adotará a abordagem iterativa proposta por Guimarães e Portugal (2010) para estimar modelos lineares com grande dimensão de efeitos fixos, tais como o modelo AKM. Para entender o funcionamento da abordagem, assumo o seguinte modelo, cujo interesse é obter estimativas de MQO para β :

$$y = X\beta + D_1\alpha_1 + D_2\alpha_2 + D_3\alpha_3 + \epsilon, \quad (3.45)$$

onde D_i são as variáveis indicadoras e α_i os efeitos fixos correspondentes. O algoritmo segue o Teorema de Frisch-Waugh-Lovell (FWL), em que y e X são regredidos contra cada D_i , gerando os resíduos u_y e u_x , e, então, u_y será regredido contra u_x . Correia (2014) desenvolveu um algoritmo de iteração de ponto fixo para implementar a abordagem no *software stata*. De maneira prática, o algoritmo faz uma regressão linear absorvendo (*demean*) os efeitos fixos indicados. A estratégia de iteração de ponto fixo é aplicada, alternando-se entre a estimação de β e α_i nas seguintes equações de solução:

$$\begin{aligned} \beta &= (X'X)^{-1}X'(y - D_1\alpha_1 - D_2\alpha_2 - D_3\alpha_3), \\ \alpha_1 &= (D_1'D_1)^{-1}D_1'(y - X\beta - D_2\alpha_2 - D_3\alpha_3), \\ \alpha_2 &= (D_2'D_2)^{-1}D_2'(y - X\beta - D_1\alpha_1 - D_3\alpha_3), \\ \alpha_3 &= (D_3'D_3)^{-1}D_3'(y - X\beta - D_1\alpha_1 - D_2\alpha_2). \end{aligned} \quad (3.46)$$

onde $(D_i' D_i)^{-1} D_i'$ é obtido como uma média de grupos dos resíduos da regressão de y sobre X . A estimação de β é realizada pela regressão de y (transformado) sobre X . O Apêndice H descreve o algoritmo desenvolvido por Correia (2014). Portanto, essa abordagem será adotada para estimar o modelo completo (3.34) e encontrar a solução condicional de MQO do modelo AKM, com a inclusão simultânea dos efeitos fixos de firma, trabalhador e arranjo. Guimarães e Portugal (2010) apontaram as vantagens de subtrair (*demean*) a influência dos efeitos fixos sobre as variáveis e utilizar somente os resíduos (FWL), em relação à inclusão de todos os efeitos diretamente. As vantagens estariam associadas ao aumento na velocidade de convergência das regressões do primeiro estágio, à possibilidade de testar especificações do modelo com as variáveis transformadas sem precisar lidar com os múltiplos efeitos fixos, e à redução da capacidade computacional necessária para grandes bancos de dados. Tais vantagens estão subjacentes aos métodos denominados de condicionais (*absorbed*).

Versões alternativas do modelo serão estimadas pelo método de Efeitos Fixos (*Fixed Effects* – FE) para a comparação com os métodos condicionais que controlam simultaneamente todos os efeitos fixos. O estimador de Efeitos Fixos reconhece a existência de correlação entre as heterogeneidades não observadas e os regressores do modelo e permite encontrar estimativas não enviesadas. O estimador FE realiza uma transformação *within* indivíduo na equação de interesse para eliminar os fatores não observados fixos no tempo. A transformação *within* subtrai da equação estimada a média temporal de cada variável calculada por indivíduo. Por sua vez, o estimador de Efeitos Aleatórios (*Random Effects* – RE) estima o modelo na presença dos fatores não observados no termo de erro, sob a hipótese de que esses fatores são ortogonais às covariadas. O estimador RE corrige a correlação serial no termo de erro decorrente da presença dos fatores não observados. Ambos os estimadores assumem exogeneidade estrita das variáveis explicativas condicionais aos fatores não observados (WOOLDRIDGE, 2002):

$$E(\varepsilon_{it}^{FE, RE} | \mathbf{x}_i, FE_i, FE_j) = 0 \quad (3.47)$$

A principal diferença entre os estimadores FE e RE é que o primeiro assume $E(\mathbf{FE}_{i,j} | \mathbf{x}_i) \neq 0$, enquanto o último requer a não existência de correlação entre os fatores não observados e as covariadas. A limitação do estimador de FE é que o modelo não permite incluir variáveis constantes no tempo, a exemplo do gênero. O teste de Hausman testa a hipótese de correlação

entre os fatores não observados e as covariadas com base na diferença entre as estimativas de FE e RE. Como FE é consistente quando existe a correlação especificada, enquanto RE é inconsistente, uma diferença estatisticamente significativa é uma evidência contra o estimador RE (WOOLDRIDGE, 2002). Portanto, este estudo realizou o teste de Hausman para avaliar qual o método mais apropriado de estimação do modelo de interesse, entre o FE e o RE. As estatísticas do teste indicaram a rejeição da hipótese nula a um nível de significância de 1%, sugerindo o estimador de FE como o mais adequado. Nesse sentido, o estudo deve apresentar apenas os resultados do estimador FE.

O método consistente denominado intra trabalhador–firma propõe estimar o modelo completo com um efeito fixo combinado de trabalhador e firma (*matching*). Nesse caso, será adotado o estimador FE tradicional. Versões incompletas da equação salarial serão estimadas, como no modelo AKM, para avaliar o poder relativo de cada efeito fixo na variação salarial. O modelo do segundo estágio será estimado inicialmente pelo método POLS, para a comparação com as versões estimadas com os outros métodos. Como a variável dependente no segundo estágio é o efeito dos arranjos estimado no primeiro estágio, sua estimativa depende da quantidade de indivíduos utilizados para a sua identificação. Nesse sentido, o termo de erro do segundo estágio pode conter parte da variância da estimação dos efeitos dos arranjos.

Portanto, o método de Mínimos Quadrados Generalizados Factível (*Feasible Generalized Least-Squares* – FGLS) será adotado para lidar com a possível heterocedasticidade no segundo estágio. Por outro lado, o uso de variáveis explicativas estimadas no primeiro estágio (efeitos do trabalhador e da firma) implica na inclusão de uma parte dos resíduos da estimação no termo de erro do segundo estágio. Nesse caso, os erros–padrão serão inválidos (FALLY; PAILLACAR; TERRA, 2010). Portanto, os erros–padrão serão computados através do método de *Bootstrap*.

Por fim, para identificar o efeito das economias de aglomeração urbana sobre os salários locais será estimado um modelo de variável instrumental (IV) com o método de Mínimos Quadrados em Dois Estágios (*Two-Stage Least Squares* – 2SLS). O método será necessário para lidar com a endogeneidade da densidade do emprego no segundo estágio. A variável instrumental deve satisfazer às seguintes condições: exogeneidade e existência de correlação parcial do instrumento com a variável endógena (WOOLDRIDGE, 2002). Portanto, a estimação com o método 2SLS assume:

$$Cov(z_{a,t}, v_{a,t}) = 0 \quad (3.48)$$

$$Cov(z_{a,t}, employment_density_{a,t}) \neq 0 \quad (3.49)$$

onde z é a variável instrumental e v é o termo de erro do segundo estágio. A condição (3.49) pode ser testada, enquanto a condição de exogeneidade do instrumento (3.48) deve ser garantida. Este estudo utilizará o *Bartik IV* como instrumento para a densidade do emprego em cada arranjo. Nesse sentido, a estimação do modelo com o método 2SLS será realizada em dois estágios:

$$employment_density_{a,t} = \theta_0 + \theta_1 Bartik_{a,t} + \theta_2 x_{a,t} + \vartheta_{a,t} \quad (3.50)$$

$$\kappa_a = \gamma_0 + \gamma_1 \widehat{employment_density}_{a,t} + \gamma_2 x_{a,t} + v_{a,t} \quad (3.51)$$

onde $x_{a,t}$ são as demais variáveis exógenas do modelo. A equação (3.50) define o primeiro estágio do modelo 2SLS ou a forma reduzida para a endógena, a qual estima um OLS para a parcela da densidade do emprego explicada pelo choque exógeno de demanda (θ_1), definido pelo *Bartik IV*. O segundo estágio, dado pela equação (3.51), busca identificar o efeito das economias de aglomeração a partir da variável instrumental utilizada. Os dois estágios são estimados conjuntamente através do método 2SLS. Quando existe apenas um instrumento, os estimadores IV e 2SLS serão idênticos. As hipóteses do estimador 2SLS são as mesmas do estimador POLS (exogeneidade e posto pleno), porém a variável instrumental é acrescida a essas hipóteses (WOOLDRIDGE, 2002).

Portanto, os modelos deste trabalho serão estimados a partir dos métodos tradicionais de painel e dos métodos específicos desenvolvidos para lidar com os múltiplos efeitos fixos em bancos de dados combinados de trabalhador e firma. Ademais, estimações adicionais com métodos específicos serão necessárias para tratar dos problemas de endogeneidade ou para corrigir potenciais problemas na variância dos estimadores decorrentes do uso de variáveis estimadas. O próximo capítulo descreve os resultados encontrados com a estimação dos respectivos modelos e as estatísticas descritivas do banco de dados utilizado.

4 RESULTADOS

Este capítulo apresenta as características da amostra e os resultados das estimações do modelo proposto nesta tese. Portanto, serão apresentadas as estimativas dos efeitos de localização sobre os salários e os resultados decorrentes de sua decomposição, no segundo estágio, em efeitos de firma e trabalhador. Além disso, o estudo realizou alguns exercícios de robustez cujos resultados serão expostos ao longo do capítulo.

4.1 CARACTERÍSTICAS DA AMOSTRA

As estatísticas descritivas apresentadas nesta seção são baseadas na amostra dos micro dados da RAIS-MTE (2002-2014) utilizada para a estimação dos resultados dessa tese. A descrição detalhada do processo de geração da amostra está disponível na seção de metodologia. A Tabela 4.1 apresenta as estatísticas descritivas das variáveis utilizadas na estimação da equação de salários, correspondente ao primeiro estágio do modelo.

Tabela 4.1 – Estatísticas descritivas das variáveis do primeiro estágio (2002-2014)

Variável	Outros Municípios				Arranjos			
	Média	Desvio	Min.	Max.	Média	Desvio	Min.	Max.
<i>lwreal_hr</i>	2,28	0,82	0,70	6,45	2,73	0,93	0,70	6,87
<i>age</i>	38,76	9,18	18	65	38,90	8,95	18	65
<i>age2</i>	1.586,23	731,95	324	4.225	1.593,01	718,44	324	4.225
<i>exp</i>	107,36	87,06	0	585,9	101,88	88,60	0	599,7
<i>exp2</i>	19.105,58	28.240,35	0	343.278,8	18.229,63	28.658,20	0	35.9640,1
<i>size_firmsect</i>	0,74	1,65	0	76,95	1,04	1,86	0	65,49
	Observações		Frequência (%)		Observações		Frequência (%)	
<i>Educ</i>								
< ensino médio		399.584		49,46		1.692.314		32,29
ensino médio		283.303		35,07		1.942.707		37,07
superior incompleto		25.104		3,11		261.497		4,99
superior		99.829		12,36		1.344.055		25,65
<i>Gender</i>								
Homem		618.368		76,55		3.689.793		70,41
Mulher		189.452		23,45		1.550.780		29,59
<i>Tamanho da firma</i>								
< 100 trabalhadores		414.063		51,3%		1.863.699		35,6%
100 ≤ trab. < 500		182.323		22,6%		1.409.403		26,9%
≥ 500 trabalhadores		211.434		26,2%		1.967.471		37,5%
Observações (N x T)		807.820		13,4		5.240.573		86,6

Fonte: Elaboração própria, a partir da amostra da RAIS (2002-2014).

Nota: A amostra total possui 6.048.393 observações.

A amostra sugere características mais favoráveis nos mercados de trabalho dos arranjos populacionais. A média e o valor máximo do logaritmo do salário real hora (*lwreal_hr*) é maior nos arranjos do que nos municípios que não pertencem a um arranjo. No entanto, os arranjos possuem maior dispersão salarial (desvio), sugerindo a concentração de trabalhadores de alta e baixa remuneração nessas áreas. Os arranjos possuem maior parcela de trabalhadores com escolaridade igual ou superior ao ensino médio completo (67,7%) do que os municípios que não pertencem a um arranjo (50,5%). Portanto, os arranjos possuem maior salário real médio, a preços de 2014, ao mesmo tempo em que concentram os trabalhadores mais qualificados. Essa relação preliminar sugere a necessidade de explicar a maior produtividade dos arranjos, condicional às heterogeneidades individuais.

A idade média dos trabalhadores na amostra é de 39 anos e existe maior dispersão (desvio) da experiência de trabalho nos arranjos do que fora deles. Os trabalhadores do gênero masculino são maioria na amostra, em torno de 70%. Nos arranjos, predominam os trabalhadores com ensino médio completo (37%), enquanto fora dos arranjos os trabalhadores com escolaridade abaixo do ensino médio são maioria (49%). Por sua vez, as firmas que estão nos arranjos populacionais apresentam um tamanho 4% maior em relação à média do seu setor, enquanto fora dos arranjos as firmas possuem tamanho 26% menor do que a média. Em números absolutos é possível perceber o mesmo comportamento, já que nos arranjos as firmas de maior tamanho (número de empregos ≥ 500) predominam, enquanto os municípios que não pertencem a arranjos concentram as firmas de menor tamanho (< 100 empregos). Essa última evidência sugere a importância de investigar o prêmio urbano condicional às heterogeneidades das firmas, particularmente se as firmas maiores são as mais produtivas.

A Tabela 4.2 apresenta a parcela de trabalhadores da amostra que mudam de firma e de arranjo, a parcela de firmas que mudam de arranjo e a distribuição dos trabalhadores na amostra por tamanho do arranjo e entre os setores produtivos, excluindo o setor público. Os arranjos concentram 86,6% dos trabalhadores da amostra, dos quais 59% estão nos grandes arranjos. Essa proporção é compatível com a distribuição de empregos da RAIS para o mesmo período (Apêndice C). Para a estimação do efeito de localização em cada arranjo e do possível prêmio salarial urbano, os municípios que não pertencem a um arranjo foram assumidos como a categoria de referência. A indústria de transformação concentra a maior parcela de trabalhadores da amostra (30,7%) e será assumida como categoria de referência nas estimações que incluam o setor.

Tabela 4.2 – Distribuição dos trabalhadores por tamanho do arranjo populacional e setor e mobilidade dos trabalhadores e firmas **continua**

Variável	Obs.	Freq. (%)
<i>Tamanho do arranjo (LMA)</i>		
Arranjo pequeno (177)	248.820	4,11
Arranjo médio (161)	1.423.823	23,54
Arranjo grande (27)	3.567.930	58,99
Total de trabalhadores nos arranjos	5.240.573	86,64
<i>Setor de atividade econômica</i>		
1. Agricultura, pecuária, silvicultura e exploração florestal	176.879	2,92
2. Pesca	1.643	0,03
3. Indústrias extrativas	67.992	1,12
4. Indústrias de transformação	1.855.156	30,67
5. Produção e distribuição de eletricidade, gás e água	198.540	3,28
6. Construção	157.591	2,61
7. Comércio e reparação de veículos automotores, objetos pessoais e domésticos	839.551	13,88
8. Alojamento e alimentação	100.447	1,66
9. Transporte, armazenagem e comunicações	580.469	9,6
10. Intermediação financeira, seguros, previdência complementar e serviço relacionado	498.466	8,24
11. Atividades imobiliárias, aluguéis e serviços prestados as empresas	857.658	14,18
12. Educação	165.953	2,74
13. Saúde e serviços sociais	306.399	5,07
14. Outros serviços coletivos, sociais e pessoais	239.300	3,96
15. Serviços domésticos	515	0,01
16. Organismos internacionais e outras instituições extraterritoriais	1.834	0,03
<i>Mobilidade do Trabalhador (i)</i>		
Trabalhador muda de LMA	178.289	38,32
Trabalhador muda de firma	286.389	61,54
Total de trabalhadores na amostra	465.261	100
<i>Mobilidade da Firma (j)</i>		
Firma muda de LMA	16.704	6,14
Total de firmas na amostra	271.951	100
Combinações de trabalhador-firma (ij)	1.081.990	100
Total de Observações na Amostra (N x T)	6.048.393	100

Fonte: Elaboração própria, a partir da amostra da RAIS (2002-2014).

Nota: *Labor Market Areas* – LMA são os arranjos populacionais.

A estratégia de identificação deste estudo requer mobilidade dos trabalhadores entre firmas e arranjos e das firmas entre arranjos, ao longo do painel. A Tabela 4.2 mostra que as condições de identificação são satisfeitas para o estudo. Nesse sentido, 38% dos trabalhadores da amostra realizaram alguma mudança de firma, enquanto 61,5% se moveram entre arranjos ou para algum arranjo. A mudança também pode representar os trabalhadores que saíram de algum arranjo. Por sua vez, 6,1% das firmas da amostra apresentaram mobilidade entre arranjos, incluindo aquelas que saíram ou entraram em algum arranjo. Portanto, as estatísticas descritivas sugerem que a amostra permite identificar os três efeitos fixos simultaneamente, do trabalhador, da firma e do arranjo.

A Tabela 4.3 apresenta as estatísticas descritivas das variáveis utilizadas na estimação do segundo estágio do modelo. A amostra contemplou apenas 362 arranjos do total de 365

definidos neste estudo, a partir da classificação do IBGE (2015). Os três arranjos não contemplados eram de tamanho pequeno (população ≤ 100 mil habitantes) e não apresentaram um número de empregos suficiente para ter representatividade na amostra gerada. A estimativa do efeito fixo dos arranjos sugere um prêmio salarial urbano médio de 2% nos respectivos mercados de trabalho locais. As estimativas dos efeitos fixos médios dos trabalhadores e das firmas, nos respectivos arranjos, sugerem maior heterogeneidade (desvio) dos efeitos de firma do que dos efeitos individuais. Essa observação pode indicar a presença de firmas mais heterogêneas nos arranjos do que os trabalhadores. Por sua vez, aproximadamente 40% da distribuição regional das observações dos arranjos está concentrada na macro região Sudeste, a qual será assumida como região de referência nas estimações econométricas. Os arranjos fronteiros e os turísticos-litorâneos representam 6,1% e 4,4% das observações no painel dos arranjos, respectivamente.

Tabela 4.3 – Estatísticas descritivas das variáveis do segundo estágio (2002–2014)

Variável	Média	Desvio	Min.	Máx.
<i>fe_lma</i>	0,02	0,27	-1,15	1,05
<i>ldensemp_lma</i>	2,41	1,75	-4,42	7,05
<i>larea_lma</i>	7,40	1,25	4,30	11,90
<i>fe_i_lma</i>	0,04	6,37	-12,53	159,18
<i>fe_j_lma</i>	0,71	15,23	-689,25	167,50
Observações			Frequência (%)	
<i>Region</i>				
Norte		341	7,28	
Nordeste		923	19,72	
Centro-Oeste		401	8,57	
Sudeste		1.859	39,71	
Sul		1.157	24,72	
<i>Frontier</i>		287	6,13	
<i>Touristic</i>		208	4,44	
Arranjos (LMA)		362	100	
Observações (LMA x T)		4.681	100	

Fonte: Elaboração própria, a partir da amostra da RAIS (2002-2014).

A Tabela 4.3 mostra, ainda, o logaritmo da densidade do emprego nos arranjos, cuja média foi de 2,41. A Figura 4.1 apresenta a relação entre os salários médios reais de 2002 a 2014, a preços de 2014, e a densidade do emprego nos arranjos populacionais do Brasil. O ajuste linear mostra uma correlação positiva entre as variáveis, sugerindo salários crescentes quanto maior a densidade. O percentual da variação salarial dos arranjos explicada pela densidade do emprego foi de 32%, sendo 30% apenas nos grandes arranjos²⁸. A estimativa encontrada no

²⁸ O Apêndice I apresenta as características populacionais, o número de empregos formais, o salário real e a densidade dos arranjos populacionais, juntamente com a classificação dos grandes arranjos.

estudo de Combes, Duranton e Gobillon (2008) para essa variação foi de 51%, nas áreas de emprego da França em 1998. Portanto, essa correlação positiva pode ser tomada como um fato estilizado nos trabalhos que tratam dos efeitos das economias de aglomeração urbana nos mercados de trabalho locais.

Figura 4.1 – Salário real e densidade do emprego nos arranjos (2002–2014)

Painel 1: Arranjos Populacionais

Painel 2: Grandes Arranjos Populacionais

Fonte: Elaboração própria a partir da RAIS (2002-2014).

Nota: Salário real deflacionado pelo IPCA (base: dezembro/2014).

Por outro lado, a estimativa da elasticidade dos salários em relação à densidade do emprego nos arranjos populacionais do Brasil foi estimada em 6,7%, para o período de 2002 a 2014. Combes, Duranton e Gobillon (2008) encontraram uma elasticidade de 4,9% para a França. Nesse sentido, embora a variação salarial local explicada pela densidade do emprego tenha se mostrado menor no Brasil do que na França, o potencial de aumento dos salários locais, mensurado pela elasticidade da densidade do emprego, se mostrou maior no Brasil.

4.2 PRÊMIO SALARIAL URBANO DOS ARRANJOS POPULACIONAIS

Essa seção apresenta os resultados da estimação econométrica da equação de salários correspondente ao primeiro estágio do modelo. As tabelas do Apêndice J apresentam os resultados completos das estimações deste capítulo. Inicialmente, são apresentadas as estimativas do prêmio salarial urbano médio de todos os arranjos populacionais do Brasil, no período de análise (2002–2014). A Tabela J.1 apresenta a estimação da equação de salários (3.34) substituindo os efeitos fixos de cada arranjo por uma *dummy* indicando se o trabalhador

está em um arranjo. As colunas de 1 a 4 apresentam as estimativas do prêmio médio para a amostra principal deste estudo. Os resultados dessa tabela foram estimados utilizando os métodos POLS e Condicional.

As evidências encontradas mostram um prêmio salarial urbano médio de 24% (POLS) para os trabalhadores dos arranjos populacionais, ao considerar as características observadas dos trabalhadores e do emprego. Por sua vez, o prêmio é reduzido para 2,2% ao incluir os efeitos não observados dos indivíduos e para 1,3% ao incluir os efeitos de firmas, ambos incluídos separadamente. Ao considerar simultaneamente as heterogeneidades não observadas dos trabalhadores e firmas o prêmio urbano dos arranjos desaparece. Esse resultado preliminar sugere que uma grande parcela do prêmio urbano seria explicada pelas características não observadas específicas dos indivíduos e das firmas que estão localizados nos arranjos. Como robustez, os resultados foram estimados com uma amostra que não realizou qualquer seleção entre as firmas que se movem entre arranjos (Colunas 5–8), ao contrário da amostra principal. Além disso, as colunas de 9 a 12 apresentam os resultados estimados para uma amostra que exclui os *outliers* de salário²⁹. Novamente, o prêmio desapareceu após o controle dos efeitos individuais e de firma. A magnitude do prêmio sem os efeitos fixos variou entre 26% (sem seleção de firmas) e 22% (sem *outliers*), sugerindo a robustez dos resultados.

A Tabela J.2 estima a mesma equação de salários, incluindo os efeitos fixos de cada um dos 365 arranjos separadamente em substituição à *dummy* de arranjo. A inclusão propõe considerar a substancial heterogeneidade existente entre os diversos mercados de trabalho locais do Brasil na identificação dos respectivos prêmios urbanos. Inicialmente, foram incluídas *dummies* para o tamanho dos arranjos (colunas 1–5). Ao considerar apenas as características observadas dos trabalhadores e do emprego e os efeitos temporais, foi encontrado um prêmio salarial urbano de 29,4% para os grandes arranjos, de 17% para os médios e de 7,4% para os arranjos pequenos (POLS). Ao controlar separadamente os efeitos de firma e indivíduo o prêmio é reduzido significativamente, mas ainda existe um ganho salarial atribuído aos arranjos com magnitude entre 1% e 2% para os pequenos, de 1,6% para os médios e de 1,7% a 2,7% para os grandes (colunas 2 e 3). O prêmio desaparece para os três grupos de arranjos após o controle da heterogeneidade não observada das firmas e trabalhadores (colunas 4 e 5).

²⁹ Foram excluídos os extremos da distribuição do salário real hora, correspondendo aos valores 1% maiores e menores.

Os resultados iniciais sugerem um alto poder explicativo dos atributos não observados das firmas e trabalhadores sobre os ganhos salariais atribuídos diretamente à localização. O controle de cada grupo de efeito individualmente reduz substancialmente a magnitude do prêmio urbano, mas não é suficiente para dissipá-lo. No entanto, o prêmio desaparece ao serem considerados ambos os efeitos simultaneamente. Portanto, essas evidências mostram a importância do controle simultâneo dos efeitos de firma e trabalhador para a obtenção de estimativas consistentes, tal como sugerido pelo modelo AKM. Os resultados estimados, como robustez, para a amostra aleatória em firmas (Tabela J.3) e para a amostra sem *outliers* (Tabela J.4) foram consistentes com as evidências encontradas. A magnitude do prêmio foi de 6,1% a 7% para os pequenos arranjos, de 16% a 18% para os médios e de 28% a 32% para os grandes. Além disso, o prêmio também desapareceu entre os grupos de arranjos após o controle simultâneo dos efeitos de firma e trabalhador.

Por outro lado, o comportamento das características observadas dos trabalhadores se mostrou consistente com as estimativas da literatura de determinação de salários individuais. Os trabalhadores com maior escolaridade apresentam maior retorno salarial, assim como os mais experientes. O retorno à educação, em ciclos escolares, é reduzido significativamente após o controle dos efeitos fixos da firma, do trabalhador ou de ambos. Os trabalhadores formais com ensino superior completo apresentaram retorno entre 141% (POLS) e 14% (FEij), em relação aos trabalhadores com escolaridade até o ensino médio incompleto. O retorno ao ensino médio completo variou entre 41,6% (POLS) e 0,004% (CONDij). A idade e a experiência apresentam retorno decrescente (formas quadráticas) e as mulheres experimentam salários em torno de 34% abaixo dos salários dos homens. As firmas que possuem tamanho maior (número de empregos) do que a média do seu setor geram efeitos positivos sobre os salários individuais, variando entre 0,004% e 2,8%. Apenas 3 setores, entre os 16 considerados, apresentaram salário acima da média da indústria de transformação em todos os modelos: eletricidade, gás e água; intermediação financeira; e indústria extrativa (Tabela J.2).

A Tabela 4.4 apresenta as estimativas econométricas dos efeitos de localização nos grandes arranjos, por ordem de tamanho populacional (2010), obtidas para as diferentes especificações da equação de salários (3.34). Os resultados completos estão disponíveis nas colunas de 6 a 12 da Tabela J.2 do Apêndice.

Tabela 4.4 – Efeitos de localização nos grandes arranjos sobre o salário-hora real:
primeiro estágio **continua**

<i>ln (w real hora)</i>	(I) POLS	(II) COND i	(III) COND j	(IV) FE (i, j)	(V) COND(i,j)
Sorocaba	0,3381*** (0,0073)	0,0211 (0,0174)	-0,0052 (0,0344)	-0,0697*** (0,0065)	-0,0649** (0,0321)
Ribeirão Preto	0,2468*** (0,0087)	0,0302** (0,0119)	-0,0038 (0,0234)	-0,0142 (0,0088)	-0,0126 (0,0167)
Campo Grande	0,0853*** (0,0114)	0,0053 (0,0123)	0,0307* (0,0185)	-0,0078 (0,0109)	-0,0118 (0,0185)
Cuiabá	0,1343*** (0,0130)	0,0197 (0,0226)	0,0139 (0,0292)	-0,0033 (0,0113)	0,0115 (0,0218)
Florianópolis	0,2888*** (0,0076)	0,0197 (0,0152)	0,1717*** (0,0293)	0,0852*** (0,0081)	0,0882*** (0,0178)
Aracaju	-0,0605*** (0,0130)	-0,0069 (0,0180)	-0,0020 (0,0409)	0,0166 (0,0119)	0,0253 (0,0200)
São José dos Campos	0,4181*** (0,0062)	0,0255** (0,0129)	0,0361* (0,0216)	-0,0322*** (0,0056)	-0,0276 (0,0182)
João Pessoa	-0,1755*** (0,0107)	-0,0157 (0,0168)	0,0689 (0,0441)	0,0448*** (0,0141)	0,0376 (0,0374)
Teresina	-0,1341*** (0,0119)	0,0359* (0,0207)	-0,0098 (0,0260)	-0,0059 (0,0096)	0,0036 (0,0254)
Maceió	-0,0579** (0,0130)	-0,0029 (0,0190)	0,0415 (0,0361)	0,0538*** (0,0115)	0,0504 (0,0335)
Natal	-0,1173*** (0,0107)	-0,0236 (0,0155)	0,0071 (0,0344)	-0,0055 (0,0120)	-0,0084 (0,0303)
São Luís	-0,0350*** (0,0130)	0,0325 (0,0214)	0,0424 (0,0400)	-0,0213 (0,0198)	0,0033 (0,0317)
Vitória	0,2162*** (0,0074)	0,0468*** (0,0146)	0,0281 (0,0373)	0,0261*** (0,0086)	0,0378 (0,0290)
Baixada Santista	0,4178*** (0,0075)	0,0574*** (0,0137)	0,0156 (0,0184)	-0,0123** (0,0057)	-0,0072 (0,0137)
Manaus	0,2714*** (0,0095)	0,0711** (0,0329)	-0,1198* (0,0709)	-0,0245*** (0,0092)	-0,0198 (0,0377)
Campinas	0,4381*** (0,0041)	0,0317*** (0,0104)	0,0275 (0,0233)	-0,0430*** (0,0035)	-0,0334** (0,0153)
Belém	0,1078*** (0,0095)	0,0478*** (0,0159)	0,0539 (0,0486)	0,0956*** (0,0105)	0,0912*** (0,0259)
Goiânia	0,1203*** (0,0077)	0,0277 (0,0206)	0,0331 (0,0399)	-0,0091 (0,0079)	-0,0089 (0,0285)
Curitiba	0,2859*** (0,0046)	0,0535*** (0,0103)	0,0326 (0,0218)	-0,0063 (0,0060)	0,0055 (0,0195)
Porto Alegre	0,3491*** (0,0044)	0,0258*** (0,0100)	0,0359 (0,0220)	-0,0170*** (0,0043)	-0,0150 (0,0165)
Fortaleza	-0,0538** (0,0066)	0,0020 (0,0105)	-0,1061*** (0,0288)	-0,0190*** (0,0062)	-0,0174 (0,0119)
Brasília	0,3922*** (0,0059)	0,0916*** (0,0286)	0,0759*** (0,0215)	-0,0121*** (0,0031)	0,0020 (0,0232)
Salvador	0,1137*** (0,0062)	0,0054 (0,0108)	0,0591** (0,0279)	-0,0196*** (0,0064)	-0,0053 (0,0152)
Recife	-0,0800*** (0,0054)	0,0252** (0,0109)	-0,0217 (0,0282)	-0,0265*** (0,0033)	-0,0289* (0,0163)
Belo Horizonte	0,2344*** (0,0042)	0,0112 (0,0080)	0,0071 (0,0162)	-0,0026 (0,0043)	0,0034 (0,0113)
Rio de Janeiro	0,2778*** (0,0035)	0,0829*** (0,0090)	0,0540** (0,0224)	-0,0045 (0,0040)	-0,0007 (0,0140)
São Paulo	0,4239*** (0,0025)	0,0427*** (0,0099)	0,0331** (0,0149)	-0,0238*** (0,0022)	-0,0160 (0,0128)

<i>ln (w real hora)</i>	conclusão				
	(I)	(II)	(III)	(IV)	(V)
	POLS	COND i	COND j	FE (i, j)	COND(i,j)
Observações (N x T)	6.048.393	6.048.393	6.048.393	6.048.393	6.048.393
R^2	0,606	0,923	0,794	0,113	0,943
R^2 Ajustado	0,606	0,917	0,784		0,935
<i>Within</i> R^2		0,437	0,311	0,379	0,360

Fonte: Elaboração própria, a partir da RAIS (2002-2014).

Notas: Erro padrão entre parênteses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Resultados completos na Tabela J.2.

A coluna I da Tabela 4.4 mostra os resultados da estimação da equação salarial pelo método *POLS*, que inclui as características observadas dos trabalhadores e do emprego, os efeitos de localização em cada um dos 365 arranjos e as *dummies* de tempo, sem controlar os efeitos fixos de firma e trabalhador. As colunas FEi e FEj da Tabela J.2 do Apêndice apresentam os resultados da estimação pelo método de Efeitos Fixos (FE)³⁰, tradicional para modelos em painel. As estimativas dos modelos de Efeitos Fixos, que incluem apenas os efeitos do trabalhador (FEi) ou da firma (FEj), foram idênticas às estimativas dos modelos condicionais que incluem os mesmos efeitos. Nesses casos, o estudo focará apenas nos resultados dos modelos condicionais. O estimador de Efeitos Fixos controla a correlação entre os efeitos fixos não observados e os regressores do modelo e o viés decorrente da omissão de tais efeitos. Por sua vez, a identificação dos efeitos de localização, com esse estimador, requer a existência de suficiente mobilidade dos trabalhadores entre os arranjos ou com os municípios que não pertencem aos arranjos.

Na Tabela 4.4, a coluna IV (FEij) apresenta as estimativas obtidas através da transformação intra combinação dos efeitos de trabalhador-firma, realizada na especificação que inclui o termo de *matching*. As unidades de análise nessa especificação são as combinações (*spells*) existentes entre as firmas e o trabalhador (*i, j*), a partir das quais é realizada a transformação *within group*, com o estimador de Efeitos Fixos. Dentro de cada *spell*, os efeitos fixos de firma e indivíduo não variam, de modo que a transformação *within* vai eliminar tais efeitos. Neste sentido, quaisquer variáveis que sejam constantes dentro de cada *spell* não serão identificadas (ANDREWS; SCHANK; UPWARD, 2006). As especificações condicionais propostas no modelo AKM adotam a primeira diferença ao invés de desvios da média para eliminar os efeitos fixos na estimação.

³⁰ As variáveis que são fixas ou possuem variação constante no tempo, tais como a idade e o gênero do trabalhador, serão retiradas das especificações de estimação por efeitos fixos. Os setores foram excluídos nas estimações dos modelos que incluem efeitos fixos de firmas, mas o tamanho relativo da firma foi mantido.

A coluna II apresenta os resultados da estimação condicional (COND), quando são controlados (*absorbed*) apenas os efeitos fixos dos indivíduos (*i*). A coluna III, por sua vez, apresenta os resultados para o modelo que inclui apenas os efeitos de firmas (*j*). Ambos os modelos, incluem os efeitos fixos de localização dos arranjos e as características observadas dos trabalhadores. Por fim, a coluna V apresenta os resultados da abordagem de estimação condicional, que inclui simultaneamente os efeitos fixos da firma e do trabalhador (*absorbed*), além dos efeitos de localização como variáveis na regressão. Essa coluna representa o modelo condicional com efeito fixo de trabalhador sendo estimado primeiro.

A variação salarial explicada pelo modelo (R^2 ajustado) foi de 60,6%, considerando somente as características observadas do trabalhador e do emprego (POLS). Ao incluir os fatores não observados nesse modelo, os resultados sugerem que o efeito fixo dos indivíduos é mais relevante para explicar a variação dos salários individuais do que o efeito das firmas. A inclusão da heterogeneidade individual não observada aumenta substancialmente a parcela da variação de salários individuais explicada para 91,7% (COND_i). Ao incluir apenas os efeitos de firma o modelo explica 78% da variação salarial (COND_j). A inclusão de ambos os efeitos não observados, da firma e do trabalhador, mantendo os demais controles, responde por cerca de 93% da variação dos salários reais no mercado de trabalho formal do Brasil (COND_{ij}), o que sugere a importância das heterogeneidades não observadas para os diferenciais salariais.

A estimação dos efeitos de localização assumiu os municípios (AMCs) que não pertencem a um arranjo como a categoria de referência. Desse modo, as estimativas dos efeitos representam uma média do ganho ou perda salarial nos mercados de trabalho aglomerados em relação aos não aglomerados. O modelo POLS foi estimado com a inclusão dos efeitos fixos de cada um dos 365 arranjos (LMAs), os quais definem os efeitos de localização sobre os salários reais individuais. A Figura 4.2 apresenta a distribuição espacial do prêmio salarial urbano, do salário-hora real (em log) e dos efeitos fixos dos trabalhadores e das firmas em cada arranjo. O prêmio urbano é estimado pelo método POLS a partir dos efeitos fixos de cada um dos 362 arranjos³¹. Quando o efeito fixo é positivo e estatisticamente significativo, o arranjo apresenta um prêmio urbano, ou seja, existem ganhos de aglomeração sobre os salários dos trabalhadores locais. Os maiores salários estão concentrados entre os arranjos das macrorregiões do Norte, Sul e Sudeste, além do arranjo de Brasília.

³¹ A amostragem não foi representativa para três arranjos populacionais pequenos.

O prêmio salarial urbano apresentou uma distribuição espacial similar à distribuição dos salários, ou seja, os maiores níveis salariais também correspondem aos maiores prêmios, os quais permaneceram concentrados em torno das macrorregiões Sul e Sudeste do país. As evidências sugerem que as áreas de emprego dessas macrorregiões pagam, em média, os maiores salários relativos. Por sua vez, a distribuição espacial dos efeitos dos trabalhadores e das firmas sobre os salários reais mostra que existe uma concentração dos maiores efeitos individuais em torno das áreas de emprego do Sul, do Sudeste e do Norte do Brasil, além de Brasília. Os efeitos das firmas são maiores em torno das áreas de emprego das regiões Norte, Nordeste e Centro-Oeste do país.

Figura 4.2 – Distribuição espacial do salário-hora real (log), prêmio salarial urbano e efeitos fixos do trabalhador e da firma nos arranjos populacionais

Fonte: Elaboração própria a partir da RAIS (2002-2014).

Notas: Salário real deflacionado pelo IPCA (base: dezembro/2014); Os arranjos sem o preenchimento da escala de cores são aqueles que não apresentaram um prêmio urbano (efeito fixo menor ou igual zero); Os efeitos estatisticamente não significantes foram excluídos.

As estimativas dos efeitos de localização dos grandes arranjos populacionais são apresentadas na Figura 4.3 (e na Tabela 4.4).

Figura 4.3 – Efeito da localização sobre o salário-hora real dos indivíduos nos grandes arranjos populacionais

Fonte: Elaboração própria, a partir da RAIS (2002-2014).

Nota: A cor cinza representa os efeitos estatisticamente significantes; As AMCs que não pertencem a algum arranjo representam a categoria de referência para os efeitos dos arranjos.

Na figura, os efeitos estatisticamente significantes são representados pela cor cinza. Nesse sentido, os resultados sugerem a existência de um prêmio urbano sobre os salários dos 19 dentre os 27 grandes arranjos populacionais (POLS), cujas maiores magnitudes estão concentradas nas macrorregiões do Sul e Sudeste. O maior prêmio foi encontrado no arranjo de Campinas (43,8%) e o menor no arranjo de Campo Grande (8,5%). Na região Nordeste, uma das mais pobres do país, apenas o arranjo de Salvador apresentou um prêmio salarial urbano (11,4%).

Por sua vez, quando os atributos não observados dos trabalhadores e firmas são considerados (CONDij), apenas 2 arranjos apresentaram um prêmio salarial urbano: Belém (9,1%) e Florianópolis (8,8%). No entanto, ambos os arranjos apresentaram uma redução no prêmio em relação à estimativa do POLS (Belém era 10,8% e Florianópolis era 28,9%). Os resultados sugerem que as principais fontes dos ganhos de aglomeração no Brasil podem ser as habilidades individuais e a maior produtividade das firmas. Nesse sentido, o *sorting* espacial dos trabalhadores e firmas parece explicar uma grande parcela dos ganhos atribuídos diretamente às economias de aglomeração nos mercados de trabalho locais.

Ao controlar apenas os efeitos fixos individuais na equação de salários (CONDi), ainda permaneceu um prêmio urbano, em relação ao POLS, em 12 grandes arranjos, porém em menor magnitude. O maior prêmio, após o controle dos atributos não observados individuais, foi do arranjo de Brasília (9,2%), seguido do Rio de Janeiro (8,3%) e o menor foi do arranjo de Recife (2,5%). Por sua vez, os arranjos de Recife e Teresina somente passaram a apresentar um prêmio após o controle dos efeitos individuais. O prêmio médio estimado no POLS contém os atributos não observados das firmas e trabalhadores. Nesse sentido, se o efeito de localização torna-se positivo (prêmio) ao eliminar os efeitos individuais, há indícios de que os atributos não observados dos indivíduos seriam negativamente correlacionados com os salários locais. Essa evidência pode sugerir que os trabalhadores do local não possuem as características produtivas melhor remuneradas no mercado de trabalho local, ou seja, o local pode concentrar trabalhadores de baixa habilidade produtiva.

Quando apenas os efeitos das firmas são controlados (CONDj), 7 arranjos ainda mantiveram o prêmio urbano, em menor magnitude em relação aos POLS. O maior prêmio foi do arranjo de Florianópolis (17,1%) e o menor foi o de Campo Grande (3,1%). A partir das estimativas encontradas para os efeitos fixos dos arranjos é possível perceber que alguns arranjos mantêm

um prêmio apenas se os efeitos individuais não são controlados, enquanto outros mantêm apenas quando os efeitos das firmas não são considerados. Uma explicação possível para esse comportamento é a desigual distribuição de firmas e trabalhadores com heterogêneos atributos produtivos não observados. Nesse sentido, os resultados podem sugerir que alguns mercados de trabalho locais concentram trabalhadores de maior habilidade enquanto outros concentram firmas mais produtivas. Por outro lado, existem arranjos que apresentam um prêmio urbano ao controlar os efeitos da firma ou do trabalhador, mas o prêmio desaparece após o controle de ambos. Nesses casos, é possível que o prêmio resulte da melhor correspondência (*match*) entre as habilidades individuais e as necessidades produtivas das firmas locais.

Portanto, uma classificação possível para o prêmio salarial urbano dos grandes arranjos, a partir dos *sorting* das heterogeneidades de firmas e trabalhadores seria: (i) Prêmio urbano resulta das habilidades produtivas não observadas dos indivíduos nas aglomerações urbanas e, portanto, desaparece após o controle dos efeitos individuais; (ii) Prêmio urbano resulta da maior produtividade das firmas locais e, portanto, desaparece após o controle dos efeitos das firmas; e (iii) Prêmio urbano resulta de melhor *matching* entre firmas e trabalhadores nos mercado de trabalho locais e, portanto, desaparece após o controle simultâneo dos efeitos da firma e do trabalhador. Portanto, os grandes arranjos classificados no grupo das habilidades individuais são: Salvador, Campo Grande e Florianópolis. No grupo das firmas produtivas estão: Manaus, Belém, Vitória, Ribeirão Preto, Baixada Santista, Campinas, Curitiba e Porto Alegre. Por último, no grupo de melhor *matching* estão: São Paulo, São José dos Campos, Brasília e Rio de Janeiro.

Por fim, existe um grupo de grandes arranjos cujo controle do efeito da firma ou do trabalhador foi suficiente para eliminar o prêmio salarial urbano: Goiânia, Cuiabá, Belo Horizonte e Sorocaba. Nesse caso, é possível que os atributos dos trabalhadores sejam melhor remunerados em relação aos não arranjos, o que é compatível com as firmas locais pagando um salário maior pelos atributos dos seus trabalhadores. Nesse sentido, o ganho previamente existente (POLS) nessas áreas pode resultar somente de uma maior remuneração dos atributos produtivos em relação aos municípios que não pertencem aos arranjos. Por outro lado, nos arranjos de Recife e Teresina, cujo prêmio apareceu somente após o controle dos atributos individuais, é possível que a quantidade de trabalhadores de baixa habilidade seja desproporcionalmente maior do que o número de firmas produtivas existentes, o que explicaria o efeito bruto negativo (POLS).

Como teste de robustez, os modelos foram estimados considerando uma amostra aleatória em firmas móveis e uma amostra sem *outliers* de salários (Tabela J.3 e Tabela J.4 – Apêndice J). Os resultados para ambas as amostras foram consistentes com as estimativas obtidas na amostra principal. Na amostra principal, a parcela da variação salarial explicada pelos componentes observados foi 60,6%, ampliando para 78% com a inclusão dos efeitos da firma e para 92% após a inclusão dos efeitos individuais. Na amostra aleatória em firmas, 59,6% da variação salarial foi explicada pelos componentes observados do trabalhador e do emprego, 79% da variação foi explicada após a inclusão dos efeitos da firma e 92% após a inclusão dos efeitos individuais. Na amostra sem *outliers*, 59,3% da variação salarial foi explicada pelos componentes observados, ampliando para 78% com a inclusão dos efeitos da firma e 91% com a inclusão dos efeitos individuais. As Figuras J.1 e J.2 do Apêndice J mostram o prêmio urbano entre os grandes arranjos nas respectivas amostras de robustez.

A amostra aleatória em firmas apresentou um prêmio urbano nos 19 grandes arranjos (POLS), os mesmos identificados na amostra principal, cujo maior prêmio foi o do arranjo de São José dos Campos (52,4%) e o menor foi o de Belém (9,8%). Após o controle dos efeitos individuais e das firmas apenas 4 arranjos apresentaram um prêmio: Rio de Janeiro (2,6%), Maceió (7,8%), Florianópolis (9,2%) e Belém (9,3%), dos quais somente Maceió e o Rio de Janeiro não foram identificados na amostra principal. A amostra sem os *outliers* salariais também apresentou um prêmio urbano no mesmo grupo de arranjos, com o maior valor no arranjo de Campinas (42,2%) e o menor no arranjo de Campo Grande (7,9%). Após o controle dos efeitos individuais e das firmas apenas 2 arranjos apresentaram um prêmio: Belém (9,1%) e Florianópolis (8,5%), os mesmos identificados na amostra principal. Em geral as estimativas da amostra aleatória em firmas são maiores, como resultado da menor consistência na identificação dos efeitos de localização e das firmas. Por sua vez, as estimativas da amostra sem os *outliers* salariais são menores, compatíveis com a exclusão de valores extremos. No entanto, em ambas as amostras os resultados são consistentes com a amostra principal.

Portanto, os resultados mostram uma grande heterogeneidade na distribuição dos atributos produtivos não observados das firmas e dos trabalhadores entre os arranjos populacionais do Brasil, no período de 2002 a 2014. As evidências sugerem que uma parte significativa dos efeitos positivos da localização sobre os salários reais nos mercados de trabalho locais seria explicada pelo *sorting* espacial associado às heterogeneidades não observadas das firmas e

dos trabalhadores, particularmente nas macrorregiões do Sul e Sudeste. Por fim, com base na variação salarial explicada, as evidências sugerem uma maior contribuição dos atributos individuais não observados para a determinação de salários individuais, do que os atributos específicos da firma. Esse resultado é consistente com os estudos que evidenciam a importância das habilidades individuais para a determinação de salários.

4.3 DECOMPOSIÇÃO DOS EFEITOS DE LOCALIZAÇÃO

Esta seção apresenta os resultados da estimação do segundo estágio do modelo. O segundo estágio corresponde a uma decomposição dos efeitos da localização, estimados no primeiro estágio com a equação de salários individuais. Os efeitos foram estimados pelo modelo POLS sem o controle dos efeitos individuais e de firma, para captar os efeitos brutos da aglomeração. Esses efeitos são definidos como os salários locais dos arranjos (LMAs). A decomposição segue o modelo AKM, mas inclui os controles típicos da literatura da Economia Urbana. O objetivo da decomposição é identificar a contribuição dos efeitos puros das economias de aglomeração (densidade) e do *sorting* espacial dos individuais e das firmas para os salários locais. O *sorting* será mensurado pelos efeitos fixos dos indivíduos e firmas, estimados a partir da equação de salários pelo método condicional (CONDij).

A estimação do modelo em dois estágios é necessária, já que a estimação em um único estágio (individual) não permite calcular a variância dos choques locais e realizar inferência com os erros-padrão das variáveis agregadas. A estimação de um único estágio não distingue os choques locais dos choques individuais. Contudo, o segundo estágio deve corrigir a possível heterocedasticidade que surge da utilização dos efeitos de localização estimados com diferentes números de observações para cada arranjo, no primeiro estágio (COMBES; DURANTON; GOBILLON, 2008).

Este estudo estima os modelos tradicionais da literatura para comparar com os resultados estimados após as modificações propostas. O modelo POLS foi estimado sem a inclusão dos componentes não observados dos trabalhadores e firmas apenas para a comparação com os resultados da literatura prévia. O primeiro modelo estimado, a partir das modificações sugeridas na estratégia de identificação deste estudo, incluiu os componentes não observados médios dos trabalhadores e firmas na decomposição dos salários locais. A segunda modificação realizada corresponde à estimação de um modelo de variável instrumental

conjuntamente com a inclusão dos componentes não observados médios. O terceiro modelo utiliza a área iluminada dos arranjos para medir a densidade. Por fim, a especificação principal deste estudo estima o modelo de decomposição dos salários locais utilizando uma variável instrumental para a densidade, a área iluminada em substituição à área total e os componentes não observados dos indivíduos e firmas.

A Tabela 4.5 apresenta os resultados da decomposição dos efeitos fixos dos 362 arranjos, estimados a partir da equação (3.35) do segundo estágio do modelo. O segundo estágio realiza a decomposição dos efeitos fixos dos arranjos em componentes associados às características locais e às heterogeneidades não observadas dos trabalhadores e firmas, ao longo do painel (2002–2014).

Tabela 4.5 – Decomposição dos efeitos de localização dos arranjos: segundo estágio

FE_LMA	(1) POLS	(2) POLS	(3) POLS	(4) POLS	(5) FGLS	(6) POLS-BS
<i>ldensemp_lma</i>	0,0421*** (0,0105)	0,0417*** (0,0104)	0,0258* (0,0156)	0,0259* (0,0155)	0,0256*** (0,0006)	0,0365** (0,0147)
<i>larea_lma</i>	0,0214 (0,0152)	0,0204 (0,0150)	0,0084 (0,0154)	0,0083 (0,0154)	0,0057*** (0,0008)	0,0656*** (0,0188)
<i>(mean) i</i>			0,2536 (0,1800)	0,2503 (0,1793)	0,2825*** (0,0045)	0,0053 (0,1300)
<i>(mean) f</i>		0,0336* (0,0196)		0,0079 (0,0063)	0,0039*** (0,0011)	0,0012 (0,0426)
<i>frontier</i>	-0,0764* (0,0455)	-0,0770* (0,0449)	-0,0820** (0,0413)	-0,0821** (0,0412)	-0,0712*** (0,0023)	-0,0983* (0,0593)
<i>touristic</i>	-0,0072 (0,0389)	-0,0074 (0,0390)	0,0053 (0,0267)	0,0051 (0,0269)	0,0233*** (0,0031)	0,0364 (0,0465)
Norte	-0,0282 (0,1050)	-0,0303 (0,1032)	-0,0021 (0,0897)	-0,0029 (0,0896)	0,0146*** (0,0048)	0,0751 (0,0697)
Nordeste	-0,3811*** (0,0445)	-0,3796*** (0,0440)	-0,2910*** (0,0721)	-0,2918*** (0,0721)	-0,2780*** (0,0023)	-0,2126*** (0,0486)
Sul	-0,0383 (0,0353)	-0,0360 (0,0346)	-0,0406 (0,0281)	-0,0401 (0,0279)	-0,0408*** (0,0013)	0,0095 (0,0339)
Centro-Oeste	-0,0636 (0,0734)	-0,0635 (0,0722)	-0,0502 (0,0601)	-0,0504 (0,0600)	-0,0416*** (0,0024)	-0,0134 (0,0491)
<i>dummies de tempo</i>	Sim	Sim	Sim	Sim	Sim	Sim
<i>constante</i>	-0,0991 (0,1159)	-0,0884 (0,1159)	0,0637 (0,1569)	0,0641 (0,1571)	0,0887*** (0,0075)	-0,5040*** (0,1845)
Observações (<i>LMA x T</i>)	4681	4681	4681	4681	4681	4681
<i>R</i> ² Ajustado	0,502	0,509	0,600	0,600		0,225

Fonte: Elaboração própria, a partir da RAIS (2002-2014).

Notas: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; Erro-padrão robusto com *clusters* de LMAs; Colunas 1-5 assumem peso analítico proporcional ao número de indivíduos em cada LMA.

A coluna 1 apresenta os resultados do modelo estimado apenas com as variáveis observadas dos arranjos (*baseline*). As colunas 2 e 3 incluem as médias dos efeitos não observados das firmas e dos trabalhadores, respectivamente, enquanto a coluna 4 inclui ambos os efeitos. As

efeitos de localização do que os efeitos das firmas, com base na inclinação da reta. Esse comportamento é consistente com os resultados encontrados para a variação salarial explicada após a inclusão de cada componente.

A inclusão das heterogeneidades não observadas das firmas e trabalhadores ampliou a parcela explicada da variação de salários locais, mas ainda existe uma parcela residual não explicada relativamente grande (40%). Esse resíduo pode ser reduzido se forem incluídos mais controles para as características locais dos arranjos, ao custo de gerar maiores problemas de endogeneidade. O estudo de Combes, Duranton e Gobillon (2008) incluiu como atributos locais, no segundo estágio do seu modelo, o percentual de municípios em cada área de emprego da França com: costa litorânea; montanhas; lagos e água; e patrimônio cultural e arquitetônico. O modelo do correspondente estudo obteve uma variação salarial explicada de 72%. O Brasil não dispõe de um banco de dados nacional com atributos específicos locais e cobertura municipal, o que limita a inclusão de controles locais. No entanto, o modelo incluiu *dummies* para indicar se o arranjo litorâneo é do tipo turístico (*touristic*) ou se o arranjo possui integração com os municípios da fronteira (*frontier*).

Os resultados mostraram um efeito negativo nos salários médios dos arranjos que possuem integração fronteiriça, sugerindo que não existem vantagens produtivas de facilidade de acesso aos mercados vizinhos, ou tais vantagens não são aproveitadas. Por sua vez, os arranjos litorâneos com função turística apresentaram um ganho salarial (0,023) em relação aos demais arranjos, somente no modelo FGLS. Os ganhos salariais locais nos arranjos turísticos podem ser gerados não somente da renda do setor de turismo, mas também da localização privilegiada em relação a rios navegáveis e portos atuando como dotação produtiva, a exemplo do arranjo de Itajaí em Santa Catarina. Em geral, os resultados indicaram que os arranjos da macrorregião Nordeste apresentam menores salários médios, relativamente ao Sudeste. Essa evidência é compatível com a distribuição espacial do prêmio urbano no Brasil, apresentada na seção anterior.

Os efeitos de área (efeitos de escala) variaram entre 0,6% e 6,6%, mas somente se mostraram estatisticamente significantes nos modelos FGLS e *Bootstrap*. A magnitude do efeito de área corresponde, aproximadamente, à metade do efeito da densidade, exceto no modelo *Bootstrap* cujo efeito passa a ser quase o dobro da densidade. Combes, Duranton e Gobillon (2008)

encontraram um efeito de área que correspondeu a quase um terço do efeito da densidade, variando entre 1,1% e 2,2%, nas áreas de emprego da França. Os resultados sugerem que um aumento na população através do aumento da densidade tem um efeito maior sobre a produtividade local do que o aumento da população decorrente da maior área disponível, mantendo a densidade constante. Os efeitos da área representam um limite à expansão urbana e, portanto, podem refletir em alguma medida o preço da terra. Nesse sentido, os efeitos positivos sobre os salários podem sugerir que a maior disponibilidade de terra para a produção implica em maior produtividade local, em parte porque o custo da terra será menor.

A densidade do emprego, principal variável de interesse no estudo, apresentou efeitos positivos sobre os salários locais variando entre 2,6% e 4,2%. O efeito mede os ganhos salariais diretamente atribuídos às economias de aglomeração urbana e suas externalidades positivas. As estimativas dos efeitos da densidade sobre os salários locais na literatura empírica internacional variam entre 4% e 8%. O estudo relativamente mais recente do Combes, Duranton e Gobillon (2008) encontrou um efeito entre 3,2% e 3,7% para as áreas de emprego da França, o qual foi definido pelos autores como um *lower bound* da literatura empírica existente. O estudo de Barufi, Haddad e Nijkamp (2016) encontrou efeitos da densidade do emprego sobre os salários, entre os setores locais do Brasil, variando entre 4,6% e 9%, no período de 2004, 2008 e 2012.

Os resultados encontrados para a densidade do emprego sugerem a importância dos efeitos não observados da firma e do trabalhador para a magnitude dos efeitos de aglomeração. Ao incluir os efeitos do trabalhador e da firma, o efeito da densidade é reduzido para quase metade, de 4,2% para 2,6% (Colunas 1 e 4), cujo menor valor é inferior ao *lower bound* da literatura empírica. Isso não quer dizer, contudo, que o *sorting* espacial dos atributos não observados de indivíduos e firmas seja independente dos efeitos de aglomeração. A literatura da Economia Urbana sugere a atração e manutenção das firmas e trabalhadores mais produtivos nos mercados de trabalho mais densos. Portanto, o *sorting* é, em parte, o resultado das economias de aglomeração urbana. No entanto, as evidências encontradas permitem decompor a parcela do *sorting* existente nos efeitos puros da aglomeração.

Os efeitos brutos de aglomeração podem ser obtidos a partir da decomposição dos efeitos médios dos trabalhadores e firmas, através das equações (3.36) e (3.37), incluindo a densidade do emprego e os demais controles utilizados na decomposição dos efeitos de localização. A

Tabela 4.6 apresenta os resultados da decomposição dos efeitos individuais e das firmas. A Tabela completa se encontra no Apêndice J (Tabela J.5). A primeira coluna da Tabela 4.6 corresponde à coluna 4 da Tabela 4.5, a qual controla os atributos não observados da firma e do trabalhador na decomposição dos efeitos de localização (LMA). A estimativa do efeito bruto de aglomeração corresponde à soma do efeito líquido ou puro da densidade (2,6%) com os efeitos da densidade sobre os atributos não observados dos trabalhadores (6,4%) e das firmas (1,1%). Portanto, as estimativas sugerem um efeito bruto da aglomeração em torno de 10,1%. Novamente, os efeitos não observados individuais respondem por quase todos os ganhos atribuídos diretamente à densidade.

Tabela 4.6 – Decomposição dos efeitos de localização e dos efeitos médios da firma e do trabalhador nos arranjos: segundo estágio

	(1) POLS (FE LMA)	(2) POLS (FEi)	(3) FGLS (FEi)	(4) POLS (FEj)	(5) FGLS (FEj)
<i>ldensemp_lma</i>	0,0259*	0,0643***	0,0669***	0,0114*	0,0116***
	(0,0155)	(0,0085)	(0,0007)	(0,0060)	(0,0004)
<i>(mean) i</i>	0,2503				
	(0,1793)				
<i>(mean) f</i>	0,0079				
	(0,0063)				
Observações (LMA x T)	4681	4681	4681	4681	4681
R ² Ajustado	0,600	0,312		0,007	

Fonte: Elaboração própria, a partir da RAIS (2002-2014).

Notas: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; Erro-padrão robusto com *clusters* de LMA; Resultados assumem peso analítico proporcional ao número de indivíduos em cada LMA. Resultados completos na Tabela J.5.

Como teste de robustez, os modelos foram estimados com a amostra aleatória em firmas (Tabelas J.6 e J.7 – Apêndice J). O efeito puro da densidade sobre os salários locais foi estimado em 5%, no método POLS, sendo reduzido para 4,8%, após a inclusão dos atributos não observados individuais e das firmas. Por sua vez, a soma do efeito puro da densidade (4,8%) com os efeitos da densidade sobre os atributos não observados dos trabalhadores (8,6%) e das firmas (4,7%) gerou um efeito bruto densidade de 18,1%. Na amostra principal deste estudo, que possui seleção em firmas móveis, o efeito puro da densidade foi de 2,6% enquanto o efeito bruto foi de 10,1%. Portanto, com uma amostra que não garante a condição de mobilidade das firmas entre os arranjos para a identificação consistente de ambos os efeitos fixos, os efeitos da aglomeração são sobrestimados.

Por fim, o modelo de decomposição foi estimado utilizando os salários locais médios dos arranjos, ao invés do diferencial salarial previsto pelos efeitos fixos locais. Essa versão é

estimada para mostrar como a magnitude dos possíveis efeitos de aglomeração se altera se os dados agregados são adotados sem a etapa anterior que controla os atributos individuais na equação de salários. A Tabela 4.7 mostra os resultados dessa estimação. A primeira mudança verificada, em relação à estimação com os efeitos de localização do primeiro estágio, é que a variação salarial explicada pelo modelo (R^2) se amplia substancialmente. Nesse sentido, o modelo POLS sem os efeitos individuais e de firma explica 74,5% da variação salarial local. A variação salarial explicada é ampliada para 75% com a inclusão dos efeitos de firma e para 80,2% após a inclusão dos efeitos individuais ou de ambos. A maior parcela da variação salarial explicada no modelo estimado em dois estágios foi 60%. No entanto, o modelo agregado não permite avaliar se os efeitos da densidade resultam apenas das economias de aglomeração, uma vez que as características individuais e do emprego que explicam a produtividade local não foram controladas.

Tabela 4.7 – Decomposição dos salários reais médios dos arranjos: segundo estágio

ln(wreal_hr)	(1) POLS	(2) POLS	(3) POLS	(4) POLS	(5) FGLS	(6) POLS-BS
<i>ldensemp_lma</i>	0,0860*** (0,0129)	0,0855*** (0,0127)	0,0641*** (0,0196)	0,0642*** (0,0196)	0,0408*** (0,0014)	0,0676*** (0,0200)
<i>larea_lma</i>	0,0832*** (0,0188)	0,0818*** (0,0186)	0,0657*** (0,0192)	0,0656*** (0,0192)	0,0382*** (0,0018)	0,1071*** (0,0257)
<i>(mean) f</i>		0,0454* (0,0261)		0,0108 (0,0081)	0,0072*** (0,0015)	0,0001 (0,0699)
<i>(mean) i</i>			0,3410 (0,2379)	0,3365 (0,2371)	0,7029*** (0,0110)	0,0099 (0,1990)
<i>frontier</i>	-0,0470 (0,0624)	-0,0478 (0,0614)	-0,0545 (0,0564)	-0,0546 (0,0562)	-0,0395*** (0,0061)	-0,1148* (0,0686)
<i>touristic</i>	-0,0312 (0,0702)	-0,0315 (0,0704)	-0,0144 (0,0492)	-0,0147 (0,0495)	0,0248*** (0,0087)	-0,0478 (0,0529)
Norte	0,0115 (0,1122)	0,0087 (0,1097)	0,0467 (0,0916)	0,0455 (0,0914)	0,0989*** (0,0080)	0,1464* (0,0797)
Nordeste	-0,4205*** (0,0559)	-0,4184*** (0,0551)	-0,2992*** (0,0944)	-0,3003*** (0,0944)	-0,1327*** (0,0060)	-0,1908*** (0,0654)
Sul	-0,0436 (0,0436)	-0,0406 (0,0427)	-0,0468 (0,0340)	-0,0461 (0,0338)	-0,0306*** (0,0034)	0,0217 (0,0398)
Centro-Oeste	-0,0276 (0,0976)	-0,0275 (0,0959)	-0,0096 (0,0801)	-0,0098 (0,0800)	0,0573*** (0,0072)	-0,0450 (0,0589)
<i>dummies de tempo</i>	Sim	Sim	Sim	Sim	Sim	Sim
<i>constante</i>	1,2993*** (0,1450)	1,3137*** (0,1446)	1,5183*** (0,1993)	1,5188*** (0,1995)	1,7829*** (0,0170)	1,0353*** (0,2600)
Observações	4681	4681	4681	4681	4681	4681
R^2 Ajustado	0,745	0,750	0,802	0,802		0,471

Fonte: Elaboração própria, a partir da RAIS (2002-2014).

Notas: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; Erro-padrão robusto com *clusters* de LMAs; Colunas 1-5 assumem peso analítico proporcional ao número de indivíduos em cada LMA.

O efeito das economias de aglomeração urbana sobre os salários médios locais, estimado a partir da densidade do emprego, variou entre 4,1% e 8,6%. O menor valor do efeito estimado com os salários agregados correspondeu à maior estimativa encontrada na estimação em dois estágios. Portanto, o efeito da densidade urbana deve ser sobrestimado se o modelo utiliza os salários agregados locais sem controlar previamente os atributos individuais e do emprego.

Os resultados mostraram que os efeitos fixos individuais foram os componentes mais importantes para explicar a variação dos salários reais individuais e dos salários locais no mercado de trabalho formal do Brasil, entre 2002 e 2014. No entanto, mesmo com a grande contribuição dos atributos não observados, ainda existe um efeito puro das economias de aglomeração urbana, associado à densidade do emprego nos arranjos populacionais. Por sua vez, o uso da área territorial oficial pode subestimar os efeitos de densidade dos arranjos, se a área efetivamente habitada for muito inferior à área total, o que será tratado na próxima seção.

4.3.1 Densidade do Emprego e Luminosidade Noturna

Esta seção apresenta os resultados da estimação do modelo do segundo estágio com a área iluminada dos arranjos. A área iluminada corresponde à proporção de *pixels* iluminados, em cada arranjo, multiplicada por sua área oficial do ano de 2013 (Equação 3.40). A Tabela 4.8 apresenta as estatísticas de iluminação do Brasil e dos arranjos.

Tabela 4.8 – Distribuição da área oficial e da área iluminada no Brasil: 2007 e 2013

Categoria	Área (km²)	Sharelight 2007	Sharelight 2013	Light 2007	Light 2013
Grandes	136.496,79	0,5442	0,5524	74.108,66	75.227,20
Médios	1.040.496,44	0,1205	0,1292	125.207,87	134.332,89
Pequenos	534.807,57	0,0900	0,0954	48.170,43	51.062,52
Arranjos	1.711.800,80	0,1448	0,1524	247.486,96	260.622,62
Outros municípios	6.791.005,35	0,0568	0,0634	385.737,78	430.677,42
Brasil	8.502.806,15	0,0745	0,0813	633.224,74	691.300,04

Fonte: Elaboração própria a partir das imagens de satélite do DMSP-OLS (NCEI, 2013) e IBGE (2013).

Notas: *Sharelight* é a parcela de *pixels* luminosos em cada área e *light* é a área iluminada (km²) obtida a partir do *share*.

A partir das imagens de satélite extraídas do DMSP-OLS (NCEI, 2013), é possível constatar que apenas 8,1% da área total do Brasil apresentou alguma luminosidade noturna no ano de 2013. O percentual é maior do que no ano de 2007 (7,4%). Por sua vez, a parcela iluminada da área total dos arranjos em 2013 foi de 15,2%, correspondendo a quase o dobro da parcela

iluminada da área total do Brasil. No ano de 2007, o percentual de iluminação da área total dos arranjos foi de 14,5%. Em geral, os arranjos possuem uma maior proporção de iluminação noturna do que a média do Brasil, mesmo entre os pequenos. A parcela de luminosidade noturna variou de 9,5%, entre os pequenos, para 55,2% entre os grandes arranjos populacionais em 2013. Os dados são consistentes com a distribuição da luminosidade no Brasil e em outros países, encontrada em estudos da literatura internacional (HENDERSON; STOREYGARD; WEIL, 2012).

Os dados de luminosidade noturna permitem observar dois importantes aspectos. Primeiro, os arranjos possivelmente representam as áreas mais dinâmicas do país, relativamente à produção, distribuição dos empregos e concentração urbana. Além disso, os dados mostram que existe uma grande diferença entre as extensões territoriais oficiais e a distribuição populacional e urbana, a partir das imagens de luminosidade noturna disponíveis, o que motiva o uso da área iluminada como medida mais adequada para calcular a densidade urbana. Nesse sentido, a área com luminosidade noturna foi utilizada para calcular a densidade do emprego nos arranjos e estimar os respectivos efeitos das economias de aglomeração urbana. Se a área administrativa oficial subestima a densidade e possivelmente os seus efeitos sobre os salários locais, é esperado um aumento nos efeitos das economias de aglomeração com o uso da nova medida de área.

A Tabela 4.9 apresenta os resultados da decomposição dos efeitos de localização, no segundo estágio do modelo, após substituir a área oficial pela área iluminada dos arranjos (2013) no cálculo da densidade do emprego. Os resultados mostraram uma elevação na magnitude dos efeitos da densidade do emprego sobre os salários locais nos arranjos. Os efeitos puros da aglomeração (densidade) variaram entre 2,9% e 5,4%, com o uso da área iluminada, enquanto os resultados anteriores, calculados a partir da área oficial total, variaram entre 2,6% e 4,2%. Portanto, como era previsto, o uso da área oficial parece subestimar os efeitos da aglomeração urbana. Para testar a robustez dos resultados, utilizou-se a área iluminada dos arranjos do ano de 2007 (Tabela J.8 – Apêndice J). Os resultados também mostraram maiores efeitos da densidade sobre os salários locais dos arranjos, em relação ao uso da área total, variando entre 2,6% e 5,2%.

**Tabela 4.9 – Decomposição dos efeitos de localização dos arranjos:
segundo estágio com a área iluminada (2013)**

FE_LMA	(1) POLS	(2) POLS	(3) POLS	(4) POLS	(5) FGLS	(6) POLS-BS
<i>ldensemplight13</i>	0,0543*** (0,0182)	0,0539*** (0,0179)	0,0306 (0,0244)	0,0308 (0,0243)	0,0290*** (0,0009)	0,0533** (0,0211)
<i>larea_lma</i>	0,0095 (0,0178)	0,0085 (0,0175)	0,0023 (0,0148)	0,0021 (0,0148)	-0,0007 (0,0008)	0,0445*** (0,0150)
<i>(mean) f</i>		0,0345* (0,0202)		0,0077 (0,0060)	0,0040*** (0,0011)	0,0013 (0,0419)
<i>(mean) i</i>			0,2612 (0,1840)	0,2580 (0,1833)	0,2998*** (0,0046)	0,0053 (0,1285)
<i>frontier</i>	-0,1298** (0,0517)	-0,1298** (0,0508)	-0,1163*** (0,0447)	-0,1164*** (0,0446)	-0,1104*** (0,0023)	-0,1353** (0,0591)
<i>touristic</i>	0,0078 (0,0410)	0,0074 (0,0410)	0,0142 (0,0268)	0,0140 (0,0269)	0,0285*** (0,0031)	0,0542 (0,0456)
Norte	-0,0760 (0,0887)	-0,0776 (0,0870)	-0,0327 (0,0810)	-0,0336 (0,0808)	-0,0089** (0,0038)	0,0501 (0,0660)
Nordeste	-0,3828*** (0,0447)	-0,3811*** (0,0442)	-0,2899*** (0,0730)	-0,2907*** (0,0730)	-0,2743*** (0,0025)	-0,2091*** (0,0478)
Sul	-0,0499 (0,0381)	-0,0474 (0,0372)	-0,0495* (0,0298)	-0,0490* (0,0296)	-0,0463*** (0,0015)	0,0035 (0,0344)
Centro-Oeste	-0,0849 (0,0761)	-0,0845 (0,0748)	-0,0659 (0,0614)	-0,0660 (0,0614)	-0,0556*** (0,0030)	-0,0391 (0,0499)
<i>dummies de tempo</i>	Sim	Sim	Sim	Sim	Sim	Sim
<i>constante</i>	-0,0760 (0,1126)	-0,0653 (0,1124)	0,0819 (0,1505)	0,0824 (0,1507)	0,1151*** (0,0073)	-0,4442*** (0,1666)
Observações	4681	4681	4681	4681	4681	4681
R^2	0,491	0,498	0,595	0,595		0,220
R^2 Ajustado	0,489	0,496	0,593	0,593		0,216

Fonte: Elaboração própria, a partir da RAIS (2002-2014).

Notas: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; Erro-padrão robusto com *clusters* de LMAs; Colunas 1-5 assumem peso analítico proporcional ao número de indivíduos em cada LMA.

Em geral, as estimativas das demais covariadas do modelo não sofreram alterações significativas em relação aos resultados da seção anterior. Por sua vez, os efeitos de escala, estimados a partir do logaritmo da área total³², foram reduzidos. Por fim, a partir da medida de área iluminada dos arranjos, o modelo foi estimado com a amostra aleatória em firmas. As estimativas dos efeitos da densidade variaram entre 6,1% e 6,7%, com a área iluminada do ano de 2007, e entre 6,3% e 7,2% ao usar a área iluminada do ano de 2013 (Tabelas J.9 e J.10 – Apêndice J). Novamente, os resultados da amostra sem seleção em firmas móveis foram maiores do que os resultados da amostra principal, o que é explicado pela identificação inconsistente dos efeitos de localização e de firma nessa amostra aleatória.

³² Optou-se por usar a área iluminada apenas para lidar com a subestimação dos efeitos de densidade, mantendo a área oficial como medida dos efeitos de escala, uma vez que a área iluminada pode ser endôgena.

Os resultados apresentados nessa seção evidenciam a subestimação dos efeitos das economias de aglomeração, ao usar a densidade calculada a partir da área territorial oficial local. A subestimação ocorre sempre que a área efetivamente ocupada por habitações e/ou empresas for menor do que a área administrativa total de uma localidade. A área com iluminação noturna dos arranjos, calculada a partir das imagens de satélites, se mostrou uma boa medida para a área efetivamente habitada e para corrigir o problema da subestimação dos efeitos da densidade. Por sua vez, a densidade do emprego é endógena aos salários locais, o que requer uma estratégia de identificação específica para estimar os resultados de interesse. Portanto, a próxima seção apresenta a estratégia utilizada e os resultados encontrados com a correção da endogeneidade do modelo.

4.3.2 Densidade do Emprego e Variável Instrumental

Os estudos da Economia Urbana utilizam a densidade para identificar os efeitos das economias de aglomeração sobre a produtividade local. No entanto, os estudos empíricos mostram que existe uma relação de causalidade reversa entre os salários locais e a densidade urbana. A endogeneidade resulta dos choques locais que afetam os salários reais, os quais motivam as decisões de localização e, portanto, a densidade. Se existem choques sobre a produtividade local, a exemplo dos choques tecnológicos, é possível que a densidade esteja capturando parte dos efeitos desses choques. Nesse sentido, a endogeneidade na densidade gera estimativas inconsistentes dos efeitos de aglomeração urbana sobre os salários locais. A literatura em Economia Urbana utiliza variáveis instrumentais para lidar com a endogeneidade da densidade, a exemplo da defasagem temporal da população.

Este trabalho fará uso de uma variável instrumental tradicional na Economia do Trabalho, o *Bartik IV* setorial, originalmente concebido por Bartik (1991). O *Bartik IV* estima qual seria o crescimento setorial previsto a partir do crescimento nacional do respectivo setor, baseado na parcela inicial do emprego setorial local. Nesse sentido, o crescimento adicional do emprego local em relação ao crescimento previsto, a partir da sua estrutura setorial, seria explicado pelos choques locais. O crescimento nacional do emprego em um setor é utilizado como *proxy* para o crescimento da sua demanda de consumo. Nesse sentido, o crescimento exógeno da demanda nacional é utilizado para estimar a demanda por emprego local. Os avanços no uso do instrumento pela literatura empírica sugeriram a exclusão do emprego do próprio local no cálculo do crescimento do emprego nacional. A exclusão é necessária para garantir a

exogeneidade do instrumento, já que um setor pode ser espacialmente concentrado em um único local ou em poucos locais, implicando em uma alta correlação dos choques locais com o emprego nacional.

Os detalhes sobre o instrumento e os principais estudos na literatura empírica que o utilizaram estão disponíveis no capítulo de Metodologia. A equação (3.39) apresenta a variável *Bartik* utilizada como instrumento para a densidade do emprego neste estudo. O estudo utilizará duas versões do instrumento *Bartik* setorial. A primeira versão (*bartik*) é calculada a partir do crescimento do emprego nacional em cada um dos 25 subsetores de atividade econômica do IBGE (2 dígitos), disponíveis na RAIS (2002-2014). A segunda versão do instrumento (*bartikpub*) é baseada no estudo de Faggio e Overman (2014), o qual utilizou o *Bartik share* do setor público como instrumento para o crescimento do emprego público, visando estimar o seu efeito sobre as mudanças no emprego privado e em outros resultados do mercado de trabalho. Portanto, a primeira versão do instrumento estima o crescimento do emprego local a partir do crescimento do emprego setorial nacional e a segunda estima o crescimento do emprego local a partir do crescimento nacional do emprego público. Ambas as versões excluem o emprego de cada arranjo específico no cálculo do crescimento nacional.

A Tabela 4.10 apresenta as estatísticas descritivas da variável instrumental e da variável endógena (densidade do emprego).

Tabela 4.10 – Estatísticas descritivas da variável endógena e do instrumento *Bartik*

Variável	Média	Desvio-padrão	Mínimo	Máximo
<i>ln(densemp_lma)</i>	2,40	1,75	-4,42	7,05
<i>ln(bartik)</i>	8,77	1,89	-1,49	15,67
<i>ln(bartikpub)</i>	6,89	1,78	-0,79	13,69
<i>employment_lma</i>	89.448,67	405.861,90	165	7.730.891
<i>Bartik</i>	47.022,09	276.617,30	-3,84	6.413.838
<i>Bartikpub</i>	8.993,82	47.339,20	0,45	882.111,5
Correlações	<i>ln(densemp_lma)</i>		p-valor	
<i>ln(bartik)</i>	0,6995		0,0000	
<i>ln(bartikpub)</i>	0,5288		0,0000	

Fonte: Elaboração própria, a partir da RAIS (2002-2014).

O instrumento *Bartik* setorial (*bartik*) parece estimar um valor mais aproximado para a distribuição dos empregos nos arranjos (*employment_lma*) do que o *Bartik* do setor público (*bartikpub*), embora indique uma redução no número de empregos em alguns arranjos (número de empregos negativo). Do mesmo modo, a correlação entre a variável endógena

(*Indensemp_lma*) e o instrumento *Bartik* setorial (69,9%) é maior do que a correlação com o *Bartik* do setor público (52,9%). Portanto, as estatísticas descritivas indicam um melhor ajuste da versão setorial do instrumento *Bartik* à variável endógena.

A Figura 4.5 apresenta um ajuste linear entre o emprego médio observado nos arranjos, no período de 2002 a 2014, e o emprego estimado pelo *Bartik* setorial e pelo *Bartik* do setor público, com base no período inicial escolhido (1991). Novamente, é possível perceber um melhor ajuste linear do emprego estimado pelo *Bartik* setorial (Painel 1 e 2) do que o emprego previsto pelo *Bartik* do setor público (Painel 3 e 4). Portanto, a análise dos resultados do modelo de variável instrumental deve focar sobre o *Bartik* setorial.

Figura 4.5 – Emprego médio observado e estimado (*Bartik* IV) dos arranjos: 2002-2014

Painel 1: Emprego observado e estimado (*Bartik*) nos arranjos populacionais

Painel 2: Emprego observado e estimado (*Bartik*) nos arranjos populacionais (sem outliers)

Painel 3: Emprego observado e estimado (*Bartikpub*) nos arranjos populacionais

Painel 4: Emprego observado e estimado (*Bartikpub*) nos arranjos populacionais (sem outliers)

Fonte: Elaboração própria, a partir da RAIS (2002-2014).

Nota: Os outliers são definidos pelos três maiores arranjos (363-365).

A Tabela 4.11 apresenta os resultados do modelo de decomposição dos efeitos dos arranjos utilizando o instrumento *Bartik* para a densidade do emprego. A equação de decomposição (3.35) foi estimada em dois estágios simultâneos pelo método 2SLS. O primeiro estágio corresponde à estimação da densidade do emprego prevista pelo instrumento *Bartik* e demais exógenas do modelo (Equação 3.50). Esse estágio permite testar a hipótese (3.49) de covariância parcial entre o instrumento e a variável endógena. O segundo estágio, correspondente à equação (3.51), utiliza a densidade do emprego prevista pelo instrumento para estimar o seu efeito sobre os salários locais dos arranjos.

**Tabela 4.11 – Decomposição dos efeitos de localização dos arranjos (segundo estágio):
Bartik IV para a densidade do emprego**

FE_LMA	(1) POLS	(2) 2SLS (<i>lbartik</i>)	(3) POLS	(4) 2SLS (<i>lbartik</i>)	(5) POLS	(6) POLS	(7) 2SLS (<i>lbartikpub</i>)
<i>ldensemp_lma</i>	0,0423*** (0,0105)	0,0249** (0,0119)	0,0261* (0,0156)	0,0264** (0,0120)	0,0436*** (0,0105)	0,0273* (0,0158)	0,0285** (0,0124)
<i>larea_lma</i>	0,0212 (0,0153)	0,0571*** (0,0140)	0,0082 (0,0154)	0,0603*** (0,0138)	0,0197 (0,0152)	0,0071 (0,0152)	0,0544*** (0,0137)
(<i>mean</i>) <i>i</i>			0,2499 (0,1792)	0,0052*** (0,0009)		0,2478 (0,1784)	0,0052*** (0,0009)
(<i>mean</i>) <i>f</i>			0,0079 (0,0063)	0,0012* (0,0007)		0,0081 (0,0063)	0,0012* (0,0007)
<i>frontier</i>	-0,0754* (0,0455)	-0,1214** (0,0504)	-0,0816** (0,0413)	-0,1224** (0,0511)	-0,0763* (0,0454)	-0,0829** (0,0414)	-0,1457*** (0,0452)
<i>touristic</i>	-0,0069 (0,0389)	0,0303 (0,0484)	0,0053 (0,0269)	0,0336 (0,0473)	-0,0112 (0,0393)	0,0014 (0,0272)	0,0118 (0,0469)
Norte	-0,0246 (0,1089)	0,0724 (0,0716)	0,0007 (0,0929)	0,0682 (0,0716)	-0,0352 (0,1088)	-0,0075 (0,0936)	0,0301 (0,0669)
Nordeste	-0,3819*** (0,0445)	-0,2176*** (0,0354)	-0,2925*** (0,0722)	-0,2294*** (0,0341)	-0,3818*** (0,0445)	-0,2932*** (0,0720)	-0,2421*** (0,0347)
Sul	-0,0383 (0,0354)	0,0114 (0,0334)	-0,0401 (0,0280)	0,0098 (0,0333)	-0,0360 (0,0355)	-0,0382 (0,0282)	0,0163 (0,0329)
Centro-Oeste	-0,0630 (0,0736)	-0,0271 (0,0480)	-0,0499 (0,0601)	-0,0295 (0,0479)	-0,0588 (0,0741)	-0,0467 (0,0606)	-0,0133 (0,0466)
<i>dummies de tempo</i>	Sim	Sim	Sim	Sim	Sim	Sim	Sim
<i>constante</i>	-0,0990 (0,1162)	-0,4160*** (0,1191)	0,0641 (0,1573)	-0,4401*** (0,1171)	-0,0939 (0,1160)	0,0668 (0,1562)	-0,4027*** (0,1189)
Observações	4623	4623	4623	4623	4511	4511	4511
R^2	0,505	0,213	0,602	0,235	0,508	0,604	0,251
R^2 Ajustado	0,503	0,210	0,601	0,231	0,506	0,602	0,247
Estatísticas do primeiro estágio - 2SLS							
<i>Bartik</i> IV		0,7923***		0,7911***			0,6557***
F(20,356)		218,73					
F(22,356)				200,71			
F(22,347)							186,85
Prob. > F		0,0000		0,0000			0,0000
R^2 Ajustado		0,915		0,916			0,736

Fonte: Elaboração própria, a partir da RAIS (2002-2014).

Notas: * $p < 0,10$, ** $p < 0,05$, *** $p < 0,01$; Erro-padrão robusto com *clusters* de LMAs; Modelos POLS assumem peso analítico proporcional ao número de indivíduos em cada LMA; Resultados completos do primeiro estágio (2SLS) na Tabela J.11 do Apêndice J.

Inicialmente, para permitir a comparação entre os modelos, foi necessário excluir alguns arranjos da estimação com a variável instrumental. Esse procedimento buscou retirar os arranjos que apresentaram uma estimativa negativa para o número de empregos (2), os arranjos correspondentes aos municípios isolados recém-criados (4), os quais não possuíam emprego em 1991, e os arranjos que não apresentaram emprego no setor público (8) em 1991. Esses arranjos geram estimativas negativas ou de valor zero para o emprego e, portanto, não permitem a aplicação do logaritmo. Nesse sentido, o número de observações é compatível entre colunas de 1 a 4, da Tabela 4.11, e entre as colunas de 5 a 7.

As estatísticas do primeiro estágio, no final da tabela, mostram que ambas as versões do *Bartik* IV apresentam correlação estatisticamente significativa com a densidade do emprego. O instrumento, conjuntamente com as demais covariadas exógenas do modelo, apresenta um alto poder explicativo para a variação na densidade do emprego (R^2). Portanto, os supostos da existência de covariância significativa com a variável endógena e da exogeneidade, com base no crescimento nacional, garantem a validade da identificação do modelo. As colunas 1 e 2 mostram os modelos POLS e 2SLS, respectivamente, sem a inclusão dos componentes não observados de indivíduos e firmas. Os resultados mostram uma mudança substancial no efeito da densidade sobre os salários locais com o uso do instrumento *Bartik* setorial. A estimativa do efeito de aglomeração foi reduzida de 4,2% para 2,5%, sugerindo que a endogeneidade tem forte impacto sobre a identificação do modelo.

Por sua vez, ao incluir os efeitos não observados das firmas e indivíduos, os resultados mostram uma mudança marginal no efeito da densidade, em ambas as versões do *Bartik* IV (Colunas 4 e 7), em relação ao modelo POLS com similares controles (Colunas 3 e 6). O *Bartik* setorial encontrou um efeito de 2,64% para a densidade do emprego, enquanto o modelo endógeno mostrou um efeito de 2,61%. Por sua vez, o *Bartik* do setor público estimou um efeito de 2,8% para a densidade, enquanto o modelo POLS encontrou um efeito de 2,7%. Portanto, os resultados sugerem que a endogeneidade não comprometeu a identificação dos efeitos de aglomeração sobre a produtividade local nos arranjos populacionais, uma vez que os componentes não observados são incluídos no modelo.

Se os efeitos da firma e do trabalhador já estiverem controlando, em alguma medida, os possíveis efeitos dos choques locais, isso poderia justificar os aumentos apenas marginais no efeito da densidade com o uso do instrumento. Nesse sentido, uma explicação possível para a

substancial redução no efeito de aglomeração com o uso do instrumento (Colunas 1 e 2), relativamente aos modelos que incluem os efeitos individuais e das firmas (Colunas 3 e 4), é que os componentes não observados correspondem a uma parcela do efeito endógeno da densidade. Esses resultados sugerem que a endogeneidade da densidade pode ser explicada pelo *sorting* espacial dos trabalhadores e das firmas com heterogêneas distribuições de produtividades.

Combes, Duranton e Gobillon (2008) utilizaram defasagens temporais da densidade e do mercado potencial, além de outras variáveis históricas, como instrumentos para a densidade do emprego nos mercados de trabalho locais da França. Os resultados mostraram uma mudança marginal nos efeitos da densidade, os quais foram reduzidos de 3,2% para 3% após o uso dos instrumentos. Ao controlar os efeitos individuais na equação salarial do primeiro estágio, os efeitos da densidade foram reduzidos para 2,9% no modelo de variável instrumental. Na literatura nacional, Barufi, Haddad e Nijkamp (2016) utilizaram a densidade populacional defasada (1940), a distância ao litoral e a exploração da cana de açúcar, ouro e café no período colonial como variáveis instrumentais para a densidade do emprego nas áreas REGIC. O efeito da densidade sobre os salários setoriais locais, após o uso dos instrumentos, variou entre 3,8% e 9,7%, os quais não se alteraram significativamente em relação às estimativas do modelo POLS (4,6% a 9,4%).

O modelo de variável instrumental também foi estimado com a amostra aleatória em firmas (Tabelas J.12 e J.13 – Apêndice J). Os efeitos da densidade foram reduzidos de 5% para 3,8% com o uso do instrumento, no modelo que não inclui os componentes não observados. Após incluir os componentes não observados, o efeito foi reduzido de 4,8% para 3,7% com o *Bartik* setorial e de 4,9% para 3,5% com o *Bartik* do setor público. Novamente, os efeitos da densidade nos modelos 2SLS com e sem a inclusão dos componentes não observados são aproximadamente equivalentes, sugerindo que a endogeneidade pode resultar do *sorting*.

Por outro lado, a seção anterior apresentou evidências de que o uso da área administrativa oficial para o cálculo da densidade do emprego subestimava os efeitos de aglomeração sobre os salários locais. Nesse sentido, os modelos foram estimados com a medida de área obtida a partir das imagens de luminosidade noturna. Contudo, ainda assim, persistia o problema da endogeneidade da densidade urbana. Nesse sentido, a Tabela 4.12 apresenta os resultados da

estimação do modelo de decomposição salarial com a densidade da área iluminada sendo instrumentalizada pelo *Bartik IV*. As estatísticas do primeiro estágio, apresentadas no final da tabela, mostram que o *Bartik IV* atende ao suposto de covariância parcial com a densidade da área iluminada, estabelecido pela significância estatística do parâmetro, bem como apresenta substancial poder explicativo para explicar a variação dessa densidade (R^2). As colunas 1 e 2 da tabela mostram os resultados da estimação dos modelos POLS e 2SLS sem a inclusão do componentes não observados das firmas e trabalhadores.

**Tabela 4.12 – Decomposição dos efeitos de localização dos arranjos (segundo estágio):
Bartik IV e densidade do emprego da área iluminada (2013)**

FE_LMA	(1) POLS	(2) 2SLS (<i>lbartik</i>)	(3) POLS	(4) 2SLS (<i>lbartik</i>)	(5) POLS	(6) POLS	(7) 2SLS (<i>lbartikpub</i>)
<i>ldensemplight13</i>	0,0545*** (0,0182)	0,0470** (0,0222)	0,0310 (0,0243)	0,0499** (0,0222)	0,0560*** (0,0183)	0,0323 (0,0246)	0,0530** (0,0235)
<i>larea_lma</i>	0,0094 (0,0179)	0,0408*** (0,0137)	0,0020 (0,0148)	0,0431*** (0,0136)	0,0078 (0,0180)	0,0008 (0,0148)	0,0364*** (0,0131)
(mean) <i>i</i>			0,2578 (0,1832)	0,0053*** (0,0009)		0,2561 (0,1827)	0,0053*** (0,0009)
(mean) <i>f</i>			0,0077 (0,0060)	0,0013* (0,0007)		0,0079 (0,0061)	0,0013* (0,0007)
<i>frontier</i>	-0,1290** (0,0517)	-0,1371*** (0,0477)	-0,1161*** (0,0447)	-0,1390*** (0,0485)	-0,1316** (0,0519)	-0,1189*** (0,0450)	-0,1627*** (0,0422)
<i>touristic</i>	0,0081 (0,0410)	0,0473 (0,0498)	0,0142 (0,0269)	0,0518 (0,0487)	0,0044 (0,0414)	0,0109 (0,0271)	0,0297 (0,0481)
Norte	-0,0731 (0,0916)	0,0646 (0,0702)	-0,0305 (0,0833)	0,0599 (0,0701)	-0,0843 (0,0915)	-0,0391 (0,0844)	0,0201 (0,0659)
Nordeste	-0,3834*** (0,0448)	-0,2065*** (0,0368)	-0,2913*** (0,0731)	-0,2181*** (0,0355)	-0,3833*** (0,0447)	-0,2918*** (0,0728)	-0,2287*** (0,0370)
Sul	-0,0501 (0,0382)	0,0063 (0,0337)	-0,0491* (0,0297)	0,0043 (0,0335)	-0,0483 (0,0383)	-0,0477 (0,0298)	0,0097 (0,0339)
Centro-Oeste	-0,0846 (0,0762)	-0,0383 (0,0474)	-0,0658 (0,0615)	-0,0415 (0,0473)	-0,0815 (0,0767)	-0,0635 (0,0619)	-0,0276 (0,0465)
<i>dummies de tempo</i>	Sim	Sim	Sim	Sim	Sim	Sim	Sim
<i>constante</i>	-0,0759 (0,1129)	-0,3995*** (0,1150)	0,0825 (0,1509)	-0,4235*** (0,1130)	-0,0707 (0,1127)	0,0855 (0,1499)	-0,3866*** (0,1160)
Observações	4623	4623	4623	4623	4511	4511	4511
R^2	0,491	0,204	0,596	0,227	0,494	0,597	0,237
R^2 Ajustado	0,489	0,201	0,594	0,223	0,491	0,595	0,233
Estatísticas do primeiro estágio 2SLS							
<i>Bartik IV</i>		0,4202***		0,4183***			0,3517***
F(20,356)		147,20					
F(22,356)				158,00			
F(22,347)							153,11
Prob. > F		0,000		0,000			0,000
R^2 Ajustado		0,737		0,741			0,575

Fonte: Elaboração própria, a partir da RAIS (2002-2014).

Notas: * $p < 0,10$, ** $p < 0,05$, *** $p < 0,01$; Erro-padrão robusto com *clusters* de LMAs; Modelos POLS assumem peso analítico proporcional ao número de indivíduos em cada LMA; Resultados completos do primeiro estágio (2SLS) na Tabela J.14 do Apêndice J.

Os resultados mostram uma redução de 5,4% para 4,7% na estimativa do efeito da densidade da área iluminada, entre os modelos POLS e 2SLS. Ao incluir os componentes não observados, as estimativas apresentam um crescimento de 3,1% para 4,9%, entre os modelos POLS e 2SLS (Colunas 3 e 4). Por sua vez, comparando os modelos 2SLS com e sem a inclusão dos componentes não observados, é possível perceber estimativas muito próximas do efeito da densidade da área iluminada (4,9% e 4,7%), o que reforça as evidências encontradas, as quais sugerem que uma parte da endogeneidade resulta do *sorting* espacial das heterogeneidades não observadas.

Ao analisar os modelos que incluem a variável instrumental e os componentes não observados, os resultados sugerem um efeito de 2,6% para a densidade do emprego sobre a produtividade local, enquanto a densidade da área iluminada apresentou um efeito de 4,9%. Portanto, o uso da área iluminada praticamente duplica o efeito da aglomeração urbana sobre a produtividade local, mesmo nos modelos que utilizam variável instrumental. Nesse sentido, os resultados desta tese sugerem que existe um efeito puro da aglomeração urbana sobre o diferencial salarial entre as principais áreas de emprego do Brasil, em todos os modelos considerados.

4.4 CONSIDERAÇÕES FINAIS

Esta tese estimou um modelo de decomposição salarial em dois estágios para identificar os efeitos de aglomeração sobre a produtividade local nos arranjos populacionais. A Tabela 4.13 resume os principais resultados encontrados, comparando com o estudo de Combes, Duranton e Gobillon (2008). O modelo foi inicialmente estimado com as especificações tradicionais da literatura, para mostrar como os resultados se alteram após as modificações sugeridas neste estudo. Nesse sentido, o modelo POLS sem a inclusão dos componentes não observados dos indivíduos e firmas foi assumido como uma referência inicial. Em seguida, os componentes não observados foram incluídos, reduzindo a estimativa dos efeitos de densidade de 4,2% para 2,6%. O uso da densidade calculada a partir da área iluminada ampliou o efeito para 3,1%. Por fim, o modelo principal deste estudo, IV com a área iluminada e os componentes não observados, gerou um efeito de 4,9%.

Tabela 4.13 – Estimativas dos efeitos de aglomeração (densidade do emprego)

Modelo	Literatura (%)	Tese (%)
1. POLS	4 a 8	4,2
2. POLS com FEi	3,2 a 3,7	2,58
3. POLS com FEi e FEj		2,59
4. POLS com área iluminada		5,4
5. POLS com área iluminada, FEi e FEj		3,1
6. 2SLS		2,5
7. 2SLS com FEi	3,02	
8. 2SLS com FEi e FEj		2,6
9. 2SLS com área iluminada		4,7
10. 2SLS com área iluminada, FEi e FEj		4,9

Fonte: Elaboração própria a partir das estimativas obtidas com a RAIS (2002-2014) e do estudo de Combes, Duranton e Gobillon (2008).

Nota: Os modelos estimados por Combes, Duranton e Gobillon (2008) incluíram o efeito de indivíduo no primeiro estágio.

A decomposição dos salários locais mostrou que os componentes não observados dos trabalhadores e firmas exercem um papel relevante na explicação dos diferenciais salariais entre os arranjos, reduzindo as estimativas dos efeitos puros da aglomeração. No entanto, o *sorting* dos atributos não observados também resulta dos efeitos da aglomeração e, portanto, pode ser considerado como uma parte desses efeitos. O *sorting* se mostrou responsável por quase metade dos efeitos de aglomeração, sugerindo a importância da distribuição espacial de trabalhadores e firmas com heterogêneos níveis de produtividade para a explicação dos diferenciais salariais entre os mercados de trabalho locais.

Portanto, esse trabalho contribui com as literaturas nacional e internacional da Economia Urbana e da Economia do Trabalho ao estimar modelos que controlam simultaneamente os efeitos das firmas, dos trabalhadores e da localização do emprego. As evidências sugerem a importância do *sorting* espacial baseado nos atributos não observados dos indivíduos e firmas para explicar os diferenciais salariais e de produtividade entre as áreas de emprego e os efeitos de aglomeração. Por sua vez, o estudo sugere uma medida mais adequada para calcular a área dos mercados de trabalho urbanos e as suas respectivas densidades. As evidências encontradas com a densidade do emprego calculada a partir da área iluminada mostram que os efeitos de aglomeração podem ser subestimados se as áreas administrativas oficiais de cada localidade são adotadas. Portanto, é necessário levar em conta a diferença entre as áreas administrativas e as áreas efetivamente ocupadas por firmas e habitações, para estimar os efeitos da densidade urbana. Se a diferença entre as áreas é muito grande, como mostram as imagens de satélite da iluminação noturna, as estimativas existentes estão subestimando os efeitos da aglomeração.

5 CONCLUSÃO

O objetivo deste trabalho de tese foi avaliar a contribuição da heterogeneidade não observada das firmas e dos trabalhadores para a determinação de salários individuais e dos salários locais (efeitos de localização) no mercado de trabalho formal do Brasil. As heterogeneidades não observadas foram definidas pelos atributos não observados e produtivos, os quais podem apresentar maior remuneração nas áreas mais densas. As áreas de emprego da análise foram definidas a partir dos arranjos populacionais do IBGE (2015), os quais são definidos a partir da integração urbana e dos movimentos pendulares para trabalho e estudo. Os arranjos concentraram 82% dos empregos formais (RAIS) e 66,4% da população em 2010, com base nas AMCs definidas neste estudo. O referencial teórico mostrou que os efeitos positivos da aglomeração urbana sobre a produtividade local (salários reais) podem ser confundidos com a concentração dos trabalhadores e firmas mais produtivos nos mercados de trabalho locais densos, caracterizando um processo de *sorting* espacial entre localizações.

O modelo de Combes, Duranton e Gobillon (2008) foi combinado com o modelo AKM de decomposição salarial para estimar os resultados de interesse deste trabalho. Ambos os modelos sugerem uma estimação em dois estágios, a qual foi utilizada para identificar os salários ou produtividade local (primeiro estágio) e encontrar a contribuição dos componentes locais observados e não observados para a decomposição dos salários locais (segundo estágio). Para encontrar os efeitos de localização dos arranjos e a contribuição dos efeitos da firma e do trabalhador para a determinação individual dos salários reais, o primeiro estágio do modelo estimou equações de salários utilizando um painel de micro dados com informações combinadas de trabalhadores e firmas (RAIS, 2002–2014). O painel foi construído de modo a garantir as condições de identificação simultânea dos efeitos fixos individuais, das firmas e dos arranjos populacionais. Nesse sentido, o painel garantiu um nível mínimo de mobilidade para as firmas entre os arranjos (6,14%) e para os trabalhadores entre as firmas (61,5%) e os arranjos populacionais (38,3%).

A amostra principal foi composta por 6.048.393 observações, correspondentes a 465.261 trabalhadores, 271.951 firmas e 1.081.990 combinações únicas de trabalhadores e firmas, ao longo do painel balanceado (2002–2014). Do total de trabalhadores na amostra, 86,6% pertencem a algum dos 365 arranjos e 13,4% estão nos municípios que não constituem um

arranjo. Entre os trabalhadores que estão em um arranjo, 4,1% estão em arranjos pequenos (população ≤ 100 mil habitantes), 23,5% em arranjos médios ($100\text{mil} < \text{pop.} \leq 750\text{mil}$) e 59% em grandes arranjos (população $> 750\text{mil}$). O modelo do primeiro estágio foi estimado com as características observadas dos trabalhadores e do emprego e os efeitos fixos dos arranjos, dos trabalhadores e das firmas, além das *dummies* de tempo. A estratégia empírica recorreu aos métodos e avanços da literatura da Economia do Trabalho, para lidar com múltiplos efeitos fixos em grandes bancos de micro dados pareados de trabalhador-firma, baseados no estudo de Abowd, Kramarz e Margolis (1999). Os resultados do primeiro estágio mostraram que as heterogeneidades não observadas individuais exerceram maior poder explicativo para a variação dos salários individuais do que os atributos não observados das firmas.

A variação salarial explicada pelo modelo (R^2) que considera apenas as características observadas individuais e do emprego, além da localização e do tempo, foi de 60,6%. Ao incluir somente os efeitos fixos das firmas, o modelo ampliou a variação salarial explicada para 78,4%. A inclusão apenas dos atributos individuais não observados elevou a variação salarial explicada pelo modelo para 91,7%. Por sua vez, a inclusão simultânea dos efeitos não observados dos trabalhadores e das firmas, além das demais covariadas observadas, explicou 93,5% da variação dos salários reais individuais. Portanto, os atributos individuais não observados respondem por uma grande parcela da variação salarial entre os indivíduos no mercado de trabalho formal, sendo mais relevantes do que os atributos específicos das firmas.

Os trabalhadores localizados nos arranjos obtiveram um prêmio salarial urbano de 23,7%, no modelo POLS do primeiro estágio, em relação aos trabalhadores que não estão em um dos arranjos. O prêmio médio foi maior para os trabalhadores localizados nos grandes arranjos (29,4%). O primeiro estágio encontrou um prêmio salarial urbano (efeitos fixos de localização positivos e significantes) em 19 dentre os 27 grandes arranjos. O prêmio urbano foi maior nos arranjos das macrorregiões do Sul e Sudeste do Brasil. Após controlar os efeitos fixos da firma e do trabalhador, apenas 2 grandes arranjos apresentaram um prêmio salarial urbano.

O estudo sugeriu um padrão de classificação dos arranjos em três grupos a partir do comportamento dos respectivos prêmios urbanos em relação aos componentes não observados dos trabalhadores e das firmas. O primeiro grupo corresponde aos arranjos cujo prêmio resulta dos atributos não observados individuais e por isso desaparece após o controle dos efeitos fixos individuais, a exemplo de Salvador e Florianópolis. O segundo grupo sugere que o

prêmio resulta dos atributos não observados da firma e, portanto, desaparece após o controle dos efeitos da firma, a exemplo de Manaus, Vitória, Campinas e Porto Alegre. Por último, o terceiro grupo sugere que o prêmio resulta do melhor *matching* entre os atributos não observados das firmas e dos trabalhadores, o que explicaria o seu desaparecimento apenas após o controle simultâneo dos efeitos das firmas e trabalhadores, a exemplo de Brasília, São Paulo e Rio de Janeiro.

O segundo estágio do modelo realizou a decomposição dos salários locais dos 362 arranjos populacionais, mensurados pelos respectivos efeitos fixos de localização estimados no primeiro estágio. A decomposição incluiu as características observadas dos arranjos e os componentes médios não observados dos indivíduos e firmas, estimados no primeiro estágio. A densidade do emprego foi incluída para captar os possíveis efeitos puros das economias de aglomeração urbana, limpos dos componentes não observados. Os resultados da decomposição mostraram que os componentes não observados dos indivíduos exercem maior poder explicativo para a variação de salários locais (produtividade) dos arranjos do que os componentes das firmas. A variação salarial explicada pelo modelo que incluiu apenas as características observadas dos arranjos foi de 50,2%, sendo ampliada marginalmente para 50,9% com a inclusão dos efeitos das firmas. O modelo com maior poder explicativo foi o que incluiu os efeitos individuais, ampliando a parcela explicada dos salários locais para 60%. A inclusão de ambos os efeitos manteve esse último percentual de variação explicada.

O efeito puro das economias de aglomeração urbana sobre a produtividade dos arranjos, representado pela densidade do emprego, apresentou estimativas positivas e estatisticamente significantes variando entre 2,6% e 4,2%. O maior efeito corresponde ao modelo POLS com as características observadas locais, enquanto o menor efeito corresponde ao modelo que incluiu os componentes não observados dos trabalhadores e firmas. A literatura empírica internacional encontrou efeitos de aglomeração variando entre 4% e 8%. No entanto, o estudo do Combes, Duranton e Gobillon (2008) encontrou estimativas em torno de 3%, sugerindo essa magnitude como um *lower bound* da literatura empírica. Portanto, as evidências nesta tese sugerem um valor menor do que o *lower bound* da literatura empírica prévia (3%), o que pode ser explicado pela inclusão no modelo dos componentes médios não observados das firmas e dos trabalhadores. A estimativa do efeito bruto de aglomeração, obtida pela soma dos efeitos

puros (2,6%) com os efeitos da densidade sobre os componentes não observados individuais (6,4) e das firmas (1,1%), foi de 10,1%.

Por sua vez, o estudo testou a hipótese de que os efeitos da aglomeração seriam subestimados se a área administrativa oficial, utilizada no cálculo da densidade urbana, fosse inferior à área efetivamente habitada e/ou ocupada pelas empresas. O estudo utilizou a área com luminosidade noturna, obtida a partir das imagens de satélite (NCEI, 2013), para definir a área efetivamente ocupada e mensurar a densidade do emprego nos arranjos. A área iluminada do Brasil, em 2013, correspondeu a somente 8,1% da sua área total (em km²). Por sua vez, 15,2% da área total dos arranjos apresentou luminosidade noturna, cujo percentual representa quase o dobro da luminosidade encontrada para a área total do Brasil. Nos grandes arranjos a parcela iluminada da sua área total é ainda maior (55,2%). Os modelos foram estimados com a densidade do emprego da área iluminada e confirmaram a hipótese de que os efeitos da densidade eram subestimados com o uso da área oficial. As evidências mostraram um efeito da densidade da área iluminada variando entre 2,9% e 5,4%, representando um aumento de 0,3 e 1,2 pontos percentuais, respectivamente, em relação aos efeitos mínimo e máximo estimados com a área administrativa oficial.

A literatura da Economia Urbana sugere a existência de causalidade reversa entre a densidade urbana e os salários locais, a qual gera endogeneidade na identificação dos efeitos da aglomeração. A literatura empírica costuma utilizar defasagens temporais da população ou variáveis geográficas como instrumento para a densidade. Este trabalho utilizou o instrumento *Bartik*, tradicional na Economia do Trabalho. O instrumento estima o emprego local a partir do crescimento setorial nacional, baseado em período inicial prévio, excluindo o emprego de cada respectivo arranjo do cálculo do crescimento. A demanda nacional é o canal exógeno através do qual o instrumento prevê a demanda por emprego local, sem os choques que causam a endogeneidade. Duas versões do instrumento *Bartik* foram adotadas: o *Bartik* setorial e o *Bartik* do setor público. Ambas as versões do instrumento atenderam às condições de covariância parcial com a densidade do emprego e significância estatística para a estimação da densidade, mas o *Bartik* setorial mostrou uma maior correlação com a endógena.

Os resultados estimados com o instrumento *Bartik* setorial, sem incluir os componentes não observados, mostraram uma redução no efeito da densidade de 4,2% (POLS) para 2,5% (2SLS). Por sua vez, ao comparar os modelos POLS e 2SLS que incluíam os componentes

não observados, a mudança estimada entre os modelos foi apenas marginal, passando de 2,61% para 2,64%. As evidências sugerem que uma parte da endogeneidade da densidade parece ser originada dos componentes não observados individuais e das firmas. Nesse sentido, se o modelo incluir tais componentes, a endogeneidade da densidade torna-se apenas um problema residual. Por fim, o modelo 2SLS foi estimado considerando a densidade calculada com a área iluminada e o instrumento *Bartik*. Essa é a principal especificação deste trabalho e buscou encontrar o efeito da densidade sobre os salários locais considerando as lacunas existentes nos modelos da Economia Urbana. Portanto, o modelo principal buscou lidar com o controle dos atributos não observados das firmas e dos trabalhadores, a endogeneidade da densidade e a subestimação dos efeitos de aglomeração com o uso da área administrativa.

O modelo 2SLS estimado com a densidade da área iluminada e o instrumento *Bartik* setorial encontrou um efeito da densidade de 4,7%, sem incluir os componentes não observados, e de 4,9% após a inclusão dos componentes. Nesse sentido, o efeito exógeno da densidade com a área iluminada é praticamente duplicado em relação aos modelos 2SLS que utilizam a área oficial para medir a densidade do emprego (2,5% e 2,6%). As evidências encontradas neste trabalho sugerem a importância de propor medidas mais adequadas para mensurar os efeitos da aglomeração, considerando as direções dos possíveis vieses existentes com a escolha de cada medida. A crescente disponibilidade recente de micro dados ou mesmo de macro dados, a exemplo das imagens de satélite, permite redefinir as medidas tradicionalmente usadas na literatura para uma identificação mais robusta dos resultados de interesse, com base nos respectivos fundamentos teóricos.

Este trabalho contribuiu com evidências para as literaturas nacional e internacional da Economia Urbana e da Economia do Trabalho. O estudo adicionou evidências sobre o papel dos atributos não observados dos trabalhadores e firmas na determinação de salários individuais e locais. A estratégia empírica é particularmente nova Brasil, considerando a estimação de um modelo com controle simultâneo de múltiplos efeitos fixos (indivíduos, firmas e localização), a partir de um grande banco de micro dados longitudinal e pareado entre trabalhadores e firmas no mercado de trabalho formal (RAIS). Além disso, a decomposição dos salários locais em componentes não observados mostrou como as estimativas dos efeitos da densidade podem ser alteradas. Por sua vez, o estudo sugere uma medida mais adequada para definir a área de uma localidade e mostra evidências para o Brasil

do uso da nova medida calculada a partir da proporção da luminosidade noturna local. Novamente, as evidências mostram alterações nas estimativas dos efeitos da densidade com o uso da área iluminada.

As evidências podem contribuir para a investigação dos componentes que geram as desigualdades espaciais persistentes entre os heterogêneos mercados de trabalhos locais do Brasil. Os resultados deste trabalho sugerem a importância de entender o funcionamento dos mercados de trabalho mais dinâmicos e, a partir desse diagnóstico, formular políticas que permitam estimular o desenvolvimento das áreas economicamente excluídas. Entre as evidências encontradas neste trabalho, existe um *sorting* espacial dos trabalhadores e firmas mais produtivos para os mercados de trabalho densos. O *sorting* pode ser explicado pela maior remuneração a determinados tipos de atributos produtivos não observados. Esse aspecto pode ser investigado com maior profundidade considerando o perfil dos trabalhadores e as características das firmas e dos setores presentes nas áreas de emprego mais dinâmicas do país. Esse aspecto é particularmente importante, pois o *sorting* tende a concentrar os melhores trabalhadores e firmas nas mesmas áreas de emprego, o que pode ampliar ainda mais as desigualdades espaciais previamente existentes.

Extensões possíveis deste trabalho incluem investigar a distribuição setorial e ocupacional dos componentes não observados dos indivíduos e firmas e sua relação com as economias de aglomeração urbana. O estudo dos efeitos de aglomeração para os trabalhadores que entram e saem do mercado de trabalho formal, além daqueles que estão fora desse mercado, se mostra uma extensão relevante. Além disso, o estudo pode investigar como os efeitos de aglomeração atuam sobre os trabalhadores com mais de um emprego. Outras extensões possíveis são considerar os componentes dinâmicos dos efeitos de aglomeração, a exemplo do *learning* e do melhor *matching* das vagas preenchidas nas grandes cidades, e avaliar os efeitos de aglomeração operando no nível das firmas. Além disso, um espaço ainda aberto na literatura da Economia Urbana é entender os principais canais através dos quais o efeito da densidade opera, gerando um aumento na produtividade local.

REFERÊNCIAS

- ABOWD, J. M.; CREECY, R. H.; KRAMARZ, F. Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data. **Longitudinal Employer-Household Dynamics Technical Papers**, Center for Economic Studies, US Census Bureau, n. 6. 2002. Disponível em: <<ftp://ftp2.census.gov/ces/tp/tp-2002-06.pdf>>. Acesso em: 6, Jan. 2015.
- ABOWD, J. M.; KRAMARZ, F.; MARGOLIS, D. N. High Wage Workers and High Wage Firms. **Econometrica**, v. 67, n. 2, p. 251-333, 1999.
- ABOWD, J. M.; KRAMARZ, F.; MARGOLIS, D. N.; TROSKE, K. R. The relative importance of employer and employee effects on compensation: a comparison of France and the United States. **Journal of the Japanese and International Economies**, v. 15, n. 4, p. 419-436, 2001.
- AGÊNCIA NACIONAL DO PETRÓLEO, GÁS NATURAL E BIOCOMBUSTÍVEIS – ANP. **Anuário estatístico brasileiro do petróleo, gás natural e biocombustíveis**. Rio de Janeiro: ANP, 2016. Disponível em: <http://www.anp.gov.br/wwwanp/images/publicacoes/Anuario_Estatistico_ANP_2016.pdf>. Acesso em: 20 Jan. 2017.
- AMARAL, P. V.; LEMOS, M.; SIMÕES, R.; CHEIN, F. Regional imbalances and market potential in Brazil. **Spatial Economic Analysis**, v. 5, n. 4, p. 463-482, 2010.
- ANDERSSON, F.; BURGESS, S.; LANE, J. I. Cities, matching and the productivity gains of agglomeration. **Journal of Urban Economics**, v. 61, p. 112–128, 2007.
- ANDREWS, M.; SCHANK, T.; UPWARD, R. Practical fixed-effects estimation methods for the three-way error-components model. **The Stata Journal**, v. 6, n. 4, p. 461-481, 2006.
- AUTOR, D. H.; DUGGAN, M. G. The Rise in the Disability Rolls and the Decline in Unemployment. **The Quarterly Journal of Economics**, v. 118, n. 1, p. 157-206, 2003.
- AZZONI, C. R. Personal income distribution within states and income inequality between states in Brazil: 1960, 70, 80 and 91. In: HEWINGS, Geoffrey et al. (Eds.). **Understanding and Interpreting Economic Structure**. Springer-Verlag Berlin Heidelberg, 1999. 372 p.
- BACOLOD, M.; BLUM, B. S.; STRANGE, W. C. Elements of skill: traits, intelligences, education, and agglomeration. **Journal of Regional Science**, v. 50, n. 1, p. 245-280, 2010.
- BARROS, R. P.; FRANCO, S.; MENDONÇA, R. A Recente Queda na Desigualdade de Renda e o Acelerado Progresso Educacional Brasileiro da Última Década. In: BARROS, R. P.; FOGUEL, M. N.; ULYSSEA, G. (Orgs.). **Desigualdade de Renda no Brasil: uma análise da queda recente**. Brasília: IPEA, v. 2, 2007. 552 p.
- BARROS, R. P.; HENRIQUES, R.; MENDONÇA, R. Pelo fim das décadas perdidas: educação e desenvolvimento sustentado no Brasil. In: HENRIQUES, Ricardo (Org.). **Desigualdade e pobreza no Brasil**. Rio de Janeiro: IPEA, 2000. 740 p. Disponível em: <<http://www.ipea.gov.br/sites/000/2/livros/desigualdadepobrezabrasil/capitulo14.pdf>>. Acesso em: 06 Ago. 2011.

BARTIK, Timothy J. The effects of metropolitan job growth on the size distribution of family income. **Journal of Regional Science**, v. 34, n. 4, p. 483-501, 1994.

BARTIK, Timothy J. **Who Benefits from State and Local Economic Development Policies?** Kalamazoo, MI: W.E. Upjohn Institute for Employment Research. 1991. 354 p.

BARUFI, A. M. B. **Agglomeration economies and labour markets in Brazil**. PhD Dissertation (Doctorate in Economics) – School of Economics, Business and Accounting, University of São Paulo, São Paulo. 2015. 171 p.

BARUFI, A.M.B.; HADDAD, E.A.; NIJKAMP, P. Industrial scope of agglomeration economies in Brazil. **The Annals of Regional Science**, v. 56, n. 3, p. 707–755, 2016.

BLANCHARD, O. J.; KATZ, L.F. Regional evolutions. **Brookings Papers on Economic Activity**, v. 1992, n. 1, p. 1-75, 1992.

BLEAKLEY, H.; LIN, J. Portage and Path Dependence. **Quarterly Journal of Economics**, v. 127, n. 2, p. 587-644, 2012.

CAIXA. **Programa de Integração Social - PIS**. Disponível em: <<http://www.caixa.gov.br/voce/social/beneficios/pis/index.asp>>. Acesso em: 4 Nov. 2012.

CAMERON, A. C.; TRIVEDI P. K. **Microeconometrics using Stata**. Stata Press, College Station, 2009. 692 p.

CAMPOS, F. M.; SILVEIRA NETO, R. M. A importância da dimensão do mercado de trabalho para os diferenciais de participação e salários entre gêneros: Uma análise empírica para os centros urbanos brasileiros. In: ENCONTRO NACIONAL DE ECONOMIA, 37, 2009, Foz do Iguaçu. **Anais...** Foz do Iguaçu: ANPEC, 2009. Disponível em: <http://www.anpec.org.br/encontro2009/inscricao.on/arquivos/000-c6f925639a711d35a1187_c1b9c5483f3.pdf>. Acesso em: 16, Jun. 2011.

COMBES, P. P.; DURANTON, G.; GOBILLON, L. Spatial wage disparities: Sorting matters. **Journal of Urban Economics**, n. 63, p. 723–742, 2008.

COMBES, P. P.; DURANTON, G.; GOBILLON, L. The identification of agglomeration economies. **Journal of Economic Geography**, v.11, n. 2, p. 253-266, 2011.

COMBES, P. P.; DURANTON, G.; GOBILLON, L.; PUGA, D.; ROUX, S. The productivity advantages of large cities: Distinguishing agglomeration from firm selection. **Econometrica**, v. 80, n. 6, p. 2543-2594, 2012a.

COMBES, P. P.; DURANTON, G.; GOBILLON, L.; ROUX, S. Sorting and local wage and skill distributions in France. **Journal of Regional Science and Urban Economics**, n. 42, p. 913–930, 2012b.

CARD, David. Immigrant Inflows, Native Outflows and the Local Labor Market Impacts of Higher Immigration. **Journal of Labor Economics**, v. 19, n. 1, p. 22-64, 2001.

CORREIA, Sergio. **Explanation of the HDFE iteration with 3 FEs**. 2014. Disponível em: <http://fmwww.bc.edu/repec/bocode/r/reghdfe_explanation.pdf>. Acesso em: 3 Jan. 2015.

CRUZ, B. O.; NATICCHIONI, P. Falling urban wage premium and inequality trends: evidence for Brazil. In: ENCONTRO NACIONAL DE ECONOMIA - ANPEC, 40, 2012, Porto de Galinhas. **Anais...** Porto de Galinhas: ANPEC, 2012. Disponível em: <http://www.anpec.org.br/encontro_2012.htm>. Acesso em: 20 Dez. 2012.

CRUZ, Bruno de Oliveira... [et al.] (Orgs.). **Economia regional e urbana: teorias e métodos com ênfase no Brasil**. Brasília: IPEA, 2011. 406 p.

DAUTH, W.; FINDEISEN, S.; MORETTI, E.; SUEDEKUM, J. Spatial wage disparities: Workers, firms, and assortative matching. In: ANNUAL NORTH AMERICAN MEETINGS OF THE REGIONAL SCIENCE ASSOCIATION INTERNATIONAL, 63, 2016, Minneapolis. **Proceedings...** Minneapolis: NARSC, 2016.

DIAMOND, Rebecca. The determinants and welfare implications of US workers' diverging location choices by skill: 1980–2000. **The American Economic Review**, v. 106, n. 3, p. 479–524, 2016.

DIXIT, A.; STIGLITZ, J. Monopolistic competition and optimum product diversity. **American Economic Review**, v. 67, p. 297–308, 1977.

DONALDSON, D.; STOREYGARD, A. The View from Above: Applications of Satellite Data in Economics. **Journal of Economic Perspectives**, v. 30, n. 4, p. 171–198, 2016.

EDLUND, L.; MACHADO, C.; SVIATSCHI, M. M. Bright Minds, Big Rent: Gentrification and the Rising Returns to Skill. **US Census Bureau Center for Economic Studies Paper**, working paper, n. 16-36R, novembro, 2016. Disponível em: <<https://ssrn.com/abstract=2871597>>. Acesso em: 01 Dez. 2016.

EHRL, P.; MONASTERIO, L. M. Historical trades, skills and agglomeration economies. **Munich Personal RePEc Archive (MPRA) Working Paper**, n. 69829, April, 2016. Disponível em: <<https://mpra.ub.uni-muenchen.de/69829/>>. Acesso em: 18 Nov. 2016.

ENTORF, H.; GOLLAC, M.; KRAMARZ, F. New Technologies, Wages, and Worker Selection. **Journal of Labor Economics**, v. 17, n. 3, p. 464–491, 1999.

FAGGIO, G.; OVERMAN, H. The effect of public sector employment on local labour markets. **Journal of Urban Economics**, v. 79, p. 91–107, 2014.

FALCÃO, N.; SILVEIRA NETO, R. Concentração espacial de capital humano e externalidades: o caso das cidades brasileiras. In: 35º Encontro Nacional de Economia, Recife, 2007. **Anais...** Recife: ANPEC, 2007.

FALLY, T.; PAILLACAR, R; TERRA, C. Economic geography and wages in Brazil: Evidence from micro-data. **Journal of Development Economics**, v. 91, ed. 1, p. 155–168, 2010.

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

FREGUGLIA, R. S. Os efeitos da migração sobre os salários e o processo de assimilação dos trabalhadores no Brasil. Brasília: IPEA, **Boletim Mercado de Trabalho - Conjuntura e Análise**, n. 48, fev, 2009. Disponível em: <http://www.ipea.gov.br/portal/images/stories/PDFs/mercadodetrabalho/3Nota_Ricardo.pdf>. Acesso em: 17, Nov. 2011.

FREGUGLIA, R. S.; MENEZES-FILHO, N. A. Inter-regional wage differentials with individual heterogeneity. **The Annals of Regional Science**, v. 49, n.1, p. 17-34, 2012.

FREGUGLIA, R. S.; MENEZES-FILHO, N.; SOUZA, D. B. Diferenciais Salariais Inter-regionais, Inter-industriais e Efeitos Fixos Individuais: Uma Análise a Partir de Minas Gerais. **Estudos Econômicos**, São Paulo, v. 3, n. 1, p. 129-150, 2007.

FREGUGLIA, R. S.; PROCÓPIO, T. S. Efeitos da mudança de emprego e da migração interestadual sobre os salários no Brasil formal: Evidências a partir de dados em painel. **Pesquisa e Planejamento Econômico (PPE)**, v. 43, n. 2, 2013.

FUJITA, M.; KRUGMAN, P.; VENABLES, A. **Economia espacial**. São Paulo: Futura, 2002.

FUJITA, M.; KRUGMAN, P.; VENABLES, A. **The Spatial Economy: Cities, Regions, and International Trade**. Cambridge: Massachusetts Institute of Technology (MIT Press). 1999.

FUJITA, M.; THISSE J. **Economics of agglomeration: Cities, industrial location, and regional growth**. Cambridge: Cambridge University Press. 2002.

GAURE, S. OLS with multiple high dimensional category variables. **Computational Statistics and Data Analysis**, v. 66, p. 8–18, 2013.

GLAESER, E.; MARÉ, D. Cities and Skills. **National Bureau of Economic Research (NBER)**, Working Papers, n. 4728. 1994.

GLAESER, E.; MARÉ, D. Cities and Skills. **Journal of Labor Economics**, v. 19, n. 2, p. 316-342, 2001.

GLAESER, E.; RESSEGER, M. G. The complementarity between cities and skills. **Journal of Regional Science**, v. 50, n. 1, p. 221–244, 2010.

GOUX, D.; MAURIN, E. Persistence of Interindustry Wage Differentials: A Reexamination Using Matched Worker-Firm Panel Data. **Journal of Labor Economics**, vol. 17, n. 3, p. 492-533, 1999.

GRAHAM, D. J.; MELO, P. S.; JIWATTANAKULPAISARN, P.; NOLAND, R. B. Testing for causality between productivity and agglomeration economies. **Journal of Regional Science**, v. 50, n. 5, p. 935-951, 2010.

GUIMARÃES, P.; PORTUGAL, P. A Simple Feasible Alternative Procedure to Estimate Models with High-Dimensional Fixed Effects. **The Stata Journal**, v. 10, n. 4, p. 628-649, 2010.

HENDERSON, J. V.; STOREYGARD, A.; WEIL, D. Measuring Growth from Outer Space. **American Economic Review**, v. 102, n. 2, p.994-1028, 2012.

HEUERMAN, D.; HALFDANARSON, B.; SUEDEKUM, J. Human capital externalities and the urban wage premium: Two literatures and their interrelations. **Urban Studies**, v. 47, n. 4, p. 749-767, 2010.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. **Arranjos Populacionais e Concentrações Urbanas do Brasil**. Rio de Janeiro, 2015. Disponível em: <www.ibge.gov.br/apps/arranjos_populacionais/2015>. Acesso em 10 Out. 2015.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. **Censo Demográfico**, 2000; 2010. Disponível em: <<http://www.ibge.gov.br/home/estatistica/populacao/censo2010/default.shtm>>. Acesso em: 08 Out. 2015.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. **Classificação Nacional de Atividades Econômicas – CNAE (versão 1.0)**. Disponível em: <<http://cnae.ibge.gov.br/busca-online-cnae.html?view=estrutura>>. Acesso em: 04 Jan. 2017.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. **Downloads: Áreas Territoriais**. 2013. Disponível em: <http://downloads.ibge.gov.br/downloads_geociencias.htm>. Acesso em: 20 Out. 2015.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA-IBGE. **Estimativas de População**, 2008. Disponível em: <<http://www.ibge.gov.br/home/estatistica/populacao/estimativa2008/default.shtm>>. Acesso em: 30 Out. 2015.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. **Índice Nacional de Preços ao Consumidor Amplo - IPCA**. Disponível em: <<http://www.sidra.ibge.gov.br/snipc/tabelaIPCA.asp?o=3&i=P>>. Acesso em: 05 Maio. 2012.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. **Pesquisa Nacional por Amostra de Domicílios (PNAD)**. Sistema IBGE de Recuperação Automática (SIDRA). 2004; 2014. Disponível em: <<https://sidra.ibge.gov.br/pesquisa/pnad/geral/pesquisa-basica>>. Acesso em: 20 dez. 2016.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA – IBGE. **Regiões de Influência das Cidades-REGIC 2007**. Rio de Janeiro: IBGE, 2008. 201 p. Disponível em: <<http://www.ibge.gov.br/home/geociencias/geografia/regic.shtm>>. Acesso em: 17 Nov. 2014.

JACOBS, J. **The Economy of Cities**. New York: Vintage, 1969.

KOMATSU, B. K.; MENEZES-FILHO, N. A. Salário Mínimo e Desigualdade Salarial: um estudo com densidades contrafactuais nas Regiões Metropolitanas brasileiras. **Pesquisa e Planejamento Econômico**, v. 45, n. 3, dez., 2015.

KRUGMAN, P. Increasing returns and economic geography. **Journal of Political Economy**, n. 99, p. 484–99, 1991.

KRUGMAN, P. Increasing returns, monopolistic competition, and international trade. **Journal of International Economics**, v. 9, p. 469–79, 1979.

KRUGMAN, P. Space: the final frontier. **Journal of Economic Perspective**, v. 12, n. 2, p. 161-174, 1998.

LANGONI, C. G. **Distribuição de Renda e Desenvolvimento Econômico no Brasil**. 3. ed. Rio de Janeiro: FGV, 2005. 280p.

LEONARD, J. S.; VAN AUDENRODE, M. Persistence of Firm and Individual Wage Components. **Working Paper**, Recherche en Politique Economique, Universite Laval, n. 9607, 1996. Disponível em: <<http://www.ecn.ulaval.ca/w3/recherche/cahiers/1996/9607.pdf>>. Acesso em: 10 Jan. 2015.

LOWE, Matt. Night Lights and ArcGIS: A Brief Guide. **Working Paper MIT**, January, 2014. Disponível em: <<http://economics.mit.edu/files/8945>>. Acesso em: 01 Out. 2016.

MARSHALL, A. **Principles of Economics**. 8th ed. London: Macmillan, 1920.

MATTOS, Fernando A. M. Emprego público no Brasil: Aspectos históricos, inserção no mercado de trabalho nacional e evolução recente. Brasília: **Texto para Discussão-IPEA**, n. 1582, 2011. Disponível em: <http://www.ipea.gov.br/portal/images/stories/PDFs/TDs/td_1582.pdf>. Acesso em: 10 Out. 2016.

McCAFFREY D. F.; LOCKWOOD, J. R.; MIHALY, K.; SASS, T. R. A Review of Stata Routines for Fixed Effects Estimation in Normal Linear Models. **The Stata Journal**, v. 12, n. 3, p. 406-432, 2012.

MELO, P. C.; GRAHAM, D. J.; NOLAND, R. B. A meta-analysis of estimates of urban agglomeration economies. **Regional Science and Urban Economics**, v. 39, n. 3, p. 332-342, 2009.

MENEZES, T.; AZZONI, C. Convergência de salários entre as regiões metropolitanas brasileiras: custo de vida e aspectos de demanda e oferta de trabalho. **Pesquisa e Planejamento Econômico (PPE)**, v. 36, n. 3, 2006.

MENEZES-FILHO, N. A.; FERNANDES, R.; PICCHETTI, P. Educação e Queda Recente da Desigualdade no Brasil. In: BARROS, R. P.; FOGUEL, M. N.; ULYSSEA, G. (Orgs.). **Desigualdade de Renda no Brasil: uma análise da queda recente**. Brasília: IPEA, v. 2, 552p. 2007. Disponível em: <<http://www.ipea.gov.br/sites/000/2/livros/desigualdaderendanobrasilv2/Livrocompleto.pdf>> Acesso em: 26 Abr. 2012.

MENEZES-FILHO, N. A.; MUENDLER, M. A.; RAMEY, G. The Structure of Worker Compensation in Brazil, with a Comparison to France and the United States. **Review of Economics and Statistics**, v. 90, n. 2, p. 324-346, 2008.

MICHALOPOULOS, S.; PAPAIOANNOU, E. Pre-Colonial Ethnic Institutions and Contemporary African Development. **Econometrica**, v. 81, n. 1, p. 113-152, 2013.

MINCER, J. **Schooling, experience, and earnings**. New York: Columbia University Press, 1974.

MINISTÉRIO DO TRABALHO E EMPREGO - MTE. **Registros administrativos: Relação Anual de Informações Sociais-RAIS e Cadastro Geral de Empregados e Desempregados-CAGED**. Brasília: MTE, SPPE/DES/CGET, 2010. 17p.

MINISTÉRIO DO TRABALHO E EMPREGO - MTE. **Relação Anual de Informações Sociais e Migração-RAIS-Migra**. Programa de Disseminação de Estatísticas do Trabalho (PDET). Disponível em: <http://www.mte.gov.br/pdet/o_pdet/produtos/BD_estatisticas.asp#raismigra>. Acesso em: 18 Ago. 2013.

MITTAG, N. New methods to estimate models with large sets of fixed effects with an application to matched employer-employee data from Germany. **Research Data Centre (FDZ) Method Report**, Institute for Employment Research (IAB), Germany, n. 1. 2012. Disponível em: <http://doku.iab.de/fdz/reporte/2012/MR_01-12_EN.pdf>. Acesso em: 3 Jan. 2015.

MORETTI, Enrico. Local Multipliers. **American Economic Review**, v. 100, n. 2, p. 373-77, 2010.

NATIONAL CENTERS FOR ENVIRONMENTAL INFORMATION – NCEI. **Global DMSP-OLS Nighttime Lights Time Series 1992 – 2013** (Version 4). National Oceanic and Atmospheric Administration – NOAA, 2013. Disponível em: <<https://ngdc.noaa.gov/eog/dmsp/download/V4composites.html>>. Acesso em: 01 Out. 2016.

OUAZAD, Amine. **Program for the Estimation of Two-Way Fixed Effects**. 2007. Disponível em: <<http://personal.lse.ac.uk/ouazad/>>. Acesso em: 6 jan. 2015.

PUGA, D. The magnitude and causes of agglomeration economies. **Journal of Regional Science**, v. 50, n. 1, p. 203-219, 2010.

RAUCH, J. Productivity gains from geographic concentration of Human Capital: Evidence from Cities. **Journal of Urban Economics**, n.34, p. 380-400, 1993.

REIS, E.J.; PIMENTEL, M.; ALVARENGA, A.I.; SANTOS, M.C.H. Áreas mínimas comparáveis para os períodos intercensitários de 1872 a 2000. In: SIMPÓSIO BRASILEIRO DE CARTOGRAFIA HISTÓRICA, 1, 2011, Paraty. **Anais...** Paraty: Centro de Referência em Cartografia Histórica da Universidade Federal de Minas Gerais, 2011. Disponível em: <https://www.ufmg.br/rededemuseus/crch/simposio/REIS_EUSTAQUIO_JOSE_R_ET_AL.pdf>. Acesso em: 01 Nov. 2015.

ROCHA, R. M.; SILVEIRA NETO, R. M.; GOMES, S. M. F. P. O. Maiores cidades, maiores habilidades produtivas: ganhos de aglomeração ou atração de habilidosos? Uma análise para as cidades brasileiras. In: ENCONTRO REGIONAL DE ECONOMIA – ANPEC NORDESTE, 16, 2011, Fortaleza. **Anais...** Fortaleza: ANPEC, 2011. Disponível em:

<http://www.bnb.gov.br/content/aplicacao/eventos/forumbnb2011/docs/2011_maiores_cidades.pdf>. Acesso em: 15 Dez. 2011.

ROSENTHAL, S. S.; STRANGE, W. C. Evidence on the nature and sources of agglomeration economies. **Handbook of Regional and Urban Economics**, v. 4, p. 2119-2171, 2004.

ROSENTHAL, S. S.; STRANGE, W. C. The Attenuation of Human Capital Externalities. **Journal of Urban Economics**, v. 64, p. 373-389, 2008.

SAVEDOFF, D. Os diferenciais regionais de salários no Brasil: segmentação *versus* dinamismo da demanda. **Pesquisa e Planejamento Econômico**, v. 20, n. 3, 1990.

SERVO, L.; AZZONI, C. Education, cost of living and regional wage inequality in Brazil. **Papers in Regional Science**, Springer, v. 81, n. 2, p. 157-175, 2002.

SILVA, D. L. G.; SANTOS, G. F.; FREGUGLIA, R. S. Distribuição espacial dos efeitos de aglomeração sobre os retornos à educação no Brasil entre 1995 e 2008. **Pesquisa e Planejamento Econômico**, v. 46, n. 2, p. 7-46, 2016.

SILVEIRA NETO, R.; AZZONI, C. Disparidades Regionais de Renda no Brasil: Qual o Papel das Amenidades? In: ENCONTRO REGIONAL DE ECONOMIA - ANPEC NORDESTE, 9, 2004, Fortaleza. **Anais...** Fortaleza: ANPEC, 2004. Disponível em: < <http://www.bnb.gov.br/edicoes-antiores2>>. Acesso em: 10 Jan. 2017.

SMALL, C.; ELVIDGE, C. D.; BALK, D.; MONTGOMERY, M. Spatial scaling of stable night lights. **Remote Sensing of Environment**, v. 115, n. 2, p. 269-280, 2011.

SMALL, C.; POZZI, F.; ELVIDGE, C. D. Spatial analysis of global urban extent from DMSP-OLS night lights. **Remote Sensing of Environment**, v. 96, n. 3, p. 277-291, 2005.

SUTTON, P.; ROBERTS, D.; ELVIDGE, C.; MEIJ, H. A comparison of nighttime satellite imagery and population density for the continental United States. **Photogrammetric Engineering and Remote Sensing**, v. 63, n. 11, p. 1303-1313, 1997.

VENABLES, Anthony J. Productivity in cities: self-selection and sorting. **Journal of Economic Geography**, v. 11, n. 2, p. 241-251, 2011.

WELCH, R. Monitoring urban population and energy utilization patterns from satellite data. **Remote sensing of Environment**, v. 9, n. 1, p. 1-9, 1980.

WOOLDRIDGE, J. M. **Econometric analysis of cross section and panel data**. Cambridge: Massachusetts Institute of Technology – MIT Press, 2002. 752p.

YANKOW, J. J. Why do cities pay more? An empirical examination of some competing theories of the urban wage premium. **Journal of Urban Economics**, vol. 60, p. 139-161, 2006.

APÊNDICE

APÊNDICE A – Distribuição do Emprego Formal e Informal no Brasil

Tabela A.1 – Ocupados por posição na ocupação: Brasil e regiões (2004; 2014)

Posição na Ocupação	Ocupados (em mil)		Ocupados (%)	
	2004	2014	2004	2014
Formal	33085	48140	38.81	48.41
Norte	1706	2752	26.11	35.07
Nordeste	5518	8741	24.15	34.12
Centro-Oeste	2488	4106	40.22	52.37
Sudeste	17370	24080	48.79	56.68
Sul	6003	8461	42.67	54.04
Informal	20397	19003	23.93	19.11
Norte	1719	1806	26.30	23.01
Nordeste	6068	6314	26.55	24.64
Centro-Oeste	1733	1557	28.01	19.86
Sudeste	8326	7091	23.39	16.69
Sul	2551	2235	18.13	14.27
Empregador	3501	3728	4.11	3.75
Norte	237	208	3.63	2.65
Nordeste	652	621	2.85	2.42
Centro-Oeste	286	328	4.62	4.18
Sudeste	1601	1804	4.50	4.25
Sul	725	767	5.15	4.90
Conta-Própria	18740	21171	21.98	21.29
Norte	1733	2154	26.52	27.45
Nordeste	6260	6384	27.39	24.92
Centro-Oeste	1221	1527	19.74	19.47
Sudeste	6628	8084	18.62	19.03
Sul	2898	3022	20.60	19.30
Outros	9523	7405	11.17	7.45
Total	85247	99447	100	100
Norte	6535	7848	100	100
Nordeste	22853	25620	100	100
Centro-Oeste	6186	7841	100	100
Sudeste	35604	42481	100	100
Sul	14069	15657	100	100

Fonte: Elaboração própria a partir da PNAD (2004; 2014).

Nota: A categoria formal compõe os empregados e trabalhadores domésticos com carteira de trabalho assinada e a categoria informal é composta pelos trabalhadores sem carteira.

Apêndice B – Organização Inicial dos Micro Dados da RAIS

A Tabela B.1 mostra as observações excluídas em cada etapa da limpeza inicial das bases anuais derivadas dos micros dados da RAIS (2002-2014). A etapa 1 excluiu os trabalhadores de tempo parcial (carga horária semanal contratada inferior à 20 horas), os registros sem remuneração ou sem PIS e os trabalhadores do setor público. A etapa 2 excluiu os trabalhadores que apresentavam mais de cinco vínculos empregatícios. A etapa 3 excluiu os vínculos cujo registro do gênero era diferente da média de cada trabalhador. Para manter apenas o vínculo principal de cada trabalhador, optou-se por excluir os registros cuja carga horária semanal contratada fosse inferior à média dos demais vínculos (etapa 4) e aqueles com data de admissão maior do que a média dos demais vínculos (etapa 5), ou seja, os vínculos mais atuais. Por fim, para aqueles trabalhadores que ainda permaneceram com mais de um vínculo, optou-se por gerar um número aleatório para cada vínculo mantendo aleatoriamente apenas um deles para cada trabalhador. A última coluna da Tabela B.1 mostra que, em média, 73% dos micro dados foram mantidos após os procedimentos iniciais de limpeza.

Tabela B.1 – Etapas iniciais de limpeza da RAIS: 2002–2014

Ano	RAIS	Etapa 1	Etapa 2	Etapa 3	Etapa 4	Etapa 5	Etapa 6	RAIS Limpa	RAIS (%)
2002	28.683.913	7.952.803	6.485	37.335	58.655	115.377	11.577	20.501.681	71,5
2003	29.544.927	8.111.654	2.171	35.658	59.702	118.792	14.035	21.202.915	71,8
2004	31.407.576	8.367.376	2.021	40.944	62.885	124.379	10.147	22.799.824	72,6
2005	33.238.617	8.943.268	1.553	37.326	68.286	128.049	13.298	24.046.837	72,3
2006	35.155.249	9.465.617	2.022	35.302	73.605	137.200	14.287	25.427.216	72,3
2007	37.607.430	10.102.695	1.216	34.917	77.192	144.012	15.013	27.232.385	72,4
2008	39.441.566	10.279.873	1.102	35.444	82.869	153.134	14.837	28.874.307	73,2
2009	41.207.546	10.789.083	764	28.723	88.735	148.985	15.240	30.136.016	73,1
2010	44.068.355	11.109.374	784	29.104	101.378	187.912	20.862	32.618.941	74,0
2011	46.310.631	11.511.555	1.330	30.261	114.109	209.679	27.174	34.416.523	74,3
2012	47.458.712	11.404.203	3.186	33.215	126.106	205.778	21.456	35.664.768	75,1
2013	48.948.433	11.835.592	1.162	34.744	133.649	214.847	13.853	36.714.586	75,0
2014	49.571.510	12.025.587	2.008	35.762	142.340	213.000	14.195	37.138.618	74,9
Total	512.644.465	131.898.680	25.804	448.735	1.189.511	2.101.144	205.974	376.774.617	73,5

Fonte: Elaboração própria a partir da RAIS (2002-2014).

Apêndice C – Representatividade Amostral dos Arranjos Populacionais

**Tabela C.1 – Empregos da RAIS e representatividade dos arranjos na amostra:
2002-2014**

Arranjo	RAIS 2002-2014 (%)	Amostra	Amostra (%)	Código
Grandes Arranjos	57,02	3.567.930	58,99	1-27
Sorocaba/SP	0,45	41.908	0,69	339
Ribeirão Preto/SP	0,52	35.120	0,58	340
Campo Grande/MS	0,57	19.454	0,32	341
Cuiabá/MT	0,60	16.264	0,27	342
Florianópolis/SC	0,90	44.310	0,73	343
Aracaju/SE	0,59	21.746	0,36	344
São José dos Campos/SP	0,60	71.342	1,18	345
João Pessoa/PB	0,73	24.947	0,41	346
Teresina/PI	0,60	23.638	0,39	347
Maceió/AL	0,59	19.917	0,33	348
Natal/RN	0,80	28.870	0,48	349
São Luís/MA	0,70	21.295	0,35	350
Vitória/ES	1,16	58.361	0,96	351
Baixada Santista/SP	0,83	61.848	1,02	352
Manaus/AM	1,08	36.178	0,60	353
Campinas/SP	1,49	176.092	2,91	354
Belém/PA	1,07	36.160	0,60	355
Goiânia/GO	1,55	49.897	0,82	356
Curitiba/PR	2,64	157.363	2,60	357
Porto Alegre/RS	2,41	172.651	2,85	358
Fortaleza/CE	1,91	97.495	1,61	359
Brasília/DF	2,75	109.701	1,81	360
Salvador/BA	2,41	120.374	1,99	361
Recife/PE	2,27	123.607	2,04	362
Belo Horizonte/MG	4,18	220.306	3,64	363
Rio de Janeiro/RJ	7,44	486.953	8,05	364
São Paulo/SP	16,19	1.292.133	21,36	365
Médios Arranjos	21,14	1.423.823	23,54	178-338
Pequenos Arranjos	3,74	248.820	4,11	1-177
Total dos Arranjos	81,91	5.240.573	86,64	1-365
Outros Municípios	18,09	807.820	13,36	0
Brasil	100	6.048.393	100	0-365

Fonte: Elaboração própria a partir da RAIS (2002–2014).

APÊNDICE D – Critérios de Definição dos Arranjos Populacionais

Os arranjos populacionais do IBGE (2015) foram identificados a partir de três critérios de integração. O primeiro critério estabelece uma forte intensidade relativa dos movimentos pendulares entre os municípios, mensurada a partir de um índice de integração populacional igual ou superior a 0,25, conforme a expressão abaixo:

$$II_A = \frac{X_{AB}+X_{BA}}{X_A}; II_B = \frac{X_{AB}+X_{BA}}{X_B},$$

onde, II é o índice de integração do município especificado (A ou B), X_{AB} (X_{BA}) é o número de pessoas que moram no município A (B) e trabalham ou estudam em B (A) e X_A (X_B) é o número de pessoas que moram em A (B) e trabalham e/ou estudam em A (B). O segundo critério requer forte intensidade absoluta dos movimentos pendulares para trabalho e estudo, cujo número de pessoas que se deslocam entre municípios para trabalho e estudo deve ser igual ou superior a 10 mil. O último critério estabelece que a distância entre as bordas das manchas urbanizadas de dois municípios deve ser de até 3 quilômetros, a distância máxima que pode ser percorrida a pé, a qual permite a manutenção de relações de convivência cotidianas. Se existe intenso fluxo pendular entre os arranjos fica caracterizada a primeira integração de arranjos e se existe forte movimento pendular entre essas primeiras unidades de integração são definidas as segundas integrações de arranjos.

APÊNDICE E – Distribuição Espacial e Densidade Populacional dos Arranjos Populacionais

Figura E.1 – Densidade populacional (2010) dos Arranjos: População por Km²

Fonte: Elaboração própria a partir do Censo Demográfico do IBGE (2010).

Nota: Os arranjos populacionais do IBGE (2015) foram ajustados às respectivas AMCs deste trabalho.

APÊNDICE F – Variáveis do Modelo Econométrico

Tabela F.1 – Descrição das Variáveis do Modelo

continua

Variável	Descrição
<i>ln(wreal)</i>	Logaritmo natural do salário-hora real (IPCA ano base: dezembro de 2014) do trabalhador em dezembro de cada ano (2002-2014).
Características do trabalhador	
<i>educ</i>	<i>Dummies</i> para os ciclos de escolaridade: educ1 = 1 se Analfabeto; Até o 5º ano do ensino fundamental incompleto; 5º ano completo do ensino fundamental; 6º ao 9º ano do ensino fundamental incompleto; Fundamental completo; Médio incompleto; educ2 = 1 se médio completo; educ3 = 1 se superior incompleto; educ4 = 1 se superior completo.
<i>exp</i>	Tempo de emprego, em meses, no mesmo vínculo.
<i>exp2</i>	Quadrado do tempo de emprego, em meses, no mesmo vínculo.
<i>age</i>	Idade do trabalhador: 18 a 65 anos.
<i>age2</i>	Quadrado da idade do trabalhador (18 a 65 anos).
<i>gender</i>	1 para mulher; 0 caso contrário.
Características do trabalho	
<i>sector1</i>	1 para empresas da seção de atividade A (CNAE 1.0): Agricultura, pecuária, silvicultura e exploração florestal; 0 caso contrário.
<i>sector2</i>	1 para empresas da seção de atividade B (CNAE 1.0): Pesca; 0 caso contrário.
<i>sector3</i>	1 para empresas da seção de atividade C (CNAE 1.0): Indústrias extrativas; 0 caso contrário.
<i>sector4</i>	1 para empresas da seção de atividade D (CNAE 1.0): Indústrias de transformação; 0 caso contrário.
<i>sector5</i>	1 para empresas da seção de atividade E (CNAE 1.0): Produção e distribuição de eletricidade, gás e água; 0 caso contrário.
<i>sector6</i>	1 para empresas da seção de atividade F (CNAE 1.0): Construção; 0 caso contrário.
<i>sector7</i>	1 para empresas da seção de atividade G (CNAE 1.0): Comércio, reparação de veículos automotores, objetos pessoais e domésticos; 0 caso contrário.
<i>sector8</i>	1 para empresas da seção de atividade H (CNAE 1.0): Alojamento e alimentação; 0 caso contrário.
<i>sector9</i>	1 para empresas da seção de atividade I (CNAE 1.0): Transporte, armazenagem e comunicações; 0 caso contrário.
<i>sector10</i>	1 para empresas da seção de atividade J (CNAE 1.0): Intermediação financeira, seguros, previdência complementar e serviços relacionados; 0 caso contrário.
<i>sector11</i>	1 para empresas da seção de atividade K (CNAE 1.0): Atividades imobiliárias, aluguéis e serviços prestados as empresas; 0 caso contrário.
<i>sector12</i>	1 para empresas da seção de atividade M (CNAE 1.0): Educação; 0 caso contrário.
<i>sector13</i>	1 para empresas da seção de atividade N (CNAE 1.0): Saúde e serviços sociais; 0 caso contrário.
<i>sector14</i>	1 para empresas da seção de atividade O (CNAE 1.0): Outros serviços coletivos, sociais e pessoais; 0 caso contrário.
<i>sector15</i>	1 para empresas da seção de atividade P (CNAE 1.0): Serviços domésticos; 0 caso contrário.
<i>sector16</i>	1 para empresas da seção de atividade Q (CNAE 1.0): Organismos internacionais e outras instituições extraterritoriais; 0 caso contrário.
<i>size_firmsect</i>	Tamanho relativo do estabelecimento: empregos no estabelecimento / média de empregos no setor (CNAE 1.0, 2 dígitos).
Variáveis de localização	
<i>LMA</i>	<i>Dummies</i> para cada um dos 365 arranjos populacionais (<i>Labor Market Areas</i>).
<i>small_lma</i>	= 1 se o arranjo é pequeno (população ≤ 100 mil habitantes).
<i>med_lma</i>	= 1 se o arranjo é médio (100 mil < população ≤ 750 mil habitantes).
<i>large_lma</i>	= 1 se o arranjo é grande (população > 750 mil habitantes).

		conclusão
Variável	Descrição	
<i>region</i>	<i>Dummies</i> para as macro-regiões do Brasil (Norte, Nordeste, Centro-Oeste, Sudeste e Sul).	
<i>touristic</i>	= 1 se o arranjo litorâneo possui uma função turística-veraneio.	
<i>frontier</i>	= 1 se o arranjo possui integração, entre municípios fronteiriços, com outro país.	
$\ln(\text{employment density})$	Logaritmo natural da densidade do emprego nos arranjos (emprego RAIS/área-km ²)	
$\ln(\text{land area})$	Logaritmo natural da área definida pela extensão territorial de cada arranjo (km ²)	

Fonte: Elaboração própria.

Notas: Área territorial oficial dos municípios extraída do IBGE (2013); Emprego total extraído da RAIS (MTE, 2002-2014).

APÊNDICE G – Luminosidade Noturna no Brasil

Figura G.1 – Luminosidade Noturna: 2013

Fonte: Elaboração própria a partir das imagens de satélite do *Defense Meteorological Satellite Program-DMSP, Operational Linescan System-OLS* (NCEI, 2013).

APÊNDICE H – Algoritmo de Iteração de Ponto Fixo para a Estimação de Modelos com Múltiplos Efeitos Fixos

Correia (2014) desenvolveu um algoritmo para implementar a abordagem iterativa do estudo de Guimarães e Portugal (2010). O algoritmo permite a estimação de modelos com múltiplos efeitos fixos, a exemplo do modelo AKM pareado de trabalhador e firma. Assuma o seguinte modelo de múltiplos efeitos fixos (D são variáveis indicadoras):

$$y = X\beta + D_1\alpha_1 + D_2\alpha_2 + D_3\alpha_3 + \epsilon.$$

O algoritmo no *software stata* (*Linear and instrumental-variable regression absorbing any number of high-dimensional fixed effects*) mantém y como variável dependente e inclui $D_i\alpha$ como covariada. Quando o procedimento de estimação converge, o vetor $D_i\alpha$ deve conter as estimativas dos coeficientes das variáveis indicadoras. Definindo $Z_2 := D_2\hat{\alpha}_2$ e $Z_3 := D_3\hat{\alpha}_3$ e supondo valores para a iteração, a iteração de ponto fixo será:

- (i) Calcula P_1y e $\tilde{y} := M_1y$;
- (ii) Inicia com $Z_2^{(0)} = Z_3^{(0)} = 0$;
- (iii) Até a convergência de Z_2 e Z_3 , são definidos:

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

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

- (iv) Calcula $Z_1 = P_1(y - Z_2 - Z_3)$ e, com ele, computa $y^* = y - Z_1 - Z_2 - Z_3$;
- (v) Após repetir os estágios de (i)-(iv) para cada variável, efetua a regressão das variáveis transformadas para obter $\hat{\beta}$: $y^* = X^*\hat{\beta} + e$;
- (vi) Para obter os efeitos fixos utiliza $y = X\hat{\beta} + Z_1 + Z_2 + Z_3 + e$, com $\hat{\beta}$ e e do estágio (v), computa $\tilde{e} := y - X\hat{\beta} = Z_1 + Z_2 + Z_3 + e$ e aplica o estágio anterior para obter Z_1, Z_2 e Z_3 .

APÊNDICE I – Características dos Arranjos Populacionais do Brasil

Tabela I.1 – Características dos arranjos populacionais: 2010

Arranjo	Emprego	Emp. (%)	Salário Real	Densidade	População	Pop. (%)	Código
Grandes Arranjos	25.159.323	57,09	1751,91	579,46	79.094.477	41,46	339-365
Sorocaba/SP	203.934	0,46	1735,22	652,06	762.865	0,40	339
Ribeirão Preto/SP	232.669	0,53	1638,29	230,49	777.565	0,41	340
Campo Grande/MS	253.488	0,58	2505,77	97,22	786.797	0,41	341
Cuiabá/MT	254.982	0,58	2094,20	176,88	803.694	0,42	342
Florianópolis/SC	389.517	0,88	1583,41	276,13	883.808	0,46	343
Aracaju/SE	255.867	0,58	1794,16	568,36	926.399	0,49	344
São José dos Campos/SP	272.952	0,62	2160,28	348,68	949.119	0,50	345
João Pessoa/PB	323.737	0,73	1390,01	722,87	1.050.872	0,55	346
Teresina/PI	265.569	0,60	1129,83	98,38	1.100.253	0,58	347
Maceió/AL	261.502	0,59	1432,58	701,23	1.115.485	0,58	348
Natal/RN	362.168	0,82	1455,93	998,07	1.187.899	0,62	349
São Luís/MA	342.344	0,78	1325,67	927,11	1.309.330	0,69	350
Vitória/ES	512.968	1,16	1845,37	908,77	1.582.418	0,83	351
Baixada Santista/SP	366.557	0,83	1962,56	765,46	1.604.363	0,84	352
Manaus/AM	492.429	1,12	2185,94	158,06	1.802.014	0,94	353
Campinas/SP	664.911	1,51	2335,40	829,56	2.010.600	1,05	354
Belém/PA	468.454	1,06	1517,61	1122,68	2.042.417	1,07	355
Goiânia/GO	697.343	1,58	1427,82	358,01	2.067.382	1,08	356
Curitiba/PR	1.163.776	2,64	1729,93	358,30	3.060.332	1,60	357
Porto Alegre/RS	1.038.192	2,36	1742,29	254,79	3.082.257	1,62	358
Fortaleza/CE	872.760	1,98	1257,05	1078,11	3.327.021	1,74	359
Brasília/DF	1.165.397	2,64	1602,53	158,16	3.391.005	1,78	360
Salvador/BA	1.072.216	2,43	2536,43	1107,12	3.440.462	1,80	361
Recife/PE	1.021.506	2,32	1451,19	1227,67	3.741.904	1,96	362
Belo Horizonte/MG	1.873.464	4,25	1723,46	845,27	4.728.059	2,48	363
Rio de Janeiro/RJ	3.176.075	7,21	1704,36	1979,84	11.946.398	6,26	364
São Paulo/SP	7.154.546	16,24	2034,16	2914,81	19.613.759	10,28	365
Médios Arranjos	9.359.053	21,24	1528,52	37,39	38.903.918	20,39	178-338
Pequenos Arranjos	1.637.124	3,71	1339,83	16,19	8.661.149	4,54	1-177
Total dos Arranjos	36.155.500	82,04	1540,09	73,99	126.659.544	66,40	1-365
Outros Municípios	7.912.855	17,96	1197,17	9,44	64.096.255	33,60	0
Brasil	44.068.355	100	1368,63	22,43	190.755.799	100	0-365

Fonte: Elaboração própria, a partir da RAIS (2010) e do Censo (IBGE, 2010).

Notas: emp. (%) e pop. (%) são as parcelas do emprego e da população em relação ao Brasil; densidade é a densidade populacional; emprego é a quantidade de empregos na RAIS; salário real é o salário real médio do mês de dezembro de 2010, deflacionado pelo IPCA (base: dezembro/2014).

APÊNDICE J – Tabelas e Estatísticas Adicionais dos Resultados

Tabela J.1 – Estimativa econométrica do prêmio salarial urbano médio dos arranjos populacionais do Brasil: 2002-2014 **continua**

	Amostra Principal				Amostra Aleatória em Firmas Móveis				Amostra sem <i>Outliers</i> de Salário			
<i>ln (w real hora)</i>	(1) POLS	(2) COND i	(3) COND j	(4) COND ij	(5) POLS	(6) COND i	(7) COND j	(8) COND ij	(9) POLS	(10) COND i	(11) COND j	(12) COND ij
Ensino médio comp.	0,4221*** (0,0018)	0,0384*** (0,0022)	0,1763*** (0,0044)	0,0038* (0,0020)	0,4129*** (0,0016)	0,0404*** (0,0019)	0,1824*** (0,0027)	0,0102*** (0,0020)	0,4076*** (0,0017)	0,0341*** (0,0021)	0,1745*** (0,0043)	0,0042** (0,0019)
Superior incompleto	0,9327*** (0,0036)	0,1316*** (0,0037)	0,4678*** (0,0071)	0,0503*** (0,0033)	0,9016*** (0,0035)	0,1299*** (0,0034)	0,4658*** (0,0051)	0,0595*** (0,0032)	0,9017*** (0,0034)	0,1277*** (0,0036)	0,4610*** (0,0069)	0,0534*** (0,0033)
Superior completo	1,4324*** (0,0025)	0,2798*** (0,0040)	0,8871*** (0,0077)	0,1710*** (0,0038)	1,3893*** (0,0024)	0,2637*** (0,0038)	0,8713*** (0,0068)	0,1712*** (0,0037)	1,3664*** (0,0023)	0,2780*** (0,0039)	0,8574*** (0,0074)	0,1762*** (0,0037)
Age	0,0639*** (0,0005)				0,0603*** (0,0005)				0,0640*** (0,0005)			
Age2	-0,0007*** (0,0000)				-0,0006*** (0,0000)				-0,0007*** (0,0000)			
Exp.	0,0004*** (0,0000)	0,0013*** (0,0000)	0,0019*** (0,0000)	0,0013*** (0,0000)	0,0005*** (0,0000)	0,0013*** (0,0000)	0,0019*** (0,0000)	0,0013*** (0,0000)	0,0004*** (0,0000)	0,0013*** (0,0000)	0,0019*** (0,0000)	0,0013*** (0,0000)
Exp ²	0,0000*** (0,0000)	-0,0000*** (0,0000)	-0,0000*** (0,0000)	-0,0000*** (0,0000)	0,0000*** (0,0000)	-0,0000*** (0,0000)	-0,0000*** (0,0000)	-0,0000*** (0,0000)	0,0000*** (0,0000)	-0,0000*** (0,0000)	-0,0000*** (0,0000)	-0,0000*** (0,0000)
Gender=feminino	-0,3379*** (0,0019)				-0,3313*** (0,0017)				-0,3188*** (0,0018)			
Agricultura e pecuária	-0,4511*** (0,0044)				-0,4071*** (0,0041)				-0,4193*** (0,0042)			
Pesca	-0,6306*** (0,0358)				-0,4971*** (0,0363)				-0,5871*** (0,0360)			
Indústria Extrativa	0,4620*** (0,0091)				0,5386*** (0,0089)				0,3549*** (0,0081)			
Eletricidade; gás; água	0,2506*** (0,0046)				0,2539*** (0,0043)				0,2577*** (0,0044)			
Construção	-0,1571*** (0,0046)				-0,1680*** (0,0047)				-0,1575*** (0,0044)			
Comércio	-0,3498*** (0,0024)				-0,3396*** (0,0022)				-0,3461*** (0,0023)			
Alojamento e aliment.	-0,5015*** (0,0051)				-0,4965*** (0,0044)				-0,4913*** (0,0050)			

conclusão

	Amostra Principal				Amostra Aleatória em Firmas Móveis				Amostra sem <i>Outliers</i> de Salário			
<i>ln (w real hora)</i>	(1) POLS	(2) COND i	(3) COND j	(4) COND ij	(5) POLS	(6) COND i	(7) COND j	(8) COND ij	(9) POLS	(10) COND i	(11) COND j	(12) COND ij
Transporte e comunic.	-0,1820*** (0,0026)				-0,1544*** (0,0024)				-0,1742*** (0,0025)			
Intermediação financ.	0,1029*** (0,0034)				0,1505*** (0,0035)				0,1185*** (0,0032)			
Atividade Imobiliária	-0,3378*** (0,0023)				-0,2777*** (0,0024)				-0,3283*** (0,0023)			
Educação	-0,3882*** (0,0051)				-0,3264*** (0,0043)				-0,3628*** (0,0049)			
Saúde e serv. sociais	-0,3179*** (0,0036)				-0,2786*** (0,0030)				-0,3136*** (0,0034)			
Outros serv. coletivos	-0,3490*** (0,0043)				-0,3056*** (0,0037)				-0,3406*** (0,0041)			
Serviços domésticos	-0,2945*** (0,0582)				-0,5508*** (0,0528)				-0,2275*** (0,0524)			
Organismos internac.	0,2358*** (0,0245)				0,3269*** (0,0380)				0,2119*** (0,0242)			
size_firmsect	0,0311*** (0,0004)	0,0069*** (0,0009)	0,0023 (0,0019)	0,0040*** (0,0012)	0,0330*** (0,0005)	0,0055*** (0,0009)	0,0028** (0,0013)	0,0033*** (0,0012)	0,0305*** (0,0004)	0,0069*** (0,0009)	0,0023 (0,0018)	0,0040*** (0,0012)
dlma=1	0,2376*** (0,0021)	0,0223*** (0,0045)	0,0129* (0,0077)	-0,0035 (0,0064)	0,2582*** (0,0020)	0,0306*** (0,0038)	0,0171** (0,0084)	0,0027 (0,0066)	0,2240*** (0,0020)	0,0189*** (0,0045)	0,0126* (0,0075)	-0,0034 (0,0065)
Efeitos de tempo	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Constante	0,4972*** (0,0099)				0,5148*** (0,0092)				0,5711*** (0,0096)			
Observações	6048393	6048393	6048393	6048393	7158099	7158099	7158099	7158099	5927470	5927470	5927470	5927470
R ²	0,573	0,922	0,794	0,943	0,556	0,925	0,800	0,943	0,561	0,918	0,787	0,940
R ² Ajustado	0,573	0,916	0,784	0,935	0,556	0,919	0,790	0,935	0,561	0,911	0,777	0,931

Fonte: Elaboração própria, a partir da RAIS (2002-2014).

Notas: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; Erro padrão entre parênteses; Os modelos foram estimados com *clusters* de indivíduos (POLS e FEi), de firmas (FEj) e de indivíduos e firmas (FEij, COND i, COND j e COND ij).

Tabela J.2 – Resultado das estimações econométricas da equação de salários individuais (primeiro estágio)**continua**

<i>ln (w real hora)</i>	(1) POLS	(2) FE i	(3) FE j	(4) FE (i, j)	(5) COND(i,j)	(6) POLS	(7) FE i	(8) COND i	(9) FE j	(10) COND j	(11) FE (i, j)	(12) COND(i,j)
Ensino médio	0,4167*** (0,0018)	0,0354*** (0,0014)	0,1763*** (0,0044)	0,0162*** (0,0015)	0,0038* (0,0020)	0,4119*** (0,0017)	0,0353*** (0,0014)	0,0353*** (0,0021)	0,1763*** (0,0043)	0,1763*** (0,0044)	0,0159*** (0,0015)	0,0037* (0,0019)
Superior incompleto	0,9220*** (0,0036)	0,1218*** (0,0026)	0,4678*** (0,0070)	0,0571*** (0,0026)	0,0503*** (0,0033)	0,9006*** (0,0035)	0,1215*** (0,0026)	0,1215*** (0,0036)	0,4677*** (0,0069)	0,4677*** (0,0070)	0,0568*** (0,0026)	0,0502*** (0,0033)
Superior completo	1,4150*** (0,0025)	0,2663*** (0,0023)	0,8871*** (0,0076)	0,1442*** (0,0023)	0,1710*** (0,0037)	1,3767*** (0,0024)	0,2649*** (0,0023)	0,2649*** (0,0039)	0,8870*** (0,0075)	0,8870*** (0,0076)	0,1438*** (0,0023)	0,1709*** (0,0037)
Age	0,0632*** (0,0005)					0,0651*** (0,0005)						
Age2	-0,0007*** (0,0000)					-0,0007*** (0,0000)						
Exp.	0,0004*** (0,0000)	0,0013*** (0,0000)	0,0019*** (0,0000)	0,0011*** (0,0000)	0,0013*** (0,0000)	0,0005*** (0,0000)	0,0013*** (0,0000)	0,0013*** (0,0000)	0,0019*** (0,0000)	0,0019*** (0,0000)	0,0011*** (0,0000)	0,0013*** (0,0001)
Exp ²	0,0000*** (0,0000)	-0,0000*** (0,0000)	-0,0000*** (0,0000)	-0,0000*** (0,0000)	-0,0000*** (0,0000)	0,0000*** (0,0000)	-0,0000*** (0,0000)	-0,0000*** (0,0000)	-0,0000*** (0,0000)	-0,0000*** (0,0000)	-0,0000*** (0,0000)	-0,0000*** (0,0001)
Gender=feminino	-0,3379*** (0,0018)					-0,3295*** (0,0018)						
Agricultura e pecuária	-0,4355*** (0,0044)	-0,0877*** (0,0043)				-0,3868*** (0,0044)	-0,0887*** (0,0042)	-0,0887*** (0,0074)				
Pesca	-0,6370*** (0,0360)	-0,1405*** (0,0222)				-0,5597*** (0,0367)	-0,1407*** (0,0223)	-0,1407*** (0,0228)				
Indústria Extrativa	0,4724*** (0,0092)	0,1496*** (0,0057)				0,4085*** (0,0085)	0,1352*** (0,0055)	0,1352*** (0,0133)				
Eletricidade; gás; água	0,2319*** (0,0047)	0,0912*** (0,0068)				0,2746*** (0,0046)	0,0821*** (0,0069)	0,0821*** (0,0137)				
Construção	-0,1803*** (0,0046)	-0,0692*** (0,0030)				-0,1582*** (0,0045)	-0,0799*** (0,0031)	-0,0799*** (0,0073)				
Comércio	-0,3610*** (0,0024)	-0,1155*** (0,0021)				-0,3110*** (0,0023)	-0,1169*** (0,0021)	-0,1169*** (0,0045)				
Alojamento e aliment.	-0,5269*** (0,0051)	-0,1738*** (0,0058)				-0,5179*** (0,0049)	-0,1769*** (0,0057)	-0,1769*** (0,0095)				
Transporte e comunic.	-0,2072*** (0,0026)	-0,0619*** (0,0029)				-0,1721*** (0,0025)	-0,0653*** (0,0029)	-0,0653*** (0,0056)				
Intermediação financ.	0,0742*** (0,0034)	0,1370*** (0,0046)				0,0959*** (0,0035)	0,1333*** (0,0046)	0,1333*** (0,0111)				

continuação

<i>ln (w real hora)</i>	(1) POLS	(2) FE i	(3) FE j	(4) FE (i, j)	(5) COND(i,j)	(6) POLS	(7) FE i	(8) COND i	(9) FE j	(10) COND j	(11) FE (i, j)	(12) COND(i,j)
Atividade Imobiliária	-0,3747*** (0,0024)	-0,1145*** (0,0020)				-0,3542*** (0,0023)	-0,1178*** (0,0020)	-0,1178*** (0,0047)				
Educação	-0,4018*** (0,0050)	-0,1425*** (0,0064)				-0,3391*** (0,0048)	-0,1432*** (0,0064)	-0,1432*** (0,0103)				
Saúde e serv. sociais	-0,3349*** (0,0035)	-0,0493*** (0,0045)				-0,2890*** (0,0033)	-0,0511*** (0,0045)	-0,0511*** (0,0068)				
Outros serv. coletivos	-0,3746*** (0,0042)	-0,0853*** (0,0038)				-0,3339*** (0,0041)	-0,0880*** (0,0038)	-0,0880*** (0,0062)				
Serviços domésticos	-0,3126*** (0,0563)	0,0434*** (0,0122)				-0,2615*** (0,0523)	0,0713*** (0,0156)	0,0713*** (0,0491)				
Organismos internac.	0,2491*** (0,0245)	-0,0534*** (0,0072)				0,1810*** (0,0246)	-0,0900*** (0,0077)	-0,0900*** (0,0174)				
size_firmsect	0,0282*** (0,0004)	0,0088*** (0,0002)	0,0023 (0,0019)	0,0047*** (0,0002)	0,0040*** (0,0012)	0,0219*** (0,0004)	0,0086*** (0,0002)	0,0086*** (0,0010)	0,0026 (0,0019)	0,0026 (0,0019)	0,0050*** (0,0002)	0,0043*** (0,0013)
small_lma	0,0740*** (0,0038)	0,0101*** (0,0030)	0,0218* (0,0129)	0,0043 (0,0031)	0,0074 (0,0084)							
med_lma	0,1700*** (0,0023)	0,0160*** (0,0018)	0,0047 (0,0094)	-0,0030* (0,0017)	-0,0015 (0,0071)							
large_lma	0,2941*** (0,0022)	0,0270*** (0,0017)	0,0173* (0,0097)	-0,0101*** (0,0015)	-0,0064 (0,0073)							
Sorocaba/SP						0,3381*** (0,0073)	0,0211*** (0,0062)	0,0211 (0,0174)	-0,0052 (0,0344)	-0,0052 (0,0344)	-0,0697*** (0,0065)	-0,0649** (0,0321)
Ribeirão Preto/SP						0,2468*** (0,0087)	0,0302*** (0,0084)	0,0302** (0,0119)	-0,0038 (0,0234)	-0,0038 (0,0234)	-0,0142 (0,0088)	-0,0126 (0,0167)
Campo Grande/MS						0,0853*** (0,0114)	0,0053 (0,0101)	0,0053 (0,0123)	0,0307* (0,0185)	0,0307* (0,0185)	-0,0078 (0,0109)	-0,0118 (0,0185)
Cuiabá/MT						0,1343*** (0,0130)	0,0197 (0,0126)	0,0197 (0,0226)	0,0139 (0,0291)	0,0139 (0,0292)	-0,0033 (0,0113)	0,0115 (0,0218)
Florianópolis/SC						0,2888*** (0,0076)	0,0197*** (0,0070)	0,0197 (0,0152)	0,1717*** (0,0292)	0,1717*** (0,0293)	0,0852*** (0,0081)	0,0882*** (0,0178)
Aracaju/SE						-0,0605*** (0,0130)	-0,0069 (0,0146)	-0,0069 (0,0180)	-0,0020 (0,0406)	-0,0020 (0,0409)	0,0166 (0,0119)	0,0253 (0,0200)
São José dos Campos/SP						0,4181*** (0,0062)	0,0255*** (0,0051)	0,0255** (0,0129)	0,0361* (0,0216)	0,0361* (0,0216)	-0,0322*** (0,0056)	-0,0276 (0,0182)

continuação

<i>ln (w real hora)</i>	(1) POLS	(2) FE i	(3) FE j	(4) FE (i, j)	(5) COND(i,j)	(6) POLS	(7) FE i	(8) COND i	(9) FE j	(10) COND j	(11) FE (i, j)	(12) COND(i,j)
João Pessoa/PB						-0,1755*** (0,0107)	-0,0157 (0,0121)	-0,0157 (0,0168)	0,0689 (0,0442)	0,0689 (0,0441)	0,0448*** (0,0141)	0,0376 (0,0374)
Teresina/PI						-0,1341*** (0,0119)	0,0359** (0,0168)	0,0359* (0,0207)	-0,0098 (0,0259)	-0,0098 (0,0260)	-0,0059 (0,0096)	0,0036 (0,0254)
Maceió/AL						-0,0579*** (0,0130)	-0,0029 (0,0134)	-0,0029 (0,0190)	0,0415 (0,0361)	0,0415 (0,0361)	0,0538*** (0,0115)	0,0504 (0,0335)
Natal/RN						-0,1173*** (0,0107)	-0,0236** (0,0115)	-0,0236 (0,0155)	0,0071 (0,0344)	0,0071 (0,0344)	-0,0055 (0,0120)	-0,0084 (0,0303)
São Luís/MA						-0,0350*** (0,0130)	0,0325** (0,0142)	0,0325 (0,0214)	0,0424 (0,0396)	0,0424 (0,0400)	-0,0213 (0,0198)	0,0033 (0,0317)
Vitória/ES						0,2162*** (0,0074)	0,0468*** (0,0079)	0,0468*** (0,0146)	0,0281 (0,0372)	0,0281 (0,0373)	0,0261*** (0,0086)	0,0378 (0,0290)
Baixada Santista/SP						0,4178*** (0,0075)	0,0574*** (0,0065)	0,0574*** (0,0137)	0,0156 (0,0183)	0,0156 (0,0184)	-0,0123** (0,0057)	-0,0072 (0,0137)
Manaus/AM						0,2714*** (0,0095)	0,0711*** (0,0106)	0,0711** (0,0329)	-0,1198* (0,0709)	-0,1198* (0,0709)	-0,0245*** (0,0092)	-0,0198 (0,0377)
Campinas/SP						0,4381*** (0,0041)	0,0317*** (0,0035)	0,0317*** (0,0104)	0,0275 (0,0233)	0,0275 (0,0233)	-0,0430*** (0,0035)	-0,0334** (0,0153)
Belém/PA						0,1078*** (0,0095)	0,0478*** (0,0118)	0,0478*** (0,0159)	0,0539 (0,0484)	0,0539 (0,0486)	0,0956*** (0,0105)	0,0912*** (0,0259)
Goiânia/GO						0,1203*** (0,0077)	0,0277*** (0,0074)	0,0277 (0,0206)	0,0331 (0,0399)	0,0331 (0,0399)	-0,0091 (0,0079)	-0,0089 (0,0285)
Curitiba/PR						0,2859*** (0,0046)	0,0535*** (0,0050)	0,0535*** (0,0103)	0,0326 (0,0216)	0,0326 (0,0218)	-0,0063 (0,0060)	0,0055 (0,0195)
Porto Alegre/RS						0,3491*** (0,0044)	0,0258*** (0,0043)	0,0258*** (0,0100)	0,0359 (0,0219)	0,0359 (0,0220)	-0,0170*** (0,0043)	-0,0150 (0,0165)
Fortaleza/CE						-0,0538*** (0,0066)	0,0020 (0,0057)	0,0020 (0,0105)	-0,1061*** (0,0288)	-0,1061*** (0,0288)	-0,0190*** (0,0062)	-0,0174 (0,0119)
Brasília/DF						0,3922*** (0,0059)	0,0916*** (0,0041)	0,0916*** (0,0286)	0,0759*** (0,0215)	0,0759*** (0,0215)	-0,0121*** (0,0031)	0,0020 (0,0232)
Salvador/BA						0,1137*** (0,0062)	0,0054 (0,0058)	0,0054 (0,0108)	0,0591** (0,0273)	0,0591** (0,0279)	-0,0196*** (0,0064)	-0,0053 (0,0152)
Recife/PE						-0,0800*** (0,0054)	0,0252*** (0,0047)	0,0252** (0,0109)	-0,0217 (0,0283)	-0,0217 (0,0282)	-0,0265*** (0,0033)	-0,0289* (0,0163)

<i>ln (w real hora)</i>												conclusão
	(1) POLS	(2) FE i	(3) FE j	(4) FE (i, j)	(5) COND(i,j)	(6) POLS	(7) FE i	(8) COND i	(9) FE j	(10) COND j	(11) FE (i, j)	(12) COND(i,j)
Belo Horizonte/MG						0,2344*** (0,0042)	0,0112*** (0,0041)	0,0112 (0,0080)	0,0071 (0,0161)	0,0071 (0,0162)	-0,0026 (0,0043)	0,0034 (0,0113)
Rio de Janeiro/RJ						0,2778*** (0,0035)	0,0829*** (0,0035)	0,0829*** (0,0090)	0,0540** (0,0224)	0,0540** (0,0224)	-0,0045 (0,0040)	-0,0007 (0,0140)
São Paulo/SP						0,4239*** (0,0025)	0,0427*** (0,0023)	0,0427*** (0,0099)	0,0331** (0,0148)	0,0331** (0,0149)	-0,0238*** (0,0022)	-0,0160 (0,0128)
Efeitos de tempo	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Constante	0,5255*** (0,0099)	2,1932*** (0,0022)	1,9939*** (0,0088)	2,2622*** (0,0017)		0,4634*** (0,0095)	2,1890*** (0,0022)		1,9852*** (0,0096)		2,2652*** (0,0018)	
Observações	6048393	6048393	6048393	6048393	6048393	6048393	6048393	6048393	6048393	6048393	6048393	6048393
R^2	0,577	0,279	0,452	0,121	0,943	0,606	0,292	0,923	0,4560	0,794	0,113	0,943
R^2 Ajustado	0,577				0,935	0,606		0,917		0,784		0,935
<i>Within R</i> ²		0,435	0,311	0,378	0,359		0,437	0,437	0,311	0,311	0,379	0,360

Fonte: Elaboração própria, a partir da RAIS (2002-2014).

Notas: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; Erro padrão entre parênteses; As colunas de 6 a 12 foram estimadas com o efeito fixo de cada um dos 365 arranjos; Os modelos foram estimados com *clusters* de indivíduos (POLS e FEi), de firmas (FEj) e de indivíduos e firmas (FEij, CONDi, CONDj e CONDij).

Tabela J.3 – Resultado da estimação econométrica da equação de salários individuais (primeiro estágio): Amostra aleatória em firmas
continua

<i>ln (w real hora)</i>	(1) POLS	(2) FE i	(3) FE j	(4) FE (i, j)	(5) COND(i,j)	(6) POLS	(7) FE i	(8) COND i	(9) FE j	(10) COND j	(11) FE (i, j)	(12) COND(i,j)
Ensino médio	0,4071*** (0,0016)	0,0372*** (0,0012)	0,1824*** (0,0027)	0,0160*** (0,0012)	0,0102*** (0,0020)	0,4063*** (0,0015)	0,0372*** (0,0012)	0,0372*** (0,0019)	0,1824*** (0,0027)	0,1824*** (0,0027)	0,0159*** (0,0012)	0,0102*** (0,0020)
Superior incompleto	0,8929*** (0,0035)	0,1221*** (0,0025)	0,4658*** (0,0049)	0,0615*** (0,0024)	0,0595*** (0,0032)	0,8667*** (0,0034)	0,1220*** (0,0025)	0,1220*** (0,0033)	0,4658*** (0,0049)	0,4658*** (0,0051)	0,0615*** (0,0024)	0,0595*** (0,0032)
Superior completo	1,3724*** (0,0024)	0,2534*** (0,0022)	0,8713*** (0,0067)	0,1497*** (0,0021)	0,1712*** (0,0037)	1,3388*** (0,0023)	0,2525*** (0,0022)	0,2525*** (0,0038)	0,8713*** (0,0067)	0,8713*** (0,0068)	0,1496*** (0,0021)	0,1712*** (0,0037)
Age	0,0595*** (0,0005)					0,0618*** (0,0005)						
Age2	-0,0006*** (0,0000)					-0,0006*** (0,0000)						
Exp.	0,0005*** (0,0000)	0,0013*** (0,0000)	0,0019*** (0,0000)	0,0011*** (0,0000)	0,0013*** (0,0000)	0,0005*** (0,0000)	0,0013*** (0,0000)	0,0013*** (0,0000)	0,0019*** (0,0000)	0,0019*** (0,0000)	0,0011*** (0,0000)	0,0013*** (0,0001)
Exp ²	0,0000*** (0,0000)	-0,0000*** (0,0000)	-0,0000*** (0,0000)	-0,0000*** (0,0000)	-0,0000*** (0,0000)	0,0000*** (0,0000)	-0,0000*** (0,0000)	-0,0000*** (0,0000)	-0,0000*** (0,0000)	-0,0000*** (0,0000)	-0,0000*** (0,0000)	-0,0000*** (0,0000)
Gender=feminino	-0,3296*** (0,0017)					-0,3252*** (0,0016)						
Agricultura e pecuária	-0,3855*** (0,0041)	-0,0833*** (0,0039)				-0,3494*** (0,0041)	-0,0838*** (0,0039)	-0,0838*** (0,0064)				
Pesca	-0,5040*** (0,0361)	-0,0898*** (0,0199)				-0,4291*** (0,0360)	-0,0890*** (0,0196)	-0,0890*** (0,0268)				
Indústria Extrativa	0,5597*** (0,0090)	0,1577*** (0,0056)				0,4430*** (0,0082)	0,1404*** (0,0054)	0,1404*** (0,0114)				
Eletricidade; gás; água	0,2250*** (0,0044)	0,1121*** (0,0082)				0,2890*** (0,0042)	0,1084*** (0,0081)	0,1084*** (0,0173)				
Construção	-0,1970*** (0,0047)	-0,0731*** (0,0035)				-0,1700*** (0,0045)	-0,0769*** (0,0035)	-0,0769*** (0,0060)				
Comércio	-0,3524*** (0,0022)	-0,1194*** (0,0021)				-0,3096*** (0,0021)	-0,1198*** (0,0021)	-0,1198*** (0,0030)				
Alojamento e aliment.	-0,5268*** (0,0045)	-0,1606*** (0,0055)				-0,5084*** (0,0043)	-0,1625*** (0,0055)	-0,1625*** (0,0106)				
Transporte e comunic.	-0,1840*** (0,0024)	-0,0574*** (0,0032)				-0,1575*** (0,0024)	-0,0592*** (0,0031)	-0,0592*** (0,0045)				
Intermediação financ.	0,1272*** (0,0035)	0,1521*** (0,0055)				0,1607*** (0,0036)	0,1502*** (0,0055)	0,1502*** (0,0102)				

continuação

<i>ln (w real hora)</i>	(1) POLS	(2) FE i	(3) FE j	(4) FE (i, j)	(5) COND(i,j)	(6) POLS	(7) FE i	(8) COND i	(9) FE j	(10) COND j	(11) FE (i, j)	(12) COND(i,j)
Atividade Imobiliária	-0,3205*** (0,0024)	-0,1086*** (0,0022)				-0,3006*** (0,0023)	-0,1100*** (0,0021)	-0,1100*** (0,0038)				
Educação	-0,3451*** (0,0042)	-0,1184*** (0,0057)				-0,2983*** (0,0040)	-0,1185*** (0,0057)	-0,1185*** (0,0081)				
Saúde e serv. sociais	-0,2958*** (0,0029)	-0,0291*** (0,0041)				-0,2508*** (0,0028)	-0,0296*** (0,0041)	-0,0296*** (0,0055)				
Outros serv. coletivos	-0,3339*** (0,0037)	-0,0644*** (0,0036)				-0,2981*** (0,0036)	-0,0652*** (0,0036)	-0,0652*** (0,0050)				
Serviços domésticos	-0,5521*** (0,0522)	-0,0344 (0,0240)				-0,5102*** (0,0447)	-0,0331 (0,0256)	-0,0331 (0,0443)				
Organismos internac.	0,2931*** (0,0383)	-0,0454*** (0,0121)				0,3077*** (0,0379)	-0,0587*** (0,0124)	-0,0587*** (0,0161)				
size_firmsect	0,0302*** (0,0005)	0,0074*** (0,0002)	0,0028** (0,0013)	0,0038*** (0,0002)	0,0033*** (0,0012)	0,0246*** (0,0004)	0,0073*** (0,0002)	0,0073*** (0,0010)	0,0029** (0,0013)	0,0029** (0,0014)	0,0038*** (0,0002)	0,0034*** (0,0012)
small_lma	0,0615*** (0,0036)	-0,0014 (0,0046)	0,0258 (0,0160)	0,0101 (0,0062)	0,0109 (0,0099)							
med_lma	0,1798** (0,0023)	0,0266*** (0,0029)	0,0107 (0,0104)	0,0041 (0,0034)	0,0050 (0,0077)							
large_lma	0,3205*** (0,0021)	0,0471*** (0,0028)	0,0205* (0,0105)	-0,0021 (0,0031)	-0,0001 (0,0074)							
Sorocaba/SP						0,3715*** (0,0078)	0,0694*** (0,0090)	0,0694*** (0,0146)	-0,0137 (0,0296)	-0,0137 (0,0297)	-0,0677*** (0,0129)	-0,0668** (0,0293)
Ribeirão Preto/SP						0,2761*** (0,0081)	0,0586*** (0,0115)	0,0586*** (0,0156)	-0,0092 (0,0339)	-0,0092 (0,0340)	-0,0114 (0,0187)	-0,0134 (0,0257)
Campo Grande/MS						0,1205*** (0,0110)	0,0104 (0,0158)	0,0104 (0,0175)	-0,0164 (0,0241)	-0,0164 (0,0242)	-0,0011 (0,0225)	-0,0141 (0,0213)
Cuiabá/MT						0,1603*** (0,0119)	0,0208 (0,0176)	0,0208 (0,0192)	0,0301 (0,0356)	0,0301 (0,0362)	0,0000 (0,0196)	0,0302 (0,0299)
Florianópolis/SC						0,3142*** (0,0073)	0,0162 (0,0099)	0,0162 (0,0136)	0,1843*** (0,0348)	0,1843*** (0,0349)	0,0887*** (0,0170)	0,0916*** (0,0229)
Aracaju/SE						-0,0314*** (0,0108)	0,0389** (0,0160)	0,0389** (0,0189)	0,0467 (0,0514)	0,0467 (0,0515)	0,0183 (0,0194)	0,0220 (0,0287)
São José dos Campos/SP						0,5245*** (0,0068)	0,0747*** (0,0086)	0,0747*** (0,0127)	0,0489** (0,0236)	0,0489** (0,0237)	-0,0119 (0,0121)	-0,0128 (0,0212)

continuação

<i>ln (w real hora)</i>	(1) POLS	(2) FE i	(3) FE j	(4) FE (i, j)	(5) COND(i,j)	(6) POLS	(7) FE i	(8) COND i	(9) FE j	(10) COND j	(11) FE (i, j)	(12) COND(i,j)
João Pessoa/PB						-0,1707*** (0,0088)	-0,0017 (0,0178)	-0,0017 (0,0198)	0,0923* (0,0497)	0,0923* (0,0478)	0,0462* (0,0275)	0,0226 (0,0418)
Teresina/PI						-0,0967*** (0,0101)	0,0125 (0,0225)	0,0125 (0,0237)	-0,0638* (0,0360)	-0,0638* (0,0358)	-0,0356* (0,0195)	-0,0347 (0,0381)
Maceió/AL						-0,0564*** (0,0099)	0,0117 (0,0167)	0,0117 (0,0181)	0,0691 (0,0531)	0,0691 (0,0532)	0,0788*** (0,0242)	0,0779* (0,0414)
Natal/RN						-0,1076*** (0,0091)	0,0193 (0,0149)	0,0193 (0,0199)	0,0180 (0,0513)	0,0180 (0,0513)	0,0290 (0,0302)	0,0185 (0,0437)
São Luís/MA						-0,0036 (0,0111)	0,0708*** (0,0183)	0,0708*** (0,0216)	-0,0265 (0,0833)	-0,0265 (0,0830)	-0,0398 (0,0382)	-0,0205 (0,0356)
Vitória/ES						0,2421*** (0,0068)	0,0719*** (0,0112)	0,0719*** (0,0137)	0,0280 (0,0470)	0,0280 (0,0474)	0,0339* (0,0177)	0,0451 (0,0310)
Baixada Santista/SP						0,4380*** (0,0068)	0,0691*** (0,0106)	0,0691*** (0,0133)	0,0425 (0,0265)	0,0425 (0,0273)	-0,0098 (0,0117)	-0,0041 (0,0170)
Manaus/AM						0,3195*** (0,0085)	0,1059*** (0,0157)	0,1059*** (0,0255)	-0,0757 (0,0703)	-0,0757 (0,0701)	0,0147 (0,0182)	-0,0035 (0,0397)
Campinas/SP						0,4852*** (0,0046)	0,0678*** (0,0056)	0,0678*** (0,0092)	0,0113 (0,0280)	0,0113 (0,0280)	-0,0400*** (0,0071)	-0,0362** (0,0168)
Belém/PA						0,0981*** (0,0078)	0,0394*** (0,0149)	0,0394*** (0,0165)	0,0794 (0,0498)	0,0794 (0,0503)	0,1056*** (0,0221)	0,0935*** (0,0337)
Goiânia/GO						0,1438*** (0,0068)	0,0185* (0,0110)	0,0185 (0,0150)	0,0565* (0,0303)	0,0565* (0,0305)	0,0043 (0,0164)	0,0015 (0,0352)
Curitiba/PR						0,3044*** (0,0041)	0,0642*** (0,0074)	0,0642*** (0,0096)	0,0165 (0,0281)	0,0165 (0,0289)	-0,0158 (0,0115)	-0,0052 (0,0229)
Porto Alegre/RS						0,4167*** (0,0043)	0,0443*** (0,0071)	0,0443*** (0,0101)	0,0181 (0,0218)	0,0181 (0,0221)	-0,0157* (0,0088)	-0,0134 (0,0167)
Fortaleza/CE						-0,1164*** (0,0058)	-0,0011 (0,0097)	-0,0011 (0,0152)	-0,1157*** (0,0364)	-0,1157*** (0,0372)	-0,0253* (0,0143)	-0,0182 (0,0178)
Brasília/DF						0,4102*** (0,0053)	0,1445*** (0,0066)	0,1445*** (0,0190)	0,0797*** (0,0232)	0,0797*** (0,0232)	-0,0005 (0,0063)	0,0089 (0,0240)
Salvador/BA						0,1516*** (0,0056)	0,0287*** (0,0081)	0,0287*** (0,0105)	0,0599* (0,0342)	0,0599* (0,0355)	-0,0191 (0,0135)	0,0010 (0,0208)
Recife/PE						-0,0236*** (0,0051)	0,0418*** (0,0080)	0,0418*** (0,0112)	-0,0076 (0,0293)	-0,0076 (0,0292)	-0,0222*** (0,0069)	-0,0232 (0,0165)

												conclusão
<i>ln (w real hora)</i>	(1) POLS	(2) FE i	(3) FE j	(4) FE (i, j)	(5) COND(i,j)	(6) POLS	(7) FE i	(8) COND i	(9) FE j	(10) COND j	(11) FE (i, j)	(12) COND(i,j)
Belo Horizonte/MG						0,2467*** (0,0038)	0,0275*** (0,0058)	0,0275*** (0,0081)	-0,0005 (0,0192)	-0,0005 (0,0195)	0,0031 (0,0093)	0,0118 (0,0155)
Rio de Janeiro/RJ						0,3026*** (0,0033)	0,1185*** (0,0054)	0,1185*** (0,0077)	0,0515* (0,0268)	0,0515* (0,0271)	0,0138* (0,0083)	0,0259* (0,0153)
São Paulo/SP						0,4647*** (0,0024)	0,0764*** (0,0038)	0,0764*** (0,0071)	0,0416*** (0,0161)	0,0416*** (0,0163)	-0,0128*** (0,0047)	-0,0096 (0,0126)
Efeitos de tempo	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Constante	0,5441*** (0,0092)	2,1269*** (0,0028)	1,9255*** (0,0084)	2,1852*** (0,0027)		0,4799*** (0,0087)	2,1161*** (0,0028)		1,9191*** (0,0091)		2,1881*** (0,0030)	
Observações	7158099	7158099	7158099	7158099	7158099	7158099	7158099	7158099	7158099	7158099	7158099	7158099
R ²	0,562	0,254	0,435	0,112	0,943	0,597	0,272	0,926	0,437	0,800	0,104	0,943
R ² Ajustado	0,562				0,935	0,596		0,920		0,790		0,935
Within R ²		0,436	0,327	0,399	0,386		0,437	0,437	0,327	0,327	0,399	0,386

Fonte: Elaboração própria, a partir da RAIS (2002-2014).

Notas: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; Erro padrão entre parênteses; As colunas de 6 a 12 foram estimadas com o efeito fixo de cada um dos 365 arranjos; Os modelos foram estimados com *clusters* de indivíduos (POLS e FEi), de firmas (FEj) e de indivíduos e firmas (FEij, CONDi, CONDj e CONDij).

Tabela J.4 – Resultado da estimação econométrica da equação de salários individuais (primeiro estágio): Amostra sem outliers (lnw)**continua**

<i>ln (w real hora)</i>	(1) POLS	(2) FE i	(3) FE j	(4) FE (i, j)	(5) COND(i,j)	(6) POLS	(7) FE i	(8) COND i	(9) FE j	(10) COND j	(11) FE (i, j)	(12) COND(i,j)
Ensino médio	0,4025*** (0,0017)	0,0313*** (0,0014)	0,1745*** (0,0043)	0,0166*** (0,0014)	0,0042** (0,0019)	0,3993*** (0,0016)	0,0313*** (0,0014)	0,0313*** (0,0021)	0,1745*** (0,0042)	0,1745*** (0,0043)	0,0163*** (0,0014)	0,0041*** (0,0019)
Superior incompleto	0,8917*** (0,0034)	0,1182*** (0,0025)	0,4610*** (0,0068)	0,0602*** (0,0025)	0,0534*** (0,0033)	0,8741*** (0,0034)	0,1179*** (0,0025)	0,1179*** (0,0035)	0,4609*** (0,0067)	0,4609*** (0,0069)	0,0599*** (0,0025)	0,0533*** (0,0032)
Superior completo	1,3504*** (0,0023)	0,2648*** (0,0023)	0,8573*** (0,0074)	0,1494*** (0,0023)	0,1762*** (0,0037)	1,3173*** (0,0023)	0,2635*** (0,0023)	0,2635*** (0,0038)	0,8572*** (0,0073)	0,8572*** (0,0074)	0,1490*** (0,0023)	0,1761*** (0,0036)
Age	0,0633*** (0,0005)					0,0655*** (0,0005)						
Age2	-0,0007*** (0,0000)					-0,0007*** (0,0000)						
Exp.	0,0004*** (0,0000)	0,0013*** (0,0000)	0,0019*** (0,0000)	0,0011*** (0,0000)	0,0013*** (0,0000)	0,0005*** (0,0000)	0,0013*** (0,0000)	0,0013*** (0,0000)	0,0019*** (0,0000)	0,0019*** (0,0000)	0,0011*** (0,0000)	0,0013*** (0,0003)
Exp ²	0,0000*** (0,0000)	-0,0000*** (0,0000)	-0,0000*** (0,0000)	-0,0000*** (0,0000)	-0,0000*** (0,0000)	0,0000*** (0,0000)	-0,0000*** (0,0000)	-0,0000*** (0,0000)	-0,0000*** (0,0000)	-0,0000*** (0,0000)	-0,0000*** (0,0000)	-0,0000*** (0,0002)
Gender=feminino	-0,3189*** (0,0018)					-0,3121*** (0,0017)						
Agricultura e pecuária	-0,4041*** (0,0042)	-0,0761*** (0,0041)				-0,3568*** (0,0042)	-0,0772*** (0,0041)	-0,0772*** (0,0073)				
Pesca	-0,5924*** (0,0361)	-0,1341*** (0,0227)				-0,5192*** (0,0368)	-0,1342*** (0,0227)	-0,1342*** (0,0228)				
Indústria Extrativa	0,3635*** (0,0082)	0,1436*** (0,0058)				0,3259*** (0,0078)	0,1291*** (0,0056)	0,1291*** (0,0133)				
Eletricidade; gás; água	0,2396*** (0,0045)	0,0940*** (0,0067)				0,2805*** (0,0044)	0,0853*** (0,0068)	0,0853*** (0,0132)				
Construção	-0,1796*** (0,0044)	-0,0703*** (0,0029)				-0,1569*** (0,0043)	-0,0806*** (0,0031)	-0,0806*** (0,0072)				
Comércio	-0,3570*** (0,0023)	-0,1133*** (0,0021)				-0,3080*** (0,0023)	-0,1146*** (0,0021)	-0,1146*** (0,0045)				
Alojamento e aliment.	-0,5156*** (0,0050)	-0,1667*** (0,0057)				-0,5067*** (0,0048)	-0,1693*** (0,0057)	-0,1693*** (0,0095)				
Transporte e comunic.	-0,1983*** (0,0025)	-0,0614*** (0,0029)				-0,1625*** (0,0025)	-0,0645*** (0,0029)	-0,0645*** (0,0056)				
Intermediação financ.	0,0910*** (0,0032)	0,1375*** (0,0045)				0,1131*** (0,0033)	0,1341*** (0,0045)	0,1341*** (0,0108)				

continuação

<i>ln (w real hora)</i>	(1) POLS	(2) FE i	(3) FE j	(4) FE (i, j)	(5) COND(i,j)	(6) POLS	(7) FE i	(8) COND i	(9) FE j	(10) COND j	(11) FE (i, j)	(12) COND(i,j)
Atividade Imobiliária	-0,3635*** (0,0023)	-0,1139*** (0,0020)				-0,3431*** (0,0022)	-0,1168*** (0,0020)	-0,1168*** (0,0047)				
Educação	-0,3760*** (0,0049)	-0,1370*** (0,0062)				-0,3157*** (0,0047)	-0,1376*** (0,0062)	-0,1376*** (0,0101)				
Saúde e serv. sociais	-0,3300*** (0,0034)	-0,0468*** (0,0045)				-0,2848*** (0,0032)	-0,0484*** (0,0045)	-0,0484*** (0,0067)				
Outros serv. coletivos	-0,3653*** (0,0040)	-0,0838*** (0,0037)				-0,3249*** (0,0039)	-0,0862*** (0,0037)	-0,0862*** (0,0062)				
Serviços domésticos	-0,2469*** (0,0512)	0,0530*** (0,0120)				-0,2050*** (0,0502)	0,0806*** (0,0155)	0,0806* (0,0474)				
Organismos internac.	0,2261*** (0,0240)	-0,0406*** (0,0073)				0,1539*** (0,0242)	-0,0790*** (0,0079)	-0,0790*** (0,0192)				
size_firmsect	0,0277*** (0,0004)	0,0088*** (0,0002)	0,0023 (0,0018)	0,0047*** (0,0002)	0,0040*** (0,0012)	0,0217*** (0,0004)	0,0086*** (0,0002)	0,0086*** (0,0010)	0,0026 (0,0019)	0,0026 (0,0019)	0,0050*** (0,0002)	0,0042*** (0,0012)
small_lma	0,0678*** (0,0036)	0,0079*** (0,0030)	0,0207 (0,0132)	0,0040 (0,0031)	0,0073 (0,0085)							
med_lma	0,1598*** (0,0023)	0,0141*** (0,0018)	0,0062 (0,0092)	-0,0026 (0,0016)	-0,0009 (0,0072)							
large_lma	0,2777*** (0,0021)	0,0227*** (0,0017)	0,0160* (0,0093)	-0,0104*** (0,0015)	-0,0067 (0,0073)							
Sorocaba/SP						0,3264*** (0,0072)	0,0186*** (0,0061)	0,0186 (0,0172)	-0,0126 (0,0331)	-0,0126 (0,0331)	-0,0709*** (0,0064)	-0,0660** (0,0311)
Ribeirão Preto/SP						0,2336*** (0,0085)	0,0244*** (0,0082)	0,0244** (0,0117)	-0,0054 (0,0231)	-0,0054 (0,0231)	-0,0158* (0,0088)	-0,0143 (0,0164)
Campo Grande/MS						0,0791*** (0,0111)	0,0035 (0,0102)	0,0035 (0,0123)	0,0309 (0,0207)	0,0309 (0,0207)	-0,0080 (0,0114)	-0,0123 (0,0187)
Cuiabá/MT						0,1297*** (0,0125)	0,0140 (0,0123)	0,0140 (0,0236)	0,0019 (0,0261)	0,0019 (0,0262)	-0,0138 (0,0110)	0,0050 (0,0208)
Florianópolis/SC						0,2710*** (0,0074)	0,0162** (0,0068)	0,0162 (0,0150)	0,1621*** (0,0280)	0,1621*** (0,0281)	0,0814*** (0,0079)	0,0848*** (0,0175)
Aracaju/SE						-0,0732*** (0,0123)	-0,0156 (0,0144)	-0,0156 (0,0181)	0,0045 (0,0391)	0,0045 (0,0394)	0,0140 (0,0128)	0,0227 (0,0212)
São José dos Campos/SP						0,4022*** (0,0060)	0,0219*** (0,0051)	0,0219* (0,0131)	0,0323 (0,0209)	0,0323 (0,0209)	-0,0319*** (0,0055)	-0,0272 (0,0178)

continuação

<i>ln (w real hora)</i>	(1) POLS	(2) FE i	(3) FE j	(4) FE (i, j)	(5) COND(i,j)	(6) POLS	(7) FE i	(8) COND i	(9) FE j	(10) COND j	(11) FE (i, j)	(12) COND(i,j)
João Pessoa/PB						-0,1842*** (0,0104)	-0,0217* (0,0122)	-0,0217 (0,0171)	0,0652 (0,0451)	0,0652 (0,0450)	0,0489*** (0,0143)	0,0411 (0,0385)
Teresina/PI						-0,1198*** (0,0118)	0,0322* (0,0169)	0,0322 (0,0211)	-0,0107 (0,0278)	-0,0107 (0,0279)	-0,0041 (0,0097)	0,0059 (0,0265)
Maceió/AL						-0,0569*** (0,0127)	-0,0024 (0,0133)	-0,0024 (0,0190)	0,0369 (0,0351)	0,0369 (0,0351)	0,0507*** (0,0119)	0,0475 (0,0337)
Natal/RN						-0,1173*** (0,0105)	-0,0267** (0,0117)	-0,0267* (0,0156)	0,0115 (0,0323)	0,0115 (0,0323)	-0,0093 (0,0113)	-0,0099 (0,0299)
São Luís/MA						-0,0356*** (0,0128)	0,0171 (0,0140)	0,0171 (0,0201)	0,0255 (0,0364)	0,0255 (0,0366)	-0,0269 (0,0200)	-0,0016 (0,0280)
Vitória/ES						0,2097*** (0,0071)	0,0457*** (0,0078)	0,0457*** (0,0142)	0,0359 (0,0379)	0,0359 (0,0380)	0,0325*** (0,0087)	0,0428 (0,0292)
Baixada Santista/SP						0,3892*** (0,0071)	0,0441*** (0,0064)	0,0441*** (0,0128)	-0,0021 (0,0185)	-0,0021 (0,0187)	-0,0180*** (0,0057)	-0,0132 (0,0136)
Manaus/AM						0,2534*** (0,0093)	0,0779*** (0,0106)	0,0779*** (0,0267)	-0,1140* (0,0623)	-0,1140* (0,0623)	-0,0180** (0,0085)	-0,0119 (0,0332)
Campinas/SP						0,4217*** (0,0040)	0,0290*** (0,0035)	0,0290*** (0,0104)	0,0249 (0,0225)	0,0249 (0,0225)	-0,0421*** (0,0034)	-0,0325** (0,0152)
Belém/PA						0,1008*** (0,0093)	0,0409*** (0,0116)	0,0409*** (0,0160)	0,0527 (0,0495)	0,0527 (0,0497)	0,0951*** (0,0105)	0,0910*** (0,0259)
Goiânia/GO						0,1184*** (0,0076)	0,0255*** (0,0074)	0,0255*** (0,0198)	0,0347 (0,0371)	0,0347 (0,0371)	-0,0095 (0,0077)	-0,0089 (0,0266)
Curitiba/PR						0,2718*** (0,0044)	0,0496*** (0,0049)	0,0496*** (0,0102)	0,0336 (0,0214)	0,0336 (0,0216)	-0,0046 (0,0059)	0,0066 (0,0191)
Porto Alegre/RS						0,3303*** (0,0043)	0,0280*** (0,0042)	0,0280*** (0,0096)	0,0555** (0,0270)	0,0555** (0,0270)	-0,0075* (0,0042)	-0,0046 (0,0145)
Fortaleza/CE						-0,0473*** (0,0065)	-0,0101* (0,0058)	-0,0101 (0,0107)	-0,1227*** (0,0329)	-0,1227*** (0,0329)	-0,0277*** (0,0066)	-0,0278** (0,0139)
Brasília/DF						0,3654*** (0,0057)	0,0848*** (0,0041)	0,0848*** (0,0278)	0,0643*** (0,0208)	0,0643*** (0,0208)	-0,0134*** (0,0031)	0,0015 (0,0228)
Salvador/BA						0,0954*** (0,0058)	0,0013 (0,0057)	0,0013 (0,0108)	0,0596** (0,0269)	0,0596** (0,0274)	-0,0192*** (0,0064)	-0,0046 (0,0144)
Recife/PE						-0,0749*** (0,0053)	0,0213*** (0,0046)	0,0213** (0,0105)	-0,0102 (0,0302)	-0,0102 (0,0302)	-0,0247*** (0,0033)	-0,0264* (0,0150)

	conclusão											
<i>ln (w real hora)</i>	(1) POLS	(2) FE i	(3) FE j	(4) FE (i, j)	(5) COND(i,j)	(6) POLS	(7) FE i	(8) COND i	(9) FE j	(10) COND j	(11) FE (i, j)	(12) COND(i,j)
Belo Horizonte/MG						0,2199*** (0,0040)	0,0056 (0,0040)	0,0056 (0,0079)	0,0010 (0,0149)	0,0010 (0,0150)	-0,0047 (0,0043)	0,0004 (0,0110)
Rio de Janeiro/RJ						0,2511*** (0,0033)	0,0757*** (0,0034)	0,0757*** (0,0090)	0,0539** (0,0211)	0,0539** (0,0211)	-0,0044 (0,0039)	-0,0004 (0,0138)
São Paulo/SP						0,3994*** (0,0024)	0,0365*** (0,0022)	0,0365*** (0,0098)	0,0264* (0,0142)	0,0264* (0,0143)	-0,0250*** (0,0022)	-0,0170 (0,0126)
Efeitos de tempo	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Constante	0,5972*** (0,0096)	2,2006*** (0,0021)	2,0144*** (0,0084)	2,2610*** (0,0017)		0,5246*** (0,0093)	2,1967*** (0,0022)		2,0067*** (0,0090)		2,2640*** (0,0018)	
Observações	5927470	5927470	5927470	5927470	5927470	5927470	5927470	5927470	5927470	5927470	5927470	5927470
R^2	0,565	0,264	0,437	0,111	0,940	0,593	0,276	0,919	0,444	0,787	0,104	0,940
R^2 Ajustado	0,565				0,931	0,593		0,912		0,777		0,931
<i>Within R</i> ²		0,432	0,307	0,376	0,358		0,433	0,433	0,307	0,307	0,376	0,358

Fonte: Elaboração própria, a partir da RAIS (2002-2014).

Notas: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; Erro padrão entre parênteses; As colunas de 6 a 12 foram estimadas com o efeito fixo de cada um dos 365 arranjos; Os modelos foram estimados com *clusters* de indivíduos (POLS e FEi), de firmas (FEj) e de indivíduos e firmas (FEij, CONDi, CONDj e CONDij).

Figura J.1 – Efeito da localização sobre o salário-hora real dos indivíduos nos grandes arranjos: amostra aleatória em firmas

Fonte: Elaboração própria, a partir da RAIS (2002-2014).

Nota: A cor cinza representa os efeitos estatisticamente significantes; As AMCs que não pertencem a algum arranjo representam a categoria de referência para os efeitos dos arranjos.

Figura J.2 – Efeito da localização sobre o salário-hora real dos indivíduos nos grandes arranjos: amostra sem *outliers* salariais

Fonte: Elaboração própria, a partir da RAIS (2002-2014). Nota: A cor cinza representa os efeitos estatisticamente significantes; As AMCs que não pertencem a algum arranjo representam a categoria de referência; *Outliers* representam a parcela 1% inferior e superior da distribuição de salários.

Tabela J.5 – Decomposição dos efeitos de localização e dos efeitos médios do trabalhador e da firma dos arranjos (segundo estágio)

	(1) POLS (FE LMA)	(2) POLS (FEi)	(3) FGLS (FEi)	(4) POLS (FEj)	(5) FGLS (FEj)
<i>ldensemp_lma</i>	0,0259* (0,0155)	0,0643*** (0,0085)	0,0669*** (0,0007)	0,0114* (0,0060)	0,0116*** (0,0004)
<i>larea_lma</i>	0,0083 (0,0154)	0,0514*** (0,0144)	0,0508*** (0,0011)	0,0303*** (0,0069)	0,0296*** (0,0007)
<i>(mean) i</i>	0,2503 (0,1793)				
<i>(mean) f</i>	0,0079 (0,0063)				
<i>frontier</i>	-0,0821** (0,0412)	0,0221 (0,0672)	0,0293*** (0,0061)	0,0166 (0,0322)	0,0150*** (0,0036)
<i>touristic</i>	0,0051 (0,0269)	-0,0493 (0,0670)	-0,0379*** (0,0069)	0,0064 (0,0206)	0,0139*** (0,0038)
Norte	-0,0029 (0,0896)	-0,1030 (0,0699)	-0,0958*** (0,0064)	0,0620 (0,0653)	0,0598*** (0,0045)
Nordeste	-0,2918*** (0,0721)	-0,3555*** (0,0421)	-0,3492*** (0,0025)	-0,0466** (0,0185)	-0,0394*** (0,0016)
Sul	-0,0401 (0,0279)	0,0093 (0,0329)	0,0111*** (0,0024)	-0,0666*** (0,0209)	-0,0659*** (0,0014)
Centro-Oeste	-0,0504 (0,0600)	-0,0528 (0,0562)	-0,0468*** (0,0039)	-0,0031 (0,0405)	0,0066* (0,0037)
<i>dummies de tempo</i>	Sim	Sim	Sim	Sim	Sim
<i>constante</i>	0,0641 (0,1571)	-0,6422*** (0,1115)	-0,6526*** (0,0087)	-0,3181*** (0,0512)	-0,3264*** (0,0058)
Observações	4681	4681	4681	4681	4681
<i>R</i> ² Ajustado	0,600	0,312		0,007	

Fonte: Elaboração própria, a partir da RAIS (2002-2014).

Notas: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; Erro-padrão robusto com *clusters* de LMA; Resultados assumem peso analítico proporcional ao número de indivíduos em cada LMA.

**Tabela J.6 – Decomposição dos efeitos de localização dos arranjos (segundo estágio):
Amostra aleatória em firmas**

FE_LMA	(1) POLS	(2) POLS	(3) POLS	(4) POLS	(5) FGLS	(6) POLS-BS
<i>ldensemp_lma</i>	0,0503*** (0,0107)	0,0502*** (0,0107)	0,0478*** (0,0109)	0,0478*** (0,0109)	0,0500*** (0,0004)	0,0447*** (0,0098)
<i>larea_lma</i>	0,0335** (0,0138)	0,0334** (0,0138)	0,0313** (0,0136)	0,0313** (0,0136)	0,0326*** (0,0005)	0,0612*** (0,0130)
<i>(mean) f</i>		0,0038 (0,0039)		-0,0006 (0,0025)	-0,0007** (0,0003)	-0,0005 (0,0026)
<i>(mean) i</i>			0,0300 (0,0196)	0,0302 (0,0199)	0,0308*** (0,0008)	0,0075 (0,0287)
<i>frontier</i>	-0,0493 (0,0540)	-0,0482 (0,0535)	-0,0358 (0,0526)	-0,0358 (0,0525)	-0,0289*** (0,0043)	-0,0495 (0,0478)
<i>touristic</i>	-0,0434 (0,0374)	-0,0435 (0,0374)	-0,0403 (0,0358)	-0,0403 (0,0358)	-0,0288*** (0,0023)	0,0093 (0,0342)
Norte	-0,0429 (0,1115)	-0,0434 (0,1113)	-0,0412 (0,1092)	-0,0411 (0,1093)	-0,0251*** (0,0032)	0,0789 (0,0693)
Nordeste	-0,3674*** (0,0476)	-0,3668*** (0,0475)	-0,3592*** (0,0465)	-0,3592*** (0,0465)	-0,3588*** (0,0012)	-0,2502*** (0,0327)
Sul	-0,0011 (0,0354)	-0,0012 (0,0353)	-0,0039 (0,0347)	-0,0039 (0,0347)	-0,0052*** (0,0009)	0,0239 (0,0333)
Centro-Oeste	-0,0309 (0,0698)	-0,0312 (0,0697)	-0,0312 (0,0685)	-0,0312 (0,0686)	-0,0118*** (0,0042)	0,0131 (0,0496)
<i>dummies tempo</i>	Sim	Sim	Sim	Sim	Sim	Sim
<i>constante</i>	-0,2210** (0,1111)	-0,2189* (0,1114)	-0,1923* (0,1125)	-0,1924* (0,1126)	-0,2075*** (0,0045)	-0,4782*** (0,1016)
Observações	4700	4700	4700	4700	4700	4700
R^2	0,525	0,526	0,540	0,540		0,299
R^2 Ajustado	0,523	0,524	0,538	0,538		0,296

Fonte: Elaboração própria, a partir da RAIS (2002-2014).

Notas: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; Erro-padrão robusto com *clusters* de LMA; Resultados das colunas 1-5 assumem peso analítico proporcional ao número de indivíduos em cada LMA.

Tabela J.7 – Decomposição dos efeitos de localização e dos efeitos médios do trabalhador e da firma dos arranjos (segundo estágio): Amostra aleatória em firmas

	(1) POLS (FE LMA)	(2) POLS (FEi)	(3) FGLS (FEi)	(4) POLS (FEj)	(5) FGLS (FEj)
<i>ldensemp_lma</i>	0,0478*** (0,0109)	0,0860*** (0,0224)	0,0853*** (0,0004)	0,0475* (0,0272)	0,0460*** (0,0005)
<i>larea_lma</i>	0,0313** (0,0136)	0,0738*** (0,0204)	0,0734*** (0,0007)	0,0377 (0,0274)	0,0375*** (0,0005)
<i>(mean) i</i>	0,0302 (0,0199)				
<i>(mean) f</i>	-0,0006 (0,0025)				
<i>frontier</i>	-0,0358 (0,0525)	-0,4521 (0,7300)	-0,1034*** (0,0249)	-0,3026 (0,4495)	0,1023*** (0,0272)
<i>touristic</i>	-0,0403 (0,0358)	-0,1004 (0,0856)	-0,0958*** (0,0033)	0,0439 (0,0638)	0,0420*** (0,0026)
Norte	-0,0411 (0,1093)	-0,0592 (0,0971)	-0,0185*** (0,0069)	0,1231 (0,0944)	0,1334*** (0,0035)
Nordeste	-0,3592*** (0,0465)	-0,2735*** (0,0596)	-0,2775*** (0,0015)	-0,1634 (0,1566)	-0,0549*** (0,0055)
Sul	-0,0039 (0,0347)	0,0929 (0,0609)	0,0887*** (0,0017)	0,0242 (0,0466)	0,0252*** (0,0010)
Centro-Oeste	-0,0312 (0,0686)	0,0104 (0,0671)	0,0062*** (0,0023)	0,0760 (0,0510)	0,0972*** (0,0044)
<i>dummies de tempo</i>	Sim	Sim	Sim	Sim	Sim
<i>constante</i>	-0,1924* (0,1126)	-0,9556*** (0,2688)	-0,9455*** (0,0051)	-0,5545 (0,3410)	-0,5467*** (0,0045)
Observações	4700	4700	4700	4700	4700
R^2	0,540	0,047		0,006	
R^2 Ajustado	0,538	0,043		0,001	

Fonte: Elaboração própria, a partir da RAIS (2002-2014).

Notas: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; Erro-padrão robusto com *clusters* de LMA; Resultados assumem peso analítico proporcional ao número de indivíduos em cada LMA.

**Tabela J.8 – Decomposição dos efeitos de localização dos arranjos:
segundo estágio com a área iluminada dos arranjos (2007)**

FE_LMA	(1) POLS	(2) POLS	(3) POLS	(4) POLS	(5) FGLS	(6) POLS-BS
<i>ldensemplight07</i>	0,0523*** (0,0189)	0,0519*** (0,0187)	0,0282 (0,0249)	0,0284 (0,0248)	0,0256*** (0,0009)	0,0487** (0,0215)
<i>larea_lma</i>	0,0105 (0,0182)	0,0095 (0,0178)	0,0035 (0,0150)	0,0033 (0,0149)	0,0018** (0,0008)	0,0440*** (0,0148)
<i>(mean) f</i>		0,0349* (0,0203)		0,0076 (0,0060)	0,0041*** (0,0011)	0,0012 (0,0420)
<i>(mean) i</i>			0,2644 (0,1853)	0,2612 (0,1846)	0,2997*** (0,0042)	0,0053 (0,1292)
<i>frontier</i>	-0,1376*** (0,0521)	-0,1375*** (0,0513)	-0,1213*** (0,0455)	-0,1215*** (0,0454)	-0,1185*** (0,0024)	-0,1416** (0,0597)
<i>touristic</i>	0,0076 (0,0415)	0,0073 (0,0416)	0,0138 (0,0270)	0,0137 (0,0271)	0,0295*** (0,0033)	0,0533 (0,0458)
Norte	-0,0760 (0,0883)	-0,0775 (0,0866)	-0,0336 (0,0801)	-0,0345 (0,0800)	-0,0203*** (0,0034)	0,0445 (0,0659)
Nordeste	-0,3836*** (0,0451)	-0,3819*** (0,0446)	-0,2896*** (0,0735)	-0,2904*** (0,0735)	-0,2744*** (0,0024)	-0,2111*** (0,0479)
Sul	-0,0528 (0,0383)	-0,0502 (0,0374)	-0,0521* (0,0298)	-0,0515* (0,0296)	-0,0492*** (0,0015)	0,0050 (0,0347)
Centro-Oeste	-0,0894 (0,0766)	-0,0888 (0,0753)	-0,0698 (0,0614)	-0,0700 (0,0614)	-0,0604*** (0,0028)	-0,0405 (0,0497)
<i>dummies de tempo</i>	Sim	Sim	Sim	Sim	Sim	Sim
<i>constante</i>	-0,0735 (0,1127)	-0,0626 (0,1125)	0,0849 (0,1499)	0,0853 (0,1500)	0,1118*** (0,0070)	-0,4278** (0,1672)
Observações	4681	4681	4681	4681	4681	4681
R^2	0,485	0,493	0,592	0,593		0,215
R^2 Ajustado	0,483	0,491	0,590	0,591		0,211

Fonte: Elaboração própria, a partir da RAIS (2002-2014).

Notas: * $p < 0,10$, ** $p < 0,05$, *** $p < 0,01$; Erro-padrão robusto com *clusters* de LMAs; Colunas 1-5 assumem peso analítico proporcional ao número de indivíduos em cada LMA.

Tabela J.9 – Decomposição dos efeitos de localização dos arranjos: segundo estágio com a área iluminada dos arranjos (2007) e amostra aleatória em firmas

FE_LMA	(1) POLS	(2) POLS	(3) POLS	(4) POLS	(5) FGLS	(6) POLS-BS
<i>ldensemplight07</i>	0,0650*** (0,0199)	0,0646*** (0,0200)	0,0611*** (0,0202)	0,0611*** (0,0202)	0,0609*** (0,0008)	0,0675*** (0,0166)
<i>larea_lma</i>	0,0186 (0,0164)	0,0186 (0,0164)	0,0173 (0,0159)	0,0173 (0,0159)	0,0187*** (0,0006)	0,0333** (0,0131)
<i>(mean) f</i>		0,0038 (0,0043)		-0,0008 (0,0028)	-0,0013*** (0,0003)	-0,0005 (0,0026)
<i>(mean) i</i>			0,0318 (0,0207)	0,0322 (0,0210)	0,0309*** (0,0009)	0,0076 (0,0293)
<i>frontier</i>	-0,1175* (0,0664)	-0,1162* (0,0658)	-0,0997 (0,0637)	-0,0998 (0,0637)	-0,1027*** (0,0019)	-0,0973** (0,0467)
<i>touristic</i>	-0,0238 (0,0409)	-0,0240 (0,0409)	-0,0217 (0,0391)	-0,0217 (0,0391)	-0,0190*** (0,0022)	0,0338 (0,0357)
Norte	-0,0939 (0,0922)	-0,0943 (0,0920)	-0,0897 (0,0905)	-0,0895 (0,0906)	-0,0926*** (0,0028)	0,0469 (0,0681)
Nordeste	-0,3696*** (0,0490)	-0,3690*** (0,0489)	-0,3609*** (0,0476)	-0,3609*** (0,0476)	-0,3613*** (0,0019)	-0,2415*** (0,0318)
Sul	-0,0166 (0,0390)	-0,0167 (0,0390)	-0,0190 (0,0382)	-0,0190 (0,0382)	-0,0127*** (0,0014)	0,0180 (0,0349)
Centro-Oeste	-0,0563 (0,0739)	-0,0567 (0,0739)	-0,0558 (0,0723)	-0,0557 (0,0724)	-0,0568*** (0,0020)	-0,0192 (0,0492)
<i>dummies de tempo</i>	Sim	Sim	Sim	Sim	Sim	Sim
<i>constante</i>	-0,1880* (0,1068)	-0,1860* (0,1071)	-0,1594 (0,1080)	-0,1596 (0,1080)	-0,1689*** (0,0045)	-0,4031*** (0,0991)
Observações	4700	4700	4700	4700	4700	4700
R^2	0,502	0,503	0,519	0,519		0,288
R^2 Ajustado	0,500	0,501	0,517	0,517		0,284

Fonte: Elaboração própria, a partir da RAIS (2002-2014).

Notas: * $p < 0,10$, ** $p < 0,05$, *** $p < 0,01$; Erro-padrão robusto com *clusters* de LMAs; Colunas 1-5 assumem peso analítico proporcional ao número de indivíduos em cada LMA.

Tabela J.10 – Decomposição dos efeitos de localização dos arranjos: segundo estágio com a área iluminada dos arranjos (2013) e amostra aleatória em firmas

FE_LMA	(1) POLS	(2) POLS	(3) POLS	(4) POLS	(5) FGLS	(6) POLS-BS
<i>ldensemplight13</i>	0.0666*** (0.0190)	0.0662*** (0.0191)	0.0627*** (0.0193)	0.0628*** (0.0193)	0.0630*** (0.0008)	0.0721*** (0.0162)
<i>larea_lma</i>	0.0180 (0.0160)	0.0179 (0.0159)	0.0167 (0.0155)	0.0167 (0.0155)	0.0173*** (0.0005)	0.0344*** (0.0133)
<i>(mean) f</i>		0.0037 (0.0042)		-0.0009 (0.0028)	-0.0013*** (0.0003)	-0.0004 (0.0026)
<i>(mean) i</i>			0.0312 (0.0205)	0.0316 (0.0208)	0.0299*** (0.0009)	0.0073 (0.0289)
<i>frontier</i>	-0.1095* (0.0653)	-0.1082* (0.0648)	-0.0924 (0.0627)	-0.0925 (0.0627)	-0.0906*** (0.0017)	-0.0895* (0.0459)
<i>touristic</i>	-0.0239 (0.0403)	-0.0242 (0.0403)	-0.0219 (0.0385)	-0.0218 (0.0385)	-0.0188*** (0.0019)	0.0343 (0.0357)
Norte	-0.0961 (0.0919)	-0.0965 (0.0917)	-0.0917 (0.0902)	-0.0916 (0.0903)	-0.0844*** (0.0023)	0.0531 (0.0684)
Nordeste	-0.3690*** (0.0482)	-0.3684*** (0.0480)	-0.3605*** (0.0468)	-0.3605*** (0.0468)	-0.3585*** (0.0019)	-0.2399*** (0.0318)
Sul	-0.0139 (0.0386)	-0.0140 (0.0386)	-0.0163 (0.0378)	-0.0163 (0.0379)	-0.0112*** (0.0014)	0.0160 (0.0345)
Centro-Oeste	-0.0529 (0.0729)	-0.0532 (0.0728)	-0.0525 (0.0714)	-0.0524 (0.0715)	-0.0487*** (0.0019)	-0.0187 (0.0497)
<i>dummies de tempo</i>	Sim	Sim	Sim	Sim	Sim	Sim
<i>constante</i>	-0.1907* (0.1067)	-0.1888* (0.1071)	-0.1625 (0.1080)	-0.1626 (0.1081)	-0.1699*** (0.0041)	-0.4220*** (0.0985)
Observações	4700	4700	4700	4700	4700	4700
R^2	0.509	0.510	0.525	0.525		0.295
R^2 Ajustado	0.507	0.507	0.523	0.522		0.292

Fonte: Elaboração própria, a partir da RAIS (2002-2014).

Notas: * $p < 0,10$, ** $p < 0,05$, *** $p < 0,01$; Erro-padrão robusto com *clusters* de LMAs; Colunas 1-5 assumem peso analítico proporcional ao número de indivíduos em cada LMA.

Tabela J.11 – Resultados do primeiro estágio do modelo 2SLS: Bartik IV

Idensemp_lma	2SLS (lbartik)	2SLS (lbartik)	2SLS (lbartikpub)
IV	0,7923*** (0,0366)	0,7911*** (0,0366)	0,6557*** (0,0535)
<i>larea_lma</i>	-0,9169*** (0,0413)	-0,9199*** (0,0410)	-0,7656*** (0,0572)
<i>(mean) i</i>		-0,0085*** (0,0013)	-0,0132*** (0,0035)
<i>(mean) f</i>		-0,0009 (0,0016)	-0,0003 (0,0018)
<i>frontier</i>	-0,7045*** (0,1420)	-0,7007*** (0,1415)	-1,3562*** (0,1587)
<i>touristic</i>	0,2516** (0,1224)	0,2463** (0,1212)	0,0580 (0,2719)
Norte	0,3223* (0,1763)	0,3291* (0,1757)	-0,7979*** (0,2478)
Nordeste	0,0607 (0,0661)	0,0761 (0,0645)	-0,7494*** (0,1309)
Sul	0,0802 (0,0692)	0,0826 (0,0692)	0,1050 (0,1283)
Centro-Oeste	0,0161 (0,1097)	0,0202 (0,1097)	-0,4063** (0,1859)
<i>dummies de tempo</i>	Sim	Sim	Sim
<i>constante</i>	2,4726*** (0,2748)	2,5011*** (0,2736)	3,8823*** (0,3753)
Observações	4623	4623	4511
R^2	0,915	0,916	0,737
R^2 Ajustado	0,915	0,916	0,736
F(20,356)	218,73		
F(22,356)		200,71	
F(22,347)			186,85
Prob. > F	0,00	0,00	0,00

Fonte: Elaboração própria, a partir da RAIS (2002-2014).

Notas: * $p < 0,10$, ** $p < 0,05$, *** $p < 0,01$; Erro-padrão robusto com *clusters* de LMA.

**Tabela J.12 – Decomposição dos efeitos de localização dos arranjos (segundo estágio):
Bartik IV para a densidade do emprego - Amostra aleatória em firmas**

FE_LMA	(1) POLS	(2) 2SLS (<i>lbartik</i>)	(3) POLS	(4) 2SLS (<i>lbartik</i>)	(5) POLS	(6) POLS	(7) 2SLS (<i>lbartikpub</i>)
<i>ldensemp_lma</i>	0,0505*** (0,0107)	0,0374*** (0,0109)	0,0479*** (0,0109)	0,0372*** (0,0106)	0,0514*** (0,0107)	0,0488*** (0,0109)	0,0346** (0,0138)
<i>larea_lma</i>	0,0336** (0,0139)	0,0622*** (0,0126)	0,0314** (0,0137)	0,0594*** (0,0126)	0,0325** (0,0138)	0,0304** (0,0136)	0,0544*** (0,0130)
(mean) <i>i</i>			0,0301 (0,0199)	0,0071* (0,0031)		0,0299 (0,0197)	0,0054* (0,0029)
(mean) <i>f</i>			-0,0006 (0,0025)	-0,0004 (0,0013)		-0,0005 (0,0026)	0,0006* (0,0003)
<i>frontier</i>	-0,0488 (0,0541)	-0,0864* (0,0472)	-0,0354 (0,0526)	-0,0839* (0,0457)	-0,0486 (0,0538)	-0,0352 (0,0523)	-0,1231** (0,0498)
<i>touristic</i>	-0,0428 (0,0374)	0,0071 (0,0328)	-0,0398 (0,0358)	0,0068 (0,0326)	-0,0464 (0,0377)	-0,0432 (0,0361)	-0,0116 (0,0320)
Norte	-0,0434 (0,1143)	0,0633 (0,0706)	-0,0414 (0,1120)	0,0646 (0,0702)	-0,0514 (0,1142)	-0,0492 (0,1120)	0,0111 (0,0703)
Nordeste	-0,3681*** (0,0477)	-0,2697*** (0,0329)	-0,3600*** (0,0465)	-0,2674*** (0,0330)	-0,3681*** (0,0477)	-0,3600*** (0,0465)	-0,2840*** (0,0344)
Sul	-0,0014 (0,0354)	0,0286 (0,0313)	-0,0042 (0,0347)	0,0264 (0,0310)	0,0005 (0,0354)	-0,0023 (0,0348)	0,0367 (0,0307)
Centro-Oeste	-0,0308 (0,0698)	-0,0099 (0,0501)	-0,0311 (0,0686)	-0,0088 (0,0496)	-0,0279 (0,0701)	-0,0283 (0,0690)	-0,0018 (0,0517)
<i>dummies de tempo</i>	Sim	Sim	Sim	Sim	Sim	Sim	Sim
<i>constante</i>	-0,2223** (0,1113)	-0,4699*** (0,0967)	-0,1938* (0,1128)	-0,4453*** (0,0979)	-0,2187** (0,1109)	-0,1906* (0,1125)	-0,4018*** (0,1061)
Observações	4647	4647	4647	4647	4532	4532	4532
R^2	0,527	0,301	0,541	0,310	0,529	0,543	0,323
R^2 Ajustado	0,525	0,298	0,539	0,307	0,527	0,541	0,320

Fonte: Elaboração própria, a partir da RAIS (2002-2014).

Notas: * $p < 0,10$, ** $p < 0,05$, *** $p < 0,01$; Erro-padrão robusto com *clusters* de LMAs; Modelos POLS assumem peso analítico proporcional ao número de indivíduos em cada LMA.

Tabela J.13 – Resultados do primeiro estágio do modelo 2SLS: Bartik IV
Amostra aleatória em firmas

ldensemp_lma	2SLS (lbartik)	2SLS (lbartik)	2SLS (lbartikpub)
IV	0,7894*** (0,0361)	0,7895*** (0,0355)	0,6583*** (0,0533)
larea_lma	-0,9089*** (0,0416)	-0,9202*** (0,0407)	-0,7594*** (0,0577)
(mean) i		0,0207** (0,0096)	0,0276** (0,0110)
(mean) f		-0,0030 (0,0019)	-0,0043*** (0,0007)
<i>frontier</i>	-0,6648*** (0,1452)	-0,6623*** (0,1404)	-1,3360*** (0,1634)
<i>touristic</i>	0,2603** (0,1232)	0,2562** (0,1215)	0,0761 (0,2755)
Norte	0,3276* (0,1763)	0,3373* (0,1745)	-0,8250*** (0,2488)
Nordeste	0,0440 (0,0667)	0,0581 (0,0663)	-0,7699** (0,1299)
Sul	0,0696 (0,0700)	0,0658 (0,0692)	0,0863 (0,1280)
Centro-Oeste	-0,0199 (0,1124)	-0,0090 (0,1110)	-0,5057** (0,1987)
<i>dummies de tempo</i>	Sim	Sim	Sim
<i>constante</i>	2,4399*** (0,2702)	2,5312*** (0,2707)	3,8431*** (0,3768)
Observações	4647	4647	4532
R^2	0,914	0,916	0,737
R^2 Ajustado	0,914	0,916	0,736
F(20, 359)	221,71		
F(22, 359)		210,40	
F(22, 349)			199,67
Prob. > F		0,000	0,000

Fonte: Elaboração própria, a partir da RAIS (2002-2014).

Notas: * $p < 0,10$, ** $p < 0,05$, *** $p < 0,01$; Erro-padrão robusto com *clusters* de LMA.

**Tabela J.14 – Resultados do primeiro estágio do modelo 2SLS:
Bartik IV e densidade do emprego da área iluminada (2013)**

ldensemplight13	(1) (lbartik)	(2) (lbartik)	(3) (lbartikpub)
IV	0,4202*** (0,0258)	0,4183*** (0,0256)	0,3517*** (0,0295)
<i>larea_lma</i>	-0,1394*** (0,0354)	-0,1414*** (0,0351)	-0,0725* (0,0401)
<i>(mean) i</i>		-0,0058*** (0,0009)	-0,0082*** (0,0019)
<i>(mean) f</i>		-0,0025 (0,0020)	-0,0020 (0,0025)
<i>frontier</i>	-0,0409 (0,1133)	-0,0392 (0,1128)	-0,4070*** (0,0982)
<i>touristic</i>	-0,2295* (0,1277)	-0,2345* (0,1271)	-0,3062 (0,1933)
Norte	0,3364** (0,1440)	0,3399** (0,1433)	-0,2400 (0,1667)
Nordeste	-0,2032** (0,0790)	-0,1861** (0,0790)	-0,6537*** (0,1073)
Sul	0,1520*** (0,0567)	0,1537*** (0,0566)	0,1809** (0,0759)
Centro-Oeste	0,2476*** (0,0917)	0,2501*** (0,0912)	0,0507 (0,1082)
<i>dummies de tempo</i>	Sim	Sim	Sim
<i>constante</i>	0,9595*** (0,1983)	0,9880*** (0,1968)	1,7781*** (0,2617)
Observações	4623	4623	4511
R^2	0,739	0,742	0,577
R^2 Ajustado	0,737	0,741	0,575
F(20,356)	147,20		
F(22,356)		158,00	
F(22,347)			153,11
Prob. > F	0,000	0,000	0,000

Fonte: Elaboração própria, a partir da RAIS (2002-2014).

Notas: * $p < 0,10$, ** $p < 0,05$, *** $p < 0,01$; Erro-padrão robusto com *clusters* de LMA.